THE UNIVERSITY OF READING

Molecular study of the mechanisms controlling the induction of the hexonate/hexuronate-utilization genes of *Salmonella enterica* (serotype Enteritidis) upon exposure to egg white

A Thesis Submitted to the Faculty of Life Sciences in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

By

Abdulameer M. Ghareeb

MSc. Genetic Engineering and Biotechnology

Supervisor: Professor Simon C. Andrews



Division of Microbiology

School of Biological Sciences University of Reading Reading United Kingdom

June 2018

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged

Abdulameer M. Ghareeb

Signed:

Date:

Abstract

Salmonella enterica serovar Enteritidis (SE) is an important Salmonella serotype that causes significant human infection through its contamination of poultry meat and eggs. Identifying processes that confer resistance to egg white (EW) might explain, and help combat, the ability of SE to survive in the harsh conditions of EW. The study described herein builds upon pervious work which shows that a set of hexonate/hexuronate (Hex) utilisation genes (dgoRKADT, uxuAB-uxaC and SEN1433-6 genes) are the most strongly induced when SE is exposed to EW. This observation is a surprise since no evidence for the presence of Hex substrates in EW is available, and these Hex utilisation (hex) genes are not know to have any role in EW survival. To study the regulation of the above 'hex' genes in response to EW, lacZ transcriptional fusions were generated to each of the potential promoter regions. The resulting transcriptional fusion data showed that seven of the fusions have activity markedly above that of the vector control, but two (dgoT; SEN2979) have weak activity, suggesting no promoter is present. To test the role of hexonates in regulating expression of the *hex* genes, four distinct hexonate compounds were employed (D-galactonic acid; D-mannono-1,4-Lactone; L-(+)-gulonic acid γ -lactone and gluconate). All four could act as sole carbon and energy source for SE at 42 °C (hen body temperature). The hexonates induced distinct regulatory responses in the expression of the various hex genes, indicating that hex gene expression is controlled in response to hexonates, as expected, and that this response involves multiple regulatory pathways. However, the data are inconsistent with any role for hexonates in induction of hex genes by EW.

EW, as expected, caused a major inhibition of *S*E growth, even when added at low levels (0.05%). In addition, the response of four *hex* genes (*sen1436, sen1432, dgoR* and *sen2977*) to EW was tested, and all four gave major induction effects (13-61 fold), confirming the previous report of EW induction of these genes. EW filtrate had little impact on EW-dependent *hex* gene induction, as did the provision of iron, temperature (30-42 °C) or pH (7-9). This finding indicated that an EW protein(s) of >10 kDa is responsible for the EW induction effect. Thus, four major EW proteins (albumin, conalbumin, ovomucoid and lysozyme) were tested for their ability to induce SEN1436 and a very strong induction effect (48 fold) was seen with lysozyme, suggesting this protein is primarily responsible for the EW-induction of the *hex* genes. Furthermore, three other *lacZ* fusions (SEN1432, *dgoR* and SEN2977) tested were also strongly induced by lysozyme (19-, 13- and 14-fold, respectively). This effect was confirmed with human lysozyme and with non-commercial sources of hen egg lysozyme. Thus, the results strongly suggest that lysozyme is the key factor in EW induction of *hex* gene expression; this is a novel finding.

The SEN1432 and *dgoR* genes, encoding GntR-like regulators, were inactivated to determine their role in *hex* gene control. The deletions caused a moderate increase in the expression of the SEN1432- and SEN1436-, and *dgoR-lacZ* fusions, but no major effect on EW or lysozyme induction. Complementation largely reversed the expression effects of the mutations. Thus, the results indicate that neither DgoR nor SEN1432 are involved in the induction of the *hex* genes by EW lysozyme. The membrane-damaging antibiotic, polymyxin B (PMB), also caused a major induction of the *hex* genes, although not so great as that of lysozyme. Experiments with *pmrA* and *phoP* global-regulatory mutants showed that the PMB effect is controlled by the PhoPQ and PmrAB systems, but that the response to lysozyme is only slightly dependent on these regulators. This conclusion was supported by complementation with *pmrAB*.

Thus, the control of the *hex* genes by the PmrAB and PhoPQ systems is complex, and involves additional factors. These results clearly show that the *hex* genes are subject to PMB induction and that this is largely controlled by PmrAB-PhoPQ. However, the response to lysozyme is only partly controlled by these factors indicating the involvement of another regulator. The results are consistent with a role for the observed *hex* gene induction by lysozyme in preserving the integrity of the cell envelope.

Poster/Oral Presentations

- Oral Presentation: 1st Postgraduate Symposium 9 June 2015, URS building, The University of Reading. UK.
- Oral Presentation: 2nd Postgraduate Symposium 26 May 2016, URS building, The University of Reading. UK
- Poster: 10th International Biometals Symposium 10–15 July 2016, Dresden, Germany.
- Poster: First Iraqi Cultural and Scientific Conference (ICSCR) 18 July 2016, Reading. UK.
- Poster: Microbiology Research Away Day 19 July 2016 Building no. 22, London Road Campus, The University of Reading. UK.
- Poster: The annual Conference of Microbiology Society 3–6 April 2017, EICC in Edinburgh. UK.
- Poster: The Seventh Congress of the International BioIron Society, Biennial World Meeting. 7– 11 May 2017, UCLA Meyer & Renee Luskin Conference Center Los Angeles, California, USA.
- Poster: The 6th International Symposium on Metallomics, 14–17 August 2017.University of Vienna, Austria.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Professor Simon C. Andrews for the continuous support of my PhD study, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. His great humanity that encouraged me to overcomes all the circumstances that I encountered during the study. I could not have imagined having a better advisor and mentor for my study and will keep proud that I am one of his students.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr Sheila MacIntyre, Professor Martin J Woodward, for their encouragement and insightful comments which incented me to widen my research from various perspectives.

My great thanks to the French team (Dr. Florence Baron and Dr. Sophie Jan) for provision of the EWF and wild type strain *Salmonella enterica* serovar Enteritidis PT4.

My sincere thanks to Dr. Chris Smith (Department of Chemistry /University of Reading) for assistance in providing the hexonate compounds and Professor John Gunn (Department of Microbiology/The Ohio State University), who provided me the *pmrAB* mutants strains.

Thanks are also given to Kevan Hanson, Jane, Jordan Kirby for their technical support. Finally, I should not forget to thank the all members of the Centre for Sterilization Services (CSS) for their hard work in preparing my all needs.

Thanks are due to the Ministry of the Higher Education and Scientific Research/ Iraq and Iraqi Cultural Attache for the financial support they have provided during my study. Thanks for Institute for genetic engineering and biotechnology for nominated me for this study.

I would also like to thank all the current and previous members of Prof. Simon Andrew's iron research lab, especially; Fawzi, Rana, Aida, Salem, Chris, Afrah, Shymaa, Arvind and Dhama for their scientific criticized discussions and being incessantly supporting and helping me to overcome difficulties during my PhD.

Dedication

I would like to dedicate this humble effort to:

My parents, they are springs of the strength and ambition in my life.

My dear wife T. Zedan, having you in my life completes me, you supported me in the most difficult circumstances with great motivation, and all your family.

My brother Haidar and his wife N. Zedan, I cannot express my thanks to him, he is you nger than me but greater in everything.

My brothers, sisters and them precious families for their counting support.

My friends whose become as a part of my family.

To the beloved who shone like the sun in my life, my son Saif.

Table of contents

Chapter 1. Introduction	. 1
1.1 Salmonella spp	. 1
1.1.1 Characteristics	. 1
1.1.2 Classification and nomenclature	. 1
1.2 Salmonella pathogenicity	. 5
1.3 Salmonella and food poisoning	. 8
1.4 Salmonella enterica serovar Enteritidis (SE)	11
1.5 Egg structure and composition	14
1.5.1 Albumen (egg white) proteins	16
1.5.1.1 Ovalbumin	17
1.5.1.2 Ovotransferrin	18
1.5.1.3 Lysozyme	18
1.5.1.4 Ovomucoid	20
1.5.1.5 Ovomucin	20
1.5.1.6 Other proteins	20
1.5.2 Carbohydrates in egg white	22
1.6 Sugar acids (hexonates) and Salmonella	23
1.7 Iron and Salmonella	26
1.8 Egg white antimicrobial activity	30
1.9 SE survival in egg white	33
1.9.1 Genetic response of SE to egg white exposure	34
1.9.2 Induction of genes encoding hexonate/ hexuronate catabolism systems by exposure of SE to egg- white medium	36
1.9.2.1 The <i>dgo</i> genes	39
1.9.2.2 The <i>uxu-uxa</i> genes	42
1.9.2.3 The SEN1433-6 genes	44
1.9.2.4 <i>ybhC</i> gene	46
1.10 Aims and objectives	47
Chapter 2. Materials and Methods	49
2.1. Materials	49
2.1.1 Chemicals	49
2.1.2 DNA marker	49
2.1.3 Protein marker	49

2.1.4 Restriction and polymerization enzymes	50
2.1.5 Bacterial media	50
2.1.5.1 Luria-Bertani broth and agar	50
2.1.5.2. Super Optimal Broth (SOC)	51
2.1.5.3. M9 minimal medium	
2.1.5.4 Media sterilisation.	
2.1.6 Antibiotics	51
2.1.7 Bacterial Strains, Plasmids and Primers	
2.2 Methods	57
2.2.1. Chemically competent cells - preparation and transformation	57
2.2.2 Extraction and purification of nucleic acids	57
2.2.2.1 Plasmid miniprep	58
2.2.2.2 Total DNA extraction	59
2.2.3 Determination of DNA concentration	60
2.2.4 Polymerase Chain Reaction (PCR) protocol	60
2.2.5 Colony PCR	61
2.2.6 Agarose gel electrophoresis	
2.2.7 Purification of PCR products	
2.2.8 Restriction digestion	63
2.2.9 PCR extraction from agarose gel	63
2.2.10 Ligation of vector with PCR product	64
2.2.11 lacZ fusions construction	66
2.2.12 β-Galactosidase assay	67
2.2.13 Preparation competent cells of SE	69
2.2.14 Electroporation	69
2.2.15 Gene inactivation procedure	
2.2.15.1 The Red disruption system	
2.2.15.2 Primer design	
2.2.15.3 PCR amplification of CAT cassette	
2.2.15.4 Induction and preparation of host cell	
2.2.15.5 Electroporation with linear DNA	
2.2.15.6 Elimination of the chloramphenicol resistance cassette	
2.2.16 Phenotypic studies	

2.2.17 Protein work	73
2.2.17.1 Prepare egg white and egg white filtrate	73
2.2.17.2 Lysozyme purification process	74
2.2.17.3 Protein quantification	75
2.2.17.4 Polyacrylamide gel electrophoresis (SDS)	76
2.2.17.8 Protein overexpression	77
2.2.18 Hexonate preparation and synthesis.	78
Chapter 3: Generation and preliminary analysis of 'hex' gene lacZ transcriptional fusions	79
3.1 Introduction	79
3.1.1 Aim of this chapter	80
3.2 Generation of hex gene lacZ transcriptional fusions.	80
3.3 Identification of promoter sites	89
3.4 Preliminary expression analysis of the <i>hex</i> genes	93
3.4.1 Activity of <i>lacZ</i> fusions in <i>E.coli</i> TOP10	93
3.4.2 Activity of <i>lacZ</i> fusions in <i>S</i> . Enteritidis PT4	97
3.5 Conclusions and Discussion	100
Chapter 4. Egg-white factors influencing expression of SE hex genes	105
4.1 Introduction	105
4.1.1 Effect of EW on <i>hex</i> gene expression in SE	105
4.1.2 Energy/carbon sources in EW and changes in energy metabolism upon exposure to EW	106
4.1. Aims of this chapter	107
4.2 Can SE grow on the four available Hex compounds (gluconate, D-mannono-1,4-lactone, gulonat lactone and/or D-galactonate)?	e γ- 107
4.2.1 Growth on glucose and glycerol at 37 and 42 °C	107
4.2.2 Growth on hexonates at 37 and 42 °C	109
4.2.3 Impact of hexonates on hex gene expression in SE	115
4.2.3.1 Do the SE transformants grow well with a hexonate as sole carbon/energy source?	115
4.2.3.2 Effect of D-galactonate of <i>hex</i> gene expression	116
4.2.3.3 Effect of D-mannonate of <i>hex</i> gene expression	118
4.2.3.4 Effect of D-gluconate of <i>hex</i> gene expression	119
4.2.3.5 Effect of L-gulonate of <i>hex</i> gene expression	121
4.3 Effect of egg white and EW components on <i>hex-lacZ</i> gene expression in SE PT4	122
4.3.1 Effect of egg white on the growth of SE.	122

4.3.2 The effect of the EW, EW filtrate and EW proteins on hex gene induction.	. 125
4.3.3 The effect of specific EW proteins on <i>hex</i> gene induction	. 129
4.3.3.1 Use of commercial egg-white proteins	. 129
4.3.3.2 The effect of lysozyme-free egg-white protein on <i>hex</i> gene induction	. 133
4.3.4 Comparing the effect of lysozyme from chicken egg with that from human milk lysozyme, and inactivated lysozyme	heat . 137
4.3.5 The effect of temperature, pH and iron on EW-mediated induction of hex gene expression	. 138
4.4 Discussion	. 141
4.4.1 Effect of hexonates on growth and expression	. 142
4.4.2 The role of lysozyme in inducing the <i>hex</i> genes in EW	. 145
4.4.4 Conclusion	. 147
Chapter 5: Role of the hex gene regulators, SEN1432 and DgoR	. 150
5.1 Introduction	. 150
5.2 Generation of SE deletion mutant	. 153
5.2.1 SEN1432 and <i>dgoR</i> gene knock out	. 153
5.2.1.1 Primers design	. 155
5.2.1.2 Replacement of SEN1436 and <i>dgoR</i> in SE PT4 with a Cm ^R cassette	. 157
5.2.1.3 Confirmation of the deletion mutants	. 160
5.2.1.4 Removal of the Cm ^R cassette from the single-gene knockout strains	. 162
5.2.2 Phenotypic analysis of SE mutants	. 163
5.2.3 Complementation of the SEN1432 and <i>dgoR</i> mutants	. 165
5.3. Discussion	. 170
Chapter 6: The Role of the Two-Component Regulators, PmrAB and PhoPQ, in mediating the <i>hex</i> get response to lysozyme	ne . 172
6.1 Introduction	. 172
6.1.1 Aims of this chapter	. 180
6.2 Transformation <i>hex-lacZ</i> fusions into <i>Salmonella phoP</i> and <i>pmrA</i> mutant strains	. 180
6.3 Phenotypic analysis of ST transformants	. 181
6.4 Is the effect of the <i>pmrA</i> mutation on <i>hex</i> gene expression reversed by <i>pmrAB</i> complementation?	. 184
6.4.1 Primer design and the amplification of <i>pmrAB</i>	. 184
6.4.2 Cloning of <i>eptA-pmrAB</i> into pSU18	. 185
6.4.3 Effect of complementation of the <i>ST pmrA</i> and <i>phoP</i> mutant, with pSU18- <i>eptA-pmrAB</i> , on <i>hex</i> gene induction by lysozyme and PMB	; . 187
6.5 Discussion	. 191

Chapter 7: Attempted overexpression of SEN1432 and dgoD	194
7.1 Amplification of SE genes of interest	194
7.2 Cloning amplified genes into the intermediate vector pJET1.2./blunt.	196
7.3 Cloning amplified genes into overexpression vector pET21a	199
7.4 Small-scale overexpression of SEN1432.	201
7.5 Discussion	205
Chapter 8: General discussion	207
8.1. Introduction	207
8.2 Exploring the potential promoter regions of <i>hex</i> genes	208
8.3 EW factors influencing expression of SE hex genes	209
8.3.1 Utilisation the hexonates as energy/carbon sources and the effect of hexonates on expression	210
8.3.2 Effect of EW on growth and <i>hex</i> gene expression in SE	212
8.3.2.1 The role of lysozyme in inducing the <i>hex</i> genes in EW	214
8.4 Role of the <i>hex</i> gene regulators, SEN1432 and DgoR	215
8.5 The two-component sensor-regulators, PmrAB and PhoPQ, mediate the response of the <i>hex</i> genes polymyxin B, and have a minor role in the response to lysozyme/EW	to 217
8.6 Attempted overexpression of SEN1432 and <i>dgoD</i>	219
8.7 Suggested future work	220
9. References	222
Appendix 1	244
Appendix 2	249
Appendix 3	253
Appendix 4	257
Appendix 5	258
Appendix 6	274
Appendix 7	275
Appendix 8	276
Appendix 9	283
Appendix 10	287
Appendix 11	288
Appendix 12	289

List of Figures:

Figure 1.1: Diagram of Salmonella species, sub species, and groups of serovars
Figure 1.2: Food ingredients associated with nontyphoidal Salmonella food poisoning outbreaks in
Europ 10
Figure 1.3: Transmission Electron Micrograph of Salmonella enterica
Figure 1.5: Egg production
Figure 1.6: Hexonate/hexuronate structures
Figure 1.7: D-Galactonate degradation
Figure 1.8: L-Idonate degradation
Figure 1.9: Siderophores of Salmonella Typhimurium. B. Chemical structures and ions (m/z) of enterobactin, salmochelin and degradation products
Figure 1.10: Siderophores of Salmonella Typhimurium
Figure 1.11: Scheme of the catecholate siderophore transport systems of S. enterica. DHBS, 2,3- dihydroxybenzoylserine
Figure 1.12: Schematic representation of the dgo cluster of SE PT4
Figure 1.13: Schematic representation of the uxuAB-uxaC operon of SE PT4
Figure 1.14: Schematic representation of the SEN1432-6 cluster of SE PT4
Figure 2.1:1kb ladder (Thermo scientific Gene Ruler TM). DNA marker was used to estimate the size and quantity of DNA
Figure 2.2: Protein marker used to estimate the size and quantity of DNA
Figure 2.3: Map of the pJET1.2/blunt
Figure 2.4: Map of pRS1274 lacZ transcriptional vector
Figure 2.5: Physical maps of the transcriptional pRS-SEN1436-lacZ fusion vectors generated during this study
Figure 2.6: Structure of three commercial hexonates used in this work and synthesised D-galactonic acid
Figure 3.1: Schematic representation of the organisation of the egg-white induced hex genes of SE PT4.
Figure 3.2: Gel electrophoresis of purified PCR products
Figure 3.3: Plasmids DNA extraction of potential pJET1.2 clones carrying promoter fusion fragments. 84
Figure 3.4: Electrophoretic analysis of pJET1.2 clones by double digestion with EcoRI and BamHI 85
Figure 3.5: Cloning steps of potential promoter regions
Figure 3.6: Gel electrophoresis of gel-extracted gene fragments for cloning into pRS1274
Figure 3.7: Plasmids DNA extractions of potential pRS1274 clones
Figure 3.8: Potential pRS1274 clones analysed by double digestion with EcoRI and BamHI

Figure 3.9: Analysis of the potential promoter regions using BPROM	
Figure 3. 10: Expression of hex genes in L-broth using E. coli TOP10 as host.	
Figure 3.11 Summary of β-galactosidase activity (average and maximum levels) of all nine lac	Z fusions. 95
Figure 3.12: Confirmation of transformants of SE by plasmid isolation followed by double dig EcoRI and BamHI.	estion with 98
Figure 3.13: β-galactosidase activity of three pRS1274 lacZ-fusion constructs in SE (SEN1436 and dgoR).	5, SEN2977 99
Figure 3.14: Comparison of β-galactosidase activity (average and maximum levels) of three pl constructs in SE and E. coli TOP10	RS1274 100
Figure 3.15: Schematic diagram showing predicted organisation of S. enterica PT4 hex gene tr based upon the promoter fusion activity	anscription
Figure 4.1. SE growth in 0.1-1.6% glucose, or 0.4% glycerol, in M9 minimal medium at 37 (A °C (B)) and 42 109
Figure 4.2. SE growth in 0.1-1.6% gluconate in M9 minimal medium at 37 (A) and 42 °C (B).	111
Figure 4.3. SE growth in 0.1-1.6% L-(+)-gulonic acid γ-lactone in M9 minimal medium at 37 °C (B).	(A) and 42
Figure 4.4. SE growth in 0.1-1.6% galactonate (A) or mannonate (B) in M9 minimal medium a	at 42 °C. 114
Figure 4.5. Comparison of hexonate and hexuronate utilisation in E. coli K-12 and SE.	115
Figure 4.6. Effect of lacZ-fusion vectors on growth of SE in 0.4% glycerol or D-galactonate in minimal medium.	M9 117
Figure 4.7: β-galactosidase activity (maximum levels were taken at 6-10 h growth) of lacZ fus Salmonella enterica serovar Enteritidis (strain PT4-P125109) in the presence 0.49 galactonate in M9 minimal medium.	ions in % D- 119
Figure 4.8: β-galactosidase activity (maximum levels at 6-10 h growth) of lacZ fusions in Salmenterica serovar Enteritidis (strain PT4-P125109) in the presence 0.4% D-manno	nonella nate 120
Figure 4.9: β-galactosidase activity (maximum levels at 6-10 h growth) of lacZ fusions in Salmenterica serovar Enteritidis (strain PT4-P125109) in the presence 0.4% gluconate minimal medium.	nonella in M9 121
Figure 4.10: β-galactosidase activity (maximum levels at 6-10 h growth) of lacZ fusions in Sal enterica serovar Enteritidis (strain PT4-P125109) in the presence 0.4% L-gulonat minimal medium.	monella e in M9 122
Figure 4.11: The effect of up 10% v/v egg white on growth of SE in 0.4% glycerol M9 minima at 42 °C.	al medium 124
Figure 4.12: The effect of egg white concentration in different temperatures on growth of SE i glycerol M9 minimal medium.	n 0.4% 125
Figure 4.13: Effect of EW on expression of sen1436-lacZ in wildtype SE in M9 medium at 42	°C 127

Figure 4.14:	Effect of EW filtrate on expression of sen1436-lacZ in wildtype SE in M9 medium at 42	°C. 128
Figure 4.15:	Effect of EW and EW filtrate on expression of sen1432, sen2977 and dgoR in wildtype S in M9 medium at 42 °C	SE 129
Figure 4.16:	Effect of lysozyme, albumin, conalbumin and ovomucoid on expression of SEN1436 in wildtype SE in M9 medium at 42 °C.	131
Figure 4.17:	SDS-PAGE (12% acrylamide) analysis of the commercial egg white proteins	132
Figure 4.18:	Anti-lysozyme Western blot analysis of the commercial egg white proteins	133
Figure 4.19:	Graph showing β -galactosidase activity at 45 minutes for SEN1436 lacZ fusions in Salmonella enterica serovar Enteritidis (strain PT4-P125109) in the presence different concentrations (all numbers is indicted to mg/ml) of egg white proteins	134
Figure 4.20:	Effect of lysozyme and lysozyme/mucin-free EW (LMFEW) on the induction of SEN1432, dgoR and SEN2977 in wildtype SE in M9 medium at 42 °C	36, 135
Figure 4.21:	Elution profile of the fractionation of mucin-free EW protein by cation-exchange chromatography (A) and SDS-PAGE analysis (B) of resulting fractions	136
Figure 4.22:	Induction of the SEN1436 lacZ fusion in SE PT4 in the purified fractions.	137
Figure 4.23:	Effect of human and chicken lysozyme, and inactivated lysozyme, on SEN1436-lacZ expression in SE PT4.	139
Figure 4.24:	Effect of temperature, pH and combining EW with EWF on SEN1436-lacZ induction in PT4	SE 141
Figure 4.25:	Graph showing the effect of iron on SEN1436-lacZ expression in SE wildtype in the presence 0.1% of EW.	142
Figure 5.1: 7	The components of the lambda Red recombineering system.	150
Figure 5.2: 7	The pKD3 plasmid linear templates	151
Figure 5.3: S	Schematic representation of the λ Red recombineering technique	151
Figure 5.4: C	Genetic map of the SEN1432 region	153
Figure 5.5: 0	Genetic map of the dgoR region in the SE PT4 genome	154
Figure 5.6: I	Primer locations for amplifying the cat gene of in pKD3.	155
Figure 5.7: I	Primer locations for inactivation of the SEN1432 gene.	155
Figure 5.8: 0	Genetic map of the dgoR gene	156
Figure 5.9: N	Map of the pKD3 vector	157
Figure 5.11:	Electrophoretic analysis of pJET1.2 clones contains the CmR cassette PCR fragments fro pKD.	om 158
Figure 5.12:	Gel electrophoresis of PCR confirmation products of SEN1432 and dgoR genes using wildtype chromosomal DNA as template	159
Figure 5.13:	Colony PCR to confirm the (ΔSEN1432):: <i>cat</i> mutation.	160

Figure 5.14:	Colony PCR to confirm of the $(\Delta dgoR)$::cat mutation 1	60
Figure 5.15:	Colony PCR to confirm the loss of the CmR cassette from the SEN1432 and dgoR mutant	s. 62
Figure 5.16:	Comparison of the expression of SEN1432 and SEN1436 in the wildtype and Δ SEN1432 strain in response to the of presence egg white, lysozyme and egg white filtrate	.63
Figure 5.17:	Comparison of the expression of dgoR in the wildtype and Δ dgoR strain in response to the presence of egg white, lysozyme and egg white filtrate	e 64
Figure 5.18:	Gel electrophoresis of PCR amplification products of the dgoR and SEN1432 genes 1	65
Figure 5.19:	Restriction map of pSU18 1	66
Figure 5.20:	Electrophoretic analysis of pSU18 plasmid following double digestion with BamHI and EcoRI.	66
Figure 5.21:	Electrophoretic analysis of pSU18 clones suspected to carry dgoR and SEN1432 PCR fragments	.67
Figure 5.22:	Electrophoretic analysis of double-digested (BamHI/EcoRI) pSU18-dgoR and -SEN1432 candidates.	.67
Figure 5.23:	Effect of complementation of the dgoR and SEN1432 mutations on SEN1432-lacZ and dgoR-lacZ expression. 1	.68
Figure 5.24:	Schematic of GntR family protein domains 1	70
Figure 6.1a:	A typical Gram-negative bacterial envelope components 1	72
Figure 6.1b.	General Structure of Salmonella LPS 1	73
Figure 6.2a:	A model of the activation and interaction of the PhoPQ and PmrAB TCSs in Salmonella s	рр. 75
Figure 6.2b:	Structure of SE LPS and targets for modification mediating resistance to PM 1	76
Figure 6.3: A	Antibacterial mechanisms of polymyxin 1	78
Figure 6.4: I	Electrophoretic analysis of pRS1274 hex gene fusions in ST 1	80
Figure 6.5: (Graph showing expression of hex genes in phoP and pmrA mutants in response to lysozyme and polymyxin	e .82
Figure 6.6: S	Schematic representation of the pmrAB (basRS) genes of ST JSG210 1	.83
Figure 6.7: 0	Gel electrophoresis of PCR amplification products of the eptA-pmrAB genes1	84
Figure 6.8: 0	Gel electrophoresis of pJET-pmrAB 1	85
Figure 6.9: I	Electrophoretic analysis of putative pSU18-eptA-pmrAB clones1	86
Figure 6.10:	Electrophoretic analysis of putative pSU18-eptA-pmrAB clones double digested with BamHI and EcoRI	.86
Figure 6.11:	Graph showing effect of pmrAB complementation on the expression of three hex genes in phoP and pmrA mutant backgrounds, in response to lysozyme and polymyxin	.87

Figure 6.12. Summary of the major regulatory effect exerted by PMB and lysozyme through PmrAB a PhoPQ.	ind 189
Figure 7.1: Gel electrophoresis of PCR amplification products of SEN1432 and dgoD genes 1	194
Figure 7.2: Gel electrophoresis of purified pJET-dgoD21R & 28R and pJET-SEN1432-21R & 28R1	195
Figure 7.3: Gel electrophoresis of pJET-SEN1432-21R & 28R digested with HindIII & NdeI 1	197
Figure 7.4: Gel electrophoresis of pJET- dgoD-21R & 28R digested with HindIII & NdeI 1	197
Figure 7.5: Gel electrophoresis of pET21a digested with XhoI, or HindIII & NdeI 1	198
Figure 7.6: Gel electrophoresis of potential pET21a-SEN1432-21R & and pET21a-SEN1432-28R isolates digested with HindIII & NdeI	199
Figure 7.7: SDS-PAGE analysis SEN1432 overexpression from of BL21/DE3-(pET21a-SEN1432-21R (#5) following overexpression	t) 201
Figure 7.8: SDS-PAGE analysis SEN1432 overexpression from of BL21/DE3 (pET21a-SEN1432-28R	:). 202
Figure 7.9: Codon usage analysis of SEN1432 and ompC generated by GenScript	203

List of tables:

Table1.1: Proteins of egg white	17
Table 1.2: Nutrient composition of avian eggs	22
Table 1.3: Egg-white induced genes from the dgo cluster involved in D-galactonate metabolisim inSalmonella enterica serovar Enteritidis (strain PT4-P125109).	41
Table 1.4: Egg-white induced genes from the uxuAB-uxaC operon involved in mannonate utilisation in o S. Enteritidis (strain PT4-P125109).	f 43
Table 1.5: Egg-white induced genes from the SEN1432-6cluster involved in hexonate utilisation in of S. Enteritidis (strain PT4-P125109).	45
Table 1.6: Egg-white induced putative pectin esterase gene of S. Enteritidis (strain PT4-P125109)	47
Table 2.1: Antibiotics used in this study.	52
Table 2.2: Strains used in this study	53
Table 2.3 Plasmids used in this study.	53
Table 2.4 Designed primers.	55
Table 2.5 Sequencing primers.	55
Table 2.6 Primers used for knockouts.	55
Table 2.7 Primers for amplification whole genes of regulators.	56
Table 2.8 Designed primers for sequencing cloned pmrAB.	56
Table 2.9 Primers to amplify genes of interest (SEN1432 and dgoA) for over-production	56

Table 3.1 Summary of predicted promoters.	92
Table 4.1: Proteins of egg white	131
Table 7.1: The primers combination used to amplify each gene and the gene product during overexpress	sion. 195
	-/-

Abbreviations

Amp	Ampicillin		
Amp ^R	Ampicillin resistance		
APS	Ammonium persulphate		
Cam	Chloramphenicol		
Cm ^R	Chloramphenicol resistant		
EDTA	Ethylene diaminotetra acetic acid		
EW	Egg white		
FWF	Egg white filtrate		
	Egg white mulate		
h	Hour(s)		
hex	Hexonates genes		
Hex	Hexonates compunds		
IPTG	Irsopropyl-β-D-thiogalactoside		
LBA	Luria-Bertani agar medium		
LB	Luria-Bertani broth medium		
min	Minute(s)		
OD	Optical density		
OD ₆₀₀	Optical density at 600 nanometer		
PAGE	Polyacrylamide gel electrophoresis		
SE	Salmonella enterica serovar Enteritidis PT4-P125109		
TEMED	N, N, N', N', Tetramethyl ethylene diamine		
X-gal	5-Bromo-4-chloro-3-indolyl-ß-D-galactoside		

Chapter 1. Introduction

1.1 Salmonella spp.

1.1.1 Characteristics

Salmonella is a Gram-negative, facultatively anaerobic, flagellated, rod-shaped, non-spore forming and regularly motile bacterium belonging to the Enterobacteriaceae family. It is about 2-3 x 0.4-0.6 μ m in size forming colonies about 2-4 mm diameter. Optimal growth temperature of most Salmonella serotypes is 35-37 °C with capability to grow at a range of temperatures from 5 to 47 °C (Pui *et al.*, 2011). However, a number of serotypes have the potential to grow at temperatures as low as 2 °C or as high as 54 °C. The preferred environment is neutral pH 6.5-7.5 with possibility of growth at pH between 4 and 9. Salmonella growth needs high water activity between 0.99 and 0.94 but it is able to survive in dried foods where water content is less than 0.2 (Pui *et al.*, 2011).

The biochemical characteristics of *Salmonella* indicate that they are able to reduce nitrates to nitrites, produce hydrogen sulphide on triple-sugar iron agar, and they are usually able to use citrate as the sole carbon source, are non-lactose fermenting and D-glucose is fermented with the production of mixed acids and usually hydrogen gas. Other carbohydrates usually fermented are L-arabinose, maltose, D-mannitol, D-mannose, L-rhamnose, D-sorbitol (except *S. enterica* subsp. indica), trehalose, D-xylose and dulcitol. *Salmonella* is oxidase negative, catalase positive, indole and Voges-Proskauer (VP) negative, methyl red and Simmons citrate positive and urea negative (Neidhardt, 2005).

1.1.2 Classification and nomenclature

The first recognition of this genus was in 1885 with the identification of *Salmonella choleraesuis* (later known as *Salmonella enterica*) by Daniel Elmer Salmon, a veterinary pathologist, and his assistant (Theobald Smith) who were working on a The United States Department of Agriculture

research program on the cause of hog cholera (Berge et al., 2004). The genus of Salmonella consists of two species: S. enterica and S. bongori (also referred to as subsp. V). S. enterica is divided into six sub-species; *enterica*, *salamae*, *arizonae*, *diarizonae*, *houtenae* and *indica*, referred to these sub-species in Latin numerals from I-VI respectively (Figure 1.1) (Le Minorb and Popoff, 1987; Brenner et al., 2000; Hurley et al., 2014). A number of nomenclature systems have been applied over time for classification of Salmonella serotypes. Salmonella enterica serovars can be classified based on antigenic reaction with specific antibodies directed against surface antigens according to the Kauffmann-White scheme (Pui, et al., 2011). There are three major antigens employed: somatic (oligosaccharide) antigens (O), flagellar antigens (H) and capsular antigens (K). They are composed of heat stable lipopolysaccharide of the outer membrane (O), heat labile proteins associated with flagella (H) and heat sensitive polysaccharide of the capsule (K), respectively. Agglutination by antibodies specific for the various O antigens is employed to group Salmonellae into the 6 serogroups: A, B, C1, C2, D and E. According to Pui et al. (2011), flagellar H antigen is highly specific for Salmonella serotyping, more so than somatic antigens which often have the disadvantages of cross-reactivity with other genera.



Figure 1.1: Diagram of *Salmonella* **species, sub species, and groups of serovars**, figure adapted from Hurley *et al.*, 2014.

Salmonella enterica serovars can be also classified as three groups according to their ability to infect a wide variety of hosts as follows: The first group is **unrestricted serovars** including serovars (e.g. *S.* Typhimurium and *S.* Enteritidis) which infect nearly all animals and cause enteric disease. The second group of serovars (e.g. *S.* Dublin) is **host adapted.** These serovars prefer specific hosts causing severe systemic infection in these hosts, but less effectively infect other hosts. While the third group is **host restricted**, which are firmly restricted to one very specific host and only cause systemic infection; serovar examples include Typhi in humans and Gallinarum in birds (Singh, 2013).

Previously, Salmonella was named based on the original place of isolation, such as Salmonella London and Salmonella Indiana. Subsequently, this system was replaced by a phage typing system based on susceptibility toward different selected bacteriophages (Pui, et al., 2011). This system was employed successfully in epidemiological investigations of Salmonella outbreaks sources and resulted in 200 identified serotypes within S. Typhimurium and 27 phage types within S. Enteritidis (Porwollik et al., 2005). Therefore, a nomenclature system was adopted, designated the provisional or definitive phage type number (depending on the reliability of the designation), which depends on phage susceptibility (Akiba *et al.* 2011). The phage type prevalence in different geographical areas is varied. For example, phage type 1 (PT1) is common in Baltic countries and Russia, whereas Phage type 4 (PT4) is most often seen in Western European countries. PT8 is frequently found in the United States. PT4 was mainly responsible for large epidemics of SE infection in the UK in the late 1980s and was also a major cause for egg-borne SE outbreaks over ten years from 1992-2002 (Gillespie et al., 2005). However, a decline in human S. Enteritidis PT4 infection was recorded in England and Wales from 1997 due to industry control programmes in the poultry sector, including vaccination of layer flocks (Cogann and Humphrey, 2003). Other studies indicate that the greatest increases have occurred in S. Enteritidis PT1- and PT14b-related infections since 2000 (Gillespi and Elson. 2005; Janmohamed et al., 2011). PT4, PT8, and PT13a comprise the majority of Entertidis infections worldwide (Zhensheng et al., 2009).

The last main classification is based on phylogeny generates phylogenetic trees by comparison of 16S rRNA or other gene sequences. According to this system, there are 2500 recognised serotypes belong to only two species, based on differences in 16S rRNA sequence analysis: 2463 serotypes for *S. enterica* and 20 serotypes for *S. bongori*. Now, relevant organizations like WHO and CDC used this system (Pui *et al.*, 2011; Silva *et al.*, 2012). Typically, *S.* Enteritidis, *S.* Typhimurium

and *S*. Heidelberg are the three most frequent serotypes recovered from humans (Boyen *et al.*, 2008). Recently whole genome sequencing (WGS) and single nucleotide polymorphisms (SNP) have been used as bacterial strain discrimination techniques in monitoring *Salmonella* epidemiology (Dallman *et al.* 2018).

There are a wide variety of methods commercially available for *Salmonella* detection and identification. These include the use of antibodies to *Salmonella* antigens (Enzyme-Linked Immuno-Sorbent Assay [ELISA], immuno-chromatography, antibody coated dipsticks, latex agglutination), and polymerase chain reaction (PCR). However, these techniques do not have the ability to detect cell number less than $(10^4-10^5 \text{ cells/ml})$, so a pre-enrichment step is needed to reach detectable numbers of cells which means identification within one day is not possible (Berge *et al.*, 2004).

1.2 Salmonella pathogenicity

Salmonella is a significant public health concern around the world. Infections by Salmonella are responsible for more than half a million deaths each year worldwide, 16 million cases of typhoid fever and 1.3 million gastroenteritis cases according to WHO estimates (Pui *et al.*, 2011). Indeed, these huge numbers have an economic effect as diagnosis and treatment are expensive as are studies that monitor outbreaks and research on anti-Salmonella approaches. Monitoring and control programs have been set up in many countries with varying success (with a 30% decrease obtained from 2007 to 2014 in the EU) to reduce Salmonella contamination on the farm, such as national control programs implemented in the European Union (EFSA, 2012). The number of human cases and outbreaks has decreased in recent years, and efforts in the poultry industry have contributed in decreasing the flock prevalence. However, Salmonella enterica still remains a major bacterial pathogen causing a significant fraction of human foodborne disease (EFSA, 2012).

According to the clinical symptoms, four main human disease patterns are caused by *Salmonella*: enteric fever, gastroenteritis, bacteraemia and other complications of nontyphoidal salmonellosis including the chronic carrier state.

Enteric fever is caused by *Salmonella* serovars Typhi and Paratyphi which are responsible for typhoid and paratyphoid fever, respectively. These tend to infect humans through ingestion of food or water contaminated with human waste. The disease is characterized by fever, headache and diarrhoea. 10% of patients may relapse, die or encounter serious complications such as gastrointestinal bleeding and intestinal perforation while proper antibiotic treatment will enable recovery within 10-14 days (Connor and Schwartz, 2005).

Gastroenteritis is caused by at least 150 *Salmonella* serotypes, which also called nontyphoidal salmonellosis or enterocolitis, including *S*. Typhimurium and *S*. Enteritidis. The disease is caused via ingestion of water or food contaminated with animal waste. It is characterized by nausea, vomiting, diarrhea, headache, chills and fever up to 39 °C. The symptoms can be mild to severe and may last between 5-7 days (Chimalizeni *et al.*, 2010).

Bacteraemia is one of the nontyphoidal salmonellosis complications: About 8% of untreated cases of salmonellosis result in bacteraemia if the pathogen passes the intestinal barrier and enters the blood stream. It has been associated with highly invasive serotypes like Cholearaesuis or Dublin (Wood *et al.*, 2008).

Chronic carrier state: nontyphoidal serotypes persist in the gastrointestinal tract from 6-12 weeks, thus, salmonellosis can be spread by chronic carriers who potentially infect many individuals, especially those who work in food-related industries (Pui *et al.*, 2011).

S. enterica has large chromosomal regions, known as *Salmonella* pathogenicity islands (SPI), which contains virulence genes. Five pathogenicity islands (SPI-1 to -5) have been found in a

range of serovars of S. enterica, with a further five islands with characteristics of SPIs identified in the complete genome of S. enterica serovar Typhi (Makham et al., 2003). Two of these pathogenicity islands (SPI-1 and SPI-2) encode two type III secretion systems (T3SS). The SPI-1 T3SS is mainly involved in host invasion while the SPI-2 T3SS plays a role in survival within the host cell (Desin *et al.*, 2009). The adherence of bacteria to the cell surface is essential for bacterial invasion and survival inside the host cells (Thiagarajan et al., 1996). The initial step in the colonization of host tissue and an essential stage in the pathogenesis of salmonellosis involves the fimbriae. These are an important factor for adhesion to different cell surfaces, and survival and persistence in the host (Thiagarajan et al., 1996). S. enterica has numerous cell surface structures involved in the process of infection such as type 1 (SEF21), thin aggregative (SEF17), SEF14, long polar (LPF) and plasmid-encoded (PEF) fimbriae, and flagella (Cogan et al., 2004). Three kinds of fimbrial gene are more important in pathogenicity and the attachment of SE to intestinal epithelium (sef14, 17, 21) (Salehi et al., 2011). SEF14 fimbriae are present in a few serovars including SE and closely related serovars suggesting that SEF14 fimbriae may affect serovarspecific virulence traits (Collighan and Woodward, 2001). SEF14 fimbriae contribute to the adherence of the pathogen to chicken ovarian granulosa cells. However, there are specific antibodies for these fimbriae in egg-yolk which reduce the invasion and colonization in the first stages of egg infection (Thiagarajan et al., 1996).

Salmonella serotypes contain virulence plasmids of different sizes and genetic composition. However, all contain a preserved region called the *spv* operon (~8 kb) which is important for the survival and multiplication of the bacteria inside particular organs such as the liver and the spleen (Castila *et al.*, 2006). According to reports, there is an increasing resistance of *Salmonella* towards common antibiotics. For example, strains have been detected that have multiple antibiotic resistances (MAR) in many countries such as UK, USA and Saudi Arabia (e.g. *S.* Typhimurium phage type 104-DT104) (Rankin and Shelley, 1998; Yoke-Kqueen, *et al.*, 2008).

WHO consider that *S*E was the leading cause of food-borne salmonellosis in 2008 which induces salmonellosis in humans characterized by diarrhea, fever, headache, abdominal pain, nausea and vomiting (CDC, 2007; Shah *et al.*, 2011). Furthermore, *S*E is also reported in cases of invasive and extra-intestinal infections such as septicaemia, arthritis, endocarditis, meningitis and urinary tract infections (Kobayashi *et al.*, 2009).

1.3 Salmonella and food poisoning

Survival of Salmonella in various environments for long periods contributes to infection transmission. Salmonella infection can be transmitted to humans through other vectors such as rats, flies and birds (Newell et al., 2010). Furthermore, there is also the possibility for transfer of contamination through the food production chain (Bouchrif et al., 2009). Any contaminated raw materials that come into contact with food processing equipment can cause infection (Wong et al., 1998). Salmonella infections are a concern in the poultry industry with infection of poultry leading to meat and egg contamination (Cox and Pavic, 2010). The most common foods associated with salmonellosis are foods of animal origin, such as egg, poultry, pork, beef and raw dairy products (Peris et al., 2010). An outbreak highlighted the emerging challenge of controlling Salmonella in different food environments such as high concentration of salt, low water levels and high temperature (Shachar and Yaron, 2006). Cross-contamination through carrying ice cream in a tanker which previously carried contaminated raw egg was the largest recorded outbreak of SE as it led to over 200,000 illnesses in several states (Hope et al., 2002). In the United States, contaminated eggs have been estimated to result in 180,000 to over 660,000 illnesses each year costing around \$150-870 million each year (Cox and Pavic, 2010). According to Kamelia et al.

(2011), poultry infections by *SE* have increased in Egypt especially the egg industry which has significant economic effects. Egg products were associated with 45% of the salmonellosis outbreaks occurring from 1999 to 2008 (Gantois *et al.*, 2008). Guard (2001) also report that most outbreaks are associated with consumed contaminated eggs.

In general, the consumption of liquid egg products has increased as it used in various food products like sausage and pasta, and sterilization is an important concern in particular when used for uncooked food (Baron *et al.*, 1997). An effective way to destroy most microorganisms is pasteurization, however the egg contains thermos-sensitive protein which is coagulated at around 60 °C. Therefore, pasteurization treatment of egg is difficult and requires incubation between 40 and 48 °C for a period of 1 to 5 days (Baron *et al.*, 2010). This aspect is one of the main concerns of the egg products industry and may contribute to the prevalence of outbreaks related to consumption of eggs.

Food of plant origin, such as vegetables, fruits, and juices, are also of increasing concern (Hanning *et al.*, 2009). Outbreaks of salmonellosis have been linked to a wide variety of fresh fruits and vegetables including apple, cantaloupe, alfalfa sprout, mango, lettuce, cilantro, unpasteurized orange juice, tomato, melon, celery and parsley (Pui *et al.*, 2011). These foods, which are usually eaten without cooking, could be contaminated during production, storage or in retail outlets. Carrier handlers that have an acute infection could also be a potential source of infection.

Large scale studies using foodborne outbreak data indicate that the most common contributing factors associated with foodborne salmonellosis are: cross-contamination, inadequate cooling or refrigeration, inadequate heat treatment or contamination from food handlers (Gormley *et al.*, 2008). The analysis of contributing factors of 195 outbreaks reported in Europe with a single implicated food vehicle is provided in Figure 1.2 (Peris *et al.*, 2010). The most commonly

associated food ingredients were egg (62.6%), chicken (16.9%) and meat (11.8%). Almost all (92%) of the egg containing food items can be accounted for by either inadequate cooking (66%) or contaminated raw food (26%), and this dominates the overall analysis of contributing factors.



Figure 1.2: Food ingredients associated with nontyphoidal *Salmonella* food poisoning outbreaks in Europe, 2003-June 2011(Peris *et al.*, 2011).

Various methods have been applied to reduce the level of contamination in eggs as a food source. In general, there are two approaches, those applied before and those applied after laying. The 'before' group includes genetic selection, husbandry methods, breeding practices, disinfection practices and hen vaccination (Baron *et al.*, 2011). A previous study showed there is a relationship between hen resistance to caecal colonization and genetic background (Berthlot *et al.*, 1998). Another study comparing two lineages of hens observed significant differences in the expression of genes encoding proteins involved in the defence against *Salmonella* colonization (Sadeyen *et al.*, 2006). Thus, hen selection may be an efficient way to improve resistance to colonization by *Salmonella*. Stress can affect hen infection through weakening the immune system and reducing the influx of macrophages to the reproductive organs (Wigley *et al.*, 2005). Hen housing systems have an impact as well via factors such as the size of flock and the design of cages. Furthermore, levels of air contamination in the breeding environment are related to egg shell contamination (De

Reu *et al.*, 2005). Another method applied in the European Union is hen vaccination against *S*E which became obligatory from 2008 if the *Salmonella* prevalence exceeded 10% in laying hens. However, there is no evidence to verify its effectiveness. On the other hand, there is a wide range of anti-*Salmonella* food additives used such as organic acids including propionic, formic acid and lauric acid which lead to reduced colonisation of the digestive tract, this decreasing the contamination rate of the environment (Van Immerseel *et al.*, 2004).

The 'after' approach includes packaging, transport and storage of eggs. Advance practices for egg collection on farm, sorting, packaging, storage and delivery contribute to reducing contamination, and the most important factor of these processes is minimizing the temperature in order to limit penetration of *Salmonella* into the egg contents (Baron *et al.*, 2011).

According to WHO (2005), up to one third of the world population suffers from a foodborne infection each year. There are number of factors that may promote spread of pathogens, which is an increasing global issue. For example, the increase of food consumption of animal origin and the globalization of the food trade and international travel. Such diseases have a negative economic and public health impact. Therefore, many countries have implemented surveillance and intervention strategies as attempts to limit foodborne illnesses. These systems depend, for instance, on epidemiological analyses (outbreak investigations) and subtyping approaches to recognize the source of infection and highlight regional effects (Pires *et al.*, 2011).

1.4 Salmonella enterica serovar Enteritidis (SE).

Salmonella enterica serovar Enteritidis (SE) (Figure 1.3) is one of the serotypes of the genus Salmonella which has the ability to infect many animal species, including human beings. The disease can develop in humans causing infections varying from gastroenteritis to septicemia, leading to severe damage and possibly death (Castila *et al.*, 2006). Adult chickens are one of the

most important carriers of this serotype. Over last 25 years, *SE* has been a continuous worldwide threat to public health through contaminated eggs (EFSA, 2012).



Figure 1.3: Transmission Electron Micrograph of *Salmonella enterica*, coloured using feature-detection software; the colours do not provide any information. Source (http://fineartamerica.com/featured/salmonella-enteritidis-bacterium-ab-dowsett.html).

SE is passed to humans mainly via handling and consumption of contaminated poultry meat and eggs - most studies have identified poultry and poultry products as the major source of human infection (Shah *et al.*, 2011). According to outbreaks report from the Centers for Disease Control and Prevention, from 1985 to 1999, eggs and egg-containing products (e.g. homemade ice cream and Caesar salad dressing) were concerned as major vehicles of *SE* infection in 298 (80%) of the 371 known source outbreaks in US (Patrick *et al.*, 2004; Shah *et al.*, 2011). Lane *et al.* (2014) reviewed 67 years of surveillance data of *SE* in England and Wales as the largest and most persistent epidemic of foodborne infection associated with the consumption of contaminated chicken meat and eggs. This survey estimated >525,000 persons became ill during the course of the epidemic with 27,000 hospitalizations and 2,000 deaths. Reports from outbreak investigations

in Spain, Hungary, France, Norway and the United States implicated eggs as the source (Kottwitz et al., 2010). In Brazil, Salmonella species were responsible for 1408 (23.2%) of the 6062 of investigated outbreaks of food associated infection between 1999 and 2008 related with consumption of foods of animal origin such as poultry meat, eggs and their products (Kottwitz et al., 2010). According to WHO data from 2001 to 2005, SE was the most common isolated servor from contaminated poultry meat and egg worldwide (65% of the isolates), followed by S. Typhimurium (12%) and S. Newport (4%) (Hendriksen et al., 2011). In Africa, SE and S. Typhimurium represented 26 and 25% of the human isolates, respectively, and in Asia, Europe and Latin America, SE was the most frequent isolate (38, 87 and 31%, respectively). While, in North America, S. Typhimurium was the most frequently reported (29%) followed by SE (21%) and other Salmonella spp. (21%) (Galanis et al., 2006). The US state and territorial health departments reported 677 SE outbreaks between 1990 and 2001, which accounted for 23,366 illnesses, 1988 hospitalizations and 33 deaths. In 2006, countries within the European Union reported 1729 outbreaks caused by SE leading to 13,853 illnesses, 2714 hospitalizations and 14 deaths. The Health Protection Agency (HPA) of the UK reported 4194 cases of foodborne SE infection in 2008 (Shah et al., 2010). The last report from HPA (May 2018) referred there was a decrease in reports of SE from 6,489 (2007) to 2,356 (2016). However, there was an increase in Salmonella Typhimurium from 1,528 (2007) to 1,770 (2016) reported cases.

Advanced genomic analysis showed that field strains of *SE* are relatively genetically homogeneous. However, the main genetic differences between these strains is displayed at the level of single nucleotide polymorphisms (Shah *et al.*, 2014). Despite the limited genomic diversity, variation in phenotypic traits, including the ability to form a biofilm, growth to high cell density, production of high-molecular-mass LPS and survival within egg albumin, have been

commonly observed among field isolates of *SE* (Clavijo *et al.*, 2006; Yim *et al.*, 2010). Molecular methods, including analysis of the macrorestriction patterns of chromosomal DNA after PFGE, have been used to characterize *SE* (Zou *et al.*, 2010). For example, molecular analysis of 674 *Salmonella* isolates from 12 serotypes identified 66 different subtypes (Gaul *et al.*, 2007).

1.5 Egg structure and composition

Humans have used eggs as a food traditionally because they supply essential nutrients. In contrast with other types of egg, chicken eggs are the most important and the most consumed by humans. The average weight of a chicken egg is 58 g. It consists of water, protein and lipids (74, 12 and 11%, respectively; Beltiz *et al.*, 2009). The egg consists of three parts: shell 11%, yolk 31% and egg white 58% (Johnson and Ridlen, 2015). The shell surrounds the albumen and yolk, protecting the embryo and providing gas permeability (figure 1.4). It is a calcareous and porous shell made of calcium carbonate of 0.2-0.4 mm thicknesses (Beltiz *et al.*, 2009). There are two membranes lining the inside of the shell separated by an air space. The yolk is located at the centre of the egg, it is composed of water (48%), proteins (17%) and lipid (35%), and is very rich in vitamins supplying enough nutrients for growth of the embryo (Beltiz *et al.*, 2009).

There is a membrane around the yolk that isolates the egg white and the yolk called the vitelline membrane. The germinal disc (blastoderm) is located at the top of one side of the yolk. The egg white is an aqueous medium consisting mainly of water and protein, helping to protect the embryo. It consists of four layers that differ in viscosity (Beltiz *et al.*, 2009).

• The **chalaziferous layer** is the inner portion of the egg white which is located between the inner thin egg white and the egg yolk. It is a thin but very firm, fibrous layer of albumen closely surrounding the yolk. It branches and twists on the opposite apical sides of the yolk into two chalazae (keeping the yolk in the centre) that extend into the thick albumen. The

chalazae are fixed to the ends of the egg and function to position the embryo correctly (shown in figure 1.4).

- The **inner layer of thin albumin** lies between the chalaziferous layer and the thick, fibrous albumin layer
- The **thick albumin** is dense and fibrous due to high levels of ovomucin. It helps to centralise the embryo.
- The **outer layer of thin albumin** is in direct contact with the shell membranes and is relatively thin.



Figure 1.4: Egg structure. Source: http://www.geauga4h.org/poultry/egg_parts.htm.

The production of the bird egg consists of a series of steps that occur as the egg enters and passes along the hen's reproductive tract (oviduct) (shown in figure 1.5). The yolk of the egg enters the top of the oviduct and passes into the infundibulum where it spends about 15 minutes. A membrane is added around the yolk and, if the hen has been inseminated, fertilisation occurs in this section of the oviduct. The yolk then spends about three hours in the magnum where the egg white is formed and then one hour in the isthmus where the shell membrane is laid down. The main part of the egg shell is formed in the tubular shell gland and the shell gland pouch which takes about 20 hours (figure 1.5). The egg shell is sometimes referred to as a bio-ceramic because it is made up of calcium carbonate with an organic matrix running through it (Beltiz *et al.*, 2009).



Figure 1.5: Egg production, total time of formation take around 25 hours: 15 minutes in infundibulum, 3 hours in magnum, hours in isthmus, 20 hours in uterus, 10 minutes for laying. Source: 1 http://www.wisc.edu/ansci_repro/lec/lec1/female_hist.html.

1.5.1 Albumen (egg white) proteins

Albumen is an intracellular fluid consisting of a 10% aqueous solution of various proteins and very low amounts of other compounds (Mine *et al.*, 1995). The pH of albumen of a freshly laid egg is 7.6–7.9 at 24 °C (Beltiz *et al.*, 2009). Table 1.1 lists the most important albumen proteins. Ovalbumin (54%), ovotransferrin (12%), ovomucoid (11%), ovomucin (3.5%) and lysozyme (3.4%) are among the major proteins in egg white (Abeyrathne *et al.*, 2014). These proteins are known to have unique functions.

Protein	Total protein	MW (kDa)	Function
	(%)		
Ovalbumin	54	44.0	storage protein and major source of amino acids
Conalbumin	12	76	binds metal ions
(Ovotransferrin)			
Ovomucoid	11	28	proteinase inhibitor
Ovomucin	3.5	$5.5-8.3 \times 10^3$	inhibits viral hemagglutination
Lysozyme	3.4	14.3	N-acetylmuramidase
(Ovoglobulin G ₁)			
Ovoglobulin G ₂	4	30-45	good foam builders
Ovoglobulin G ₃	4	30-45	
Flavoprotein	0.8	32	binds riboflavin
Ovoglycoprotein	1.0	24	
Ovomacroglobulin	0.5	760-900	inhibits serine and cysteine proteinase
Ovoinhibitor	1.5	49	proteinase inhibitor
Avidin	0.05	68.3	binds biotin
Cystatin	0.05	12.7	Inhibits cysteine peptidases
(ficin inhibitor)			

Table1.1: Proteins of egg white (adapted from Belitz et al., 2009).

1.5.1.1 Ovalbumin

Ovalbumin is the major egg white protein with a molecular weight of 45 kDa. It is a monomeric phosphoglycoprotein composed of 385 amino acids and an isoelectric point of 4.5 (Abeyrathne *et al.*, 2014). It is a storage protein and major source of amino acids for the developing embryo (Mine *et al.*, 2008). Ovalbumin is the only egg white protein which contains free sulphhydryl groups (Nisbet *et al.*, 1981). Ovalbumin is a heterogeneous molecule with variation in its composition, which includes the degree of phosphorylation, glycosylation and genetic variance. The amino acid sequence and 3D structure of ovalbumin show similarities to a group of serine protease inhibitors but it does not have inhibitory activity (Abeyrathne *et al.*, 2014).

Introduction

1.5.1.2 Ovotransferrin

Ovotransferrin is the second major egg white protein. It has a molecular weight of 77.7 kDa with a pI of 6.1 and is a glycoprotein consisting of 686 amino acid residues (Mine et al., 1995). Ovotransferrin is a member of an iron-binding protein group known as transferrins. Ovotransferrin is well known to have a strong iron-binding capability (Ko et al., 2009; Wu and Acer-Lopez, 2012). Ovotransferrin is synthesized in the hen oviduct before being deposited in the albumen fraction of eggs. It comprises two similar domains, the amino (NH_2) and carboxy (COOH) terminal domains. It binds Cu(II), Al(III), Co(II), and other metals, as well as Fe(III) very tightly and specifically (Ichimura et al., 1989.) Ovotransferrin is present in apo- (iron free) and holo-(iron bound) forms, and the chemical and physical properties of these two forms differ significantly (Wu and Acer-Lopez, 2012). Ovotransferrin functions as an antimicrobial agent and transports iron to the developing embryo. Ovotransferrin binds up to two ferric iron atoms at its two similar anion-binding sites, this makes it difficult for harmful bacteria to acquire sufficient iron in the egg white. Thus ovotransferrin acts as an antimicrobial component. Considering the low concentration of iron (around 25 μ M) and the high concentration of ovotransferrin (1.7 mM) in egg white, it can be concluded that ovotransferrin is predominantly iron-free under the natural conditions of egg white and that all the iron present in egg white is bound to ovotransferrin due to the strong affinity (K_a around 10^{20} mol/l) (Thapon *et al.*, 1994; Baron *et al.*, 2016).

1.5.1.3 Lysozyme

Another important small protein found in egg white is lysozyme. The molecular weight of lysozyme is 14.4 kDa and it consists of a single polypeptide chain of 129 amino acid residues (Radziejewska *et al.*, 2008). It is a strongly basic protein with isoelectric point (pI) of 10.7 and has four disulphide bridges leading to high thermal stability (Abeyrathne *et al.*, 2014). Avian egg

white is a rich and easily available source of lysozyme (~3.5 g/L) (Pellegrini et al., 1997). This enzyme activity is essential because is provides non-specific protection mechanisms due to its ability to control the growth of bacteria (Bera et al., 2005). Lysozyme activity causes degradation of the glycosidic (1-4) β-linkage between the N-acetylglucosamine and the N-acetylmuramic acid residues of the bacterial peptidoglycan of Gram-positive bacteria (Nikaido 2003). This activity leads to destruction of bacterial cells by damaging the cell wall. This activity does not work as well with Gram-negative bacteria because of the resistance provided by the outer-membrane to the lytic action of lysozyme, which prevents entrance into the periplasm of molecules larger than 650 Da (Nikaido 2003). However, there are non-hydrolytic mechanisms of lysozyme activity against Gram-negative bacteria involving membrane disruption (Masschalck et al, 2003). In particular, induction of pore formation in the outer membrane of *Escherichia coli* has been recently recognized for lysozyme. This occurred due to the high affinity of lysozyme for the LPS monolayer through its ability to enter this monolayer whenever polysaccharide moieties exist leading to reorganization of LPS monolayer (Derdre et al., 2013; Baron et al., 2016). Therefore, it is possible that SE is affected through membrane disruption under the specific conditions of egg white. In addition, lysozyme causes inhibition against DNA and RNA synthesis in E. coli (Pellegrini et al., 2000). However, it is not surprising that pathogenic bacteria have developed mechanisms of resistance to hydrolysis by lysozyme. As suggested from Baron et al. (2016), SE show resistance to the peptidoglycan lytic activity of lysozyme due to the presence of its outer membrane and to the periplasmic lysozyme inhibitor (PliC). The *pliC* gene was identified by Callewaert et al. (2008) as knockout of this gene showed resulted in susceptibility to the antimicrobial activity of lysozyme. However, according to Baron et al. (2016), there is still a
possibility that synergetic activity of other egg white conditions (high pH and metal-ion limitation) could be supportive to the membrane disruption mechanisms of lysozyme against *S*E.

1.5.1.4 Ovomucoid

Ovomucoid is a glycoprotein with a molecular weight of 28.0 kDa and pI of 4.1. About 25% of the protein is carbohydrate attached via Asp residues. There are disulphide bridges and no free sulfhydryl groups. Ovomucoid is a well-known trypsin inhibitor (Mine *et al.*, 1995). The 3D structure has three domains which are cross-linked via disulphide bonds. The domains are homologous to pancreatic secretory trypsin inhibitor. The trypsin inhibitor reactive site is located in domain 2. Ovomucoid is very stable due to its multiple disulphide bridges (Abeyrathne *et al.*, 2014).

1.5.1.5 Ovomucin

Ovomucin is a viscous glycoprotein responsible for the gel structure of the thick egg white, consisting of two subunit types (α and β) which differ in amino acid sequence and carbohydrate level (Hiidenhovi, 2007; Abeyrathne *et al.*, 2014). It contributes about 3.5% of protein in egg white has a complex molecular weight of 5.5- 8.3 × 10³ kDa (Abdou *et al.*, 2013). There are two different forms of ovomucin: the soluble form is present in both the thick and thin part of the albumen, while insoluble ovomucin is found only in thick part. The insoluble form has 2.5-fold higher levels of β -ovomucin compared to soluble ovomucin. This higher level contributes to the greater viscosity of the insoluble ovomucin (Hiidenhovi, 2007; Abeyrathne *et al.*, 2014). Ovomucin showed obvious anti-bacterial activity against *Escherichia coli* and *Salmonella*, with an MIC of 0.05 mg/mL and 0.4 mg/mL, respectively, but showed no effect against *Staphylococcus aureus* (Shan *et al.*, 2013). In addition, ovomucin has been shown to inhibit haemagglutination by viruses (Abdou *et al.*, 2013).

1.5.1.6 Other proteins

Avidin is a strongly basic glycoprotein synthesized in the hen oviduct and deposited in the albumen fraction of eggs. It is a homo tetrameric protein (subunits of 15.6 kDa and 128 amino acid residues). Avidin is a trace component (0.05%) of egg white, it has ability to tightly and specifically bind biotin of the vitamin B group (Abdou *et al.*, 2013). Ovoglobulin is present as two forms; ovoglobulins G2 (4%) and G3 (4%), which have molecular weights of 36 and 45 kDa, respectively. The biological function of these proteins has not been clearly elucidated, but they appear to be important in the foaming capacity of egg white (Abdou *et al.*, 2013).

Ovomacroglobulin (0.5%) is the second largest egg glycoprotein after ovomucin and its molecular weight is 760-900 kDa. Ovomacroglobulin, like ovomucin, has the ability to inhibit hemagglutination (Abdou *et al.*, 2013). Ovoflavoprotein is an acidic protein with a molecular weight of 32-36 kDa, and contains a carbohydrate moiety (14%) made up of mannose, galactose and glucosamines, 7-8 phosphate groups and 8 disulphide bonds. After being transported from the blood to the egg white, most of the riboflavin (vitamin B2) is stored in the egg white bound to an apoprotein called flavoprotein. One mole of apoprotein binds to one mole of riboflavin (Abdou *et al.*, 2013). It has antimicrobial properties due to depriving microorganisms of riboflavin (Abdou *et al.*, 2013).

Cystatin is the third proteinase inhibitor in egg white (also called ficin-papain inhibitor). In contrast to ovomucin, cystatin is a small molecule (12.7 kDa) and it has no carbohydrates and a high thermal stability. Ovoglycoprotein is an acidic glycoprotein with a molecular weight of 24.4 kDa. This protein contains hexoses 13.6%, glucosamine 13.8%, and N-acetylneuraminic acid 3%. The biological functions of ovoglycoprotein are still unclear (Abdou *et al.*, 2013). Other constituents present in albumin include 0.03% lipids. Carbohydrates (approx. 1%) are partly

bound to protein (approx. 0.5%) and partly free (0.4–0.5%). Free carbohydrates include glucose (98%) and mannose, galactose, arabinose, xylose, ribose and deoxyribose, totalling 0.2–2.0 mg/100g egg white (Beltis *et al.*, 2009). The nutrient content of egg white is listed in Table 1.2.

Nutrient	Egg white (%)	Egg yolk (%)				
Protein	9.7 – 10.6	15.7 – 16.6				
Lipid	0.03	31.8 - 35.5				
Carbohydrate	0.4 - 0.9	0.2 - 1.0				
Water	84.3 - 88.8	48				
Minerals						
Sulphur	0.195	0.016				
Phosphorus	0.015-0.03	0.543-0.980				
Sodium	0.161-0.169	0.026-0.086				
Potassium	0.145-0.167	0.112-0.360				
Magnesium	0.009	0.016				
Calcium	0.008-0.02	0.121-0.262				
Iron	0.0001-0.0002	0.0053-0.011				

Table 1.2: Nutrient composition of avian eggs (adapted from Belitz et al., 2009).

Microorganisms require certain basic nutrients for growth and maintenance of metabolic functions. Foodborne microorganisms can derive energy from carbohydrates, alcohols, minerals and amino acids. An example of a pathogen with specific nutrient requirements is *S*E. The growth of *S*E may be limited by the availability of iron (Clay and Board, 1991). The addition of iron to an inoculum of *S*E in egg albumen resulted in the growth of the pathogen to higher levels compared to levels reached when a control inoculum (without iron) (Clay and Board, 1991).

1.5.2 Carbohydrates in egg white

The amounts of saccharides in eggs (in dry matter) is about 10g/kg, of which ~ 9g/kg are present in egg white and 1g/kg in the yolk. Protein bound saccharide in the form of glycoprotein occurs in the egg white at a level of about 0.2g/kg in form; galactose, mannose, glucosamine, galactosamine and lactaminic acid predominately (Velišek, 2014). The rest are free sugars, mainly monosaccharides, with some free oligosaccharides and polysaccharides. About 98% of free monosaccharides are glucose, while mannose, galactose, arabinose, xylose, ribose, and 2 deoxyribose (2-deoxy-d-erthropentose) are present of at concentration of 2-20 mg/kg (Baron *et al.*, 2016).

1.6 Sugar acids (hexonates) and Salmonella

Hexonates are straight-chain, six carbon, carbohydrate acid anion molecules. They carry a terminal carboxyl group and five hydroxylated carbons. Oxidation of the terminal aldehyde of sugars yields an aldonic acid. Eight isomers are recognised in the ChEBi data base on the basis of the orientations ('up' or 'down') of their hydroxyl groups. Each of these eight forms (altronates, fuconates, galactonates, gluconates, gulonates, mannonates and rhamnonates) exist as two alternative enantiomeric types (D and L), named according to the glyceraldehyde based designation system (Robyt, 1998). Figure 1.6 illustrates structures of four representative hexonates. Hexonates can serve as the sole sources of carbon and energy, and they commonly occur in foods as free substances and components. Frequently these acids are biologically derived from monosaccharides by oxidation of aldehyde groups (Velišek, 2014). However, there is no report on their presence in egg.

Hexuronates or hexuronic acid is a carbohydrate acid formally derived by oxidation of the hydroxyl group on carbon-6 of any aldose or ketose to a carboxylic acid. There are known forms of the hexuronates like fructuronic acids, galacturonic acids, glucuronic acids, guluronic acids, iduronic acids, mannuronic acids, and tagaturonic acids (Dictionary of Food and Nutrition, 2005; https://www.ebi.ac.uk).



Figure 1.6: Hexonate/hexuronate structures. *Modified from source:* <u>http://www.ebi.ac.uk/chebi/chebi-Ontology</u>. Bonds are shown using the wedge-dash notation.

D-Galactonate can serve as the sole source of carbon and energy for *Escherichia coli* (Deacon and Cooper, 1977). The initial step in the degradation of D-galactonate is dehydration to 2-dehydro-3-deoxy-D-galactonate by D-galactonate dehydratase. Subsequent phosphorylation by 2-dehydro-3-deoxygalactonate kinase and aldol cleavage by 2-oxo-3-deoxygalactonate 6-phosphate aldolase produces pyruvate and D-glyceraldehyde-3-phosphate, which enter central metabolism (Fig 1.7) (Szumiło, 1981; Latendresse *et al.*, 2012)



Figure 1.7: D-Galactonate degradation. Source: Robyt (1998).

L-Idonate can also serve as carbon and energy source in *E. coli*. It is catabolized via a pathway in which D-gluconate is an intermediate. L-Idonate is converted to D-gluconate by two consecutive oxidation and reduction reactions. D-Gluconate is then phosphorylated, forming 6-phosphogluconate which is an intermediate of central carbon metabolism. 6-Phosphogluconate is metabolized via the pentose phosphate or Enter-Doudoroff pathway (Bausch *et al.*, 1998) as shown in figure 1.8 (Lamble *et al.*, 2004).



Figure 1.8: L-Idonate degradation. Source: Bausch et al. (1998).

Little work has been performed on the role of hexonate utilisation in survival and colonisation of *S*E. Coward *et al.* (2012) investigated the role of a hexonate uptake and catabolism for the *S*E genomic island locus, SEN1432–SEN1436, (encoding two suspected dehydrogenases enzymes and one dehydratase enzymes; see chapter 3 for further detail) during colonization of the chicken reproductive tract and other organs following oral challenge. The deletion of these loci resulted in a decrease in bacterial load in the spleen by 14 days post infection suggesting a minor role in systemic colonization. Another study showed that several genes involved in the transport and metabolism of D-galactonate (*dgo*), D-gluconate (*gntU*, *kdgT* and *kduD*) and L-idonate (*idn*) were upregulated (2.5-3.5 fold) in *S*E which was considered indicative of its metabolism in macerated leaf tissue in cilantro and lettuce soft rot lesions (Goudeau *et al.*, 2013). Comparison of the *S*. Entertitidis PT4 and *S*. Typhimurium LT2 genomes (Thomson *et al.*, 2008) showed a PT4 specific

region ('ROD13') corresponding to the SEN1432–SEN1436 (6 kb) locus encoding one of the three hexonate-utilisation loci induced by egg white. Although absent in the LT2 strain, this locus is present in the chicken pathogen, *S*. Gallinarum. The reason for the absence of this locus in LT2 is unclear. However, the SEN1432–36 genes show sequence similarity as well synteny to the genes of the *gntII* locus of *E. coli*; these are involved in L-idonate catabolism (Bausch *et al.* 1998) suggesting a similar function for the SEN1432-36 genes.

1.7 Iron and Salmonella

One of the key obstacles to survival in both the host and non-host environment (including egg white, as highlighted above) is the lack of essential nutrients, such as iron (Ratledge and Dover, 2000). The absence of free iron makes the egg white quite inhospitable for bacterial growth (Baron et al., 2016). Iron is an essential element required for the growth of all animals, plants and most microorganisms. It plays vital roles in many important biological processes such as DNA synthesis, gene regulation and amino-acid and pyrimidine biosynthesis (Andrews, 1998). Systems for its acquisition, storage and utilisation exist in nearly all forms of life, and its absence can be causative lack of growth or loss of pathogenicity in micro-organisms (Ratledge and Dover, 2000). Iron is soluble under anaerobic conditions at physiological pH where it persists in its reduced form, favouring bacterial iron acquisition. However, under oxygenic conditions at neutral or higher pH, iron in the form of Fe³⁺ forms insoluble hydroxides, making the metal less accessible (Andrews, 1998). To obtain iron, the bacterial pathogen secretes siderophores, which chelate Fe³⁺ with high affinity and specificity, even when bound to host proteins transferrin and lactoferrin (Miethke and Marahiel, 2007). In Gram-negative bacteria, the response to iron concentration is regulated by the ferric uptake regulator (Fur) which was initially identified in E. coli (Schaffer et al., 1985; Escolar et al., 1999).

Bacterial iron acquisition is essential for *Salmonella* spp. survival and growth within its host (Andrews *et al.*, 2003). According to Kang *et al.* (2006), increasing iron concentration enhanced *SE* survival in albumen, indicating that iron limiting conditions may contribute to the bacteriostatic activity of egg albumen.

Salmonella has iron-acquisition systems for both ferric and ferrous iron which are expressed in response to iron restriction. These include two types of siderophores, enterobactin and its glucosylated derivative salmochelin. It uses enterobactin and its stable breakdown products, the linear trimeric, dimeric, and monomeric forms of 2,3-dihydroxybenzoylserine (DHBS3, DHBS2, and DHBS1, respectively). In addition, salmochelin S4 (two carbohydrate moieties added to enterobactin) and its degradation products (linear trimer salmochelin S2, the dimer salmochelin S1, and the monomers salmochelin SX) are also used (figure 1.9) (Chart *et al.*, 1993; Crouch *et al.*, 2008).



Figure 1.9: Siderophores of *Salmonella* Typhimurium. B. Chemical structures and ions (*m/z*) of enterobactin, salmochelin and degradation products. (Crouch *et al.*, 2008).

Salmochelin was the first glucosylated siderophore described. It is a C-glucosylated enterobactin produced by *Salmonella* species, uropathogenic and avian pathogenic *Escherichia coli* strains, and certain *Klebsiella* strains (Hantke *et al.*, 2003). The bacteria recover the ferri-siderophore complex through specific receptors on the outer membrane (Sood *et al.*, 2005). Some bacteria (e.g. *Klebsiella pneumoniae*) secrete modified microcins with glucosyl-enterobactin like moiety. Such microcins are taken up across the outer membrane by the same catecholate siderophore receptors (IroN, Cir, Fiu, and FepA) used for salmochelin/enterobactin uptake (Muller *et al.*, 2009). The *entABCDEFHS* gene cluster is responsible for biosynthesis and export enterobactin (figure

1.10). The production of salmochelin is dependent on the synthesis of enterobactin and the *iroBCDEN* gene cluster (Crouch *et al.*, 2008).



Figure 1.10: **Siderophores of** *Salmonella* **Typhimurium.** Genetic organization of the enterobactin (*ent*) and salmochelin (*iroA*) synthetic (black arrows), export (white arrows), and utilization loci (grey arrows) (Crouch *et al.*, 2008).

The high similarity of the siderophore systems of *E. coli* and *Salmonella* suggests that their uptake systems behave similarly, so in both cases ferric-enterobaction is transported mainly through FepA across the outer membrane (Figure 1.11) (Hantke *et al.*, 2003). The ferric-enterobactin complexes are then transported through the cytoplasmic membrane via the ABC transporter consisting of the binding periplasmic protein FepB, the membrane components FepD and FepG, and the ATPase FepC. Inside the cell, the Fes protein is required for iron release from the ferric enterochelin complex (Hantke *et al.*, 2003).



Figure 1.11: Scheme of the catecholate siderophore transport systems of *S. enterica*. DHBS, 2,3dihydroxybenzoylserine; Glc, glucose. The iro gene cluster encodes IroB, a C-glucosyltransferase; IroC, an ABC transporter; IroD and IroE, two esterases; and IroN, the outer membrane receptor for salmochelin (Hantke *et al.*, 2003).

The activity of most of these specialized transport systems requires the function of the bacterial outer membrane protein TonB (Zhou *et al.*, 1999) utilising a mechanism that is common among enterobacteriaceae (Andrews *et al.*, 2003). The TonB–ExbB–ExbD complex is required for the energy-dependent transport of ferric siderophores across the outer membrane of Gram-negative bacteria (Postle and Kadner, 2003).

Salmonella grown under iron-limiting conditions have the capability to increase the concentration of several iron-regulated outer-membrane proteins (IROMP) to augment the acquisition of the metal (Zarate-Bonilla *et al.*, 2012). *S. enterica* serovar Typhimurium expresses three outer membrane proteins of approximately 83, 78, and 74 kDa under conditions of iron starvation (Chart *et al.*, 1993; Rabsch *et al.*, 2003). FepA is the largest (IROMPs) and was identified over 30 years ago as a receptor for ferri-enterobactin. Later, the 78-kDa IROMP, IroN was identified to be an alternative ferri-enterobactin receptor. The *iroN* gene is present in all phylogenetic groups of *S*E (Rabsch *et al.*, 2003). Another type of system identified in *Salmonella* is encoded by the *feoABC*

locus and mediates the transport of iron (II) through the inner membrane (Kammler *et al.*, 1993). This anaerobically induced system does not require siderophores, as iron (II) is soluble and therefore readily enters the periplasmic space by diffusion through the porins (Zhou *et al.*, 1999).

1.8 Egg white antimicrobial activity

Avian albumen is a complex multifunctional medium promoting the growth and development of the embryo. It provides water and nutrients to the developing embryo. Eggs have efficient protective barriers preventing contamination if laid in hygienic conditions (Van Dijk *et al.*, 2008). There are various protection barriers working together to protect the embryo. These are divided into physical protection by the egg shell and chemical protection by antibodies, known as IgYs, mainly concentrated in the egg yolk and other proteins throughout the egg in the form of numerous peptides and proteins possessing antimicrobial properties (Bedrani *et al.*, 2013). With regard to egg antimicrobial proteins and peptides, they operate via three main mechanisms. Firstly, sequestration of essential nutrients from bacteria by chelating minerals (e.g. iron) using proteins like ovotransferrin, and vitamins (e.g. biotin) using proteins such as avidin. Secondly, inactivation of exogenous proteases using inhibitors such as cystatin, ovomucoid and ovoinhibitor; such proteases are necessary for microbial metabolism and invasion of host tissues. The third way is direct lytic action on microorganisms by proteins such as lysozyme which leads to the disruption of the bacterial cell wall (Gantois *et al.*, 2009).

Raw hen egg white inhibits the growth of bacteria. *Staphylococcus aureus, Shigella dysenteriae, Escherichia coli, Listeria monocytogenes, Campylobacter jejuni* and *Saccharomyces cerevisiae* (Sahin *et al.*, 2003; Wellman-Labadie *et al.*, 2009). In addition to protein factors involved in egg white immunity, there are physicochemical factors that affect the growth of bacteria such as pH, viscous structure and temperature (Baron *et al.*, 2011). The pH value of egg white rapidly

increases from 7.6 up to 9.3 a few days after laying due to loss carbon dioxide through the pores of the egg shell. The alkaline pH of egg white (8.1-9.7) reduces the growth of microorganisms as it is higher than the growth range of many bacterial species, including *Salmonella* spp. (Wellman-Labadie *et al.*, 2009). An alkaline pH mainly affects the respiratory status of bacteria leading it to suppress the systems that consume high energy, such as flagella biosynthesis (Maurer et al., 2005). Thus, Baron (1998) has shown that *SE* lacks flagella at alkaline pH, which may limit its access to nutrients. Moreover, the activity of egg-white proteins is affected by alkaline pH; e.g. ovotransferrin is more effective at higher than at lower pH, due to an acceleration of iron release under acidic conditions and a slowdown under alkaline conditions (Halbrooks *et al.*, 2005).

The viscosity of fresh egg white is around 30-fold higher than that of water (Lucisano *et al.*, 1996). The viscosity of thick egg white is 40-fold higher than thin egg white and these regions remain distinctly separate inside the shell egg for at least a few days after laying (Lang and Rha, 1982). This high viscosity may induce heterogeneity and makes motility of bacteria in egg white difficult, limiting the spread and access to nutrients required for bacterial growth. Moreover, another source of heterogeneity would be that the ferri-ovotransferrin complexes are probably not distributed uniformly within egg white (Li-Chan and Nakai, 1989; Baron *et al.*, 2017).

According to previous studies (Ruzckova, 1994; Chen *et al.*, 2005), there is a significant effect of temperature on the survival in the egg white. There is a bactericidal effect of low temperatures (below 10 °C). Reasonable growth is observed (1-4 \log_{10} CFU/ml) between 20 and 30 °C. However, bacteriostatic or bactericidal effects are observed at temperatures above 37 °C._ A bactericidal effect of egg white is reported in all cases at 42 °C. The destruction ranges from less than 2 \log_{10} CFU/ml to 3.5 \log_{10} CFU/ml for incubation times between 24 and 96 hour (Kang *et al.*, 2006; Guan *et al.*, 2006). Investigating the effect of temperature from 37-48 °C on the

survival of SE demonstrated that egg white is bactericidal at temperatures higher than 42 $^{\circ}$ C (Alabdeh *et al.*, 2011).

Other factors also contribute to antimicrobial mechanisms of egg albumen to control *SE*, including nuclease activity and the concentration of bacteria (Lu *et al.*, 2003; Kang *et al.*, 2006). Lu *et al.* (2003) identified endonuclease and exonuclease activities of egg white leading to the damage of DNA as a new bactericidal mechanism. Using *in vitro* assays it was suggested that egg albumen degrades DNA by converting supercoiled plasmid DNA to nicked and linear DNA. Moreover, intracellular plasmid DNA showed increased nicking after exposure to egg albumen which suggests the same effects on bacterial genomic DNA. However, this activity was affected by temperature in that it appeared lower at 4 and 25 °C, and higher at 37 and 42 °C; this may explain the negative effect of high temperatures (37 and 42 °C) upon bacterial survival in egg albumen (Gast and Holt, 2000). In addition, other possible enzymatic antimicrobial activities are more active at higher temperatures and thus lead to more antibacterial activity of egg white.

Kang *et al.* (2006) indicated that the initial bacterial concentration affects the bactericidal activity of egg albumen; egg white had no ability to control *S*E when bacterial concentration was higher than $\sim 10^3$ CFU/ml. Three possibilities could explain this observation: high concentrations of bacteria may saturate the antimicrobial factors; insufficient local concentration of antimicrobial factors; and killed bacteria might be releasing their contents supporting the survival of remaining bacteria.

In the yolk, the situation is different because the bacteria gain access to an environment that is rich in nutrients, and lacks inhibiting conditions and/or compounds such as lysozyme, iron-binding ovotransferrin, and an alkaline pH (Cogan *et al.*, 2004). A number of compounds such as vitamins, amino acids and fatty acids are present that may stimulate bacterial growth by the activation of alternative metabolic pathways and in this way contribute to high cell density and thus enhanced egg contamination (Morales *et al.*, 2005).

Despite all the previous factors, it is thought that the most important protection against bacterial survival is achieved by the iron restricting influence of ovotransferrin (Ahlborn and Sheldon 2005). Moreover, Baron *et al.* (2016) summarized the mains defences in egg white as iron deficiency through iron chelation by ovotransferrin and disruption of bacterial membranes through particular components such lysozyme, ovotransferrin or other antimicrobial molecules interacting with the bacterial envelope and forming pores in the bacterial cell wall (Clavijo *et al.*, 2006 and Kang *et al.*, 2006). However, the various findings highlighted above suggest that all the antibacterial activity of egg white work together to prevent contamination.

1.9 SE survival in egg white

Usually, *SE* is the only *Salmonella* serotype responsible for human infection from intact eggs (Kang *et al.*, 2006). Keller *et al.* (1997) reported that only *SE* survived in eggs after laying but the frequencies of Typhimurium serovar were higher than the Enteritidis serovar in eggs recovered from reproductive tracts before they are laid. This means that forming eggs can eliminate most of the contaminating bacteria and that *SE* has enhanced survival ability in eggs (Killer *et al.*, 1997). It is suggested that for *SE* to contaminate eggs, a specific interaction with the oviduct tissue occurs which leads to persistent oviduct colonization (Gantois *et al.*, 2008).

There are two possible routes of egg contamination by *Salmonella* known. The horizontal route involves penetration through the egg shell. While the vertical route involves direct contamination of the egg content before oviposition, as a result of *Salmonella* infection of the reproductive organs such as oviduct and the ovary (Keller *et al.*, 1995). *Salmonella* has been found on the mucosal surface and within epithelial cells, lining the oviduct in naturally and experimentally

infected hens (Gantois *et al.*, 2008). It also has the capability to cross from egg white to egg yolk through the vitelline membrane (Gast and Holt, 2000).

It is believed that the start point of the infection pathway relies on some virulence factors such as type 1 fimbriae and capsular-like lipopolysaccharide (Paker *et al.*, 2002). Evidence suggests that LPS is a significant factor in *S*E colonization of the gastrointestinal tract in the chick (Carroll *et al.*, 2004). *S*E has a specific ability to contaminate eggs and survive/grow in egg albumen at chicken body temperature (42 °C) (Hermans *et al.*, 2011). A study with 89 *Salmonella* strains from different serotypes incubated for 24 h in egg-white at 42 °C showed that the number of *S*E strains able to survive in egg white is significantly higher compared with strains belonging to other serotypes (Vylder *et al.*, 2013). Therefore, for most studies on the antimicrobial activity of egg white, *S*E is used as a model bacterium as it represents the predominant (90%) serotype responsible for foodborne diseases (salmonellosis) resulting from egg or egg product consumption (EFSA 2009).

1.9.1 Genetic response of SE to egg white exposure

Most studies aimed at investigating the ability to survive in egg white are carried out with SE. The high occurrence of this serovar in foodborne diseases can be explained by the enhanced ability of this serovar, over other *Salmonella*, to survive in egg white (Clavijo *et al.*, 2006; Gantois *et al.*, 2008; De Vylder *et al.*, 2013). Studies on SE survival in egg white have mainly focussed on the identification of specific genes that could endow SE with resistance during incubation in egg white. The main approaches used to identify such genes are directed mutagenesis (Lu *et al.*, 2003; Cogan *et al.*, 2004; Kang *et al.*, 2006), insertional mutagenesis (Clavijo *et al.*, 2006), *in vivo* expression technology (IVET) (Gantois *et al.*, 2008), and a microarray-based transposon library screening (Raspoet *et al.*, 2014a). The genes identified or suspected to be involved are

implicated in membrane structure and function, metabolism of nucleic acids and amino acids, motility, synthesis and DNA repair, invasion and pathogenicity. These results provide different explanations for the response of *S*E to the antimicrobial effects of egg white.

Raspoet et al. (2014b) showed that 16 genes from 23 induced (e.g rfbABCDFIKMNPU) in SE by hen body temperature (42 °C for 24 hour) are involved in lipopolysaccharide biosynthesis. In addition, they showed that an *rfaI* (encoding the enzyme that catalyzes an early step in lipopolysaccharide biosynthesis) mutant was unable to survive in egg white at 42 °C. Egg white can also act on the cytoplasmic membrane. During colonization of the oviduct and contamination of the forming eggs by SE, the induction of bacterial genes involved in membrane stress (uspBA) and in the monitoring of the status of the cytoplasmic membrane (hfl K) was shown (Gantois et al., 2008). These genes (uspBA) were also induced after contact with egg white (Raspoet et al., 2011). In addition, the induction of a gene (*murA*) involved in the synthesis of peptidoglycan was observed in SE during hen oviduct colonization and in contaminated eggs, suggesting a response to the permeabilization of the peptidoglycan by lysozyme (Gantois *et al.*, 2008). The motility of bacteria is also disturbed in egg white. Gantois et al. (2008) showed induction of flgG, that encodes a component of flagella, during the colonization of the oviduct and in the contaminated laid eggs. In addition, this study showed that survival of mutants lacking flagella is reduced at the 42 °C in egg white.

Another study focussing on bacterial factors needed to survive within eggs used a genomic DNA library to show that YafD and XthA (exonuclease III) provide a survival advantage to *SE* in eggs by repairing DNA damage caused by egg albumen (Lu *et al.*, 2003). Moreover, in a transposon mutant library approach, genes involved in amino acid and nucleic acid metabolism, and cell wall integrity were indicated as important for *SE* to survive in egg albumen (Clavijo *et al.*, 2006).

Raspoet *et al.* (2011) used *in vivo* expression technology (IVET) to identify *Salmonella* genes involved in the interaction with the oviduct or eggs; such genes included those involved in cell wall integrity, regulation of fimbrial operons, stress responses and motility; these were identified as highly expressed in the oviduct tissue. This expression screening method identified two universal stress protein genes, UspA (a cytoplasmic autophosphorylating serine/threonine phosphoprotein) and UspB (an anchored cytoplasmic membrane protein) as being highly expressed in the oviduct tissue and in eggs. They demonstrated that expression of these is induced after contact with egg white. Intra-oviduct inoculation of SE *uspB* and *uspBA* mutant strains showed that the mutants had a decreased ability to colonize the magnum and isthmus of the oviduct; they hypothesized *uspA* and *uspB* are involved in long term persistence of *SE* in harmful environments, such as in the oviduct and eggs, by conferring resistance against compounds that damage the bacterial cell membrane and DNA (Raspoet *et al.*, 2011).

1.9.2 Induction of genes encoding hexonate/ hexuronate catabolism systems by exposure of SE to egg-white medium.

Characteristics that have been mentioned previously mark-out the SE serotype as the most relevant model for studying the response of bacteria to the antimicrobial activities of egg white. This study is complementary to studies conducted at the Agrocampus Rennes-INRA, France (Egg & Egg Product Microbiology team) (Baron *et al.*, 2017). The aim of their study is further understanding SE behaviour towards bactericidal mechanisms of the egg at temperatures \geq 42 °C. To advance this aim, the global transcriptional response of SE was previously determined, using microarray technology, upon exposure to egg-white medium (egg-white filtrate with 10% egg-white protein; EWMM) under bactericidal conditions (45 °C, pH 9.3 - i.e. the pH of egg white several days after laying) over a 45 min time period. Results showed global expression changes of SE in response to

exposure to egg-white medium for 7, 25 and 45 min at 45 °C. This medium was used to avoid the difficulty of RNA extraction from authentic egg white to enable analysis of the global transcriptional response. Previous work has shown that this model medium is an accurate mimic of authentic egg white (Baron *et al.*, 2017). At each incubation time, expression was compared to that of the inoculum just prior to its exposure to egg-white medium.

Genes with a statistically significant \geq 2-fold change in expression were considered as differentially regulated. Thus, at 7, 25 and 45 min, 13.4% (288 induced and 277 repressed), 15.3% (304 induced and 362 repressed) and 18.7% (318 induced and 468 repressed) of genes were differentially regulated. The greatest expression effects were seen at 45 min. Over-represented categories at 45 min include: signal transduction (25.7% of genes in this category were affected; of which 77.8% were down regulated), energy metabolism (24.4%; 80.3% down regulated), motility (23.8%; 96.3% down regulated), metabolism and transport of amino acids (18.4%; 74.4% down regulated), metabolism and transport of nucleotides (27.2%; 83.3% up regulated genes), metabolism and transport of coenzymes (26.2%; with 71.1% up regulated), catabolism of secondary metabolites (22.2%; with 72.2% up regulated), inorganic ion transport (26.6%; 50% down regulated) and post-transcriptional modification (19%; 50% down regulated). Many of the genes affected by egg white exposure have functions that had already been reported to be related to egg-white survival. These are summarised below:

- 1. **Induction of biotin biosynthesis genes**, the *bioABCDF* operon, likely to be due to poor biotin availability in egg white resulting from the presence of avidin, a powerful biotin-chelating protein (Beckett, 2007).
- 2. A major iron-restriction response, induction of iron-uptake genes and repression of ironrationing genes (e.g. strong expression of the *entABCDEFHS* gene cluster which encodes the

proteins involved in biosynthesis and export of enterobactin). This reflects the low iron availability of egg white due to the high levels of iron chelating ovotransferrin.

- 3. **Down regulation of energy metabolism genes** (e.g. *napFDA*, *dmsABC*, *frdAB*, *fdoIGH*, *sdhCDAB*, *cyoABCDE* and *nuoABCDEFGHIJKLM*) which is consistent with a homeostatic adjustment of SE metabolism in response to the high pH, membrane-disruption activity and an attempt to overcome iron deficiency of egg-white medium by iron rationing.
- 4. **Induction of the Kdp potassium uptake system**, the genes specifying the high-affinity K⁺ uptake system (*kdpABCD*), which would be consistent with an alteration of turgor pressure providing the signal for the *kdp* induction in egg white medium.
- 5. **Down-regulation of amino acid biosynthesis and uptake**, genes involved in the synthesis and transport of amino acids were generally repressed (e.g. *lysC* which encodes aspartate kinase); this down-regulation suggests a reduced requirement for amino acids which probably results from the non-permissive growth conditions provided by the egg-white medium. Furthermore, the high levels of amino acids (Belitz*et al.*, 2009) found within egg white might repress expression of genes required for amino acid biosynthesis.
- 6. **Repression of motility and chemotaxis** The class II FlhC₂FlhD₂.regulated genes, that encode the flagella basal body export machinery, were down-regulated, as the class III genes (*motAB, cheAW, cheRBYZ, cheM,* SEN30590, *tcp, tsr,* and *fliB*), encoding chemotaxis proteins and structural subunits of the flagellum. The reduced expression might, at least partly, explain the inability of *SE* to propagate in egg-white medium under the conditions employed. In addition, reduced-motility could represent an energy-conserving response to the growth-inhibitory conditions presented by egg white (Zhao *et al.*, 2007)

- Repressions of a subset of virulence genes; six genes (*invH*, *invAE* and *prgIHK*) within Salmonella Pathogenicity Island 1 (SPI1) were down regulated.
- 8. Induction of a heat-shock response; the up-regulation of heat-shock proteins genes (*groEL*, *groES*, *grpE*, SEN1800 and *htpG*) and down-regulation of two cold-shock proteins; this correlates well with the temperature upshift experienced by *S*E upon transfer of the inoculum (37 °C) to the egg-white medium (45 °C).
- 9. **Induction of an envelope-stress response,** suggestive of membrane damage induced by egg white exposure (e.g. the *spy* gene which encodes a periplasmic chaperone protein).

10. Induction of hexonate/hexuronate utilization genes.

An unexpected finding of the egg-white exposure data is the high degree of induction for three distinct gene clusters involved in hexonate/hexuronate utilization: the *dgoRKADT* operon (13.59-to 31.13-fold); the *uxuAB-uxaC* operon (10.68- to 28.2-fold); and the SEN1433-6 genes (5.17- to 33.38-fold) (detailed in chapter 3). The surprise was there is no evidence indicating the presence of hexonates/hexuronates in egg white. In addition, no role for these genes in survival of *SE* in egg had been previously suggested, indicating that they may comprise a novel regulon. Thus, the reason behind the strong induction of these genes and whether they have any impact on survival in egg white is unclear and so demands investigation.

1.9.2.1 The dgo genes

The dgo genes (dgoRKADT) were 23.6 - to 31.1-fold induced (at 45 min; Table 1.3) by EW exposure. It is believed that the general function of these genes in *E. coli* is utilization of D-galactonate and 2-keto-3- deoxygalactonate (Neidhardt, 2005). The dgoT encoded permease transports D-galactonate, which is then converted to 2-deoxy-3-keto-D-galactonate by a dehydrase encoded by dgoD. After this, the glyceraldehyde 3-phosphate and pyruvate are produced by a

kinase reaction specified by *dgoK*, and the phosphorylated intermediate is then cleaved by an aldolase specified by *dgoA*. So the latter three genes *DKA* are suggested to code for enzymes involved in the conversion of D-galactonate to pyruvate and glycerldehyde-3-phosphate (Cooper, 1978; Neidhardt, 2005; Zhou and Rudd, 2013). *dgoD* and *dgoA* mutants of *E. coli* K-12 were unable to grow on D-galactonate (Cooper, 1978). The *dgoR* gene encodes a GntR/FadR-related regulator which likely acts as a D-galactonate-responsive transcriptional repressor of the *dgo* operon (Zhou & Rudd, 2013). These genes, including the regulatory *dgoR* gene, cluster at min 82.5 in *E. coli* (Neidhardt, 2005).

The schematic representation of the dgo operon (dgoRKADT) in SE PT4 is illustrated in figure (Fig. 1.12). Alignment showed that dgoT and dgoD of SE encode proteins are 100% identical to the *E. coli* equivalents, and that dgoA, dgoK and dgoR of SE encode proteins 85, 82 and 94% identical to their *E. coli* counterparts (Zhou and Rudd. 2013). A few studies showed the up regulation of genes related to hexonate metabolism in *S*. Typhimurium. A microarray experiment global expression effect caused by exposure to macerated lettuce leaf tissue showed the up regulation of several genes involved in the transport and metabolism of D-galactonate (dgo), D-gluconate (gntU, kdgT, and kduD), and L-idonate (idn) (Goudeau *et al.*, 2013). Another study using microarrays to study the effects of expression upon macrophage colonisation by *S*. Typhimurium showed up regulation of three genes of the dgo operon (dgoT, dgoK and dgoA) without any indication of the inducing signal, although it was suggested that hexonates may be an important source of carbon for intracellular bacteria (Eriksson *et al.*, 2003). Similarly, the expression data of Baron *et al.* (2017) suggest that hexonates/hexuronates may be utilised by *S*E in egg white.

Table 1.3: Egg-white induced genes from the *dgo* cluster involved in D-galactonate metabolisim in *Salmonella enterica* serovar Enteritidis (strain PT4-P125109). Strand indicted by S. Fold change indicted by FC. Expression data from Baron *et al.* (2017).

No	Gene	Aliases	Protein names	Entry name	Length	GC	S.	FC (X)
	names				a.a bp	%		(45 m.)
1	dgoT	SEN3643	D-galactonate transporter	B5QUN8_SALEP [3,903,869 <- 3,905,161]	430 1293	54.06	R	14
2	dgoD	SEN3644	D-galactonate dehydratase	DGOD_SALEP [3,905,291 <- 3,906,439] EC:4.2.1.6	382 1149	55.35	R	25
3	dgoA	SEN3645	2-dehydro-3- deoxy-6-phospho galactonate aldolase	B5QUP0_SALEP [3,906,436<- 3,907,053] EC:4.1.2.21	205 618	58.58	R	22
4	dgoK	SEN3646	2-dehydro-3-deoxy galactono kinase	B5QUP1_SALEP [3,907,037 <- 3,907,915] EC:2.7.1.58	292 879	56.31	R	31
5	dgoR	SEN3647	Galactonate operon transcriptional repressor	B5QUP2_SALEP [3,907,912 <- 3,908,601]	229 690	53.33	R	27
6	yidA	SEN3648	Uncharacterized protein	B5QUP3_SALEP [3,908,862 <- 3,909,707]	281 846	52.60	R	
7	torS	SEN3642	Two-component sensor protein	B5QUN7_SALEP [3,901,108 -> 3,903,843]	911 2736	54.24	F	



Figure 1.12: Schematic representation of the *dgo* cluster of SE PT4. The corresponding nucleotide sequence was analysed and annotated using Vector NTI. Genes that are related to hexonate/hexuronate (Hex) metabolism/control are in green (those in red are flanking genes, unrelated to Hex metabolism), direction is indicative of polarity. Sizes of open-reading frames are given in amino acids codons and bp, along with the assigned functional annotation and coordinates (Fr-To). The position in the genome is indicated in kb (see Appendix 1 for more detail).

1.9.2.2 The uxu-uxa genes

The second cluster, the uxuAB-uxaC operon, is believed to be involved in mannonate utilisation (Suvorova *et al.*, 2011). These genes were induced by egg white at levels (10.7- to 28.2-fold, at 45 min, Table 1.4) similar to those observed for the *dgo* genes. uxuA (SEN2978) encodes mannonate dehydratase that catalyzes the formation of 2-dehydro-3-deoxy-D-gluconate from mannonate. uxuB (SEN2979) encodes D-mannonate oxidoreductase while the third gene, uxaC (SEN2980), encodes a glucuronate (hexuronate) isomerase and its function is to catalyse the interconversion of D-glucuronate to D-fructuronate or D-galacturonate to D-tagaturonate (Suvorova *et al.*, 2011). In a study on the carbon nutrition of *E. coli* in the mouse intestine, the hexuronate pathway was knocked out through uxaB mutation which resulted in no effect on colonizing ability (Chang *et al.*, 2004). A schematic representation of the uxuAB-uxaC operon is shown in figure 1.13.

Table 1.4: Egg-white induced genes from the *uxuAB-uxaC* operon involved in mannonate utilisation in of *S*. Enteritidis (strain PT4-P125109). Expression data from Baron *et al.* (2017). For further details, see Table 3.1.

No	Gene	Aliases	Protein names	Entry name	Length	GC%	S.	FC (X)
	names			Position	aa-bp			(45 m.)
1	ихиА	SEN2978	Mannonate	B5QYB0/UXUA_	394 -	52.49	F	28
			dehydratase	SALEP	1185			
				[3,184,087 ->				
				3,185,271]				
				EC:4.2.1.8		_		
2	ихиВ	SEN2979	Mannonate oxido	B5QYB1_SALEP	490 -	56.62	F	19
			reductase	[3,185,382 ->	1473			
			fructuronate	3,186,854]				
			reductase	EC:1.1.1.57				
3	uxaC	SEN2980	Glucuronate	B5QYB2/UXAC_	470 -	54.00	F	11
			isomerase	SALEP	1413			
			=Uronate isomerase	[3,186,866 ->				
				3,188,278]				
				EC:5.3.1.12		_		
			Flan	king genes				
4		SEN2977	Hexuronate	B5QYA9_SALEP	434 -	52.49	R	
			transporter	[3,182,378 <-	1305			
				3,183,682]				
5		SEN2981	Uncharacteri-zed	B5QYB3_SALEP	351 -	42.80	R	
			protein	[3,188,578 <-	1056			
				3,189,633]				



Figure 1.13: Schematic representation of the *uxuAB-uxaC* **operon of** *SE* **PT4.** The corresponding nucleotide sequence was analysed and annotated using Vector NTI. *uxuAB-uxaC* genes are shown as green arrows. For further details, see Fig. 3.1.

E. coli is capable of utilizing all forms of sugar acids (hexonates, hexuronates and hexuronides) as sources of carbon and energy including hexuronate like D-glucuronate and D-galacturonate via the Ashwell catabolic pathway (Robert *et al.*, 1974; Suvorova *et al.*, 2011).

1.9.2.3 The SEN1433-6 genes

The SEN1433-5 genes form a putative operon adjacent to the functionally related and divergent SEN1436 gene (Table 1.5). They are induced by 5.17-to 33.4-fold, similar to *dgo* and *uxuAB-uxaC*. Schematic representation of this cluster is shown in Fig. 1.14. The genes of the SEN1432-6 cluster specify three enzymes (two suspected dehydrogenases and one dehydratase), likely to be involved in hexonate utilization, and a proposed hexonate transporter (Thomson *et al.*, 2008). This is confirmed using comparative genomic hybridization analysis by Betancora *et al.* (2012). A study was conducted by Coward (2012) involving the deletion of specific genomic islands, including that containing SEN1432-36, to investigate their role in SE in colonization of the chicken reproductive tract and other organs. The results showed that all tested regions appear to play a small role in infection of liver and spleen, but not in colonization of chickens.

No	Aliases	Protein names	Entry name	Length	GC	S.	FC (X)
	-		Position	aa-bp	<u>%</u>		(45 m.)
1	SEN1433	L-idonate 5-	B5R538_SALEP	347	56.90	R	5
		dehydrogenase	[1,521,017 <-	1044			
			1.522.060]				
			EC:1.1.1.264				
2	SEN1434	putative hexonate	B5R539_SALEP	469	47.94	R	6
		sugar transport	[1,522,077 <-	1410			
		protein	1,523,486]				
3	SEN1435	putative hexonate	B5R540 SALEP	255	52.86	R	7
		dehvdrogenase OR	[1.523.522 <-	768			
		gluconate 5-	1.524.2891				
		dehvdrogenase	EC: 1.1.1.69				
4	SEN1436	putative dehvdratase	B5R541 SALEP	419	53.81	F	33
		I mana se ga antes a se a	[1.524.621 ->	1260			
			1,525,880]				
5	SEN1432	Putative GntR-family	B5R537_SALEP	239	48.47	R	
		regulatory protein	[1,520,207 <-	720			
			1,520,926]				
			Flanking gene				
6	SEN1437	Aminoglycoside	B5R542_SALEP	145	52.28	R	
		N(6')-acetyl	[1,526,021 <-	438			
		transferase type 1	1,526,458]				
		• •	-				

 Table 1.5: Egg-white induced genes from the SEN1432-6cluster involved in hexonate utilisation in of S. Enteritidis

 (strain PT4-P125109). Expression data from Baron *et al.* (2017). For further details, see Table 3.1.



Figure 1.14: Schematic representation of the SEN1432-6 cluster of SE PT4. SEN1432-6 genes are shown as green arrows, direction is indicative of polarity. See Fig. 3.1 for further detail.

Thomson *et al.* (2008) demonstrated that *SE* PT4 has a specific region (ROD13) encoding five proteins displaying sequence similarities and synteny with the *gntII* locus genes of *E. coli* which are associated with the uptake and catabolism of the hexonate sugar acid L-idonate . Another study showed that *E. coli* mutants that are unable to utilize hexonates (gluconate) are unable to colonize the mouse large intestine suggesting that hexonates represent an important source of nutrients at this site (Sweeney *et al.*, 1996). Moreover, as indicated above, a transcriptomics study showed upregulation of the genes involved in the transport of gluconate and related hexonates for *S*. Typhimurium in macrophages, suggesting that hexonates may also be an important source of carbon for intracellular bacteria (Eriksson *et al.*, 2003). Note that as SEN1432 was not reported to be induced by EW (Baron *et al.*, 2017), this suggests it is constitutive and might be involved in controlling genes related to hexonate catabolism as it specifies a predicted regulator.

1.9.2.4 *ybhC* gene

In addition to the three hexonate utilisation pathways, SEN0731 was also induced, up to 5.8 fold (Table 1.6). This gene encodes a putative exported pectin-esterase, predicted to mediate conversion of pectin into pectate (poly-1, 4- α -D-galacturonate) and so may also have a function related to hexonate metabolism. *hutI* is a flanking gene (Fig. 1.15) and codes for imidazolone-5-propionate hydrolase; its function is in the histidine catabolism process yielding glutamate and formamide (it is not induced in egg white medium).

Table 1.6: Egg-white induced putative pectin esterase gene of *S*. Enteritidis (strain PT4-P125109). Expression data from Baron *et al.* (2017). For further details, see Table 3.1.

No	Gene	Aliases	Entry name	Protein	Length	GC	S.	FC (X)
	names		Position	names	aa-bp	%		(45 m.)
1	ybhC	SEN0731	B5QX57_SALEP	Possible	427	57.48	R	5
			[809,468 <- 810,751]	pectin	1284			
			EC:3.1.1.11	esterase				
			Flanking	g gene				
2	hutI	SEN0732	B5QX58	Imidazolon	407	58.66	F	
			HUTI_SALEP	epropionas	1224			
			[810,989 -> 812,212]	e				
			EC:3.5.2.7					
r								
		hutt						
					yb	nc.		
8,090	D							8,125
-								

Figure 1.15: Schematic representation of the *ybhC* **gene of** *SE* **PT4**. The *ybhC* gene is shown as a green arrow, see Appendix 4 and Fig. 3.1 for more detail.

Apart from the hexonate-related genes, several other genes involved in sugar metabolism were also induced, but to a lesser degree (3- to 5.6-fold) than the hexonate gene clusters. These include genes involved in the non-oxidative branch of pentose and glucuronate interconversion.

Interestingly, despite the observed induction in genes involved in hexonate/hexuronate (Hex) utilization, these sugars are not known to be present in egg white (Velišek, 2014). Therefore, the identity of the inducer (and its source) responsible for *dgoRKADT*, *uxuAB-uxaC* and SEN1433-6 up-regulation is unclear, although evidently these genes are not subject to any substantial catabolite repression since induction is observed despite the high glucose levels in egg white.

1.10 Aims and objectives

The aim of this work is to determine the role of the hexonate/hexuronate utilisation genes, as described above, in the survival of *S*E upon exposure to egg white. A further aim is to determine whether these genes are subject to induction by a common regulatory pathway within egg white and if so to characterise the regulatory mechanism and identify the environmental inducing signal within egg white. Further understanding of mechanisms applied by pathogenic bacteria to counter the host protective method might be contributed and help in find ways to prevent the host from pathogenic survival.

Specific objectives are as follow:

- Confirm the induction of the hexonate/hexuronate (Hex) utilisation genes in egg white.
- Determine the factors governing expression of the Hex utilisation genes in egg white.
- Identify the regulator that controls the induction of the Hex utilisations genes in egg white
- Determine the ability of SE to utilise a range of Hex as substrates for growth.
- Investigate the purpose of the *hex* gene induction in egg white.

Chapter 2. Materials and Methods

2.1. Materials

2.1.1 Chemicals

All chemicals used were of analytical grade and were from Sigma, Fisher (Fermentas & Thermo Scientific), Oxoid, Bio-Rad, Fluka and Fermentas unless otherwise stated.

2.1.2 DNA marker

1 kb DNA ladder (Gene Ruler TM) from Thermo Scientific was used to estimate the size and quantity of DNA following gel electrophoresis, using UV-induced fluorescence in the presence of Gel Red TM Nucleic Acid Gel Stain from Biotium at a final concentration of 1X from 10,000X product (Figure 4.1).



Figure 2.1:1kb ladder (Thermo scientific Gene RulerTM**). DNA marker was used to estimate the size and quantity of DNA.** Source: http://2009.igem.org/wiki/images/3/3f/Generulers_1kb_marker_ Fermentas.jpg

2.1.3 Protein marker.

Protein molecular weight markers used in this study were: Unstained Protein Molecular PageRulerTM unstained molecular weight ladder (10-200 kDa) from Fermentas, and PageRulerTM pre-stained molecular weight ladder (10-170 kDa) from Fermentas. Markers were used to determine the size and quantity of protein following SDS-PAGE analysis.



Figure 2.2: Protein marker used to estimate the size and quantity of DNA. (A & B) Fermentas unstained protein molecular marker, Fermentas PageRulerTM unstained molecular weight ladder, (C) Fermentas PageRulerTM prestained molecular weight ladder.

2.1.4 Restriction and polymerization enzymes

Restriction endonuclease (*Bam*HI, *Eco*RI, *Xho*I, *Hin*dIII and *Nde*I) and Phusion® High-Fidelity DNA polymerase were provided from Thermo Scientific. Optimal conditions were used according to the manufacturer's instructions.

2.1.5 Bacterial media.

2.1.5.1 Luria-Bertani broth and agar

LB broth was used for routine bacterial work. One litre volumes were prepared with 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl dissolved in qH_2O which was autoclaved to ensure sterility before use (Sambrook *et al.*, 2001). Any antibiotics or other sterile additives were added after sterilization and cooling to 55 °C to protect heat labile additives. Heat-labile substances were filter sterilized through a 0.22 µm Millipore filter. To prepare one liter of LB-agar, 15 g of agar was added to one litre LB-broth. The medium was mixed and dispensed into appropriate aliquots and autoclaved. The agar was then cooled to 50 °C before adding any

antibiotics or other heat labile additives, and then poured (~30 ml) into sterile Petri dishes and left to solidify before use.

2.1.5.2. Super Optimal Broth (SOC)

The nutrient-rich medium was used in transformation. SOC medium was 2% (w/v) Bacto Tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCI, 10 mM MgCl₂, 20 mM glucose (Hanahan, 1983). SOC medium was prepared and autoclaved without MgCl₂ and glucose. Stocks of 2 M MgCl₂-6H₂O and 20% glucose (both sterile filtered) were used to make the medium 10 mM in MgCl₂ and 20 mM in glucose. The final pH was 6.8 to 7.0.

2.1.5.3. M9 minimal medium

Minimal medium was the medium used for the growth of *SE* in the presence of different substrates. Minimal medium contained 10 g M9 salts (Sigma) per litre with supplements added before use: 0.2 mM CaCl₂, 2 mM MgSO₄, 0.001% vitamin B1 and 0.4% glucose/glycerol. The M9 minimal solid medium contained 1.5 w/v agar in the M9 minimal medium.

2.1.5.4 Media sterilisation.

Bacterial medium components were prepared as described by Sambrook *et al.* (2001). All media and heat stable solutions were sterilised by autoclaving at 121 °C, 20 lb/in² for 20 min. Sterilisation of heat labile solutions was achieved by filtration through a sterile 0.22 μ m membrane (Whatman). Media were solidified with 1.5% w/v agar which was added before autoclaving. Glassware used in microbiological procedures was sterilised by dry heat (150 °C for 2 to 2.25 h). For all iron-restricted growth, acid-washed glassware was used.

Materials and Methods

2.1.6 Antibiotics

Antibiotics were prepared as described (Table 2.1) with those dissolved in water being filter sterilised through a sterile 0.22 μ m membrane (Whatman) and stored at -20 °C.

Antibiotic	Mode of action	Uses	Working Strength
Ampicillin: (100	Gram negative bacterial. Inhibits	Selection and maintenance of	100 µg/mL
mg/mL stock in	cell wall peptidoglycan synthesis	<i>E. coli</i> strain carrying the β -	
nano pure H ₂ O).	at the transpeptidation step	lactamase gene	
Chloramphenicol:	Bacteriostatic, inhibits 50S	Selection and maintenance of	35 µg/mL
(50 mg/mL in	ribosomal elongation	E. coli strains to carry the cat	
ethanol).		gene	
Kanamicin: (50	Interacts with a 30S subunit of	Selection and maintenance of	35 µg/mL
mg/mL stock in	bacterial ribosomes and inhibits	E. coli strain carrying the kan	
nano pure H ₂ O).	translocation during protein	gene.	
	synthesis		

Table 2.1: Antibiotics used in this study.

2.1.7 Bacterial Strains, Plasmids and Primers

Bacterial strains, plasmids and primers used in this study are listed in Tables 2.2-2.8. For bacterial growth, a single colony was incubated 16 h at 37 °C with shaking at 250 rpm in 3 ml LB broth using 6-inch sterile test-tubes with caps. Glycerol was added to cultures (to give 20% glycerol) after growth for long term maintain of strains at -80 °C in cryovials. Primers for amplification of target genes were designed using Vector NTI 10 software (Table 2.4-2.8). After design, suitable recognition sites (e.g. *Bam*HI and *Eco*RI) were added in addition to three random nucleotides at the 5′ end to enable subsequent restriction enzyme recognition of PCR products. All oligonucleotides were ordered from Eurofins Genomics.

Table 2.2: Strains used in this study.

Strain	Genotype	Source
		(Reference)
	Salmonella Strains	
Salmonella enterica serovar Enteritidis PT4-P125109	Wild type	Sophie Jan and Florence Baron Rennes, France
JSG210	(Wild Type) ATCC 14028s	John Gunn The Ohio State University
JSG421	$pmrA$::Tn10 Δtet	John Gunn The Ohio State University
JSG425	λ-Pir <i>phoP</i> ::Tn10 Δtet	John Gunn The Ohio State University
Salmonella enterica SEN1432	ΔSEN1432	This study
Salmonella enterica ∆dgoR	$\Delta dgoR$	This study
	Escherichia coli strains	
Тор10тм	<i>E.</i> coli F., mcrA, Δ (mrr-hsdRMS-mcrBC), φ 80lacZ Δ M15, Δ lacX74, nupG, recA1, araD139, Δ (ara-leu)7697, galE15, galK16, rpsL(Str ^R), endA1, λ ⁻	Invetrogen
BW25113	F ⁻ , Δ (araD-araB)567, ΔlacZ4787(::rrnB-3), λ^- , ΔfocB740::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514	(Datasenko. and Wanner, 2000)
XL1-blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI ^q $\Delta(lacZ)$ M15 Tn10 (Tet ^R)]	Stratagene
BL21(DE3)	$F ompT hsdSB(r_B m_B) gal dcm (\lambda DE3)$	Invitrogen
BL21(DE3) Rosetta®	F- $ompT hsdSB(r_B^{-}m_B^{-}) gal dcm (\lambda DE3) pRARE (Cam^R)$	Invitrogen
BL21(DE3) Star®	F- $ompT hsdSB(r_B^- m_B^-)$ gal dcm rne131 ($\lambda DE3$)	Thermo Scientific

Table 2.3 Plasmids used in this study. All plasmid stocks were maintained at -20 °C in ultra-pure water.

Plasmid	Genotype	Source (Reference)
pJET1.2/blunt	Cloning vector, Amp ^R .	Fermentas
pRS1274	<i>lacZ</i> transcriptional fusion vector containing <i>Bam</i> HI- <i>SmaI-Eco</i> RI- <i>lacZ</i> cloning site, <i>lacZ lacY lacA</i> Amp ^R	Lab stock Simons <i>et al.</i> , 1987
pET21a(+)	Overexpression cloning vector with T7 promoter	Novagen
pSU18	Cloning vector with $lacZ\alpha$ gene, Cm^R	Bartolome et al., 1991/Lab stock

pKD3	Derived from pANTSγ, containing FRT- flanked <i>cat</i> gene from pSC140, Cm ^R	Wanner and Datsenko, 2000/ Lab stock
pCP20	Temperature sensitive plasmid (30 °C) encoding a Flp-recombinase, Amp ^R and CmR	H.Mori, Japan/ Lab stock
pKD46	Temperature sensitive replication (<i>repA101</i> ^{ts}); encodes lambda Red genes (<i>exo, bet, gam</i>); native terminator (tL3) after <i>exo</i> gene; arabinose-inducible promoter for expression (P_{araB}); encodes <i>araC</i> for repression of ParaB promoter; Amp ^R , Kan ^R this plasmid can be cured of a strain with growth at 37 – 42 °C	Lab stock
pJET1.2 plus	s target regions from Salmonella enterica se	rovar Enteritidis PT4-P125109
pJET-ybhC'	Possible pectinesterase	This study
pJET-SEN1435'	Putative hexonate dehydrogenase	This study
pJET-SEN1436'	Putative dehydratase	This study
pJET-SEN1432'	Putative GntR-family regulatory protein	This study
pJET-dgoR'	Galactonate operon transcriptional repressor	This study
pJET-dgoT'	D-galactonate transporter	This study
pJET-SEN2978'	Mannonatedehydratase	This study
pJET-SEN2977'	Hexuronate transporter	This study
pJET-SEN2979'	Mannonate oxidoreductase	This study
pRS1274 plu	s target regions from Salmonella enterica se	rovar Enteritidis PT4-P125109
pRS-ybhC-lacZ	Possible pectinesterase	This study
pRS-SEN1435-lacZ	Putative hexonate dehydrogenase	This study
pRS-SEN1436-lacZ	Putative dehydratase	This study
pRS-SEN1432-lacZ	Putative GntR-family regulatory protein	This study
pRS-dgoR-lacZ	Galactonate operon transcriptional repressor	This study
pRS-dgoT-lacZ	D-galactonate transporter	This study
pRS-SEN2978-lacZ	Mannonatedehydratase	This study
pRS-SEN2977-lacZ	Hexuronate transporter	This study
pRS-SEN2979-lacZ	Mannonate oxidoreductase	This study
pSU18 plus	target regions from Salmonella enterica service	ovar Enteritidis PT4-P125109
pSU18-SEN1432	Putative GntR-family regulatory protein	This study
pSU18-dgoR	Galactonate operon transcriptional repressor	This study
pSU-PmrAB	Two-Component System Regulator BamHI and GAATTC for EcoRI	This study

Table 2.4 Designed primers. Vector NTI 10 software was used to design primers for amplification of selected regions. Restriction sites, where present, are in green (GGATCC for *Bam*HI and GAATTC for *Eco*RI).

Name	Sequence 5' - 3'	GC%	Primer length (bp)	Tm (°C)	Fragment length (bp)
ybhC-F	GAGGGATCCATCAGCGCCTGGTTATCCACCAGC	58.33	24	64.11	446
ybhC-R	CACGAATTCTTGATCGGAAGGGATCTGATCGGG	54.17	24	63.94	
SEN1435-F	GAGGGATCCGCCCTGGCTCGTTGGTTTCTATCTT	52.0	25	61.76	504
SEN1435-R	CACGAATTCCAAAGCCCAGTCCTCGTGCAGAAC	58.33	24	62.95	
SEN1436-F	GAGGAATTCGCCCTGGCTCGTTGGTTTCTATCTT	52.0	25	61.76	504
SEN1436-R	CACGGATCCCAAAGCCCAGTCCTCGTGCAGAAC	58.33	24	62.95	
SEN1432-F	GAGGGATCCGGTGTCAACGATGCTGGTTAAAGAAC	46.15	26	59.12	420
SEN1432-R	CACGAATTCAGTTCCACTTCTGAGGGCAAACGG	54.17	24	61.56	
dgoR-F	GAGGGATCC GAGGTGATGGCGATTGGCGATCAG	58.33	24	65.41	551
<i>dgoR</i> -R	CACGAATTCCAGCGCCGAACCGGGTACGTATTT	58.33	24	65.48	
<i>dgoT</i> −F	GAGGGATCC ACTATAACAAGGGCGCGGAGCTGCT	56.0	25	64.05	434
<i>dgoT-</i> R	CACGAATTCGTTGGCGCGATCGACGTAGCAAAT	54.17	24	64.87	
SEN2978-F	GAGGGATCCCCCTACGCAGACCAGGCCGATAAT	58.33	24	63.13	557
SEN2978-R	CACGAATTCGATATGGTGTAACGCCGTTACCACGC	58.85	26	63.19	
SEN2977-F	GAGGAATTCCCCTACGCAGACCAGGCCGATAAT	58.33	24	63.13	557
SEN2977-R	CACGGATCCGATATGGTGTAACGCCGTTACCACGC	58.85	26	63.19	
SEN2979-F	GAGGGATCCGAAGAAGAGCACCGTCGTAAAGCCGA	53.85	26	64.49	399
SEN2979-R	CACGAATTCCCCCGCAGCCCAGATGCACAATAC	62.5	24	66.94	

Table 2.5 Sequencing primers.

Name	Sequence 5' - 3'	GC%	Primer length (bp)	Tm (°C)
pJET_T7 -F	TAATACGACTCACTATAGGG	40	20	45.58
pJET_RP2	AAGAACATCGATTTTCCATGGCAG	42	24	64.18
pRS1274-F	GGATTTGAACGTTGCGAA	44.44	18	49.28
pRS1274-R	AAGTTAAAATGCCGCCAG	44.44	18	48.11

Table 2.6 Primers used for knockouts.

Name	Sequence 5' - 3'	GC%	Primer length (bp)	Tm (°C)
D-dgoR-FOR	GTAAGAGAGTTCACATCGAGCACAAGGACTCTCT ATGACTCTCAATTGTGTAGGCTGGAGCTGCTTC	48	67	84.62
D-dgoR-REV	CGCAGATTGGTCGATCCCCAGTCAATTGCGATG TAGCGAGCTGTCACATATGAATATCCTCCTTAGT	48	67	87.45
PDCFO_dgoR	TGGCATGATAACGACGGTTG	50.0	20	54.33
PDCRE_dgoR	GTGTAACGCCTGCTTCTGATTG	50.0	22	54.24
D-1432-FOR	ATAAAGCACTTCAGCGACATCTTAACGGATACCC ATCTTGAGCATAAATGTGTAGGCTGGAGCTGCTTC	45	69	85.28
D-1432-REV	TCAGATATGTTAAATTGCTCTACTACTTGAGCTTG TAACCAACGGTTACATATGAATATCCTCCTTAGT	35	69	77.59
PDCFO-1432	TTCGTTTCGATTAACGGTGA	40	20	50.97
PDCRE-1432	GCACTGCCACGATTTTAAAGT	42.86	21	51.53
Table 2.7 Primers for amplification whole genes of regulators. Restriction sites, where present, are in green (GGATCC for *Bam*HI and GAATTC for *Eco*RI)

Name	Sequence 5' - 3'	GC%	Primer length (bp)	Tm (°C)
dgoRToT- For	CACGAATTCTAAGCCAGAGGAGGTGATGGCGATT	50.0	34	70.7
dgoRToT- Rev	GAGGGATCCAGGCGTGTAACGCCTGCTTCTGATT	55.9	34	73.1
1432ToT- For	CACGAATTCTGAGTTCATCACCGCGGTACGCTGG	55.9	34	73.1
1432ToT- Rev	GAGGGATCCGATTTCAGGCCGCACTGCCACGATT	58.8	34	74.3

Table 2.8 Designed primers for sequencing cloned *pmrAB*. Restriction sites, where present, are in green (GGATCC for *Bam*HI and GAATTC for *Eco*RI)

Name	Sequence 5' - 3'	GC%	Primer length (bp)	Tm (°C)
pmrAB-FOR	CCACGTGTAGTTAATGTTATCGCAA	40.0	25	55.3
pmrAB-REF	CAACATCCGCGTATCGATGAATAAA	40.0	25	59.03

 Table 2.9 Primers to amplify genes of interest (SEN1432 and dgoA) for over-production. Restriction sites,

 where present, are in green (CATATG for NdeI, AAGCTT for HindIII:)

Name	Sequence 5' - 3'	GC%	Primer length (bp)	Tm (°C)
OP- SEN1432- FOR	GAGCATATGAGCATAAAATCCATTCAAAAACAG AAT	48	36	70.7
OP- SEN1432- 21R	GTGAAGCTTTTTTGTCCCTGATGTCTCTGTAGA TTT	48	36	73.1
OP- SEN1432- 28R	GTGAAGCT1TTATTATTTTGTCCCTGATGTCTC TGTAGATTT	50.0	42	73.1
OP-dgoA- FOR	GAGCATATGAAAATAACTCACATCACCACGTAC	50.0	33	74.3
OP-dgoA- 21R	GTGAAGCTTCCACTCGGCTACCGATCCGTCA	45	31	65.28
OP-dgoA- 28R	GTGAAGCTTTTATTACCACTCGGCTACCGATTCGTCA	35	37	77.59

Materials and Methods

2.2 Methods

2.2.1. Chemically competent cells - preparation and transformation

For the preparation of competent cells, a single colony was inoculated from an agar plate into 3 ml LB-broth in a 6-inch test tube and then incubated overnight at 250 rpm and 37 °C. From this overnight culture, 0.5 ml was transferred to an Erlenmeyer flask containing pre-warmed 50 ml LB broth. The culture was incubated at 37 °C on a rotary shaker (250 rpm) and OD_{650nm} measurements were taken with a "WPA Biowave CO8000 Cell Density meter" until this reached between 0.4-0.5 (usually 2-3 h). Cells were then centrifuged for 5 min at 5000 rpm at 4 °C using pre-chilled and sterile 50 ml Falcon tubes, and then the cell pellet was re-suspended in 30 ml ice cold 100 mM MgCl₂ and incubated for 10 min on ice. Cells were again centrifuged for 5 min at 5000 rpm at 4 °C and then re-suspended in 30 ml ice cold 100 mM CaCl₂ plus 20% glycerol. Finally, 0.2 ml of cells were aliquoted in 1.5 ml Eppendorf tubes on ice, and then stored at -80 °C until use.

For transformation, aliquots were removed from the freezer, placed into an ice box and left to thaw (not more than 10 min). Competent cells were incubated with 1 μ l of plasmid DNA on ice for 30 min. The transformation mixture was then placed into a water bath (42 °C) for 45 s for heat-shock and returned to the ice for 5 min. SOC medium (250 μ l) was added to the collection tube and the cells were allowed to recover by incubation at 37 °C on a rotary shaker (225 rpm) for 1 h. 100 μ l of cells were spread onto LB agar containing appropriate antibiotics and Xgal (2 ml of a 20 mg/ml solution in DMSO added to one litre of medium; 40 μ g/ml final concentration), and plates were incubated inverted at 37 °C overnight.

Materials and Methods

2.2.2 Extraction and purification of nucleic acids

All centrifugation involved an Eppendorf mini-centrifuge used at 13,000 rpm. Purification was carried out in 1.5 ml Eppendorf tubes as described below. DNA was stored at -20 °C until required.

2.2.2.1 Plasmid miniprep

Plasmid DNA minipreps were carried out using a GeneJET plasmid miniprep (Thermo Scientific) kit to screen colonies for the correct plasmid. Firstly, transformants were streaked onto LB Amp plates; the same inoculated loop was used to make the primary smear and to inoculate a fresh 3 ml overnight culture of Amp containing LB broth, which was grown at 37 °C, 250 rpm, over-night. Dependent on the plasmid copy number, 1 ml (for high copy plasmids) to 5 ml (for low copy plasmids) of an overnight culture was used to extract DNA. In the final step, DNA was eluted into 50 µl of sterile water.

Details of the DNA isolation are as follow. The tube was centrifuged at 8000 rpm for 2 min to pellet the cells in the culture. The supernatant was then carefully discarded leaving a dry pellet and the pellet was subsequently re-suspended in 250 μ l of the Resuspension solution (50 mM Tris-HCl, 10 mM EDTA, pH 8.0) including RNase A, by vortexing to ensure a homogenous cell suspension. A 250 μ l volume of the Lysis solution (1% SDS, 0.2 M NaOH) was then added to each tube and each tube was mixed thoroughly by inverting the tube 4-6 times until the solution became viscous. A 350 μ l volume of the Neutralization solution was then added to each tube and mixed by inverting the tube 4-6 times, a white precipitate formed almost immediately (chromosomal DNA and proteins). The tubes were left to stand on ice for another 5 min. The tubes were then centrifuged at 13,000 rpm for 5 min, to pellet the white precipitate along the side of the tube. The supernatants were transferred into the GeneJET spin column by pipetting. The tubes were then centrifuged 13,000 rpm for 1 min and the flow through discarded and the column was returned back to the same tube. The solution back to the same tube.

solution were added, which included ethanol, and this was centrifuged for 1 min, the flow through discarded, and this step was repeated. Finally, the column was transferred to a fresh tube, and then 50 μ l of sterilized distilled water were added. A sample (~ 4 μ l) was electrophoresed in a 0.8% agarose gel and the concentration was measured using a NanoDrop® ND-1000 UV-Vis Spectrophotometer. DNA was stored at -20 °C.

2.2.2.2 Total DNA extraction

Chromosomal DNA was extracted and purified using Thermo Scientific GeneJET Genomic DNA Purification Kit following the protocol guidelines for Gram-negative bacteria. Around 2×10^9 bacterial cells (1 ml) were harvested from an overnight culture in a 1.5 or 2 ml microcentrifuge tube by centrifugation for 6 min at 8000 rpm. After discarding the supernatant, the pellet was resuspended in 180 µl of Digestion Solution. A 20 µl volume of Proteinase K solution was added and mixed thoroughly by vortexing or pipetting. The sample was incubated at 56 °C while vortexing occasionally until the cells were completely lysed (30 min). A 20 µl volume of RNaseA solution was then added, mixed by vortexing and the mixture incubated for 10 min at room temperature. Then 200 µl of Lysis Solution were added to the sample. This was mixed thoroughly by vortexing for 15 s until a homogeneous mixture was obtained. A 400 µl quantity of 50% ethanol was added and mixed by pipetting or vortexing. The lysate was transferred to a GeneJET Genomic DNA Purification Column inserted in a collection tube. The column was centrifuged for 1 min at 12000 rpm and the GeneJET Genomic DNA Purification Column was placed into a new 2 ml collection tube. Then 500 μ l of Wash Buffer I (with ethanol added) were added and the column was centrifuged for 1 min at 12000 rpm. The flowthrough was discarded and the purification column placed back into the collection tube. A 500 µl volume of Wash Buffer II (with ethanol added) was added to the GeneJET column which was centrifuged for 3 min at 12000 rpm. The column was transferred to a sterile 1.5 ml microcentrifuge tube and 200 µl of sterile distilled water were added to the centre of the

column. This was incubated for 2 min at room temperature and centrifuged for 1 min at 12000 rpm.

2.2.3 Determination of DNA concentration

Prior to ligation reaction, the concentration of the plasmid DNA was determined using the Nanodrop spectrophotometer. A $2 \mu l$ drop of plasmid DNA was placed onto the spectrophotometer's pedestal and the absorbance of the sample at 260 nm was used to determine DNA concentration.

2.2.4 Polymerase Chain Reaction (PCR) protocol

All PCR reactions were carried out in an Eppendorf Master cycler[®] gradient PCR machine. Primer stocks were produced by suspending primer DNA into the appropriate volume of water according to manufacturer's specifications generating a 100 μ l stock solution. This was then diluted 1 in 10 to generate a 10 μ l working stock for the PCR reaction. All reactions were performed in 0.2 ml thin-wall PCR tubes purchased from Eppendorf. DNA polymerase, MgCl₂, 10X reaction buffer and dNTP's were obtained from Invitrogen. Each reaction was made up to 50 μ l master mix that contained: 10 μ l 10X reaction buffer, 2 μ l dNTP's (2 mM), 1 μ l of each forward and reverses primer (10 pmol/ μ l), 1 unit of Phusion® High-Fidelity DNA polymerase (Fermentas), and 1 μ l of template DNA (~100 ng genomic or plasmid DNA).The PCR reactions were performed using a lid heated to 105 °C with the following steps used as the standard protocol.

Initial denaturation	-	98°C, 30 s		
Denaturation	-	98 °C, 8 s	٦	
Annealing	-	57 °C, 20 s	-	X3 cycles
Extension	-	72 °C, 15 s		
Denaturation	-	98 °C, 8 s	7	
Annealing	-	67 °C, 20 s	-	X27 cycles
Extension	-	72 °C, 15 s		
Final Extension	-	72 °C, 5 min		
Final step	-	4 °C, hold		

Note: Annealing temperature is changeable according to Tm of primers. Extension time is changeable according to expected length of PCR product in addition to extension ability of the used polymerase.

2.2.5 Colony PCR

Colony PCR was used to rapidly screen multiple colonies for successful plasmid constructs following ligation and transformation. This was achieved using primers flanking the multiple cloning regions of the selected plasmid. If the selected colony contained a plasmid construct with the desired fragment, a PCR product corresponding to the insert size would be amplified; if the plasmid did not contain the insert of interest, any fragment amplified would be of incorrect size. The protocol for colony PCR was essentially the same as that for standard PCR, with the difference that a single bacterial colony was used as the DNA template instead of purified genomic DNA. One colony was selected from an agar plate using a sterile tip. The tip was touched to a separate agar plate (so that a stock of the colony was retained), then dipped into an aliquot of sterile water 20 μ l and stirred gently. Typical reaction volumes used for colony PCR were 25 μ l reactions consisting of 2.5 μ l 10x Dream TaqTM DNA polymerase buffer, 0.5 μ l 10 mM dNTP's, 1 μ l 10 μ M forward primer, 1 μ l 10 μ M reverse primer, 14.75 μ l qH2O and 0.25 μ l Dream TaqTM DNA polymerase added on ice. Then, 5 μ l colony solution

was mixed with the reaction constituents by gently pipetting up and down taking care not to introduce too much air to the PCR reaction. Once all reactions were prepared they were placed into an Eppendorf Mastercycler[®] PCR machine to be amplified. The above conditions were used as a standard with the annealing temperature adjusted according to appropriate primer Tm. The PCR reactions were cooled to 10 °C and analysed for the appropriate plasmid insert by agarose gel electrophoresis by using 10 µl of the reaction and visualized under UV light.

2.2.6 Agarose gel electrophoresis

0.8% w/v agarose gels were prepared by using Melford Molecular Grade Agarose powder in 0.5X TBE buffer (5X solution contains 0.45 M Tris, 0.45 M borate, 0.01 M EDTA). Biotium Gel RedTM (10,000X in water) was added to a final concentration of 1X for visualization of DNA fragments. Samples were prepared by the addition of 6X loading buffer (0.25% w/v bromophenol blue, 0.25% w/v xylene cyanol FF, 15% w/v Ficoll) to a final concentration of 1X. The samples were electrophoresed for 45-60 min in gels submerged in 0.5X TBE buffer with a voltage gradient of 70 V cm⁻² in a BioRad horizontal gel tank. DNA bands were visualised using a G-Box UV transilluminator and photographs taken digitally. DNA concentration was measured on a NanoDrop® ND-1000 UV-Vis Spectrophotometer.

2.2.7 Purification of PCR products

PCR products were purified using a Thermo Scientific GeneJET PCR Purification Kit according to the manufacturer's instructions. One volume of Binding Buffer was added to the completed PCR mixture, mixed thoroughly until the colour of the mix became yellow. Two volumes of 100% isopropanol were added and mixed thoroughly. Up to 800 μ l of the solution was transferred to the GeneJET purification column. Tubes were centrifuged for 30-60 s and the flow-through was discarded. Then, 700 μ l of Wash Buffer were added to the GeneJET purification column. The mixed additional 1 min. The

column was transferred to a clean 1.5 ml microcentrifuge tube and 30 μ l of sterilized ultra-pure water were added to the centre of the GeneJET purification column membrane and this was centrifuged for 1 min. The column was incubated for 1 min at room temperature before centrifugation, and the purified DNA was stored at -20 °C.

2.2.8 Restriction digestion

Digestions with restriction endonucleases of PCR products reactions or plasmid DNA were performed for cloning purposes or to confirm the desired DNA insert was carried. Reactions varied but typically a 10 µl reaction was used for plasmid digestion which usually comprised: 4 µl of plasmid DNA, 1 µl of 10x Fast Digest buffer, 0.5 µl of each Fast Digest restriction endonuclease were added (totalling 0.5-1 µl). The reaction mixture was then made up to 10 µl using 4 or 4.5 µl of qH_2O respectively. Tubes were incubated at 37 °C water bath for 5 min. Following the reactions, the enzymes were usually inactivated by incubation at 65 °C for 5 min.

2.2.9 PCR extraction from agarose gel

All purification steps were carried out at room temperature. The gel slice containing the DNA fragment was excised using a clean scalpel or razor blade and blue light box. The gel slice was placed into a pre-weighed 1.5 ml tube and weighed. One volume (volume:gel-weight) of Binding Buffer was added to the gel slice. The mixture was incubated at 50-60 °C for 10 min or until the gel slice was completely dissolved, and then mixed by inversion every few minutes. The gel mixture was mixed briefly before loading on the column. The colour of the mix became yellow. One gel volume of 100% isopropanol was added to the GeneJET purification column. This was centrifuged for 1 min, the flow-through discarded and the column back placed into the same collection tube. A 700 μ l volume of Wash Buffer (diluted

with ethanol) were added to the GeneJET purification column. This was centrifuged for 1 min and the flow-through discarded and the column placed back into the same collection tube and centrifuged again. The column was centrifuged for an additional 1 min and was then transferred into a clean 1.5 ml microcentrifuge tube. A 50 μ l volume of Elution Buffer was added to the centre of the purification column membrane. This was incubated for 1 min at room temperature before centrifugation for 1 min. The purified DNA was stored at -20 °C.

2.2.10 Ligation of vector with PCR product

The first step in this study was determining promoter regions for genes of interest and designing primers for their amplification using Vector NTI. For this purpose, desired regions were amplified and the PCR products were initially cloned into the pJET1.2 cloning vector. PCR cloning was performed using Thermo Scientific CloneJET PCR Cloning Kit. pJET1.2/ blunt is a linearized cloning vector, which accepts inserts from 6 bp to 10 kb. The recircularized pJET1.2/blunt vector expresses a lethal restriction enzyme after transformation and so such transformants cannot propagate. As a result, only recombinant clones containing the insert appear on culture plates. PCR products and any other DNA fragment, either blunt or sticky-end, can thus be successfully cloned. The vector contains an expanded multiple cloning site and sequencing primers are included for convenient sequencing of the cloned insert (Fig. 2.3).



Figure 2.3: Map of the pJET1.2/blunt. This plasmid used for cloning PCR fragments. Source: http://www.bioinfo.pte.hu/f2/pict_f2/pJETmap.pdf.

Purified PCR products were ligated with selected plasmids according to manufacturer's instructions and then placed on ice. Reactions consisted of 10 µl of 2X Reaction Buffer, 1 µl purified PCR product, 1 µl pJET1.2/blunt cloning vector and 1 µl T4 DNA ligase in a final volume of 20 µl. The ligation mixture was vortexed briefly and centrifuged for 3-5 s. The ligation mixture was incubated at room temperature 22 °C for 5 min. A 5 µl volume of ligation mixture was used directly for transformation into chemically competent *E. coli* TOP10 cells. Transformants thus generated were grown overnight in Amp containing LB broth for plasmid isolation. Following purification of plasmids of interest, 5 µl of plasmid DNA (50-100 ng/µl) was sent for sequencing with an appropriate primer (Table 4.3) to Source Bioscience (Cambridge, UK) to determine if the insert has the expected sequence. The sequence data was analysed by NCBI alignment software.

Materials and Methods

2.2.11 lacZ fusions construction

The *lacZ* transcriptional (promoter-less) plasmid (pRS1274, figure 2.4; Simons *et al.*, 1987) was obtained for generation of transcriptional fusions. The diluted DNA was transformed into chemically competent Top10 and transformants selected on LB agar containing ampicillin. The ~450 bp *Eco*RI and *Bam*HI PCR fragments released from the pJET1.2 clones were then ligated with the digested vector. The same conditions of pJET cloning were used with pRS1274 to construct *lacZ* fusions but in this case, 3 μ l (~40-100 ng) of released fragments were ligated with 2 μ l (100-200 ng) of *Eco*RI and *Bam*HI digested cloning vector. Ligation reactions were incubated at 22 °C for 15 min. Then, ligations were transformed into chemically competent Top10 using transformation protocol as in section 4.3.1. Transformants were selected on LB medium with ampicillin and Xgal. Amp^R and Lac⁺ transformatis were selected for further analysis. Isolated plasmids were subjected to restriction digestion analysis and nucleotide sequencing using specific primers (Table 2.5).



Figure 2.4: Map of pRS1274 *lacZ* **transcriptional vector.** Map illustrates the multiple cloning site, *lacZ*, and Ampicillin resistance gene.

After sequencing, the pRS constructs were assigned unique designations for easier identification: pRS-*ybhC*-*lacZ*; pRS-SEN1435-*lacZ*; pRS-SEN1436-*lacZ*; pRS-SEN1432-*lacZ*;

pRS-*dgoR*-*lacZ*; pRS-*dgoT*-*lacZ*; pRS-SEN2978-*lacZ*; pRS-SEN2977-*lacZ*; and pRS-SEN2979-*lacZ* (see Table 2.3).

The desired plasmid constructs were also generated in silico using Vector NTI (Figure 2.5 as example and for the rest see Appendix 7).



Figure 2.5: Physical maps of the transcriptional pRS-SEN1436-*lacZ* fusion vectors generated during this study. All plasmids included pRS1274 as the vector. Inserts are in red and proximal region of fused genes are indicated by small green arrows just upstream of *lacZ*. Maps were drawn using the Vector NTI program.

The sequences obtained were compared with the sequence database using BLAST which confirmed that the inserts have the correct sequence correctly located at the desired cloning sites.

2.2.12 β-Galactosidase assay

1. The growth of cultures and collecting samples

Fifty ml of LB broth (containing 100 μ g/ml ampicillin, as required) in sterile Erlenmeyer 250 ml flask were inoculated with 0.5 ml preculture (grown overnight in 2.5 ml LB broth in 6 inch test tubes, in duplicate). Cultures were grown to stationary phase at 37 °C, 250 rpm for 24 h using a Sanyo Gallenkamp shaker. The OD at 600 nm was measured using a 'WPA Biowave CO8000 Cell Density meter' and once the OD_{650nm} was above 0.1 (~2-3 h), samples (0.5 OD

units) were collected in pre-chilled tubes every hour until growth was complete (~8 h). A final measurement and sample was taken at 24 h. Samples were centrifuged for 2 min at 13000 rpm, the residual supernatant thoroughly removed, and cell pellets were stored at -80 $^{\circ}$ C until use in the next step.

2. Cell lysis

Samples (*E. coli* transformants) were defrosted on ice and permeabilised using 100 μ l 1X Bugbuster (Novagen). Pellets were resuspended and incubated at 37 °C with shaking at 250 rpm in a Gallenkamp shaker for 30 min. Samples were centrifuged for 5 min at 13000 rpm, the supernatants were transferred into separate Eppendorf tubes. The same protocol was followed for measurement of β -galactosidase activities in *Salmonella* Enteritidis transformants. However, Bugbuster failed to fully lyse *Salmonella* and so B-PER (Thermo Scientific) was used instead as a cell lysis reagent. The required amount of B-PER was pre-warmed at 37 °C, then 100 μ l (4 ml of B-PER Reagent per gram of cell pellet) were added and the re-suspended pellet was incubated for 10-15 min at room temperature.

3. β-Galactosidase assay

Reactant solution (Buffer Z: 80 mM Na₂HPO₄.7H₂O, 45 mM Na₂HPO₄.H₂O, 10 mM KCl, 1 mM MgSO₄.7H₂O₂) contained 4 mg ml⁻¹ o-nitrophenyl- β -D-galactopyranoside (ONPG) and 5 mM DTT (aliquots of 20 ml at -20 °C). The β -galactosidase assay was performed as described by Miller *et al.* (1972). Four μ l of each sample were added to wells (in triplicate) of a microtitre plate. Using a multichannel micropipette 16 μ l of PBS were added into the same wells, then 180 μ l of the reactant solution were added to each well. PBS-only wells were included as control. The assay was monitored by immediately inserting the plate into a plate reader equipped with kinetic capacity. Readings were taken every 2 min for an hour at an A_{420nm} in a Spectra MAX 340 pc (Molecular Devices). After the assay, final absorbance was taken at 420 nm using the Endpoint programme. Then the data was exported using Excel

format and used to calculate to calculate β -galactosidase activities (nmol ONPG/min/OD unit) according to this equation: (raw Abs/minx 100 μ l/4 μ l x 1/OD used x 0.135 nmoles ONP).

2.2.13 Preparation competent cells of SE

A single colony of *Salmonella enterica serotype* Enteritidis was inoculated into 5 ml of LB medium and incubated for overnight at 37 °C, 250 rpm for 24 hours using Sanyo Gallenkamp orbital shaker (Lee and Chang, 1994; Siguret *et al.*, 1994). One ml of this culture was transferred to 100 ml of LB medium and incubated at 37 °C with vigorous shaking until the OD_{600nm} of the culture reached 0.6 (~100-120 min). The culture was divided into two 50 ml Falcon tube and chilled on ice for 30 min (to ensure that the temperature was no more than 4 °C). Cells were harvested at 5000 rpm, for 15 min, at 4 °C; the pellet was re-suspended after removing the supernatant then the suspensions were combined together. The pellet was resuspended to a final volume of 0.2 ml in ice-cold GYT (10% glycerol, 0.125% yeast extract and 0.25% tryptone; this medium was sterilised by 0.22 µm Millipore filter). Forty µl aliquots in 1.5 pre-chilled Eppendorf tubes were then prepared, and the tubes were subjected to snap freezing using liquid nitrogen before transfer to storage at -80 °C.

2.2.14 Electroporation

Electroporation was carried out following a method described by Lee and Chang (1994) and Siguret *et al.* (1994). Electroporation was performed using a Gene Pulser (Bio-Rad) in a prechilled 1 mm cuvette (at least 40 min on ice), under conditions suggested by the manufacturer (i.e. 25 μ F, 200 Ohms, 1.8 kV). A 2 μ l volume (~35 ng/ μ l) of plasmid DNA was mixed by pipetting with 40 μ l of pre- prepared of *Salmonella* Enteritidis competent cells, and the mixture was transferred immediately into a pre-chilled electroporation cuvette (volume capacity 20-90 μ l). The cuvette was wiped dry before placement in the electroporation holder, and then two red bottoms were pressed together (constant time 4 ms). Then 1 ml of pre-warmed SOC medium (see Methods 2.1.5.2) was added immediately. Cells were transferred to a polypropylene tube (17×100 mm) and incubated at 37 °C for 1 h. After this, 100 µl were taken for plating onto LB agar plate containing ampicillin and Xgal; the remaining broth was micro centrifuged at 13,000 rpm, 800 µl of supernatant were then removed and the pellet was suspended in the residual supernatant (~150 µl) before plating on to another plate. Plates were incubated overnight at 37 °C for 18-24 h. Next day, a single colony was selected for confirmation and experimental use.

2.2.15 Gene inactivation procedure

2.2.15.1 The Red disruption system

Gene knockout in SE (SEN1432 and *dgoR*) was achieved using the Wanner method (Wanner and Datsenko, 2000). This method relies upon the presence of a low copy, temperature sensitive "helper" plasmid encoding components of the homologous recombination system found in bacteriophage λ . These components are called Exo (a 5'-3' exonuclease, which processes along double-stranded DNA), Bet (a single-stranded DNA-binding protein, which is able to anneal complementary single strands) and Gam (an inhibitor of host RecBCD exonucleases). Expression of these genes is under the control of an arabinose-inducible promoter (P_{araBAD}). When cells expressing the plasmid are grown in the presence of arabinose, exogenously applied linear DNA is able to undergo homologous recombination with the bacterial chromosome. In this way, it is possible to generate an in-frame gene deletion using a PCR product.

2.2.15.2 Primer design

Primers were designed to anneal at the 4th codon and the penultimate codon of the target gene (Table 2.6), allowing generation of an in-frame deletion with minimal downstream effects. The 5' end of each primer (between 45-48 nucleotides) was a 100% match to the target gene, whereas the 3' end of each primer was designed to amplify the chloramphenicol resistance cassette encoded by pKD3.

2.2.15.3 PCR amplification of CAT cassette

The plasmid pKD3 was used as a template for PCR so that linear DNA encoding the *cat* cassette could be generated. PCR was carried out as described in section 2.2.9 and the product was purified as described in section 2.2.11.

2.2.15.4 Induction and preparation of host cell

Cells expressing pKD46 plasmid were grown in LB (containing antibiotics as appropriate) at 30 °C, 250 rpm for 4 h. At this point, arabinose was added to a final concentration of 10 mM in order to induce expression of the homologous recombination system. The cells were incubated under the same conditions for 1 h and then harvested by centrifugation at 4000 rpm for 20 min at 4 °C. The cell pellet was then aspirated and re-suspended in 1 ml ice cold water. The cells were then centrifuged at 13000 rpm for 1 min, the supernatant was removed and the pellet was re-suspended in the same volume of ice cold water. This washing process was repeated five times in total, after which the cells were re-suspended in a volume of ice cold water approximately double that of the pellet. The cells were then aliquoted into pre-chilled electroporation cuvettes and incubated on ice for 15 min prior to use.

2.2.15.5 Electroporation with linear DNA

About 2 μ g of the linear DNA was added to each electroporation cuvette and mixed by pipetting. The cell-DNA mixture was then subject to electroporation (see Methods 2.2.14) The cells were then incubated at 30 °C for 1-3 h and subsequently spread on solid media containing chloramphenicol (8 μ g/ml). The plates were then incubated at 37 °C overnight. Next day, single colonies were selected for further work and were propagated on LB-agar plates containing (34 μ g/ml) chloramphenicol.

2.2.15.6 Elimination of the chloramphenicol resistance cassette

The Cm^R cassette was removed from Cm^R substitution mutants as above as part of the strain construction process, and the Flippase (FLP) recognition target (*frt*) sites were used in order to do so. The method used to delete the antibiotic resistance gene was as described by Wanner and Datsenko (2000). Strains from which the Cm^R cassette needed to be removed were transformed with pCP20 plasmid (Table 2.3). This is an ampicillin and chloramphenicol resistant plasmid that displays temperature sensitive replication and thermal induction of FLP synthesis. The transformed cells were plated onto LB agar containing ampicillin and incubated overnight at 30 °C. A few colonies were selected, plated on LB agar and incubated overnight at 44 °C in order to delete the chloramphenicol resistance cassette from the bacterial chromosome. Single colonies were picked and streaked onto LB agar, LB agar plus ampicillin and LB agar (without any additional antibiotics) were those that had the Cm^R cassette removed and had also lost the plasmid. The deletion of the resistance cassette was confirmed by colony PCR (Methods 2.2.9) using primers indicated in Table 2.6.

2.2.16 Phenotypic studies

All bacterial strains used were grown in the appropriate medium under aerobic conditions, unless otherwise specified. To prepare an overnight culture, 5 ml of medium was inoculated with a single bacterial colony from a freshly streaked plate. Overnight cultures were grown in a shaking incubator for 16 h (stationary phase) at 250 rpm in 6-inch sterile test tubes at 37 °C. Overnight cultures were used to inoculate 50 ml of the desired medium in 250 ml sterile flasks and incubated at required temperature with 250 rpm in a shaking incubator for 24 h. Samples were taken at regular intervals to measure the optical density at 600 nm using a spectrophotometer.

Alternatively, tests were carried out using a Bioscreen C Microbiological Growth Analyser (Labsystems, Helsinki, Finland) which measures the turbidity (growth) by vertical photometry. The set-up of the experiment is based on the use of the non-standard, 100-well honeycomb micro-plates manufactured for this machine. The test organisms were grown for 16 h (overnight) in L-broth or M9 minimal medium and incubated in an orbital shaking incubator at 37 °C at 250 rpm. Equal optical density at 600 nm of cultures needed for inoculation at 1:100 dilutions was calculated using the following formula:

Volume of inoculum needed in ml = (desired OD)/(actual OD) x total volume of culture in ml. Each growth condition was performed in triplicate and 300 μ l of each pre-inoculated culture were loaded in each well. Plates were incubated at a suitable temperature for 24 h and the OD_{600nm} was measured at 60-min intervals.

2.2.17 Protein work

2.2.17.1 Prepare egg white and egg white filtrate

Egg-white has to be prepared aseptically from fresh eggs (less than one week old) (Baron *et al.*, 2015). Eggs were bought from the local supermarket (free-range eggs). First, the presence of cracks of any kind was checked to avoid potentially contaminated egg. Then, eggs were wiped with 70% ethanol and flamed into a pan covered with aluminium foil. All the material used (beakers, mixer bar from homogeniser) were sterilised before use, by autoclaving at 120 °C. Eggs (3 to 4 for each batch) were broken and the egg white split from the yolk in a sterile beaker avoiding any contamination with yolk or shell. Then, the suspension was transferred to a fresh sterile beaker for homogenisation in a sterile environment with a Silverson homogeniser for 1-2 min at 10000 rpm (the time depended on the volume of egg-white, ~50 ml per egg). To verify the sterility of the prepared suspension, 1 ml was inoculated in Tryptone soy agar and incubated for 4 h at 37 °C. The egg white was stored in at 4 °C in 50 ml sterile Falcon tubes for use within one week.

Egg white filtrate (EWF) (10 kDa cut-off) was provided by Drs Sophie Jan and Florence Baron (Agrocampus, Rennes, France). It was delivered in 16 ml sterile Falcon tubes and stored at 4 °C until use. It was prepared by ultrafiltration of three different batches of liquid egg white (from different eggs). Ultrafiltration was performed using a pilot unit (TIA, Bollène, France) equipped with an Osmonics membrane (5.57 m², 10 kDa cut-off; PW2520F, Lenntech B.V., Delft, Netherlands). Filtration was achieved according to Baron *et al.* (1997). Concentrated egg white (retentate) was circulated back to the feed tank and permeate (filtrate) was drained off, collected in a beaker, sterilized by filtration (NalgeneR filter unit, pore size <0.2 μ m, Osi, Elancourt, France), and then stored at 4 °C until use. The pH (9.3 ± 0.1) of the egg white filtrate remained unchanged.

2.2.17.2 Lysozyme purification process

The method is based on ovomucin extraction by precipitation in a first step, and further lysozyme extraction by ion exchange chromatography in the second step according to Guérin-Dubiard *et al.*, (2005). Egg white was collected from 6 to 7 eggs to get 190 ml, this volume was diluted with 570 ml of distilled water and pH adjustment to 6 with HCl 1N. The diluted suspension was stirred overnight at 4 °C to enable ovomucin precipitation. The suspension was centrifuged for 5 min at 3000g at 4°C to remove the precipitate. Once the ion exchange chromatography ready, the pH of the supernatant is adjusted 8 with 1 N NaOH and then removing the insoluble material by centrifugation at top speed for 30 min at 4°C. This suspension is mucin free egg white (MFEW).

A low-pressure chromatography system was used with a 100 ml of cation exchanger (SP sepharose) packed in a suitable column. The column was equilibrated with two column volumes of distilled water. A 100 ml of the MFEW was loaded in the column, the flow was applied at 5 ml/min. At pH 8, lysozyme and avidin are positively charged. Lysozyme and avidin fraction has been eluted by a washing step with 150 ml gradient 1 M NaCl. Avidin could be neglected because of its very low concentration (0.05%) in egg white. The resulting fractions corresponding to the UV absorbance peaks were analysed by SDS-PAGE for protein content and purity. Fractions containing protein were pooled together and stored at 4 °C. The column was then cleaned and stored in 30% ethanol.

2.2.17.3 Protein quantification

After determining the level of sample purity via gel electrophoresis, the protein concentration was then measured using two different methods, Bio-Rad protein assay or absorbance at 280 nm. The Bio-Rad protein assay (a dye-binding assay based on the Bradford method) involved use of a range of freshly prepared protein standard solutions (bovine serum albumin, 0.025 to 5

µg/ml). These and samples were combined with the Bio-Rad dye concentrate according to the manufacturer's instructions. The absorbance of each solution at 595 nm was measured and a standard curve was generated (protein concentration vs. absorbance). Reference to the standard curve then allowed the concentration of the sample to be determined.

Absorbance of protein samples was measured using a Nanodrop ND-100 spectrophotometer (Nanodrop Technologies). This method allows protein sample concentration to be estimated by monitoring the absorbance of the sample at 280 nm. Briefly, a small aliquot of the sample (2 μ l) was dispensed onto the lower half of an optical pedestal and then drawn up into a column as the upper half of the pedestal was lowered. The machine determines optical path-length and measures the absorbance of the sample, then calculates sample concentration (mg/ml) based upon an assumed extinction coefficient of the protein.

2.2.17.4 Polyacrylamide gel electrophoresis (SDS)

Polypeptide molecular weights and protein purity were estimated by using 15% polyacrylamide gels and the Bio-Rad Mini Protein II system. 15% SDS-polyacrylamide gels contained 5 ml Tris-HCI (0.5 M, pH 8.8), 10 ml 30% w/v acrylamide (Bio-Rad), 0.2 ml 10% w/v sodium dodecyl sulphate (SDS), 0.07 ml 10% w/v fresh ammonium persulphate, 0.015 ml TEMED, 4.7 ml qH₂O. The gel was cast and, once set, the stacking gel applied to the top. The stacking gel was made up of 2.5 ml Tris-HCI (0.5 M, pH 6.8), 1.5 ml 30% w/v acrylamide, 0.035 ml 10% w/v SDS, 0.01 ml 10% w/v ammonium persulphate, 0.015 ml TEMED, 4.9 ml qH₂O. SDS-loading buffer was made up of 50 mM Tris- HCl (0.5 M, pH 6.8), 10% v/v glycerol, 2% w/v SDS, 0.1% w/v bromphenol blue, 200 mM dithiothreitol (DTT) or β mercaptoethanol, and 8.85 ml qH₂O.

Materials and Methods

2.2.17.5 Western blotting

Proteins were separated by SDS-PAGE and then transferred to a nitrocellulose membrane by electroblotting at 60 V for 1 h. Following transfer, the membrane was washed in 1x TBS (20 mM Tris, 500 mM NaCl, pH 7) for 10 min on a shaking platform. The membrane was then blocked in blocking solution (1% BSA-TBS) for 1 h at room temperature or overnight at 4 °C while shaking gently. After blocking, the membrane was washed twice in TTBS (20 mM Tris, 500 mM NaCl, 0.05% Tween 20, pH 7.0) for 10 min for each wash. The membrane was then probed with the primary antibody diluted in antibody buffer (1% BSA-TTBS) and incubated for 1 h at room temperature. Following three 10 min washes in TBST, membranes were incubated for 1 h with the secondary antibody conjugated to horseradish peroxide (HRP) or alkaline phosphatase diluted in antibody buffer. Finally, membranes were washed three times for 10 min each in TBS to remove excess tween detergent followed by signal detection using BCIP tablets (Sigma). Images were visualised and recorded using a G:BOXChemi (Syngene) with GeneSys software.

2.2.17.8 Protein overexpression

In order to obtain a large amount of the SEN1432 and DgoA proteins in native form, the encoding genes were cloned directly into pET21a or pET28 (Novagen) vectors. The plasmids were transformed into the *E. coli* BL21/ λ DE3 or BL21/ λ DE3 Star or Rosetta strains before expression was induced by IPTG.

Initial small-scale protein overexpression was conducted as follows, using BL21 (λ DE3) transformants containing an overexpression pET21a vector. Strains were inoculated into 3 ml LB broth with ampicillin in sterile test tubes and incubated for 12-16 h at 37 °C on a rotary shaker (250 rpm). Then, 500 µl of overnight culture were used to inoculate 50 ml pre-warmed LB broth with ampicillin in 250 ml flasks and the cultures were incubated at 37 °C at 250 rpm.

Growth was monitored (OD_{650nm}) until the OD reached 0.5. At this point, 0.5 OD units of cells were collected in a 1.5 ml Eppendorf tube on ice which was centrifuged at 13000 rpm for 7 min to the pellet cells. The supernatant was discarded and the dry pellet was stored at -20 °C. IPTG was then added to the culture to a final concentration of 1.0 mM to induce protein expression. Each hour for 5 hours after adding IPTG, and after overnight growth, 0.5 OD units of cells were collected and treated as above. All cell pellets were then defrosted and resuspended in 100 µl of 1x SDS sample loading buffer and subjected to analysis by SDS-PAGE (Methods 2.2.17.4).

2.2.18 Hexonate preparation and synthesis.

Three commercial hexonate forms were purchased (sodium gluconate, D-mannono-1,4-lactone, and L-(+)-gulonic acid γ - lactone) from Tokyo Chemical Industry Co., Ltd. (TCI) in addition to the synthesis of D-galactonic acid from D-galactose (by Dr Chris Jones, Chemistry Department, University of Reading). Stock solutions of hexonates were prepared at 10% w/v concentration, and the pH of the gulonic and D-galactonic acid was adjusted from 2.5 to 7 using KOH, while Na-gluconate and D-mannono-1,4-lactone were already at pH ~7. All solutions were sterilized using 0.22 Millipore filters. These hexonates have the chemical structures as shown in figure 2.6.



Figure 2.6: Structure of three commercial hexonates used in this work and synthesised D-galactonic acid (Pezzotti *et al.*, 2006).

The method used for synthesis of D-galactonate was to employ the corresponding hexose (in this case, galactose), oxidise it with bromine (Br2) in water to give galactonate (Pezzotti *et al.*, 2006), as follows. Four gram from of D-galactose (22.2 mmol) were dissolved in H₂O (150 mL). Br₂ (1.14 mL, 22.2 mmol) was added and the reaction was stirred at ambient temperature for 48 h. The remaining Br₂ was removed by sparging with compressed air for 1 h. Then, the mixture was concentrated in vacuum at 35 °C (to avoid browning of the solution). The syrup formed was then made up to 22 mL with H₂O to give a 1 M solution of D-galactonic acid. So the quantitative yield was 22 mmol in 22 mL H₂O and the purity was established by NMR. The final concentration of galactonate was 1 M, which was neutralised in same volume of 1 M NaOH. The final sample was stored at -20 °C.

Chapter 3: Generation and preliminary analysis of '*hex*' gene *lacZ* transcriptional fusions.

3.1 Introduction

In order to further understand the behaviour of SE when exposed to the bactericidal conditions of egg white, Baron and co-workers studied the global transcriptional response of SE to egg-white (Baron *et al.*, 2017) using microarray technology. The resulting change in expression involved groups of genes which have functions related to survival in egg white (EW), as follows:

- 1- Up-regulated biotin biosynthesis, iron-restriction response, Kdp potassium uptake system, heat-shock response, and envelope-stress response; and
- 2- Down regulated energy-metabolism, amino acid biosynthesis and uptake, motility and chemotaxis, and a subset of virulence factors.

Surprisingly, in addition to the effects summarised above, egg-white exposure strongly induced expression of genes involved in utilisation of hexonates/hexuronates. These systems had not been previously reported to have any role in EW survival, nor had they been shown to be upregulated by EW exposure or to be subject to co-regulation. These genes were located in three distinct gene clusters: the *dgoRKADT* operon, the *uxuAB-uxaC* (SEN2978-SEN2980) operon and the SEN1432-6 locus.

3.1.1 Aim of this chapter

This chapter aims to confirm induction of these genes by EW. The first objective was to generate a series of transcriptional reporter constructs for each of the genes of interest (those anticipated to possess proximal promoters).

3.2 Generation of *hex* gene *lacZ* transcriptional fusions.

For this purpose, the transcriptional fusion vector, pRS1274, was selected for incorporation of relevant gene fragments from *SE* PT4. pRS1274 DNA was obtained from lab stocks and used for generation of further plasmid DNA for use in cloning experiments. Plasmid DNA from lab stocks was transformed into chemically competent TOP10 and eight of the Amp^R colonies. Plasmid DNA presence was confirmed (section 2.2.2.1) and by agarose gel electrophoresis. The expected size of the pRS1274 vector is 10,752 bp. The plasmid DNA was analysed further by restriction digestion (*Bam*HI, *Eco*RI and *Nde*I), to confirm identity. The pattern observed matched with the in-silico analysis.

In a previous study, pRS1274 was used to study the *yohD–yohC* intergenic regions in *S*. *enterica* serovar Typhimurium (Kenyon *et al.*, 2007), and so this vector is thus suitable for use in *Salmonella*.

Genomic DNA was extracted from *SE* PT4 using the Thermo Scientific GeneJET Genomic DNA Purification Kit (4.3.2.2) to provide a PCR template for amplification. Results showed that the quantities and purity of DNA were sufficient for amplification by PCR.

Using Vector NTI, primers were designed for amplification of putative promoter regions for the induced genes of interest (*ybhC*, SEN1435, SEN1436, SEN1432, *dgoR*, *dgoT*, SEN2978, SEN2977 and SEN2979) (Table 2.4) from SE using High Fidelity Phusion[®] DNA polymerase (section 2.2.4). Seven intergenic regions were selected, two of which contained putative divergent promoters and so primers were designed to allow cloning in both orientations. The amplified regions are summarised in Fig. 3.1 below and primer locations are indicated in the Appendix. The regions were selected to include a portion of the upstream gene (#100 bp) as well as a similar portion of the gene of interest in an attempt to ensure that the entire promoter region was included.



Figure 3.1: Schematic representation of the organisation of the egg-white induced *hex* genes of *SE* PT4. Regions are as follows: A, *dgo* cluster; B, *uxuAB-uxaC* operon; C, SEN1432-6; D, *ybhC* gene. Genes are shown as green arrows, direction is indicative of polarity. Numbered rectangles indicate amplified regions.

The next step was to PCR amplify the target regions, clone them into pJET2.1, and then to subclone them into pRS1274 to allow studies on pattern of expression to progress. A two step cloning procedure was used to since cloning of the PCR fragments into pJET2.1 is highly efficient and their subsequent subcloning into pRS1274 would be enhanced through the ability to confirm complete double digestion and sticky-end generation. The target sequences were amplified successfully with bands at approximately corresponding to the expected sizes of the target promoter fragments.

Comparing the mobilities of the observed bands with the expected sizes listed below indicates the validity of the PCR: *ybhC* (expected size 446 bp); SEN1435 (expected size 504 bp); SEN1436 (expected size 504 bp); SEN1432 (expected size 420 bp); *dgoR* (expected size 551 bp); *dgoT* (expected size 434 bp); SEN2978 (expected size 557 bp); SEN2977 (expected size 557 bp); and SEN2979 (expected size 339 bp).

Chapter 3

Following PCR, the DNA was purified using Thermo Scientific GeneJETTM PCR purification kits to remove any contaminants/enzymes (section 2.2.7) and agarose gel electrophoresis was performed again to ensure the correct bands were present (Fig. 3.2).



Figure 3.2: Gel electrophoresis of purified PCR products. Lanes 1 and 8 contain GeneRulerTM 1kb ladder (250-10,000 bp). Purified PCR products are as follow: lane 2, *ybhC*(expected size 446 bp); lane 3, SEN1435(expected size 504 bp); lane 4, SEN1436 (expected size 504 bp); lane 5, SEN1432(expected size 420 bp); lane 6, *dgoR*(expected size 551 bp); lane 7, *dgoT*(expected size 434 bp); lane 9, SEN2978(expected size 557 bp); lane 10, SEN2977(expected size 557 bp); lane 11, SEN2979 (expected size 339bp). Electrophoresis was performed on 2% agarose TBE gels at 60 V for 70 min.

Purification of the PCR products decreased the apparent DNA concentration. This was confirmed by Nanodrop analysis (for a typical sample, the concentration dropped from 1060 to 46 ng/µl after purification). However, this step enhances the downstream applications of cloning and digestion.

The Thermo Scientific CloneJET PCR Cloning Kit was used for inserting the purified PCR products into linearised pJET1.2 cloning vector. Ligation was achieved with T4 DNA ligase according to the instructions provided, with the 'Blunt end protocol'. Ligation mixtures were transformed into chemically competent TOP10 cells (section 2.2.1) and transformants selected on LB agar with Amp after overnight growth at 37 °C. pJET1.2/blunt vector has the *eco47IR* lethal gene which is disrupted by ligation of a DNA insert into the cloning site. As a result, only cells with recombinant plasmids are able to propagate.

Two to four colonies for each transformation were then selected for plasmid extraction. The isolated DNA was then analysed by agarose gel electrophoresis (Fig.3.3).



Figure 3.3: Plasmids DNA extraction of potential pJET1.2 clones carrying promoter fusion fragments. Lanes 7, 20: Fermentas GeneRulerTM 1kb ladder. Potential pJET1.2 clones are as follows: lanes 1-4, *ybh*C insert; lanes 5-6 & 8-9, SEN1435 inset; lanes 10-13, SEN1436 insert; lanes 14-15, SEN1432 insert; lanes 16-17, *dgoR* insert; lanes 18-19, *dgoT* insert; lanes 21, 22, SEN2978 insert; lanes 23, 24, pJET1.2 and SEN2977 inserts; lanes 25-26, pJET1.2 and SEN2979 insert. Electrophoresis was as in Fig. 3.9.

The insert-carrying plasmids were expected to have a lower mobility (just above the 2 kb marker) than the re-ligated vector (just below the 2 kb marker). Nearly all of the extracted plasmids had mobilities suggesting the presence of an insert (Fig 3.3). The presence of inserts in the pJET1.2 clones was tested by double digestion with *Eco*RI and *Bam*HI restriction enzymes (section 2.2.8). Digesting should linearise the plasmids to generate two fragments corresponding to the vector (~3 kb) and insert (~0.5 kb). Analysis of the digested products by agarose gel electrophoresis (Fig. 3.4) showed that the PCR products had been cloned in each case. The plasmids thus verified were designated as follows:

pJET-*ybhC* (#1-4); pJET-SEN1435 (#5,6,8,9); pJET-SEN1436 (#10-13); pJET-SEN1432 (#14,15); pJET-*dgoR* (#16,17); pJET-*dgoT* (#18,19); pJET-SEN2978 (#21,22); pJET-SEN2977 (#23,24); pET-SEN2979 (#25,26) (see Table 2.3).



Figure 3.4: Electrophoretic analysis of pJET1.2 clones by double digestion with *Eco*RI and *Bam*HI. Lane 8&23: Fermentas GeneRulerTM 1kb ladder. Lanes 1, 15, 16, 30: undigested pJET1.2 vector (constructed). Digested pJET1.2 vectors are as follow: lanes 2-5, *ybh*C insert; lanes 6, 7, 9, 10, SEN1435 insert; lanes 11-14, SEN1436 insert; lanes 17, 18, SEN1432 insert; lanes 19, 20, *dgoR* insert; lanes 21, 22, *dgoT* insert; lanes 24, 25, SEN2978 insert; lanes 26, 27, SEN2977 insert; lanes 28, 29, SEN2979 insert. Electrophoresis was as in Fig. 3.9.

Two plasmids for each PCR product were then selected for further analysis by nucleotide sequencing. DNA samples were sequenced by Source Bioscience using T7 promoter and pJET_RP2 primers (Table 2.5) in order to confirm the identity and authenticity of the inserts. The resulting data were analysed using BlastX and Vector NTI. Both the forward and reverse primers sequences showed that all plasmids submitted carry the expected inserts and that the nucleotide sequences exhibit a 100% match to the published *S*E sequence (accession no CP008928) (see Appendix 5 for more detail).

One isolate from each sequenced pair was selected for use as a source of DNA for subcloning into pRS1274 as illustrated in Figure 3.5. Plasmid DNA for each of the selected pJET clones was subjected to double digestion with *Bam*HI and *Eco*RI. The inserts thus released were separated from the vector fragment by agarose electrophoresis and then isolated from the gel (section 2.2.9).



Figure 3.5: Cloning steps of potential promoter regions. Red is target regions.

The fragments thus purified were analysed by agarose gel electrophoresis to confirm if the extractions were successful (Fig. 3.6). Results showed pure bands at approximately 446, 504, 420, 551, 434, 557 and 399 bp which correspond to the sizes of the target putative promoter fragments released from pJET-clones.



Figure 3.6: Gel electrophoresis of gel-extracted gene fragments for cloning into pRS1274. Lane 7 is GeneRulerTM 1kb ladder (250-10,000 bp). Gene fragments are as follow: lane 1, *ybhC*; lane 2, SEN1435; lane 3, SEN1436; lane 4, SEN1432; lane 5, *dgoR*; lane 6, *dgoT*; lane 8, SEN2978; lane 9, SEN2977; lane 10, SEN2979. Electrophoresis was as in Fig. 3.9.

Ligation reactions were as described in Methods (section 2.2.10). The *lacZ* transcriptional fusion vector, pRS1274, was first digested with *Bam*HI and *Eco*RI, and was then purified (section 2.2.2). Reactions included 3 μ l of digested PCR product (45 ng) and 2 μ l (100-200ng) of digested vector. The ligation reactions (5 μ l aliquots) were used to transform chemically competent TOP10 (section 2.2.1) and transformants were selected on L-plates containing Amp and Xgal. Lac⁺/Amp^R colonies were thus obtained in all cases. Four to eight such colonies were selected for plasmid DNA isolation (section 2.2.2.1). The extracted plasmids were analysed by agarose gel electrophoresis for the presence of DNA of the expected size (Fig. 3.7).



Figure 3.7: Plasmids DNA extractions of potential pRS1274 clones. (A) Lanes 13, 29: Fermentas GeneRulerTM 1kb ladder. Lanes 14, 28, cut pRS1274 plasmid. Lanes 15, 30, uncut pRS1274. Potential pRS1274 clones are as follow: lanes 1-8, *ybh*C insert; lanes 9-12, 16-19, SEN1435 insert; lanes 20-27, SEN1436 insert. (B) Lanes 13, 29, Fermentas GeneRulerTM 1kb ladder. Lanes 14, 28: cut pRS1274. Lanes 15, 30, uncut pRS1274. Potential pRS1274. Potential pRS1274. Potential pRS1274 clones are as follow: lanes 1-4, SEN1432 insert; lanes 5-8, *dgoR* insert; lane 9-12, *dgoT* insert; lanes 16-19, SEN2978 insert; lanes 20-23, SEN2977 insert; lanes 24-27, SEN2979 insert. Electrophoresis was performed on 0.7% agarose TBE gels at 60 V for 70 min.

As can be observed from Fig. 3.7, most of the extracts contained plasmid of the expected size while there are number of extracts that contained a smaller plasmid (circled) likely corresponding to pJET1.2 which might remain after extraction of fragments from the gel. This

interpretation was confirmed by restriction digestion (Fig. 3.15) and so these isolates were not used for further work. The plasmids were further analysed by double digestion with *Bam*HI and *Eco*RI (2.2.8) and then the uncut and cut DNA was electrophoresed (Fig. 3.8).

The results showed that the target fragments were released from pRS1274 vector, in all cases, successfully with bands at approximately 446, 504, 420, 551, 434, 557 and 399 bp which correspond to the sizes of the target promoter fragments. The plasmids carrying the expected insert were designated as follows: pRS-*ybhC*-*lacZ* (#1-8); pRS-SEN1435-*lacZ* (#9-12 and 16-19); pRS-SEN1436-*lacZ* (#20-27); pRS-SEN1432-*lacZ* (#1-4); pRS-*dgoR*-*lacZ* (#5-8); pRS-*dgoT*-*lacZ* (#9-12); pRS-SEN2978-*lacZ* (#16-19); pRS-SEN2977-*lacZ* (#20-23); pRS-SEN2979-*lacZ* (#24-27) (see Table 2.3).



Figure 3.8: Potential pRS1274 clones analysed by double digestion with *Eco*RI and *Bam*HI. (A) Lanes 13, 28: Fermentas GeneRulerTM 1kb ladder. Lanes 14, 29, cut pRS1274. Lanes 15, 30, uncut pRS1274. Potential pRS1274 clones are as follow: lanes 1-8: *ybhC* insert; lanes 9-12, 16-19, SEN1435 insert; lanes 20-27, SEN1436 insert. (B) Lanes 14 & 22, Fermentas GeneRulerTM 1kb ladder. Lane 29, cut pRS1274. Lanes 15, 30, uncut pRS1274. Potential pRS1274 clones are as follow: lanes 1-2, SEN1432 insert; lanes 3-4, *dgoR* insert; lanes 5-6, *dgoT* insert; lanes 7-8, SEN2978 insert; lanes 9-10, SEN2977 insert; lanes 19, 21, 28, SEN2979 insert. Electrophoresis was performed on 2% agarose gel and at 60 V for 70 min.

For each plasmid type, as confirmed above, DNA for one isolate was submitted for sequencing of the fragment inserted at the multiple cloning site. Specific primers (pRS1274-FOR and

pRS1274-REV, Table 2.5) were used and sequence contained as identical with accession no CP008928.

3.3 Identification of promoter sites

All the constructs that exhibited expression levels above the vector control were analysed using the promoter finder BPROM program (http://www.softberry.com/berry.phtml) to recognize promoters for these genes by determining the -35/-10 sites. For SEN1436 (Figure 3.21), BPROM indicated that the -35 and -10 sites are located over 153 bp upstream of the SEN1436 start codon (as defined by the NCBI database). The -10 site consists of a sequence similar to the TATAAT motif (TAAATT) at 4/6 points (upper case), whereas the -35 site matches the consensus sequence TTGACA (TTGAAT) at 4/6 positions (upper case). The -35 and -10 sites are separated by 19 bp which is close to the ideal spacing of 17 bp. This supports the hypothesis that this is a functional promoter

For SEN2977 (Figure 3.9), BPROM indicates that the -35 and -10 sites are located over 244 bp upstream of the SEN2977 start codon as defined by the NCBI database. The -10 site consists of a sequence similar to the TATAAT motif (TATCAT) at 5/6 points (upper case), whereas the - 35 site matches the consensus sequence TTGACA (TTGGCT) at 4/6 positions (upper case). The -35 and -10 sites are separated by 13 bp, which is clearly an unacceptable distance. The BPROM software showed the presence of a predicted OmpR-binding site (AAATCACA). OmpR is required for the transcriptional expression of both major outer membrane protein genes, *ompF* and *ompC*, in response to osmolarity (Lijestroem *et al*, 1988)

For *dgoT* and *dgoR* (Figure 3.21), BPROM indicated that the -10 sites are located 54 and 123 bp upstream of the *dgoT* and *dgoR* start codons, respectively. For *dgoT*, the -10 site consists of a sequence identical to the TATAAT motif (100%), whereas the -35 site matches the consensus sequence TTGACA (TTTAAA) at 4/6 positions (upper case). The -35 and -10 sites are separated by the ideal spacing of 17 bp. In addition, BPROM analysis showed that the

dgoA-dgoT intergenic region contains a potential CpxR binding site; the CpxAR system responds to envelope and pH stress by activating expression of genes including *cpxP*, *degP*, *dsbA* and *ppiA* (Danese *et al.*, 1995). For *dgoR*, analysis showed a -10 site consisting of a sequence similar to the TATAAT motif (CATAAT) at 5/6 points (upper case)), whereas the - 35 site matches the consensus sequence TTGACA (TTGTGA) at 4/6 positions (upper case). The -35 and -10 sites are separated by the ideal spacing of 17 bp. For SEN1435 and SEN2979, no promoters were predicted by BPROM, and predictions for SEN2978 and *ybhC* were a poor match to the ideal promoter. For SEN1432, indicated that the -10 sites are located over 35 bp upstream of the start codon. The -10 site consists of a sequence similar to the TATAAT motif (TATAAT) at 6/6 points (upper case), whereas the -35 site matches the consensus sequence TTGACA (TTGTTC) at 3/6 positions (upper case). The -35 and -10 sites are separated by 17 bp which is ideal spacing. This supports the hypothesis that this is a functional promoter.

SEN1436

	SEN1435
GGATCCCAAAGCCCAGTCCTCGTGCAGAACCCGTTACCAGCGCCGTTT	ICCCAGTTAAATCAAATAAAGCGGT <mark>CAT</mark> GTTGTTTCCTCACTTGTTT
AATTTGTATGACGACTATCCTTTTTTAGGTTGAATTTTCGCCCTGATA Potential -10 site (55)	AAATCAACAGTTCACCCATGAATTTGCAACAAGGATCACAAACAGCT Potential -10 site (56)
CCACATGCCGACCGCGTAATTAATATTAATTAATTAATTA	ATATTTGGTT <mark>TAAATT</mark> TAACGCAGTTTGATCGCTGTCACAGAATGGCA ♪p>
CTCGCAGCGATCCGCTGTAAAAGAAGCGTGATATAACAGCATAAAGTT SEN1436	GTAGGACAACTTACGTATATCTGTTGTATCATCCACAACGGTATGAC
ATGCGGTAAATTCGCTGAGTTAAGGA <mark>GTG</mark> AAAGTGAGTAACCTGAAAA	FTACCAACGTGAAAACGATTCTGACGGCGCCGGGCGGCATTGATTTG
GCAGTCGTTAAGATAGAAACCAACGAGCCAGGGC <mark>GAATTC</mark>	

SEN1432

ACGGCAGTTGGGCGCCGGGTAGCCCTTTGCCCTCAGAAGTGGAACT<mark>GAATTC</mark>

SEN2977

GGATCCGATATGGTGTAACGCGTTACCACGCCGGTTGCGCCAGCCTGGCGTACATCTGA	CAGCGTTACCGGGTCGTTAGGTCCGTACCAGCGCCA
GGTTTGTTTCATATCTCGTTTCCTCTTCTTGCGATAACGTCTTCGTGGTTGACC <mark>CAT</mark> TG Potential -35 site (33)	CCAGCCAACATCGAAACGTGCTTTGTAAACCCGTTC Potential -10 site (45)
TGACCCCTAAATTCAACCAAAATTTTTCTCATGTCAACCTTATTGTCTAAATTGGCTAA	CC <u>AAATCACA</u> AA <mark>TATCAT</mark> CATTCACGGTCTGCCAAT
-	ompR TF 13bp
TTTATTTATTTGATCTGTGTCAATTTTTGCTGGGTGAAAAGCATTCACCATTCAACTTG	AAATGAGTTGATGTATTTATTTCAAGAATATTAAGG
GCGGGAGTTGCCGCCAGATTTTGACCGGTCCGGATGAGAAAATATTGATTG	ATTTTTGTGATTTCAGTTTTCCCGCTACAGGTCAGA
CGGCGCGGAGCTAATGTTTTTTAACGAGGCTTTATCATGAAGATGACAAAATTAAGATG	GTGGATTATCGGCCTGGTCTGCGTAGGG <mark>GAATTC</mark>

SEN2978

dgoT

<mark>GGATCC</mark>ACTATAACAAGGGCGCGGAGCTGCTCGACTTTGTGAAAAACAAAGAAGACTTCAGCATGGACGGCGGCTTCTTTAAACCCTTAACCAAA

dgoR

GGATCC GAGGTGATGGCGATTGGCGATCAGGAAAACGACATTGCGATGATCGAATACGCCGGTATGGGCGTGGCAATGGACAACGCCATTCCGTC GGTCAAAGAGGTGGCTAACTTTGTGACTAAATCGAACCTTGAAGATGGTGTTGCCTGGGCGATTGAAAAATTTGTGCTGAACCCCGATCACTCAT yidA terminator CCGGCCATTTCCCCGCCCGATAAGGCATAGCCGCCATCGGGCAAATACGCGCCTTAACGACCCCGCACTTGCTGCGGGGTTTTTTTATGTCTTTCGTT Potential -35 site (30) Potential -10 site (38) TACGTCTTATAACGTTCCCATAACCAATTGTTGTTTTTGTGATCTAAATTGTAGTACAACATAATTATGTTGTACTACATTAATGGCATGATAAC 4 - 17 bp - 4GACGGTTGATATCACGCTAGTACTACAAAAATTGCGGCGTAATTCAGCTATCGCGGTAAAGTAAGAGAGGTTCACATCGAGGACAAAGGACTCTCTAT dgoR - 4

<mark>G</mark>ACTCTCAATAAAACCGATCGCATCGTTATCACGCTGGGCAAACAGATTGTCAGCGGTAAATACGTACCCGGTTCGGCGCTG<mark>GAATTC</mark>
ybhC



Figure 3.9: Analysis of the potential promoter regions using BPROM. Predicted -35 and -10 sites are highlighted in grey. The start codons are highlighted in green. Restriction sites generated for the lacZ fusion construction are highlighted in yellow. The predicted -10 and -35 sites are in grey, with spacing indicated. BPROM scores indicated in brackets. The predicted transcription factors labelled underlined (cpxAR in *dgoT*, ompR in SEN2977).

Table 3.1: Summary of predicted promoters.

Genes	TATAAT -10	TTGACA -35	Distance between -10 and -35
SEN1436	TAAATT	TTGAAT	19
SEN1432	TAAAAT	TTAACG	17
SEN2977	TATCAT	TTGGCT	13
SEN2978	TAAAAT	TTGACA	13
dgoT	TGTAAT	TTTAAA	17
dgoR	CATAAT	TTGTGA	17
ybhc	TATTTT	TTGCCT	21
SEN1435	No prediction		
SEN2979	No prediction		

3.4 Preliminary expression analysis of the *hex* genes.

Initially, the expression activity of all nine constructs was tested using *E. coli* as the host. Expression was determined by measurement of the β -galactosidase activity of corresponding transformants. The β -galactosidase assays (section 2.2.12) allow the activity of potential promoter regions fused with the *lacZ* gene in the promoter-less transcriptional plasmid (pRS1274) to be monitored. If the promoter is active, the *lacZ* gene is transcribed and the cell produces β -galactosidase (which normally cleaves lactose). In this assay, colourless ONPG was used instead of lactose as the substrate. β -galactosidase hydrolyzes ONPG to produces a colour change due to release of ONP (yellow) at a rate that correlates with the amount of the enzyme expressed and activity of the fused promoter.

3.4.1 Activity of *lacZ* fusions in *E.coli* TOP10

All nine constructs (pRS-*ybhC*-*lacZ*; pRS-SEN1435-*lacZ*; pRS-SEN1436-*lacZ*; pRS-SEN1432-*lacZ*; pRS-*dgoR*-*lacZ*; pRS-*dgoT*-*lacZ*; pRS-SEN2978-*lacZ*; pRS-SEN2977-*lacZ* and pRS-SEN2979-*lacZ*) were selected for expression analysis. Overnight cultures of TOP10 transformants were used to inoculate LB containing Amp, and these cultures were then grown to stationary phase and samples taken at regular intervals for β -galactosidase activity assay (Fig. 3.10).

The expression data are shown in Fig. 3.11 summarised in Fig. 3.10, and the full expression data with error bars are provided in Appendix 6. Seven of the fusions gave expression levels greater than the vector-only control (~20-700 times higher), suggesting active promoters. The activity was converted to the following units: nmol ONPG/min/OD unit cells (according the equation as described in section 2.2.12). Two (SEN1436-*lacZ* and SEN2977-*lacZ*) gave particularly high log-phase activity (maximum of 1250 and 1360 U, respectively); one (SEN1432-*lacZ*) gave moderate activity (maximum of 740 U); and four (pRS-SEN1435-*lacZ*; pRS-*dgoR-lacZ* and pRS-*ybhC-lacZ*) gave relatively weak log-phase

activities (maximum of 200, 350, 190 and 180 U respectively). Two (*dgoT-lacZ* and SEN2979*lacZ*) gave very weak activity that was only slightly above that of the vector control (~2- 15 U) suggesting that no promoter is present directly upstream of *dgoT*. Thus, *dgoT* may be expressed from the *dgoR* promoter and is likely to comprise part of an operon: *dgoRKADT*. SEN2979 encodes D-mannonate oxidoreductase and as its gene fusion is weakly expressed; this suggests that it also does not have an independent promoter so would depend upon any distal promoter associated with SEN2978 for its expression, indicating that the SEN2978 promoter may be required for expression of downstream genes (SEN2979-80).



Figure 3. 10: Expression of *hex* **genes in L-broth using** *E. coli* **TOP10 as host**. Growth was aerobic in LB with Amp at 37 °C and 250 rpm. Data given is the average from triplicate cultures each assayed in triplicate. Error bars are indicated in the Appendix as differences between the triplicate cultures. All experiments were repeated once or twice, with similar results obtained. Dashed line is growth rate; solid line is β -gal activity.

The greatest β -galactosidase activity (~500 and 700-fold greater than the control; Fig. 3.11) was observed for SEN1436- and SEN2977-*lacZ*, respectively, in the exponential growth phase. Both exhibited comparatively strong maximum expression levels of ~1250 and 1360 U respectively. SEN1436 encodes a putative hexonate dehydratase and according to microarray results (Baron *et al.*, 2017) was induced by 33-fold in egg white medium (45 min). The expression declined dramatically (~12 fold) towards the end of the exponential growth phase (6 h). Using different media should be considered as Baron *et al.* (2017) used EW media while here LB is used.



Figure 3.11 Summary of β -galactosidase activity (average and maximum levels) of all nine *lacZ* fusions. The data were taken from Fig. 3.10. Averages and maximum taken for activities at 1-8 h growth. Asterisks indicate significant difference with respect to vector control (P ≤ 0.05).

SEN2977 encodes a putative hexonate transporter (Thomson *et al.*, 2008) and was not reported as being induced by exposure to EW medium (Baron *et al.*, 2017). SEN2977-*lacZ* activity was greatest at 6 h, towards the end of the exponential -growth phase, and declined rapidly (fourfold) at 8 h in the early stationary phase.

Moderate activity was observed for SEN1432-*lacZ* of ~740 U in the exponential growth phase. SEN1432 specifies a putative GntR-family regulatory protein which could control the SEN1433-6 genes through interaction with divergent putative promoters at the SEN1436-35 intergenic region. The *lacZ* fusion data suggest that SEN1432 is well expressed and thus has an independent proximal promoter such that its expression would not depend upon any promoter associated with SEN1435. The array data (Baron *et al.*, 2017) showed no evidence of induction of SEN1432 in egg white which suggests it is constitutive. This would match its role as a regulator.

SEN1435-lacZ was one of four fusions showing relatively weak expression levels. It had highest activity at 0-3 h (200-60 U), but activity declined to ~5 U by 4 h and remained relatively low from then on. This relatively weak expression indicates that SEN1435-lacZ is repressed under the conditions employed or has a weak promoter. Note that SEN1435 appears to be the first gene of a three gene operon (SEN1435-33) and all were induced in egg white (5-7 fold). SEN2978-lacZ was also quite weakly expressed (~350 U max during exponential growth). Its expression was very similar to that of SEN1436 with two peaks in expression (2 and 5 h) followed by very low expression levels. SEN2978 is the proximal gene in the SEN2978-80 (potential) operon. SEN2978 was 28-fold induced in egg white (Baron et al., 2017) and its relatively weak expression in the results relayed here suggests that it is repressed under the conditions employed in this work or its promoter is weak. The *dgoR-lacZ* fusion was expressed at relatively low but consistent levels during the exponential growth phase (~180-190 U). ybhC-lacZ was also weakly but consistently expressed (84-170 U) during the exponential growth phase. dgoR is the first gene in the apparent dgoRKAT operon. The weak *dgoR-lacZ* expression suggests that the corresponding operon is repressed in LB (in E. coli), or that its promoter is weak, although it is subject to strong induction (14-31 fold) in egg white (in SE PT4). Likewise, the weak expression of the ybhC fusion suggests that it may also be

repressed or weakly transcribed under the conditions employed here, although is induced by up to 6 fold in egg white (Baron *et al.*, 2017).

Note that the growth patterns were similar although there was a difference in maximum OD achieved (between 4-5.2 OD units) and the growth curve for the SEN2977 fusion strain was clearly delayed with respect to the others. The reasons for these relatively modest effects are not clear.

3.4.2 Activity of *lacZ* fusions in *S*. Enteritidis PT4

All nine *lacZ* fusion plasmids were transformed into *S*E by electroporation (section 2.2.13 & 2.2.14). The identity of the resulting transformants was confirmed by re-isolation of the plasmids from the electro-transformants and double digested with *Eco*RI and *Bam*HI (Fig. 3.12). Three of the resulting *S*E transformants were selected on the basis of expression level (in *E. coli*, Fig. 3.11), one from each of the three hexonate utilisation gene clusters, for further study. The three fusions employed were SEN1436- (putative dehydratase), SEN2977- (hexuronate transporter) and *dgoR-lacZ* (repressor). These three were selected as representative genes from the three gene cluster; in addition, SEN1436 and SEN2977 showed the highest activity in previous the experiments.

Chapter 3



Figure 3.12: Confirmation of transformants of SE by plasmid isolation followed by double digestion with *Eco*RI and *Bam*HI. Lanes 11, 20: Fermentas GeneRulerTM 1kb ladder. Potential pRS1274 clones are as follow: lanes 1-2, *ybhC* insert; lanes 3-4, SEN1435 insert; lanes 5-6, SEN1436 insert; lanes 7-8, SEN1432 insert; lanes 9-10, *dgoR* insert; lanes 12-13, *dgoT* insert; lanes 14-15, SEN2978 insert; lanes 16-17, SEN2977 insert; lanes 18-19, SEN2979 insert. Electrophoresis was performed on 2% agarose gel and at 60 V for 70 min. Fragments were released from pRS1274 vector successfully with bands at approximately 446, 504, 420, 551, 434, 557 and 399 bp which correspond to the sizes of the target promoter fragments.

As shown in Fig. 3.13, SEN1436-*lacZ* exhibited highest activity with a maximum (~2200 U) achieved at the mid to late exponential growth stage. Expression increased ~fourfold from early to mid-log phase and reduced ~50% by 6-8 h. In contrast, SEN2977- and *dgoR-lacZ* expression was much lower (by ~sevenfold) with a maximum level of ~350 U, but expression was relatively consistent during growth (170-290 U). As shown in Fig 3.13, SEN1436-*lacZ* expression was ~2.5-fold higher than achieved in *E.coli* TOP10 (Fig. 3.11). Maximum expression of *dgoR-lacZ* was ~1.7 fold higher than seen in *E. coli*, whereas SEN2977-*lacZ* maximum expression was ~3.5 fold lower than in *E. coli* (Fig. 3.14). The other six fusions were tested in *SE* (data not shown) and showed very similar activity to that seen in *E.coli*. These expression differences are relatively minor and are likely to relate to the different genetic backgrounds and experimental conditions (i.e. use of B-PER in place of Bugbuster).



Figure 3.13: β -galactosidase activity of three pRS1274 *lacZ*-fusion constructs in SE (SEN1436, SEN2977 and *dgoR*). Growth was aerobic in LB (containing ampicillin) at 37 °C and 250 rpm. Dashed line;growth rate, solid line β -gal activity. Each growth was in duplicate and each culture was assayed in triplicate.



Figure 3.14: Comparison of β -galactosidase activity (average and maximum levels) of three pRS1274 constructs in SE and *E. coli* TOP10. Average values are shown for activities at 1-8 h growth. The data were taken from Fig. 3.16 and 3.19. Fold difference with respect to *E. coli* are shown with asterisks indicating significance.

3.5 Conclusions and Discussion

Previous work showed that three distinct gene clusters (the *dgoRKADT* operon, the *uxuAB-uxaC* operon and the SEN1432-6 genes) are strongly induced upon SE exposure to EW (Baron *et al.*, 2017. These three clusters are involved in utilisation of hexonates/hexuronates but have not previously been reported to possess any role in EW survival or to be up-regulated by EW and subject to co-regulation. Therefore, this chapter aimed to determine the patterns of transcriptional regulation of the genes of interest using transcriptional fusions. The first objective was to clone nine putative promoter regions of the genes of interest upstream of a promoter-less *lacZ* reporter gene, to allow promoter activity to be monitored by β -galactosidase assays. These regions were inserted into the pRS1274 *lacZYA* transcriptional fusion be used

as an alternative to the *lacZ* fusions, or to confirm results obtained with such fusions. Use of *lacZ* fusions offers the advantage of a simplistic enzyme assay for expression analysis and stability of the *lacZ* gene product. In contract, mRNA is highly unstable and assay by RT-PCR requires comparison with a second message (that may not be stable/constitutive) as well as a technically challenging two step amplification process.

The transcriptional fusion data in *E. coli* TOP10 indicated that eight of the fusions had activity markedly above that of the vector control, but one (*dgoT-lacZ*) had weak activity only slightly higher than the vector suggesting no promoter is present, although the in silico analysis showed a strongly predicted promoter.

The eight active fusions were divided into three groups on the basis of the relative expression levels during the exponential growth phase in LB. All showed peak activities during the exponential growth phase with reduced activity in the stationary phase (Fig. 3.10). SEN1436-*lacZ* and SEN2977-*lacZ* exhibited high activity. Both are divergently arranged with respect to adjacent operons (SEN2978-80 and SEN1435-33) but only SEN1436 was shown to be egg-white induced (33-fold); SEN2977 showed no induction in egg white (Baron *et al.*, 2017) which is a surprise and should be confirmed in future work within this thesis.

The SEN1432-*lacZ* fusion (encoding a putative transcriptional regulator) was moderately expressed. This indicates that SEN1432 has an independent proximal promoter despite its location at the end of the SEN1435-33 operon and its co-polarity. Thus, SEN1432 specifies a regulatory protein which could control the SEN1435 and SEN1435-33 genes through interaction with divergent putative promoters at the SEN1436-35 intergenic region. As SEN1432 was not reported to be induced by EW (Baron *et al.*, 2017), this suggests it is constitutive and might be involved in controlling genes related to hexonate catabolism.

Four fusions (SEN1435-, SEN2978-, *dgoR*- and *ybhC-lacZ*) gave relatively weak log-phase activities (maximum of 180-350 U). These results suggest that these genes are repressed under

the conditions employed in this work. Three of these fusions represent the first gene in apparent operons (SEN1435-33, SEN2978-80 and *dgoRK*-SEN3645-*dgoDT*) and so the data suggest that the corresponding operons are repressed in LB (in *E. coli*). However, all three operons were subject to strong induction (14-31 fold) in egg white (in *SE* PT4) which would be consistent with the suggested repression in L broth.

Little work has been performed on the role of hexonate utilisation in survival and colonisation of SE. Coward *et al.* (2012) investigated the role of a hexonate uptake and catabolism SE genomic island locus (SEN1432–SEN1436) in colonization of the chicken reproductive tract and other organs following oral challenge. The deletion of these loci did result in a decrease in bacterial load in the spleen by 14 days post infection suggesting a minor role in systemic colonization.

Comparison of the *S*. Enteritidis PT4 and *S*. Typhimurium LT2 genomes (Thomson et al., 2008) showed a PT4 specific region ('ROD13') corresponding to the SEN1432–SEN1436 (6 kb) locus encoding one of the three hexonate-utilisation loci induced by egg white (section 2.1). Although absent in the LT2 strain, this locus is present in the chicken pathogen, *S*. Gallinarum as well as PT4. The reason for the absence of this locus in LT2 is unclear. However, the SEN1432–36 genes show sequence similarity as well to the genes of the *gntII* locus of *E. coli*; these are involved in L-idonate catabolism (Bausch *et al.* 1998) suggesting a similar function for the SEN1432-36 genes.

Two fusions (*dgoT*- and SEN2979-*lacZ*) gave very weak activity that was only slightly above that of the vector control (Fig. 3.15) suggesting that no promoter is active directly upstream of these two genes. However, both showed induction by egg white in the array data (Baron *et al.*, 2017) and the upstream promoters (associated with *dgoR* and SEN2978, respectively) showed much higher activity, using different media could be contributed in different expression. This

indicates that these two genes are transcribed from upstream distal promoters as part of operons and they do not possess independent promoters.

Another study showed that several genes are up regulated (2.5-3.5 fold) in operons involved in the transport and metabolism of D-galactonate (*dgo*), D-gluconate (*gntU*, *kdgT*, and *kduD*), and L-idonate (*idn*) genes in *S*E that are indicative of its metabolism in macerated leaf tissue in cilantro and lettuce soft rot lesions (Goudeau *et al.*, 2013). However, the precise environmental factor inducing their expression is unclear. Interestingly, genes involved in the utilisation of gluconate and related hexonates (*gntT*, STM3134, *dgoT*, *dgoK* and *dgoA*) were up-regulated in *S*. *Typhimurium* upon macrophage colonisation. The reason for this is unclear but one suggestion was that hexonates may be an important source of carbon for intracellular bacteria (Eriksson *et al.* 2003).

The expression of three fusions (SEN1436-, SEN2977- and *dgoR-lacZ*) was also monitored in *S*E and activity levels were similar to those seen in *E. coli*. Thus, expression experiments can now be performed in *S*E using representative fusions and environmental factors relevant to those associated with egg white exposure.



Results and discussion

Chapter 4. Egg-white factors influencing expression of SE hex genes

4.1 Introduction

4.1.1 Effect of EW on *hex* gene expression in SE

The egg white (EW) exposure experiments of Baron *et al.* (2017) employed an 'EW model medium' (EWMM) composed of EW filtrate with 10% EW protein; this is a medium that closely mimics EW. The SE control against which expression changes were identified was grown overnight in TSB (a rich medium) at 37 °C, and the cells were then washed and resuspended in EWMM at ambient temperature to give the 0 time point. Then, the SE cells in EWMM were incubated at 45 °C (to mimic egg incubation temperature and hen body temperature) and samples were taken at 7, 25 and 45 min for analysis of effects of EW exposure on the transcriptome with respect to the zero time point. Incubation at 45 °C in EWMM causes a gradual killing effect for SE (and is entirely growth inhibitory) over a 24 h period, and so the 45 min incubation corresponds to the early phase of EWMM-induced cell damage/death and is thus a condition under which SE would be expected to suffer considerable stress. SE was not killed by TSB at 45 °C and was only growth-inhibited in EWMM at lower temperature. Thus, it is the combined effects of temperature and EW exposure that causes the loss of viability in EWMM at 45 °C (Baron *et al.*, 2017).

In general, upon EWMM exposure at 45 °C, the *hex* genes were weakly induced at 7 min, and strongly induced at 25 and 45 min, with little change between 25 and 45 min. Of particular interest is the observation that four of the *hex* genes (*dgoK*, *dgoR*, SEN1436, SEN2978) were more strongly induced in EWMM than any of the other ~320 EW-induced genes. Thus, the *hex* gene response to EW was greater than for any other gene, suggesting that the effect observed is of considerable physiological significance for EW exposure.

Although it is clear that the three induced *hex* gene clusters are subject to major up regulation during exposure of *S*E to EW, it remains unclear why these genes are induced, what factors in EW are responsible for their induction and what transcription factor(s) might mediate their induction in EW. In total, there were 15 genes associated with hexonate and hexuronate (Hex) metabolism genes (including *eda* and *yiaE* that are unassociated with the *dgoRKADT*, *uxuAB-uxaC* and SEN1433-6 clusters) that showed significant overall induction in EW, by up to 33 fold according to the microarray data (and by 240 fold for *dgoK* by RT-PCR; Baron *et al.*, 2017). However, previous reports had not identified any roles for these genes in the survival of *S*E in EW or shown up-regulation by EW exposure (Baron *et al.*, 2017).

Several EW-related environmental factors were interpreted as exerting a major regulatory influence on the expression profile of *S*E in EWMM (Baron *et al.*, 2017). These factors were mainly iron deficiency (mediated by Fur, RfrA and RfrB), envelope disruption (mediated by CpxAR, RpoE and PspF), high pH (mediated by CpxAR), and temperature (mediated by RpoH). The possibility that one or more of these factors might be responsible for the *hex* gene induction observed cannot be ignored, although such an effect would be novel.

4.1.2 Energy/carbon sources in EW and changes in energy metabolism upon exposure to EW

According to Guérin-Dubiard *et al.* (2010), EW contains glucose (98% of total sugar; 0.4–0.5% w/v) as the main carbohydrate, in addition to lower levels of other sugars (mannose, galactose, arabinose, xylose, ribose and deoxyribose). A key point of note is that hexonates and hexuronates are not considered to be present within EW (Guérin-Dubiard *et al.*, 2010). Therefore, the reason for the *hex* gene induction in EW and the factor stimulating their expression are obscure. It appears likely that the *hex* genes are subject to EW induction in response to some factor other than

a hexonate/hexuronate (Hex). In addition to the *hex* genes, exposure of *S*E to EW triggered a general change in expression of other carbohydrate metabolism genes: the pentose phosphate and glycolysis pathways were induced, and the TCA cycle was repressed (Baron *et al.*, 2017). Thus, Hex metabolism was not the only catabolic pathway to be affected by EW exposure.

Heterotrophic bacteria generate energy through catabolic processes and often employ respiratory pathways to yield energy from the disposal of the reducing equivalents thus liberated. However, bacteria may switch energy metabolism away from respiration and toward fermentation (where energy generation generally involves substrate-level phosphorylation) when suitable electron acceptors are not available (Peter and Jr, 1992). When *SE* was exposed to EWMM, the upregulation of glycolysis along with the down-regulation of both the TCA cycle and respiration were considered indicative of a switch from respiratory to fermentative metabolism (Baron *et al.*, 2017). The activation of the acetate kinase (*ackA*) and ethanol dehydrogenase (*adhP*) genes encoding mixed-acid fermentation enzymes further supported this suggestion. The reason for such a shift in energy metabolism was unclear although this observation does raise the possibility of a link with the observed induction of *hex* gene expression.

4.1. Aims of this chapter

In this chapter, the *lacZ* fusions created in chapter 3 were used to investigate the effect of various relevant environmental factors on the induction of the *hex* genes in *S*E. In particular, the original findings by Baron *et al.* (2017) regarding *hex* genes were re-investigated with the *lacZ* fusions to confirm the proposed induction of the *hex* genes upon EW exposure.

4.2 Can SE grow on the four available Hex compounds (gluconate, D-mannono-1,4-lactone, gulonate γ-lactone and/or D-galactonate)?

4.2.1 Growth on glucose and glycerol at 37 and 42 °C

Initially, the ability of *SE* to grow on the four available Hex compounds was tested. However, before using the Hex compounds, growth of *SE* at mammal and hen body temperatures (37 and 42 °C; Raspoet *et al.*, 2014; Baron *et al.*, 2017) with standard carbon sources was performed in M9 minimal medium. Glucose was selected as it is present in EW at 0.4-0.5%, and glycerol was used as an example of a non-fermentable carbon source that does not induce catabolite repression. Growth on these carbon sources would then be used to compare with growth on the hexonates. Note that no hexuronates were available. A range of glucose concentrations was employed to show a quantitative effect on growth, above and below the levels found in EW. All growths were performed in triplicate and all measurements produced here are the average of three treatments. Each experiment was performed twice with one representative batch of data presented.

Growth was monitored using a Bioscreen plate reader with up to two 100-well Honeycomb plates (see 2.2.16). Precultures were prepared overnight in 3 ml of 0.4% glucose M9 medium in sterile test tubes at 37 °C and 250 rpm and were used to provide a starting a OD of 0.01 in fresh medium. Then, aliquots of 300 μ L were dispensed into wells in triplicate in a Honeycomb plate. The negative control was un-inoculated medium.

SE growth was tested in ranges of glucose between 0.1% and 1.6% in the M9 minimal medium in aerobic condition at 37/42 °C and compared with 0.4% glycerol as control as shown in Fig. 4.1.



Figure 4.1. SE growth in 0.1-1.6% glucose, or 0.4% glycerol, in M9 minimal medium at 37 (A) and 42 °C (B). Growth was aerobic with continuous shaking. Data is the average of three replicates with errors bars indicating standard deviation.

As shown in Fig. 4.1a, the SE wild-type strain grew well on both glucose and glycerol. Increasing glucose caused, in general, an increase in the rate of growth and the final density achieved. Glucose and glycerol at 0.4% gave similar final densities, but the rate of growth with glycerol (growth rate ~0.45) was lower than that with glucose (growth rate ~0.55), with an ~1 h difference in time taken to achieve the same OD during log phase at 0.4% glucose/glycerol. Temperature had little notable impact on growth, except with 1.6% glucose were the final density at 42 °C (growth rate ~1.1) was impaired with respect to that at 37 °C (Fig. 4.1; growth rate ~1.5). Previous work has reported that the optimal growth temperature of most *Salmonella* serotypes is 35-37 °C, but that sub-optimal growth is achieved at 5 to 47 °C (Pui *et al.*, 2011). These reason for the major effect of temperature on growth with 1.6% glucose (e.g. OD 0.7 versus 1 at 14 h; P= 0.48) is

unclear, but might be related to higher concentrations of metabolic end products (organic acids) generated with increased glucose availability which might exert an enhanced growth inhibition effect at higher temperature, as indicted by Charalampopoulos *et al.* (2002).

In summary, these data indicate that *SE* P14 grows well at 42 °C with glycerol or glucose at 0.4%, and thus either of these can be used as positive controls for growth tests with the Hex compounds at this temperature.

4.2.2 Growth on hexonates at 37 and 42 °C

The ability of the available hexonates (D-galactonic acid; D-mannono-1,4-Lactone; L-(+)-gulonic acid γ -lactone and gluconate), at 0.1-1.6% w/v, to support growth of *SE* was tested, with glycerol acting as the control. The results show that *SE* grows well at both 37 and 42 °C on all four of the Hex compounds tested (Figs 4.2-4.8). Thus, these hexonates can act as sole carbon and energy source for *SE* growth at hen body temperature. This is consistent with previous reports for *E. coli* (showing growth on gluconate, gulonate, glucuronate, D-galactonate and D-2-oxo-3-deoxygalactonate, galacturonate, fructuronate) and for *Salmonella* (reporting growth on gluconate, normal galactonate and gulonate) (Deacon & Copper, 1977; Eisenberg & Dobrogosz, 1967; Nemoz, *et al.*, 1976; Robert-Baudouy *et al.*, 1974; Cooper, 1978; 1980).



Figure 4.2. SE growth in 0.1-1.6% gluconate in M9 minimal medium at 37 (A) and 42 °C (B). Growth was aerobic with continuous shaking. Data is the average of three replicates with errors bars indicating standard deviation.

Enhanced growth with increasing gluconate concentration was clear at 37 °C (growth rate from ~0.6 to 1.1), but at 42 °C such an effect was less apparent indicating that raising concentration from 0.1 to 1.6% has little impact on growth at 42 °C (Fig. 4.2). Growth with gluconate was superior (growth rate from ~0.5 to 1) to that achieved with the same concentration of glycerol, with a ~2.x fold higher culture density observed, indicating that gluconate is a good carbon source for SE growth. In *E. coli*, gluconate is utilised via the GntI and GntII pathways, with GntII being a heat labile subsidiary system (Gómez *et al.*, 2011) SE carries the *gntK* (gluconokinase; SEN3365), *gntU* (low-affinity gluconate transport; SEN3364) and *gntR* (gluconate utilization operon repressor; SEN3366), *gntT* (high-affinity gluconate transporter; SEN3338) genes of the GntI system, but lacks the GntII system according to the annotated genome sequence (Parkhill *et al.*, 2008). These genes are likely to be subject to catabolite repression (Rodionov *et al.*, 2000)

and induced by gluconate through GntR transcriptional control (see Fig. 4.5). The end products of the GntI pathway feed into the ED pathway (Fig. 4.5).



Figure 4.3. SE growth in 0.1-1.6% L-(+)-gulonic acid γ -lactone in M9 minimal medium at 37 (A) and 42 °C (B). Growth was aerobic with continuous shaking. Data is the average of three replicates.

With L-gulonate (Fig. 4.3), again little overall difference in growth was seen at 37 (growth rate \sim 0.7) and 42 °C (growth rate from \sim 5.2). However, at 37 °C the growth with 0.4% glycerol was similar to that with 0.4% gulonate indicating that growth on gulonate is weaker than with gluconate. Indeed, growth density was weaker with gulonate than with gluconate at all equivalent concentrations employed. A clear enhancement of growth was observed at both temperatures as gulonate levels were increased, except with 1.6% gulonate at 42 °C where a marked reduction in growth (density and rate) was observed. In addition, growth with 0.2% was greater than with 0.4%; and the reason for this is unclear. Interestingly, unlike with glycerol or glucose, there was a

raised growth at 42 cf. 37 °C with 0.1-0.8% gulonate and with 0.1-0.4% gluconate, suggesting that growth on these substrates at the indicated concentrations might better support growth at higher temperature and thus raise the optimal growth temperature. It is unclear how such an effect would be exhibited.

It is unclear by which pathway gulonate would be degraded in *S*E, and whether such capacity exists in *E. coli*, but this pathway is likely to involve one or more of the GntI system, the Dgo pathway or the SEN1433-6 pathway (Fig. 4.5). Gluconate is slightly more similar to gulonate than is galactonate, which would suggest that the GntI pathway (also present in *E. coli*) may be responsible for the observed consumption of gulonate. This question could be resolved by producing and studying the effect of mutations in the corresponding genes.

Galactonate gave very good growth for SE at 42 °C (not tested at 37 °C) (Fig. 4.4A: growth rate 1.92), stronger than same concentration of glycerol (growth rate 0.54) any of the other hexonates tested (max growth at 15 h of 1 OD unit for galactonate cf. 0.85, 0.69 0.63. for mannonate, gluconate and gulonate, respectively) or glucose (0.67 OD units at 15 h), indicating that galactonate is a good carbon source for SE. Galactonate is expected to be catabolised via the Dgo pathway (Fig. 4.5) and feed end products into the glycolytic pathway. The Dgo pathway in *E. coli* is subject to catabolite repression and is induced by D-galactonate (Deacon & Cooper, 1977; Cooper, 1978). Growth with galactonate was increased as the concentration was raised from 0.1 to 0.2%, but further increases in concentration had little impact indicating other factors limiting growth (Fig. 4.4A).



Figure 4.4. SE growth in 0.1-1.6% galactonate (A) or mannonate (B) in M9 minimal medium at 42 °C. Growth was aerobic with continuous shaking. Data is the average of three replicates with errors bars indicating standard deviation.

For mannonate, growth was greater at 0.8-1.6% than at 0.1-0.4% (Fig. 4.4B). As for the other hexonates, growth with mannonate (growth rate ~0.53) was superior to that with equivalent levels of glycerol (growth rate ~0.43). The pathway by which mannonate is consumed in *SE* is likely to be that operated by SEN2977-90 (UxuAB/UxaA; Fig. 4.5). However, this possibility remains to be proven in *SE*. The *uxuAB* genes of *E. coli* are subject to multiple regulatory control: glucose (catabolite) repression via CRP; induction by D-galacturonate via ExuR; induction by D-glucuronate via UxuR; and repression by peroxide via OxyR (Robert-Baudouy & Stoeber, 1973; Blanco *et al.*, 1986; Zeng *et al.*, 2001). The *uxaA* gene of *E. coli* is also ExuR and CRP regulated, and in addition is FNR induced anaerobically (Portalier *et al.*, 1980) ExuR is absent in *SE*, but the other four regulators are present indicating that the *uxuAB-uxaA*-SEN2977 genes might be induced in response to glucuronate via UxuR, and also subject to catabolite and peroxide repression, as

well as anaerobically induction. The absence of UxaBC as well as the GlnP and ExuT transporters in *SE* (with respect to *E. coli*) might explain the associated SEN2977 predicted-Hex transporter which is absent in *E. coli* according to no identical found in alignment result of *sen2977* sequence in *E.coli* genome (Zhou and Rudd. 2013). This transporter is predicted to deliver substrates for utilisation by UxuAB-UxaA in Fig. 4.5.



Figure 4.5. Comparison of hexonate and hexuronate utilisation in *E. coli* K-12 and SE. Transporters are in orange, enzymes in yellow, the Entner-Doudoroff (ED) pathway in green, and the pentose phosphate pathway (PPP), TCA cycle and glycolysis pathway are indicated simply in grey. Corresponding regulators are indicated above with targets in correspondingly coloured or underscored text. Note, most of the pathways above are induced by the cAMP-CRP complex. Uncertain pathways are dashed. Asterisks indicate proteins present in *S*E; hashes indicate proteins absent in *E. coli* K-12. Information was derived from the following sources: Robert-Baudouy & Stoeber, 1973; Portalier *et al.*, 1980; Blanco *et al.*, 1986; Zeng *et al.*, 2001.

Results and discussion

4.2.3 Impact of hexonates on *hex* gene expression in SE.

4.2.3.1 Do the SE transformants grow well with a hexonate as sole carbon/energy source?

The above work shows that *S*E can utilise all four of the available hexonates as carbon/energy sources and thus suggests that these hexonates would be suitable for testing the effect of hexonates on the expression of the *hex* genes of interest, using the *lacZ* fusions generated in the previous chapter. The protocol employed was as described in Methods (2.2.16), with *S*E PT4 transformants and growths at 42 °C (250 rpm) in 50 ml medium (100 μ g/ml ampicillin) in 250 ml Erlenmeyer flasks inoculated with 0.5 ml preculture.

Initially, the effect of the presence of the *lacZ* fusions plasmids on growth in M9 minimal medium with a hexonate (D-galaconate) was tested, to ensure that the presence of the plasmids did not unduly influence the growth of *S*E in the presence of hexonate as sole energy and carbon source. The results obtained are summarised in Fig. 4.6, and show that all transformants grew well and similarly with D-galactonate, and gave better growth than with the same level of glycerol, as was expected from the results above. Thus, the presence of the *lacZ* fusion plasmids should not greatly impact growth with hexonates as the carbon source.



Figure 4.6. Effect of lacZ-fusion vectors on growth of SE in 0.4% glycerol or D-galactonate in M9 minimal medium. Growth was aerobic in M9 salts medium and ampicillin, at 37 °C with continuous shaking in a BioScreen apparatus. ODs correspond to the 18 h point of growth. Growths were in triplicate (average provided with errors bars indicated as standard deviation), and the experiment was performed twice with similar results obtained, Results with glycerol are indicated by the dark blue bars, and for galactonate by the light blue bars. Asterisks indicate significant difference (P < 0.05).

4.2.3.2 Effect of D-galactonate of hex gene expression

The effect of 0.4% D-galactonate on expression of the *lacZ* fusions was compared with that of glycerol during growth at 42 °C (Fig. 4.7). One fusion (*sen2979*) that gave very weak expression in chapter 3, and also gave weak expression in *S*E (data not shown) was excluded from analysis in this chapter. One of the seven *lacZ* fusions showed induction in response to D-galactonate (list the *sen2979*), two showed no effect (less than twofold) and three showed a repression effect (*sen1436*, *sen1432*, *sen2977*). Note that the vector control did not respond to D-galactonate. The greatest induction (sixfold) effect was seen for *dgoR*. In *E. coli*, the *dgoR* gene is known to be

autoregulatory and to respond to D-galactonate. Indeed, DgoR acts as a repressor for the dgo genes, mediating their induction in response to D-galactonate (Neidhardt, 2005). The dgoR gene is the first gene in the dgoRKDA-T cluster of SE, and all of these genes are copolar and closely adjacent indicating that they are co-operonic. Although there is a 130 bp gap between dgoA and dgoT suggesting that dgoT may possess an independent promoter, the lacZ fusion data (Fig. 4.7) indicates that the dgoT-lacZ fusion is only weakly active (53-fold lower than dgoR-lacZ) and is not subject to D-galactonate induction, and may not carry its own promoter. This would suggest that the entire dgoRKDA-T cluster is indeed co-operonic. The 130 bp gap between dgoT and the upstream dgoA may explain the weaker induction of dgoT with respect to the other dgo genes in EWMM (10.8 fold, cf. 23.9-34.4 fold, respectively at 25 min; Baron *et al.*, 2017). The degree of induction observed here (Fig. 4.7) does not match that seen in EWMM (up to 28.7 fold; Baron et al., 2017) suggesting with that either: D-galactonate is not the relevant inducer in EWMM; that the lacZ fusions report lower degrees of expression than the microarray; or that the conditions used here are not sufficiently similar to those used by Baron *et al.* (2017).

The *sen1435-lacZ* fusion was only weak expressed (~180 U activity) but was modestly (although insignificantly) induced by D-galactonate, suggesting that the entire *sen1435-1434-1433-1432* operon is weakly expressed under the conditions employed and at most only modestly induced by galactonate. The *sen1432* gene is separated from the rest of the cluster by ~90 bp, and the *lacZ* fusion data suggests that it is more strongly expressed than *sen1435*, indicating that it may be independently transcribed. Sen1432 encodes a GntR-like regulator and so is expected to control the expression of the *sen1436-32* genes in response to the presence of the cognate substrate. Thus, the 47-fold higher expression of *sen1432* cf. *sen1435* is in keeping with the need for continuous availability of the regulator to mediate transcription control of the cluster. The *sen1432-lacZ*

fusion showed a significant repression effect with galactonate which suggests that it responds to a different effector. In addition, the divergent *sen1436* gene was also well expressed in glycerol but was repressed by galactonate, which further suggests (as in chapter 3) that this gene possesses an independent divergent promoter, and additionally is not subject to induction by galactonate.

The *sen2978-lacZ* fusion was well expressed in glycerol and not affected by galactonate indicating that the entire *sen2978-80 (uxaAB-uxuA)* operon is similarly expressed. In contrast, the divergent *sen2977* gene (encoding a transporter) was strongly repressed (x20) by galactonate (although was well expressed in glycerol). This would be consistent with repression by DgoR. The *ybhC* expression not effected as well suggested this gene is not under this regulater control.



Figure 4.7: β -galactosidase activity (maximum levels were taken at 6-10 h growth) of *lacZ* fusions in *Salmonella enterica* serovar Enteritidis (strain PT4-P125109) in the presence 0.4% D-galactonate in M9 minimal medium. Glycerol, dark blue bar, and 0.4% galactonate, light blue bar. Standard deviation of three values is given as error bars. Growth was aerobic at 42 °C and 250 rpm. Statistically significant difference (asterisks P= ≤ 0.05) as determined by Student's T-test and difference are indicted for all expression changes of twofold or more.

4.2.3.3 Effect of D-mannonate of hex gene expression

Mannonate had little effect on the expression of the *hex* genes. Indeed, only in one case was there a significant change in expression of twofold or more caused by mannonate (14-fold induction of *sen2977*; Fig. 4.15). The pathway by which mannonate is consumed in *SE* is likely to be that operated by SEN2977-90 (UxuAB/UxaA; Fig. 4.5). This would indicate a role for the *sen2977-uxuAB-uxaA* genes in utilisation of mannonate and/or related compounds. The induction of *sen2977* by mannonate is in contrast to its repression by galactonate and gluconate, and indicates a wide degree (280 fold) of transcriptional control for *sen2977* in response to different hexonates.

Previous work has shown that UxuR of *E. coli* represses its own expression (Ritzenthaler and Mata-Gilsinger, 1982).



Figure 4.8: β -galactosidase activity (maximum levels at 6-10 h growth) of *lacZ* fusions in *Salmonella enterica* serovar Enteritidis (strain PT4-P125109) in the presence 0.4% D-mannonate. Glycerol, dark blue bar, and D-mannonate, light blue bar. Further details are as for Fig. 4.7.

4.2.3.4 Effect of D-gluconate of *hex* gene expression

Gluconate failed to cause induction of any of the *hex* genes tested. However, it did result in significant repression of twofold or more in three cases (*sen1436*, *dgoT*, *sen2977*; Fig. 4.9). The greatest effect was seen for *sen1436* (17-fold); this gene was also 7-fold repressed by galactonate (Fig. 4.9). It is unclear which regulator might respond to gluconate, but it is possibly the same regulator (suggested as DgoR) as caused the observed repression with gluconate. The *dgoT* gene was also significant repressed with gluconate (1.6-fold), although its level of expression was relatively weak in the control (3.4 units), indicating (as suggested above) that any *dgoT* specific promoter would be weak. The *sen2977* gene was ~six-fold repressed by gluconate also and, like *sen1436*, was also down regulated by galactonate – which again suggests a common regulatory response for gluconate and galactonate. The *dgoR* gene was also significant repressed with gluconate in the *dgo* genes, mediating their induction in response to gluconate.



Figure 4.9: β-galactosidase activity (maximum levels at 6-10 h growth) of *lacZ* fusions in *Salmonella enterica* serovar Enteritidis (strain PT4-P125109) in the presence 0.4% gluconate in M9 minimal medium. Glycerol, dark blue bar, and gluconate, light blue bar. Further details are as for Fig. 4.7.

4.2.3.5 Effect of L-gulonate of hex gene expression

The expression of three hex genes was found to be significantly affected by gulonate at twofold or more (*sen1435*, *sen1432*, *dgoR*) (Fig. 4.10). No previous data on gulonate-dependent gene control in SE or *E. coli* could be found in the literature so the manner in which such control is exerted is not clear. Given that the purpose of the *sen1432-36* cluster is uncertain and yet all three corresponding fusions were induced (*sen1432*, x2.1; *sen1435*, x4; *sen1436*, x1.6) this might provide an indication of a role for these genes in gulonate utilisation with a potential role for the GntR-like *sen1432* product in controlling expression of these genes in response to gulonate. The

Chapter 4

dgoR gene was 3.5-fold repressed by gulonate indicating that the entire *dgo* operon is down-regulated by gulonate (possibly through SEN1432-mediated control, as raised above).



Figure 4.10: β-galactosidase activity (maximum levels at 6-10 h growth) of *lacZ* fusions in *Salmonella enterica* serovar Enteritidis (strain PT4-P125109) in the presence 0.4% L-gulonate in M9 minimal medium. Glycerol, dark blue bar, and L-gulonate, light blue bar. Further details are as for Fig. 4.7.

In summary, the data above indicate that the *hex* genes are indeed subject to regulatory control by hexonates, and that different hexonates elicit distinct regulatory responses suggestive of multiple regulatory pathways. Arguments for roles of DgoR, GntR and SEN1432 in mediating many of the hexonate-dependent responses observed have been provided, but these suggestions would require further experimental work with relevant regulatory mutants in order to confirm.

4.3 Effect of egg white and EW components on *hex-lacZ* gene expression in SE PT4.

The above results show that expression of several of the *hex* genes responds to the presence of hexonates in the medium, with both positive and negative effects observed. This raises the possibility that exposure of SE to EW results in release of hexonates that cause the induction of *hex* genes, as observed by Baron *et al.* (2017). However, there was no single hexonate that resulted in the strong induction of all relevant *hex* gene fusions (although an exhaustive range were not employed, due to lack of availability), and it is unclear how hexonates or hexuronates could be generated upon exposure of SE to EW. Thus, in order to further explore how EW causes the large increase in *hex* gene expression for SE, experiments involving exposure of SE *hex-lacZ* transformants to EW and EW components were performed (described below). EW, EW filtrate, EW total proteins and EWMM were prepared as described in Methods (2.2.17.1).

4.3.1 Effect of egg white on the growth of SE.

Initially, the effect of EW on the growth of SE was tested to confirm the inhibitory effect of EW on SE growth and to determine appropriate levels to employ in subsequent EW expression experiments. The concentrations of EW employed were from 0.05 to 10% in M9 medium with 0.4% glycerol, initially at 42 °C (hen body temperature). The results show (Fig. 4.11) that even a low level of EW has a major inhibitory effect on growth at 42 °C, with just 0.05% v/v EW reducing growth rate and culture density (~fivefold difference at 13.5 h).



Figure 4.11: The effect of up 10% v/v egg white on growth of SE in 0.4% glycerol M9 minimal medium at 42 °C. Growth was aerobic and carried out at 42 °C with continuous shaking in a Bioscreen apparatus (Methods 2.2.16). All data points are the average of three replicates. The experiment was repeated once and similar results were obtained. M9, control (glycerol M9 medium only); 0.05-10%, glycerol M9 slats medium with the indicated levels of EW added.

No growth was observed over 24 h when EW levels were at 0.5% or higher. This observation confirms the antimicrobial activity of EW as well observed for many bacterial species (Sahin *et al.*, 2003; Wellman-Labadie *et al.*, 2009).



Figure 4.12: The effect of egg white concentration in different temperatures on growth of SE in 0.4% glycerol M9 minimal medium. (A1) 2.5%, (A2) 10% and (A3) 15% v/v egg white at 30, 37 and 42 °C; and the effect of EW at (B1) 37 °C, (B2) 30 °C and (B3) 40 °C at 2.5, 10 and 15% v/v of EW. Conditions are as for Fig. 4.11.

To consider the impact of temperature, a similar experiment was performed at 30, 37, and 42 °C. As shown in Fig 4.12, EW inhibited growth at all three temperatures tested but was far less effective at 37 °C than at 30 and 42 °C, with no growth seen at 30 and 42 °C with the lowest EW levels used (2.5%) whereas at 37 °C growth was seen at all concentrations tested, although was reduced with respect to the EW-free control. Such an impact of temperature on the antimicrobial activity of EW is reported by Baron *et al.* (2011).

4.3.2 The effect of the EW, EW filtrate and EW proteins on hex gene induction.

The above results thus indicate that SE fails to grow in glycerol-containing M9 medium at 42 °C when low levels ($\geq 0.25\%$) of EW are included. Therefore, the effect of EW on *hex* gene expression can be tested in M9 medium at 42 °C, as described in Methods (2.2.16), using different levels of EW for short times post exposure to reduce or prevent growth. Cultures of SE transformants were grown in TSB until mid-log phase at 37 °C was achieved (OD 0.5), washed in M9 medium and then used to inoculate fresh M9 medium with EW. The cultures were incubated at 42 °C and 250 rpm, and samples were taken at 5, 25 and 45 min for expression analysis.

To begin with, a range of EW concentration were used (0.0001-10%) in M9 medium at 42 °C, with SE carrying pRS-SEN1436-*lacZ* (encoding a predicted D-galactonate dehydratase); *sen1436* was selected for further study as a representative *hex* gene that showed good expression in the previous experiments and was the most greatly induced gene in response to EWMM in the previous work of Baron *et al.* (2017).


Figure 4.13: Effect of EW on expression of *sen1436-lacZ* in wildtype SE in M9 medium at 42 °C. Growth was aerobic in 0.4% glycerol M9 medium, at 42 °C and 250 rpm, with the indicated levels of EW. Strains were SE carrying either the vector control (pRS415) or pRS-SEN1436-*lacZ*). Samples were taken for β -galactosidase assay at the indicated times post inoculation. Statistically significant difference as determined by Student's T-test (asterisks P= < 0.05) between M9 and EW. Results given are the average for xx cultures, each assayed in duplicate. The experiment was repeated once more and similar results were obtained. Error bars indicated standard deviation.

As shown in Fig 4.13, *sen1436* expression was induced by 22-61 fold with 0.01-10% EW. Very little induction was seen at 0.0001-0.001% EW, and the vector control showed no such response to EW. There was a general trend towards increased expression as the EW level was raised from 0.01 to 10% (Fig. 4.13), although the increment from 0.1 to 1% resulted in a drop in expression. The observed 61-fold induction with 10% EW is even higher that that (33-fold) reported by Baron *et al.* (2017), and is far greater than that seen above with hexonates, where a maximum 7 fold induction was observed.



Figure 4.14: Effect of EW filtrate on expression of *sen1436-lacZ* **in wildtype** *SE* **in M9 medium at 42** °C. Conditions are as for Fig. 4.19, except for the use of EW filtrate in place of EW.

The previous work by Baron *et al.* (2017) indicated that the induction of the hex genes in EWMM depended on the presence of EW proteins since EW (10 kDa cutoff) filtrate without addition of untreated-EW failed to induce the *hex* genes. Therefore, EW filtrate (10 kDa cutoff) was used in place of EW, as above, to determine whether the EW proteins of >10 kDa are responsible for the induction observed for *sen1436*. As clearly shown in Fig. 4.14, the EW filtrate gave only a very weak induction of *sen1436* expression, of just under twofold compared, compared with the expression level in the M9 medium. These effects of EW and EW filtrate were consistent as repeat experiments showed similar results (not shown).

Thus, the data strongly suggest that the EW factor that causes induction of *hex* gene expression in *SE*, is likely to be a protein of mass ≥ 10 kDa. Therefore, a test of egg white proteins individually

Chapter 4

is required for further investigation of the specific factor responsible. However, it was first considered necessary to repeat the experiments above performed with the *sen1436-lacZ* fusion using other *hex* gene fusions.



Figure 4.15: Effect of EW and EW filtrate on expression of *sen1432*, *sen2977* and *dgoR* in wildtype SE in M9 medium at 42 °C. Details are as for Figs. 4.13 and 4.14, except for the use of plasmids pRS-*sen1432-lacZ*, *sen2977-lacZ* and *dgoR-lacZ*. Samples were taken after 45 min incubation only. Statistically significant difference as determined by Student's T-test (P < 0.05).

Three other *hex* gene fusions (SEN1432, *dgoR* and SEN2977) were thus used to test the effect of EW and EW filtrate on *hex* gene induction by EW (Fig. 4.15), using concentrations at 10%. The results show that expression of all three fusions was induced by EW by 21-, 21- and 13-fold for SEN1432, *dgoR* and SEN2977, respectively, with respect to M9 only, whereas the negative control (pRS1274 empty vector) showed no effect. In contrast, all fusions showed no effect in expression level during the exposure to EW filtrate which matches the previous finding (Fig. 4.14)

and supports the conclusion that EW protein(s) of mass greater than 10 kDa is/are responsible for the induction of *hex* genes by EWMM. This experiment was repeated twice and the results were found to be reproducible.

4.3.3 The effect of specific EW proteins on *hex* gene induction.

4.3.3.1 Use of commercial egg-white proteins.

In order to identify the specific EW component responsible for the observed EW induction of *hex* gene expression, it was necessary to purchase and/or prepare the required proteins. The major proteins, as well as the minor proteins and peptides, found in egg white that exhibit confirmed or predicted antimicrobial activities in their native state are presented in Table 4.1. Four egg white proteins (albumin, conalbumin, ovomucoid, and lysozyme) are available commercially; these were purchased (from Sigma). Stock solutions of all four proteins were then prepared at 10% w/v and all solutions were sterilized using 0.22 Millipore filters. These proteins were then used to treat *S*E transformants carrying *hex* gene fusions in order to determine their effects on *hex* gene expression.

Protein	Total protein (%)	Molecular weight (kDa)
Ovalbumin	54	44.0
Conalbumin	12	76
(Ovotransferrin)		
Ovomucoid	11	28
Ovomucin	3.5	$5.5-8.3 \times 10^{6}$
Lysozyme	3.4	14.3
(Ovoglobulin G ₁)		
Ovoglobulin G ₂	4	30-45
Ovoglobulin G ₃	4	30-45
Flavoprotein	0.8	32
Ovoglycoprotein	1.0	24
Ovomacroglobulin	0.5	760-900
Ovoinhibitor	1.5	49
Avidin	0.05	68.3
Cystatin	0.05	12.7
(ficin inhibitor)		

Three different concentrations (0.01, 0.1 and 1 mg/ml) of each specific EW protein was used to examine their impact on SEN1436-*lacZ* induction (Fig. 4.16). Lysozyme gave a very strong induction effect for SEN1436 expression. The greatest induction (48 fold), with lysozyme, was seen at 7 min with 0. 1 mg/ml lysozyme. Ovomucoid II also gave a strong induction (5.6 fold), but not as large as that seen for lysozyme. The other EW proteins (albumin, conalbumin and ovomucoid III) gave little induction (Fig. 4.16). It should be noted that the effect obtained with lysozyme was ~1.25 lower than that seen with total EW, suggesting that lysozyme may not be the only factor required for high level EW induction of *hex* gene expression (although see below). Note that for the EWMM experiments of Baron *et al.* (2017), the concentration of lysozyme provided by the addition of 10% EW would be ~0.35 mg/ml, which is within the range of concentrations used here.



Figure 4.16: Effect of lysozyme, albumin, conalbumin and ovomucoid on expression of SEN1436 in wildtype SE in M9 medium at 42 °C. Conditions were as in Fig. 4.13 and 4.14 except for the use of commercially available EW proteins at 0.001-0.1% in place of EW and EW filtrate. Incubations times are indicated. Statistically significant difference as determined by Student's T-test (P < 0.05).

The finding that lysozyme gives the greatest response suggests that this protein is primarily responsible for the EWMM-induction of the *hex* genes. However, ovomucin II also gave relatively high induction. It is possible that the commercial EW proteins carry contaminants that might give misleading results. Thus, the purity of the tested proteins was examined by SDS-PAGE (Fig. 4.17). The SDS-PAGE analysis showed that Ovomucoid II contains another band of relatively high abundance that is similar in size to lysozyme (Fig. 4.17); a similar extra bad was observed for albumin. No such band could be seen in the ovomucoid III samples. This potential lysozyme contamination might explain the induction of SEN1436 by ovomucoid II but not ovomucoid III.



Figure 4.17: SDS-PAGE (12% acrylamide) analysis of the commercial egg white proteins. 10 µl of each protein at 10 µg/ml were loaded in each well. Well 1 is PageRuler Unstained Broad Range Protein Ladder (Fermentas); well 2, lysozyme; well 3, albumin; well 4, conalbumin; well 5, ovomucin II; well 6, ovomucin III.

To further investigate contamination of with traces of lysozyme, the samples were analysed by Western blotting using anti-lysozyme antibodies (Methods 2.2.17.7). As illustrated in Fig. 4.18, the resulting Western blot clearly shows an immune-reactive band in the lysozyme track, at the expected size. However, there is also a weaker band at the same migration point for the albumin and ovomucoid II, and similar bands for conalbumin and ovomucoid III but at even lower intensities. This suggests that the commercial EW proteins are contaminated with low levels of lusozyme that might explain the observed induction of SEN1436 by ovomucound II, and the weaker induction by albumin and ovomucoid III (Fig. 4.16).



Figure 4.18: Anti-lysozyme Western blot analysis of the commercial egg white proteins. See Fig. 4.17 for further details and Methods 2.2.17.7.

There is a possibility that the action of lysozyme on *hex* gene induction is enhanced by other EW factors. To test this possibility, the effect of 0.01 mg/ml lysozyme on SEN1436 expression was tested with and without additional EW proteins (Fig. 4.19). The results show no major or significant difference in SEN1436 induction in the condition where only lysozyme is used compared to where lysozyme is used with any of the other three major EW proteins (Fig. 4.19).

Chapter 4

This suggests that lysozyme is likely to be the only major EW protein that contributes to *hex* gene induction by EW.



Figure 4.19: Graph showing β -galactosidase activity at 45 minutes for SEN1436 *lacZ* fusions in *Salmonella enterica* serovar Enteritidis (strain PT4-P125109) in the presence different concentrations (all numbers is indicted to mg/ml) of egg white proteins. Growth was aerobic and carried out at 42 °C and continuous shaking. Statistically significant difference cf. M9 medium, as determined by Student's t-test.

4.3.3.2 The effect of lysozyme-free egg-white protein on *hex* gene induction.

The results above with SEN1436-*lacZ* expression and lysozyme suggest that lysozyme is the major EW protein that is responsible for the observed *hex* gene induction effect by EW proteins. To determine whether other *hex* genes are also subject to induction by lysozyme, the effect of lysozyme on expression of the SEN1436, SEN1432, *dgoR* and SEN2977 *lacZ* fusions was tested both with lysozyme and with a lysozyme-free EW preparation. The results showed that the

expression of all three fusions was increased, by 56-, 19-, 13- and 14-fold (respectively), by lysozyme (Fig. 4.20). In contrast, the fusions showed no induction effect upon exposure to an EW protein preparation that was free of both lysozyme and mucin (LMFEW; see Methods 2.2.17.2, for preparation of LMFEW). This experiment was repeated twice and similar results were obtained. The lack of any induction with EW lacking lysozyme supports the suggestion that lysozyme is the key factor in EW induction of *hex* gene expression.



Figure 4.20: Effect of lysozyme and lysozyme/mucin-free EW (LMFEW) on the induction of SEN1436, SEN1432, *dgoR* and SEN2977 in wildtype SE in M9 medium at 42 °C. Samples for assay were taken after 45 min incubation with EW factors. Other details are as described in Fig. 4.13.

Further exploration of the role of specific EW proteins in *hex* gene induction was progressed by analysing the effect of EW protein chromatographic fractions (Fig. 4.21) on the expression of the SEN1436 *lacZ* fusion (Fig. 4.22).





Figure 4.21: Elution profile of the fractionation of mucin-free EW protein by cation-exchange chromatography (A) and SDS-PAGE analysis (B) of resulting fractions. A. See Methods 2.2.17.2 for preparation of EW protein fractions. B. 10 μ l of each fraction were loaded into each well. Well 1 is PageRuler Unstained Broad Range Protein Ladder (Fermentas); well 2, Fractions 4-7; well 3, Fractions 8-21 diluted 5X; well 4, Fractions 8-21 (repeat) diluted 5X; well 5, Fractions 31-32; well 6, Fractions 42-44; well 7, Fractions 46-48; well 8, Fractions 49-52; well 9, Fractions 53-55. Bands of interest are labelled with levels in EW indicated

The results indicate that the fractions (4-32), lacking lysozyme but containing bands corresponding to ovotransferrin, albumin and ovomucoid, gave a relatively weak (up to fourfold) or no induction, whereas those containing lysozyme (42-55) gave a strong induction (up to 30

Results and discussion

Chapter 4

fold) (Fig. 4.22). It should also be noted that the 10% EW induction effect (31 fold) was similar in degree to that of pure lysozyme and of the lysozyme-containing fractions, suggesting that lysozyme is likely to be the sole EW protein that contributes to the induction of *hex* genes by EW.



Figure 4.22: Induction of the SEN1436 *lacZ* fusion in SE PT4 in the purified fractions. Details are as for Fig. 4.13 except for the use of the fractions indicated in Fig. 4.21 at 10μ g/ml concentration, EW at 10% v/v and lysozyme at 0.1 mg/ml.

4.3.4 Comparing the effect of lysozyme from chicken egg with that from human milk

lysozyme, and heat inactivated lysozyme

To determine whether lysozyme must be active in order to induce *hex* gene expression, and whether lysozyme from a difference source can also induce expression of the *hex* genes, and thus further indicate whether it is the enzymatic activity of the lysozyme that drives the *hex* gene induction effect, lysozyme was heat treated to inactivate it and lysozyme from human milk was compared to that from chicken eggs to determine whether the lysozyme affect is lysozyme-species specific. Heating inactivation was achieved by incubating the enzyme for 30 min at 100 °C. This resulted in almost complete loss solubility (not shown).

Human and hen lysozyme from commercial sources gave similar ~50-fold induction effects on SEN1436 expression. In addition, non-commercial lysozyme purified from chicken eggs (as described in Methods 2.2.17.2) also gave a 30-fold induction effect (Fig. 4.23). However, the heat inactivated (now insoluble) lysozyme (commercial, chicken egg source) showed very weak induction. Thus it can be concluded that the lysozyme effect requires an active enzyme, and that there is no marked species specificity requirement for the lysozyme effect.



Figure 4.23: Effect of human and chicken lysozyme, and inactivated lysozyme, on SEN1436*lacZ* **expression in SE PT4.** Details are as described in Fig. 4.13, except that human milk lysozyme was used (as indicated) at 1 mg/ml, as was heat (100 °C) inactivated hen-egg lysozyme (Sigma) and non-commercial chicken-egg lysozyme purified as part of this PhD, as described in Methods (2.2.17.2).

4.3.5 The effect of temperature, pH and iron on EW-mediated induction of *hex* gene expression.

The results presented above strongly indicate that lysozyme of EW is the major factor responsible for the EW induction of *hex* gene expression. However, there are other (anti-bacterial) factors associated with EW that might impact the action of lysozyme. Two important factors to consider are temperature and pH. Studies on the effect of temperature on the antibacterial activity of EW, from 37 to 48 °C, on the survival of *S*E demonstrated that EW is more harmful towards *S*E as the temperature rises, and indeed become bactericidal above 42 °C (Alabdeh *et al.*, 2011) with complete loss of viability (due to lysis) observed at 45 °C. Baron *et al.* (2015) showed that bacterial destruction was higher at pH 9.3 (the pH of EW shortly after laying) than at pH 7.8. The EW pH value dramatically and rapidly increases after laying from 7.6 up to 9.3 in a few days due to lost CO₂ through the pores of the eggshell (Sauveur, 1988). Many studies have shown that at pH \geq 8.8, bacteriostatic or bactericidal effects are observed towards different bacteria in EW, including *Salmonella*, which is in contrast to pH 7.5-8 that allow either slight bacterial growth or cause no more than a bacteriostasis effect (Kang *et al.*, 2006. Messens *et al.*, 2004).

The previous experiments with EW induction of *hex* gene expression were performed at 42 °C. Here, these experiments were repeated at lower (37 °C, non-bactericidal) and higher temperature (45 °C, as originally used by Baron *et al.*, 2017 which represents suitable conditions for examination of the response of *SE* to the bactericidal activity of EW), and at two distinct pH values, 7 and 9 (to determine if the pH-dependent bactericidal nature of the EW affects the observed EW induction of *hex* genes). A pH of 7 was the starting pH for the experiments so far described in this thesis in M9 minimal medium. The EWF in glycerol M9 medium was pH adjusted with 2.5 M HCl to decrease pH from ~9 to 7 (Methods 2.2.17.1).

As shown in Fig 4.24, temperature and pH made little difference to the EW induction of SEN1436. In addition, the inclusion of EWF along with EW, also had little impact on SEN1436 induction. Thus, the results suggest that EW protein (lysozyme) is likely the only major EW factor required for the observed EW induction of *hex* gene expression.



Figure 4.24: Effect of temperature, pH and combining EW with EWF on SEN1436-*lacZ* induction in SE PT4. Details are as in Fig. 4.13, except that three different temperatures were used (37, 42 and 45 °C) and two different pH values (7 and 9). Also, 100% EWF was used in many of the incubations (45 min) either with or without 10% EW. The condition with 10% EW in 100% EWF at pH 9 and 45 °C is a very close match the condition employed by Baron *et al.* (2017) for their global transcriptomic analysis of SE gene expression in EWMM.

The expression of one exemplar *hex* gene (SEN1436, which was selected as it gave the strongest response to EW) was tested in EW with and without 20 μ M ferric citrate. Note that 20 μ M ferric citrate was found to restore the growth of *S*E in 0.1% EW to levels similar to those achieved without EW (data not shown). The results showed (Fig. 4.25) that the induction of SEN1436 by EW is not influenced by provision of iron. Thus, it can be assumed that the EW response of the *hex* genes is unrelated to the low iron availability in EW.



Figure 4.25: Graph showing the effect of iron on SEN1436-*lacZ* expression in SE wildtype in the presence 0.1% of EW. Growth was aerobic and carried out at 42 °C with continuous shaking. Expression was measure after 45 min incubation. Conditions were 0.4% glycerol M9 medium with/without 20 μ M ferric citrate and/or 0.1% EW, as indicated

4.4 Discussion

In this chapter, SE genes involved in Hex metabolism (*dgoRKADT*, *uxuAB-uxaC* and SEN1433-6 clusters) were studied.. The high response shown suggested that the effect observed is of considerable physiological significance during EW exposure. These results confirmed the significant overall induction at 25 and 45 min upon EW exposure (Baron *et al.*, 2017). Thus, it is clear that the three induced *hex* gene clusters are subject to major up regulation during exposure of SE to EW. However, the inducer was not recognised and was particularly unclear given that the the presence of hexonates and hexuronates within EW is not recognised (Guérin-Dubiard *et al.*, 2010). Thus, in this chapter, the relevant environmental factors affecting the induction of the *hex* genes in SE were investigated through use of *lacZ* fusions (created in chapter 3) to confirm the

proposed induction of the *hex* genes upon EW exposure and to identify the specific EW factor that caused this effect. The results obtained can be divided into two main sections.

4.4.1 Effect of hexonates on growth and expression

The growth of SE in the presence of four different hexonates as carbon sources was considered. In addition, the effect of these substrates on *hex* gene expression was examined, using *lacZ* fusions. Firstly, the growth of SE at mammal and hen body temperatures (37 and 42 °C; Raspoet et al., 2014; Baron et al., 2017) with standard carbon sources (glucose, glycerol) was performed in M9 minimal medium to compare with growth on the hexonates. Glucose was selected as it is present in EW at 0.4-0.5% (Guérin-Dubiard et al., 2010), and glycerol was used as an example of a nonfermentable carbon source that does not induce catabolite repression. A range of glucose concentrations (0.1-1.6%) was employed in aerobic condition at 37/42 °C to show a quantitative effect on growth, above and below the levels found in EW and compared with 0.4% glycerol as control. The wild type SE PT4 grew well at 42 °C with glycerol or glucose at 0.4%, and thus either of these could be used as positive controls for growth tests with the Hex compounds at this temperature. Temperature had little notable impact on growth of SE, except with 1.6% glucose were the final density at 42 $^{\circ}$ C (~0.69) was impaired with respect to that at 37 $^{\circ}$ C. The ability of the four forms of available hexonates (D-galactonic acid; D-mannono-1,4-Lactone; L-(+)-gulonic acid γ -lactone and gluconate), at 0.1-1.6% w/v, to support growth of SE was tested, with glycerol acting as the control. The results showed that the growth of SE was well supported by all four hexonates at both 37 and 42 °C, although some differences in the degree of growth supported were apparent. Best growth was achieved with galactonate, followed by gluconate, then mannonate, and finally gulonate (maximum ODs of 1.1, 0.9, 0.8, 0.7, respectively, at 42 °C). Galactonate enhanced growth for SE at 42 °C more than any of the other hexonates tested (max growth at 15 h of 1 OD unit for galactonate cf. 0.85, 0.69 0.63. for mannonate, gluconate and gulonate, respectively) or glucose (0.67 OD units at 15 h), indicating that galactonate is a good carbon source for *S*E. Therefore, it was concluded that these four hexonates can act as sole carbon and energy source for *S*E growth at hen body temperature. This finding confirmed the ability of *Salmonella* to grow on galactonate and gulonate (Cooper, 1980).

SE carries the genes of the GntI system: *gntK* (gluconokinase; SEN3365), *gntU* (low-affinity gluconate transport; SEN3364), *gntR* (gluconate utilization operon repressor; SEN3366) and *gntT* (high-affinity gluconate transporter; SEN3338). However, the GntII system is absent (Parkhill *et al.*, 2008). These genes are likely to be subject to catabolite repression (Rodionov *et al.*, 2000) and induced by gluconate through GntR transcriptional control (see Fig. 4.5). The gulonate degradation pathway in *SE* is unclear. However, this pathway is likely to involve one or more of the GntI system, the Dgo pathway or the SEN1433-6 pathway (Fig. 4.5). While galactonate is expected to be catabolised via the Dgo pathway is subject to catabolite repression and is induced by D-galactonate (Deacon & Cooper, 1977; Cooper, 1978). For mannonate, the utilisation pathway in *SE* is expected to be that operated by SEN2977-90 (UxuAB/UxaA; Fig. 4.5). However, this not confirmed in *SE*.

So, since all four of the available hexonates were utilised by SE as carbon/energy sources, they were considered suitable for testing their effect on the expression of the *hex* genes of interest, using the *lacZ* fusions generated in the previous chapter. The results confirmed that the presence of the *lacZ* fusion plasmids does not greatly impact growth with hexonates as the carbon source. The *hex* gene expression results showed a varied response to the four hexonates available. All

substrates were used at 0.4% was compared with that of glycerol-M9 medium during growth at 42 °C.

For 0.4% D-galactonate, the greatest induction (sixfold) effect was seen for *dgoR* and a repression effect was seen for sen1436, sen1432 and sen2977 by 6-, 3.5- and 20-fold, respectively. The vector control did not respond to D-galactonate. It is likely that DgoR acts as a repressor for the dgo genes; dgoR is the first gene in the dgoRKDA-T cluster of SE. In E. coli, the dgoR gene is known to be autoregulatory and to respond to D-galactonate mediating the induction of the dgogenes in response to D-galactonate (Neidhardt, 2005). The results observed here are consistent with this since the dgoRKDA-T genes were the only hex genes well induced by galactonate. However, the induction level observed does not match that seen in EWMM (up to 28.7 fold; Baron et al., 2017) suggesting with that D-galactonate is either not the relevant inducer in EWMM or that the conditions used here are not sufficiently similar to those used by Baron et al. (2017) to enable the same level of induction to be acieved. Mannonate showed little effect on the expression of the hex genes. The only significant change was with SEN2977 (14-fold). This would indicate a role for the sen2977-uxuAB-uxaA genes in utilisation of mannonate and/or related compounds. In contrast, this fusion showed repression by galactonate and gluconate. While gluconate showed repression effect on most fusions tested with greatest effect was seen for sen1436 (17-fold). It is unclear which regulator might respond to gluconate, but it is possibly the same regulator as caused the observed repression with gluconate. Gulonate showed a significant effect on three fusions (sen1435, sen1432, dgoR) at twofold or more. No previous data on gulonate-dependent gene control in SE or E. coli could be found in the literature so the regulator responsible is not clear. However, three corresponding fusions were induced (sen1432, x2.1; sen1435, x4; sen1436, x1.6) which might suggest a role for these genes in gulonate utilisation with

a potential role for the GntR-like *sen1432* product in control. To summarise, it is clear that the *hex* genes are indeed subject to regulatory control by hexonates, and that different hexonates show distinct regulatory responses suggestive of multiple regulatory pathways. Arguments for roles of DgoR, GntR and SEN1432 in mediating many of the hexonate-dependent responses observed have been provided. However, these possibilities need confirmation through further investigation with relevant regulatory mutants.

4.4.2 The role of lysozyme in inducing the *hex* genes in EW

The up and down regulation in expression of several of the *hex* genes are clearly response to the presence of hexonates in the medium. This suggests the possibility that exposure of SE to EW results due to release of hexonates that cause the change in expression level of *hex* genes, as observed by Baron et al. (2017). However, it is unclear how hexonates or hexuronates could be generated upon exposure of SE to EW. Therefore, further investigation was performed to explore how EW causes the large increase in *hex* gene expression for SE. The effect of EW on the growth of SE was tested at different level (0.05-10%) at hen body temperature (42 °C) to confirm the inhibitory effect and to determine appropriate EW levels to employ in subsequent EW expression experiments. The results showed that even a low level of EW has a major inhibitory effect on growth at 42 °C, with just 0.05% v/v EW reducing growth rate and culture density (~fivefold difference at 13.5 h). This is in agreement with the well observed antimicrobial activity of EW for many bacterial species (Sahin et al., 2003; Wellman-Labadie et al., 2009). Growth was observed at 37 °C at all EW concentrations tested, but was reduced with respect to the EW-free control. On the other hand, at 30 and 42 °C, growth was totally inhibited at relatively low EW levels (2.5%). Such an impact of temperature on the antimicrobial activity of EW has been reported previously by Baron *et al.* (2011).

Initially, the effect of EW on hex gene expression was tested in M9 medium at 42 °C using different levels of EW (0.0001-10%) in M9 medium at 42 °C. SE carrying pRS-SEN1436-lacZ (encoding a predicted D-galactonate dehydratase) was selected for further study as a representative hex gene that showed good expression in the previous experiments and was the most greatly induced gene in response to EWMM in the previous work of Baron et al. (2017). The results showed that *sen1436* expression is induced by 22-61 fold with 0.01-10% EW, compared to the vector control. The induction observed with 10% EW (61-fold) is higher than (33-fold) that reported by Baron et al. (2017), and is far greater than that seen above with hexonates, where a maximum 7 fold induction was observed. The experimental conditions applied by Baron et al. (2017) showed hex gene induction depended on the presence of EW proteins since EW (10 kDa cutoff) filtrate without addition of EW failed to induce the *hex* genes. Therefore, EW filtrate (10 kDa cutoff) was tested in place of EW to confirm that the EW proteins of >10 kDa are indeed responsible for the induction observed for *sen1436*. The results showed the EW filtrate gave only a very weak induction of *sen1436* expression, of just under twofold compared, compared with the expression level in the M9 medium. Therefore, this finding strongly suggests that the EW factor causing induction of hex gene expression in SE, is likely to be a protein of mass ≥ 10 kDa. Therefore, a test of egg white proteins individually was initiated to identify the specific factor responsible. The experiment was repeated with three other hex gene fusions (sen1432, dgoR and sen2977) and the results showed that expression of all three fusions was induced by EW by 21-, 21- and 13-fold for sen1432, dgoR and sen2977, respectively using concentrations of EW at 10%. So, individual EW proteins (albumin, conalbumin, ovomucoid, and lysozyme) at three different concentrations (0.01, 0.1 and 1 mg/ml) of each were used to examine their impact on SEN1436*lacZ* induction. The results showed lysozyme gave a very strong induction effect for SEN1436

expression. The greatest induction (48 fold), with lysozyme, was seen at 7 min with 0.1 mg/ml lysozyme suggesting this protein is primarily responsible for the EWMM-induction of the *hex* genes. The lysozyme impact was confirmed in several ways (different combination of EW proteins, different source of lysozyme, heat inactivation, different pH, iron and temperatures). However, the mechanism of by which lysozyme induces the *hex* genes is unclear.

4.4.4 Conclusion

To conclude, *dgo*, *uxu/uxa* and SEN1433-6 gene induction during exposure of *SE* to EW suggested that this up regulation is due to hexonate and/or hexuronates. However, the absence of these organic acids in EW, with the recognition of lysozyme as the main inducer, allows a new hypothesis to be proposed whereby the induction observed is caused by the release of an endogenous inducer from *SE* in response to cell envelope damage elicited by lysozyme. A further understanding of the precise mechanisms could help development of new approaches towards the preservation of foods against bacterial infection.

Chapter 5: Role of the *hex* gene regulators, SEN1432 and DgoR

5.1 Introduction

Targeted gene knock out is a key approach in studies on gene function. There are various strategies for gene inactivation in bacteria. For *Salmonella* spp., the R6K-suicide plasmid, the λ Red disruption system, the suicide plasmid combined with the Red system or the temperature-sensitive plasmid carrying a *sacB* gene for negative selection (Geng *et al.*, 2011). The λ Red recombineering technology has been used extensively in *Escherichia coli* and *Salmonella typhimurium* for easy PCR-mediated generation of deletion mutants (Murphy and Campellone, 2003).

The most allelic exchange methods require the engineering of a gene disruption on a suitable plasmid, although genes can be directly disrupted in some organisms (e.g. *Saccharomyces cerevisiae*) by transformation with PCR fragments encoding a selectable marker with sufficient flanking homologous DNA. However, not all recipients are readily transformable with linear DNA due to the activity of the intracellular exonucleases that degrade linear DNA. Therefore, Datsenko and Wanner (2000) developed the simple and highly efficient Red Disruption system to directly inactivate chromosomal genes in *E. coli* K-12 using PCR products based on the phage λ -Red recombinase, which is synthesized under the control of an inducible promoter on an easily curable, low copy number plasmid, such as pKD46 or pKD20 (Geng *et al.*, 2009).

Here, the λ Red disruption system (Wanner and Datsenko, 2000) was used for single gene knockout in *Salmonella*. This method relies upon the presence of a low-copy, temperature-sensitive 'helper' plasmid encoding components of the homologous recombination system found in bacteriophage λ (pKD46). These components are called Exo (a 5'-3' exonuclease which processes along double-stranded DNA), Bet (a single-stranded DNA-binding protein which is capable of annealing complementary single strands) and Gam to inhibit host exonuclease such RecBCD and SbcCD (Fig. 5.1). Expression of these genes is under the

control of an arabinose-inducible promoter (P_{araBAD}). When cells expressing the plasmid are grown in the presence of arabinose, exogenously applied linear DNA is able to undergo homologous recombination with the bacterial chromosome. In this manner, it is possible to generate an in-frame gene deletion using a PCR product.



Figure 5.1: The components of the lambda Red recombineering system. Exo (a 5'-3' exonuclease which processes along double-stranded DNA), Bet (a single-stranded DNA-binding protein which is capable of annealing complementary single strands) and Gam to inhibit host exonuclease such RecBCD and SbcCD. From: Beth Kenkel (2016). http://blog.addgene.org/lambda-red-a-homologous-recombination-based-technique-for-genetic-engineering.

To inactivate chromosomal genes, an amplified fragment carrying an antibiotic cassette flanked by a region homologous to the target locus is electroporated into a strain that expresses the λ Red recombination system to replace the target gene with an antibiotic resistance gene, usually kanamycin or chloramphenicol resistance (Lesic and Rahme, 2008). To generate the PCR fragment, pKD3 is used as a template to amplify a chloramphenicol resistance cassette flanked by FRT sites, which allow the removal of the cassettes once inserted in the bacterial chromosome with an FLP helper plasmid, such pCP20 (Fig. 5.2). Figure 5.3 outlines the generic λ Red recombineering technique.



Figure 5.2: The pKD3 plasmid linear templates. Arrowheads show locations and orientations of priming sites. P1 & P2: priming sites, k1, k2, and kt: common test primers, *oriR* and *rgnB* show transcription origin and terminator respectively. Arrows with open arrowheads show the nearly perfect FRT site inverted repeats. The black arrows show antibiotic markers (Wanner and Datsenko, 2000).



Figure 5.3: Schematic representation of the λ Red recombineering technique. H1 and H2 refer to the homology extensions or regions. P1 and P2 refer to priming sites (Wanner and Datsenko, 2000).

5.1.1 Aim of this chapter

In this chapter, genes related to hexonate utilisation were selected for further analysis through

knock out (SEN1432 and dgoR) to investigate whether deletion of these genes has any obvious

phenotypic effect.

5.2 Generation of SE deletion mutant

5.2.1 SEN1432 and *dgoR* gene knock out

In order to determine the roles of the *hex*-gene specific regulators in controlling the response of the Hex genes to EW factors and to hexonates/hexuronates, in this chapter the SEN1432 and dgoR genes of SE PT4 were targeted for inactivation.

SEN1432 belongs to the GntR subfamily of the FadR family of transcriptional regulators (Haydon and Guest, 1991) which have similar DNA binding N-terminal winged helix-turnhelix domains (located at residues 5-78 for the SEN1432 protein; InterPro database) and a Cterminal effector-binding/oligomerization domain (residues 87-234 for the SEN1432 protein; InterPro database)SEN1432 is located at 1,520,207 to 1,520,926 bp in the SE PT4 genome under the entry name B5R537 SALEP and consists of 720 bp encoding a 239 amino acid residue primary translation product (Fig. 5.4). GntR-family transcription factors interact with DNA as dimers where they act as repressors. Binding of an inducer (usually the substrate of the metabolic pathways that the transcription factor regulates; Jain, 2015) appears to trigger a change in confirmation which releases the transcription factor from the DNA (Resch et al., 2010). The SEN1432 inducer is suspected to be a hexonate, such as gulonate (see chapter 4.2.3.5) but this remains to be proven. However, it is highly likely that the SEN1432 product regulates the SEN1435 and/or SEN1436 promoters and this controls the entire SEN1432-36 cluster in response to the cognate catabolite. Although there is no evidence that Hex is present in EW, confirming the control of SEN1432-6 by the SEN1432 product would give a better understanding of how the Hex utilization genes are induced in egg-white and would give a clearer indication of what their role might be. Note that inactivation of SEN1432 is unlikely to cause any downstream polarity effect on the expression of adjacent genes since it lies at the end of the SEN1435-2 operon.



Figure 5.4: Genetic map of the SEN1432 region. From https://biocyc.org/Salmonella enterica enterica P125109 NC_011294: SEN_RS07425

The genes of the SEN1432-6 cluster specify three enzymes (two suspected dehydrogenases and one dehydratase), likely to be involved in hexonate utilization, and a proposed hexonate transporter (Baron *et al.*, 2017). Studies conducted by Thomson *et al.* (2008) comparing the genomes of *S*. Enteritidis PT4, *S*. Gallinarum 287/91 and *S*. Typhimurium LT2 showed that the 6 kb region called ROD13, carrrying SEN1432–SEN1436 is present in *S*. Gallinarum as well as *S*. Enteritidis but not in *S*. Typhimurium LT2. This suggests that the SEN1432-6 genes may be part of the accessory genome of *Salmonella* spp. and thus subject to variation between strains according to environmental demand and evolutionary pressures (Betancor *et al.*, 2012).

dgoR also encodes a GntR-related regulator likely acting as a D-galactonate-responsive transcriptional repressor of the *dgo* operon (Cooper, 1978; Neidhardt, 2005; Zhou & Rudd, 2013). All of the genes, including the regulatory gene *dgoR*, cluster at min 83.40 (Neidhardt, 2005). *dgoR* is located 3,907,912 to 3,908,601 bp in the SE PT4 genome under entry name B5QUP2_SALEP (SEN3647) and consist of 690 bp coding a 229 residue polypeptide (Fig. 5.5). According to Baron *et al.* (2017), *dgoR* was induced by up to 28.7 fold by exposure to egg white, whereas SEN1432 (unlike other members of the putative SEN1435-2 operon) was not EW induced, indicating that it may be expressed independently of the SEN1432-5 genes. The *dgoR* gene was also induced strongly by EW (and lysozyme) in chapter 4, and its activation by replacement with a Cm^R cassette would be expected to exert a polar effect on the rest of the *dgo* operon (see Fig. 5.5) which could result in a growth defect on a subset of hexonates.



Figure 5.5: Genetic map of the *dgoR* **region in the SE PT4 genome.** https://biocyc.org/Salmonella enterica Enterica P125109/SEN_RS18935

The general function of the *dgoRKADT* operon is believed to be in the utilization of Dgalactonate with the release of glyceraldehyde 3-phosphate into the glycolytic pathway and pyruvate into the TCA cycle (Neidhardt, 2005). Note that in chapter 4, *SE* grew better on galactonate than on the other three hexonates tested. *dgoT* is inferred to encode a D-galactonate uptake system whereas *dgoA*, *dgoK* and *dgoD* are suggested to code for enzymes required for the conversion of D-galactonate to pyruvate and glyceraldehyde-3-phosphate (Walters *et al.*, 2008; Ran *et al.*, 2004; Deacon 1977; Cooper 1978).

To recall results obtained in chapter 4, SEN1432-*lacZ* showed relatively moderate expression in *S*E and was only induced by gulonate, suggesting that gulonate may act as an inducer for SEN1432 expression and thus that Sen1436 may utilise gulonate as an effector. *dgoR-lacZ* was also moderately expressed in *S*E but was induced by galactonate by sixfold (whereas SEN1432 was 3.5-fold repressed), but was either unaffected or repressed by the other there hexonates. This result is consistent with a role in mediation of galactonate repression for DgoR, and a role in utilisation of galactonate for the *dgo* gene products.

5.2.1.1 Primers design

Forward and reverse primers were designed to anneal at the 4th and penultimate codon of the target gene, respectively (Table 2.6), allowing generation of an in-frame deletion with minimal downstream effects once the Cm^R cassette is removed. The 5' end of each primer (between 45-48 nucleotides) was homologous to the target gene, whereas the 3' end of each primer was designed to amplify the chloramphenicol resistance cassette encoded by pKD3 (Fig. 5.6). Figs

5.7 and 5.8 illustrate the position of the primers used for the purposes of generating a PCR product for deletion of the SEN1432 and dgoR genes, respectively. In addition, further primers were used to confirm that the desired mutation had occurred; these were designed to primer at the flanking regions of the targeted gene.

TGTGTAGGCTGGAGCTGCTTC GAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAA TAGGAACTTCATTTAAATGGCGCGCCCTTACGCCCCGCCCTGCCACTCATCGCAGTACTG TTGTATTCATTAAGCATCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGA ATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCCATGGTGAAAAC GGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTC GTGGTATTCACTCCAGAGCGATGA CAT Cassette TTTGCTCATGGAAAACGGTGTAAC AAGGGTGAACACTATCCCCATATCACCAGGTGACGTCATCGTCATTGCCATACGTAATTCC GGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTT ATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGG TACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATA **TCAACGGTGGTATATCCAGTGATTTTTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAA** TCTCGACAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGG AACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCG GTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTAG GCGCGCCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGAACTAAGGAG GATATTCATATG

Figure 5.6: Primer locations for amplifying the cat gene of in pKD3. Grey part: target gene; Green part: Primers, expected size ~1015 bp.

TTCGTTTCGATTAACGGTGAAAAGCGCCCGGCCGGGCGCTTTGTTCTTAAAAGAGAAATT GTTATATAAAAAGCACTTCAGCGACATCTTAACGGATACCCATCTTGAGCATAAAAATC CATTCAAAAACAGAATGTTGTAAATGAAATTTATGATCAGATAAGTAGCAAACTGCTGG ACGGCAGTTGGGCGCCGGGTAGCCGTTTGCCCTCAGAAGTGGAACTGACCGCCTCATTT AACGTCAGCCGGGTCAGCGTTCGCAGCGCAGTACAGCGTTTTCGTGACCTGGGGATTGT GGTGACGCGTCAGGGCAGCGGCAGCTACGTGAGCGAAAACTTCACCCCGCAGATGTTGA GTAACGATCCCCGCCCAATCATGCA(GCGAAGAGTTTCACGATATGATGATT SEN1432 TTTCGTCAGACCGTGGAGTTCAAATC CTCGCCGTCACACACGCCACCGATGA TGACATTCGCCAGCTCGAGGAAGCATTGAACAACATGCTGATCCACAAAGGTGATTATA AAAAATACTCGGAAGCGGACTACGAGTTCCATCTGGCGATTGTCAGGGCATCGCACAAC AGCGTGTTCTACAACGTGATGAGCTCGATTAAAGACATCTATTACTACTATCTTGAAGA GCTTAACCGTGCGCTGGGTATTACCCTTGAAAGTGTGGAAGCCCATATCAAGGTCTACA TGTCGATAAAGAATCGCGATGCCAGCACGGCCGTCGAAGTGCTCAATGAAGCGATGTCA GGCAATATTATTGCGATCGAAAAAATCAAATCTACAGAGACATCAGGGACAAAA**TAACC** GTTGGTTACAAGCTCAAGTAGTAGAGCAATTTAACATATCTGAATCCGAAATAGTTGCC ATCAACTATTTAGTGACATAGTCCC<mark>ACTTTAAAATCGTGGCAGTGC</mark>

Figure 5.7: Primer locations for inactivation of the SEN1432 gene. Grey part: target gene; Green part: post deletion confirmation primers; Bold & underlined regions, sequence matching the deletion primers.

TGGCATGATAACGACGGTTG<mark>ATATCACGCTAGTACTACAAAATTGCGGCGTAATTCAGC</mark> TATCGCGGTAAAGTAAGAGAGTTCACATCGAGCACAAGGACTCTCTATGACTCTCAATA AAACCGATCGCATCGTTATCACGCTGGGCAAACAGATTGTCAGCGGTAAATACGTACCC GGTTCGGCGCTGCCAGCGGAAGCGGATCTGTGCGAGGAGTTTGAAACGTCGCGCAACAT CATTCGCGAAGTGTTTCGTTCGCTTATGGCGAAGCGGCTAATTGAAATGAAGCGCTATC GCGGCGCGTTTATCGCACCGCGTAACCAGTGGAATTATCTCGATACCGACGTGCTGCAA TGGGTGCTGGAAAATGACTACGACCCGA ATCAGCGCGATGAGCGAAATACGAAA dgoR CCTGGTGGAGCCAGCAATAGCACGCTG JGAACGGGCAACATCAAGCGATCTGG CTGAAATTGAGTCGGCGCTAAACGACATGATTGCCAACAACCAGGACCGGGAAGCGTTT AACGAGGCGGATATTCGCTATCACGAAGCAGTGTTGCAGTCGGTGCATAACCCGGTACT TGGGCGATGCGGCCAATATGCCGAAAACGCTCCAGGAACATAAGGCGCTATTCGATGCG ATACGGCATCAGGATGGCGATGCGGCAGAGCAGGCGGCATTAACCATGATCGCCAGCTC GACACGAAGGTTAAAGGAAATCACATGACAGCTCGCTACATCGCAATTGACTGGGGGATC GACCAATCTGCGCGCCTGGCTTTACCAGGGCGACAAATGCCTGGAGAGCAGGCAAGC AAGCAGGCGTTACAC

Figure 5.8: Genetic map of the *dgoR* gene. See Fig. 5.7 for details.

5.2.1.2 Replacement of SEN1436 and dgoR in SE PT4 with a Cm^R cassette

The first step in the gene inactivation was the generation of the linear DNA PCR products with the Cm^{R} cassette flanked by sequence matching the flanks of the target gene. The pKD3 plasmid was isolated and *Nde*I digestion was performed to confirm the size of the linearised plasmid (Fig. 5.9) by agarose gel electrophoresis, which showed a fragment of the expected size.



Figure 5.9: Map of the pKD3 vector. Restriction sites and other key features are indicated. https://www.addgene.org/45604/

PCR was carried out as described in section 2.2.9 using primers as in Table 2.6. As shown in Fig. 5.10, the target sequences were amplified successfully to give bands at ~1100 bp which correspond to the sizes of the target fragments. The products were purified using a Thermo Scientific GeneJETTM PCR purification kit (section 2.2.7).



Figure 5.10: Gel electrophoresis of PCR amplification products of the Cm^R cassette of pKD3. Lane 1 GeneRulerTM 1kb ladder (250-10,000 bp). PCR products are as follow: lane 2 & 3, SEN1432 specific Cm^R cassette; lane 3 & 4, *dgoR* specific Cm^R cassette. Electrophoresis was as in Fig. 5.10.

After PCR clean up, the purified PCR products were cloned, using the Thermo Scientific CloneJET PCR Cloning Kit, into the pJET1.2 cloning vector (section 2.2.10). The cloning reaction products were transformed into E. coli TOP10 and Cm^R transformants were selected for plasmid isolation (seven isolates of each). The DNA thus obtained was subject to agarose gel electrophoresis (Fig. 5.11) which showed a decreased mobility for 11 out of the 14 plasmids indicative of the presence of an insert of ~1 kb (Fig 5.11). Subsequent nucleotide sequencing (using primers in Table 2.6) confirmed that desired insert was present, in each case, and was 100% identical as expected (see Appendix 8 for further detail).



Figure 5.11: Electrophoretic analysis of pJET1.2 clones contains the Cm^R cassette PCR fragments from pKD. Lanes 1 & 9, Fermentas GeneRulerTM 1kb ladder; lanes 2-8, undigested pJET1.2 clones with SEN1432 specific PCR product; lanes 2-5, undigested pJET1.2 clones with dgoR specific PCR product. See Fig. 10 for further detail.

The pKD46 plasmid was transformed into *SE*, as described in section 2.2.13. Before transformation, the identity of pKD46 (6329 bp) was confirmed by single digest with *Bam*HI.

SE PT4 carrying the pKD46 plasmid was grown in LB (with ampicillin) and arabinose was added to a final concentration of 10 mM in order to induce expression of the homologous recombination system (section 2.2.15.4). The linear DNA (PCR products derived from pJET1.2 clones) carrying the Cm^R cassette was then electroporated into SE(pKD46) and transformant selected on Cm (8 μ g/ ml) at 42 °C (the non-permissive temperature for pKD46).

For further work, single colonies (12) were selected and propagated on L-agar plates containing Cm ($34 \mu g/ml$), and their Ap^S status (loss of pKD46) was confirmed.

5.2.1.3 Confirmation of the deletion mutants

The colonies (12 and 12 for SEN1432 and dgoR, respectively) obtained above were subject to colony PCR (as described in section 2.4.4; Figs. 5.7 and 5.8). The primers used were designed to anneal to the DNA regions ~100 bp upstream and downstream of the corresponding target gene. As shown in agarose gel electrophoresis analysis (Fig. 5.12), the target sequences were amplified successfully for the wildtype, giving bands at ~900 and 930 bp for dgoR and SEN1432 genes respectively, which correspond to the sizes of the target fragments, and indicate that the PCR was successful.



Figure 5.12: Gel electrophoresis of PCR confirmation products of SEN1432 and *dgoR* genes using wildtype chromosomal DNA as template. Lane 1 GeneRulerTM 1kb ladder (250-10,000 bp). PCR products are as follow: lanes 2 & 3, *dgoR* (expected size 900 bp); lanes 4 & 5, SEN1432 (expected size 930 bp). Electrophoresis was as in Fig. 5.10.

Figures 5.13 and 5.14 show the confirmation of the (Δ SEN1432)::*cat* and (Δ *dgoR*)::*cat* genotypes by PCR. The presence of DNA bands of the expected size in the mutants (~1200 bp) compared to DNA bands of expected size in wild type *S*E (900-930 bp) indicates that the

Chapter 5

 Cm^{R} -cassette replacement had occurred. Isolates 2, 3 and 4 for the *dgoR* inactivation did not yield any PCR product and so were discarded, and isolates 10-13 gave products matching that of the wildtype. Thus, isolates #7 and #8 were used for further work as they gave the expected PCR product. For SEN1432 mutation, all 12 isolates gave the expected PCR product.



Figure 5.13: Colony PCR to confirm the (Δ SEN1432)::*cat* mutation. Lanes 1 & 9 GeneRulerTM 1kb ladder (250-10,000 bp). PCR products are as follow: lanes 8 & 16 wildtype (expected size 930 bp); lanes 2-7 & 10-15 (Δ SEN1432)::*cat* candidates (expected size ~1200 bp). Electrophoresis was performed as in Fig. 5.10.



Figure 5.14: Colony PCR to confirm of the ($\Delta dgoR$)::*cat* mutation. Lanes 1 & 9, GeneRulerTM 1kb ladder (250-10,000 bp). PCR products are as follow: lane 10, wild type (expected size 900 bp); lanes 1-8 & 11-16 ($\Delta dgoR$)::*cat* (expected size ~1200 bp) candidates. Electrophoresis was as in Fig. 5.10.

5.2.1.4 Removal of the Cm^R cassette from the single-gene knockout strains

In this step, the chloramphenicol resistance genes were removed from the single-gene knockout strains generated above, as described in section 2.2.15.6. This is optional to some extent, but is important when additional Cm^R mutations are required and eliminates any complications associated with the presence of the *cat* gene, such as polarity effects on downstream genes. Therefore, the Cm^R cassette present in the mutant strains was removed. This was achieved by transformation with pCP20 plasmid into #7 and #8 isolates for SEN1432 and *dgoR* respectively; this is an Ap^R plasmid that exhibits Ts replication and thermal induction of FLP synthesis. The Flippase recognition target (*frt*) sites enable site-specific recombination at such sites and the subsequent loss of the Cm^R cassette.

pCP20 transformants were selected for resistance to ampicillin. Several Ap^R transformants were grown on non-selective medium at 44 °C for 48 h and then tested for loss of all antibiotic resistance. The deletion of the Cm^R gene was confirmed by colony PCR (2.2.5) and using primers indicated in Table 2.6. Fig. 5.15 illustrated the sizes of the PCR products obtained for both the putative Δ SEN1432 and Δ *dgoR* mutants. Fragments of the expected sizes were obtained (~300 bp) thus confirming the loss of the Cm^R cassette in all 11 isolates tested mutants. Isolates 6 and 5 for the Δ SEN1432 and Δ *dgoR*, respectively.



Figure 5.15: Colony PCR to confirm the loss of the Cm^R cassette from the SEN1432 and *dgoR* mutants. Lanes 1 & 9 GeneRulerTM 1kb ladder (250-10,000 bp). PCR products are as follow: lanes 2-7, putative Δ SEN1432 isolates; lane 8 (Δ SEN1432)::*cat* strain; lanes 10-14, putative Δ *dgoR* isolates; lane 15 (Δ *dgoR*)::*cat* strain; lane 16 wildtype (expected size 900 bp). Electrophoresis was performed on 1% agarose gel and at 60 V for 70 min. Lanes 2-8, SEN1432 PCR; lanes 10-16, *dgoR* PCR.

5.2.2 Phenotypic analysis of SE mutants

Following removal of the Cm^R cassette from mutants, the resulting strains (*S*. Enteritidis Δ SEN1432 and *S*. Enteritidis Δ dgoR) were used for further phenotypic analysis. To do this, the corresponding fusions pRS-SEN1436-*lacZ*, pRS-SEN1432-*lacZ* and pRS-*dgoR-lacZ* were introduced into the respective *S*E mutants and the wildtype (section 2.2.13 and 2.2.14). The expression levels of the *lacZ* fusions were then tested in both the wildtype and mutant strains to investigate the effects of the loss of the regulators on EW and lysozyme induction.

As observed from Fig. 5.16, SEN1432 deletion showed a significant impact on the expression level of SEN1432 and SEN1436 with an approximately twofold increase in expression observed in the presence of either EW or lysozyme when the regulator was absent. A similar effect was seen for SEN1432 in M9 medium or with EWF, although such an effect was not seen for SEN1436, possibly because expression levels were relatively low in addition to the of
impact of using a plasmid based promoter to determine the effect of a chromosomal knock out of a repressor on the efficiency of repression. Thus, the results are consistent with SEN1432 acting as a repressor for both SEN1432 and SEN1436. The absence of SEN1432 had no major effect on EW or lysozyme induction of SEN1432 and SEN1436. Thus, it seems unlikely that SEN1432 regulates the SEN1432-6 genes in response to the lysozyme of EW.



Figure 5.16: Comparison of the expression of SEN1432 and SEN1436 in the wildtype and Δ SEN1432 strain in response to the of presence egg white, lysozyme and egg white filtrate. Growth was aerobic and carried out at 42 °C with continuous shaking. Lysozyme was from hen egg, and was at 0.1 mg/ml. Growth medium used was 0.4% glycerol in M9 medium and strains were introduced to the same medium with 10% EW or 1 mg/ml lysozyme, fresh glycerol M9 medium (control) or to 100% EWF. Samples were harvested for β -galactosidase activity measurement at 45 min. P value ≤ 0.5 indicated with asterisks.

Figure 5.17 show that *dgoR* deletion also caused a significant impact on the expression level of the *dgo-lacZ* fusion. The results show a ~3 fold increase in *dgoR-lacZ* expression caused by lack of DgoR. This effect was observed with EW and with lysozyme, and was also seen in M9 medium as well as in EWF, although effect was less intense in EWF (~twofold). Thus, the data are consistent with DgoR acting as a repressor for the *dgoRKAD* operon. However, the degree

of induction by EW and lysozyme was not affected by the lack of DgoR which in both the wild type and mutant show a threefold induction which clearly indicates that DgoR does not mediate the *hex* gene induction effect observed with EW lysozyme. Thus, it appears likely that another regulator, outside of the set of *hex* genes considered here, is responsible for the observed lysozyme induction.



Figure 5.17: Comparison of the expression of dgoR in the wildtype and $\Delta dgoR$ strain in response to the presence of egg white, lysozyme and egg white filtrate. Details are as in Fig. 5.17.

5.2.3 Complementation of the SEN1432 and *dgoR* mutants.

To confirm that the SEN1432 and dgoR gene deletions are indeed responsible the increased gene expression observed above, it was necessary to generate complementing plasmids carrying the SEN1432 and dgoR genes. Therefore, specific primers were used to amplify the SEN1432 and dgoR genes. The forward primers were designed to include at least 100 bp

upstream of the start codon so that the promoter of each gene (as included in the lacZ fusion constructs) would be expected to be present within the amplified region (see section 2.2.4).

Agarose gel electrophoresis (Fig. 5.18) showed that the target sequences were amplified successfully with bands of the expected sizes generated: 1276 and 1082 bp for dgoR and SEN1432, respectively.



Figure 5.18: Gel electrophoresis of PCR amplification products of the *dgoR* and SEN1432 genes. Lanes 1 GeneRulerTM 1kb ladder (250-10,000 bp). PCR products are as follow: lanes 2 & 3, *dgoR* gene (expected size 1276 bp); lanes 4 & 5, SEN1432 gene (expected size 1082 bp). Electrophoresis was as in Fig. 5.10.

The resulting PCR products were subjected to double digestion with *Bam*HI and *Eco*RI (section 2.2.8) and were then purified by extraction from the gel (section 2.2.9). The purified fragments were then ligated with the medium-copy vector, pSU18 (2300 bp, Cm^R , see Fig. 5.19), which was also double digested with same enzymes, according to section 2.2.10 (Fig. 5.22). The pSU18 plasmid has the pACYC184 *oriV* with the *lac* promoter directing transcription across the multiple cloning sites (Bartolome *et al.*, 1991, see Fig. 5.21); this plasmid is compatible with the pRS1274 vector.



Figure 5.19: Restriction map of pSU18. Modified from Bartolone (1991).



Figure 5.20: Electrophoretic analysis of pSU18 plasmid following double digestion with *Bam***HI** *and Eco***RI**. Lane 1, Fermentas GeneRuler[™] 1kb ladder. Lane 2, undigested pSU18 vector; lanes 3, digested pSU18 vector. Electrophoresis was as in Fig. 5.10.

The ligation products were used to transform *E. coli* TOP10 and Cm^R transformants isolated. Plasmid DNA was then isolated and examined by (Fig. 5.21). Of the eight plasmids isolated, six had mobilities indicative of the presence of an insert, however, plasmids 5 and 6 showed mobilities consistent with no insert. Subsequent nucleotide sequencing with the forward and reverse primers showed that all plasmids (two of each type) submitted carry the expected inserts and that the nucleotide sequences exhibit a 100% match to SEN1432 and *dgoR*.



Figure 5.21: Electrophoretic analysis of pSU18 clones suspected to carry *dgoR* and **SEN1432 PCR fragments.** Lane 1, Fermentas GeneRulerTM 1kb ladder. Lanes 2-5 undigested pUS18-*dgoR* candidates; lanes 6-8, undigested pSU18-SEN1432 candidates. Electrophoresis was as in Fig. 5.10.

The plasmids were extracted from each SE transformants colonies and then analysed by double restriction digestion using *Bam*HI and *Eco*RI as described in section 2.2.8. As shown in agarose gel electrophoresis analysis (Figure 5.22), except for plasmids 5 and 6, all others showed the expected size of inserted fragment (~1.1 and 1.3 kb).



Figure 5.22: Electrophoretic analysis of double-digested (*BamHI/EcoRI*) pSU18-*dgoR* and -SEN1432 candidates. Lane 1, Fermentas GeneRulerTM 1kb ladder. Lanes 2-5, digested pSU18-*dgoR* candidates; lanes 6-7, digested pSU18-SEN1432 candidates. Electrophoresis was as in Fig. 5.10.

The resulting complementation plasmids (pSUSEN1432 and pSUdgoR) were transformed into SE strains carrying pRS-SEN1432-lacZ and pRS-dgoR-lacZ to generate double transformants. Expression studies were then performed, as before, in 0.4% glycerol M9 medium, with EW or lysozyme, in order to determine whether the complementation plasmids can reverse the raised dgoR and SEN1432 expression effects seen previously. From Fig. 5.23, expression of SEN1432 and dgoR fusions were again ~two and threefold increased by the corresponding SEN1432 and dgoR mutations, respectively. However, the inclusion of the appropriate complementation plasmids largely reversed this effect with expression levels returning to levels just above those observed in the wildtype: for dgoR expression, the 3.1-fold increase caused by the dgoR mutation was reduced to 1.25-1.5 fold by complementation; for SEN1432, the 2.1-2.4 fold increase in expression was lowered to 1.3-1.4 fold by complementation.



Figure 5.23: Effect of complementation of the *dgoR* **and SEN1432 mutations on SEN1432**-*lacZ* **and** *dgoR*-*lacZ* **expression.** Conditions were as described in Fig. 5.17 & 5.18, except for the inclusion of the complementing plasmids, pSU*dgoR* or pSUSEN1432, as indicated. This experiment was repeated twice and similar results were obtained.

5.3. Discussion

To summarise the findings above, deletion of the SEN1432 or *dgoR* genes caused a moderate increase in the expression of the SEN1432- and SEN1436-, and dgoR-lacZ fusions, of ~two and threefold, respectively, indicating a role for their GntR-like products in repression of the corresponding genes. Complementation with plasmid-borne versions of the SEN1432 and *dgoR* genes largely reversed the increased expression caused by the mutations. The regulatory mutations (SEN1432 and dgoR) did not affect induction by EW lysozyme, indicating that neither DgoR nor SEN1432 are involved in the induction of the *hex* genes by EW lysozyme. Both SEN1432 and DgoR are GntR-like transcriptional repressors (Fig. 5.24) with common structural organisations. Coward et al. (2012) investigated the role of hexonate uptake and catabolism in SE colonization of the chicken reproductive tract, the results show the deletion of the genomic island locus (SEN1432-SEN1436) decreased the bacterial load in the spleen by 14 days post infection suggesting a minor role in systemic colonization for this cluster, although its precise purpose remains unclear. Comparison of the S. Enteritidis PT4 and S. Typhimurium LT2 genomes (Thomson et al., 2008) showed a PT4 specific region ('ROD13') corresponding to the SEN1432–SEN1436 (6 kb) locus. Although absent in the LT2 strain, this locus is present in the chicken pathogen, S. Gallinarum. The reason for the absence of this locus in LT2 is unclear, but the results of Coward et al. (2012) suggest the possibility of a specific role in chicken reproductive tract (and, by inference, egg) colonisation. However, the SEN1432-36 genes show sequence similarity as well as synteny to the genes of the gntII locus of E. coli, which are absent in SE PT4 (Fig. 4.5). The GntII system is involved in L-idonate catabolism (Bausch et al. 1998) suggesting a similar function for the SEN1432-36 genes. Another study showed that several genes are upregulated (2.5-3.5 fold) in operons involved in the transport and metabolism of D-galactonate (dgo), D-gluconate (gntU, kdgT, and kduD), and L-idonate (idn) in SE in softened leaf tissue in cilantro and lettuce soft rot lesions; this finding was considered indicative of the catabolism of these hexonate substrates within these leaf tissue environments (Goudeau *et al.*, 2013).



Figure 5.24: Schematic of GntR family protein domains. DBD, DNA binding domain; EBD, effector binding domain or 'FadR-like C-terminal Domain' (FCD). wHTH, winged helix-turn-helix domain which interacts with a consensus sequence in the operator (*N* is any nucleotide and *n* is any number) (Jain, 2015).

Any further work should analyse the effects on the remaining *hex* gene *lacZ* fusions in each of the mutants to investigate more completely the regulatory influences of the GntR-like regulators. In addition, experiments with mutations in other relevant regulatory genes (e.g. *gntR*, *idnR*) should be included and analysis of the effects of the various hexonates on *hex* gene expression with each regulatory mutant should be performed to investigate how the impact of these hexonate on *hex* gene expression is affected by absence of these regulators. In this way, it should be possible to define the effectors that each responds to. Studies on the effect on *hex* gene expression of multiple deletions of the genes encoding the regulators of relevance would contribute to further understand the regulatory processes governing the expression of the *hex* genes. Isolation of the DgoR and SEN1432 proteins would enable direct DNA and ligand binding experiments to proceed which would extend and support the work with the *lacZ* fusions. Further, the use of a wider range of hexonates, and use of hexuronates, would allow a more comprehensive understanding of substrate specificities of the various Hex systems of *SE*.

Chapter 6: The Role of the Two-Component Regulators, PmrAB and PhoPQ, in mediating the *hex* gene response to lysozyme

6.1 Introduction

In addition to the known enzymatic hydrolysis activity of lysozyme against the peptidoglycan layer of Gram-positive bacteria, lysozyme also shows the ability to disrupt the bacterial membranes of Gram-negative bacteria, to inhibit the synthesis of DNA or RNA and to induce autolysin production. Therefore, lysozyme can affect Gram-negative bacteria and is able to permeate both the outer and inner membranes of *E. coli*, depolarize the cytoplasmic membrane and cause cytosol leakage (Derde *et al.*, 2015).

In egg white, the lysozyme is considered to be more effective against bacteria due to the synergistic activity of other EW components. Such synergist components potentially include the chelating activity of ovotransferrin to remove metals associated with the lipopolysaccharide moieties of the outer membrane of Gram-negative bacteria which could disrupt this membrane and allow lysozyme access to the peptidoglycan layer (Baron *et al.*, 2015).

The expression data of Baron *et al.* (2017) were consistent with a considerable membranestress response imposed by EW on SE, a stress that is likely to be caused by lysozyme, in part at least. The genes thus up-regulated, that are related to membrane-stress, include degP (a periplasmic/membrane-associated serine endoprotease that degrades abnormal proteins), Tol-Pal system genes (involved in the maintenance of cell-envelope integrity) and *ompC* (encoding an outer-membrane porin). Raspoet *et al.* (2014) indicate a role for DegP in the survival of S. Enteritidis in EW at high temperatures, further suggesting that EW induces membrane stress in SE. In addition, several peptidoglycan hydrolase genes (dacC, dacD amiC, mltA, mltD, emtA, yfhD) were induced by EW exposure, which provides another indication for an envelope-stress response. Gantois *et al.* (2008) suggest that maintenance of cell-envelope integrity is a significant feature of resistance to EW, with cell wall disruption and progressive cell lysis reported as the major mechanisms of EW-mediated bactericidal action at 45 °C for *E. coli* (Jan *et al.*, 2013; Baron *et al.*, 2017); a similar effect can be anticipated for SE.

A typical Gram-negative bacterial envelope consists of the three main layers (plasma membrane, peptidoglycan and the outer membrane) (Fig. 6.1a) and general Structure of Salmonella LPS (Fig. 6.1b). The outer membrane is anchored to the peptidoglycan layer through a set of lipoprotein molecules consisting of two layers, a phospholipid layer on the inner side and a lipopolysaccharide (LPS) layer towards the outer side. This LPS comprises side chains anchored to a core LPS. The side chains are made up of repeating oligosaccharide units. The LPS layer is also known as endotoxin and serves as a major virulence factor and PAMP (pathogen-associated molecular pattern). The outer membrane is selectively permeable owing to the presence of specialized membrane proteins called porins. The second layer is the periplasmic space (containing one or two layers of peptidoglycan), which separates the outer membrane from the third layer (cytoplasmic/plasma membrane).



Figure 6.1a: A typical Gram-negative bacterial envelope components. https://biologywise.com/gram-negative-bacteria. Illustrated by Kalyani Dhake.



Figure 6.1b. General Structure of *Salmonella* **LPS.** Glc = glucose; GlcNac = N-acetyl- glucosamine; Gal = galactose; Hep = heptose; P = phosphate; Etn = ethanolamine; AraN= 4-amino-4-deoxyarabinose; KDO = keto-deoxyoctulosonate. Ra to Re indicate incomplete forms of LPS. https://www.hindawi.com/journals/jl/2012/475153/tab1/

Studies reported in chapter 4 show that lysozyme causes strong induction of the *hex* genes upon exposure of *S*E to EW; the most likely reason for the lysozyme-dependent induction observed would appear to be the release of an endogenous inducer/signal generated by *S*E in response to cell-envelope damage. Previous work has shown that two two-component regulatory systems (PhoP-PhoQ and PmrA-PmrB), that are activated *in vivo*, are necessary for resistance to antimicrobial peptides (Fig. 6.2a & b). These regulators control the introduction of modifications to the LPS that decrease antimicrobial-peptide binding to the envelope and reduce membrane permeability (Gunn, 2008). A set of PmrAB-regulated genes has been identified, and partly characterised, that provide antimicrobial-peptide resistance and induce the resulting LPS modifications. Roland *et al.* (1993) identified PmrAB from a mutant strain associated with resistance to polymyxin B (PMB). The *pmrCAB* operon encoding this twocomponent sensor-regulator (TCS) produces three protein products: a phosphoethanolamine (pEtN) phosphotransferase (PmrC) (also known as EptA or YjdB), a response regulator (PmrA or BasR) and a sensor kinase (PmrB or BasS). PmrAB regulates over 20 confirmed genes (and possibly up to 100 genes in total) in *Salmonella*, as determined by microarray, mutagenesis and in silico analyses (Marchal *et al.*, 2004; Tamayo *et al.*, 2005).

Bacterial two-component regulatory systems (TCSs) are key factors in the ability of microorganisms to sense and respond to changing environmental conditions (Gunn, 2008). Direct PmrAB activation is thought to be mediated by PmrB which is associated with the inner-membrane through two transmembrane helices and contains a short periplasmic segment of just 30 residues that might mediate its sensory activity. Known activating signals for PmrAB in *Salmonella* are extracellular ferric iron and aluminium (Al³⁺), and low extracellular pH (e.g. pH 5.5) (Zhou, 1999). PmrAB can also be indirectly activated through the PhoPQ TCS (Gunn and Miller, 1996). PhoPQ activates the expression of *pmrD* which produces a 9.6 kDa product that regulates PmrA activity at a post-transcriptional level, as PmrD binds to and stabilizes PmrA in its phosphorylated form (Kato and Groisman, 2004). *S.* Typhimurium, PmrA-PmrB activates gene expression in response to antimicrobial peptides (AP) (including PMB) that are encountered, for instance, in the phagosomes of professional macrophages and at the surface of the intestinal mucosa, to enhance AP resistance through LPS modification (Gunn *et al.*, 2000; Tamayo *et al.*, 2002).

Activation of PmrA-PmrB provides resistance to *ST* against different type of AP including polymyxin. In addition, specific conditions in eukaryotic cell vacuoles or phagosomes like low Mg and acidic pH can activate the PmrA-PmrB regulon in *Salmonella* (Wosten *et al.*, 2000; Tamayo *et al.*, 2002).



Figure 6.2a: A model of the activation and interaction of the PhoPQ and PmrAB TCSs in *Salmonella* **spp.** From Gunn (2008). The arrow with '+', whose product binds to and stabilizes PmrA in its phosphorylated state. IM, inner membrane. Note that PmrAB is known as BasRS in *E. coli*.



Figure 6.2b: Structure of SE LPS and targets for modification mediating resistance to PM. PmrA-P activates transcription of LPS modification loci (i.e. Wzz, PmrG, CptA, *ugd*, *pbgP*, and *pmrC*). The O-antigen synthesis is controlled by products of the *wzz* gene. The PmrG and CptA proteins are responsible for the phosphorylation modification of heptose (I) and heptose (II) (blue segments), respectively. Lipid A (red part) can be phosphorylated with phosphoethanolamine (pEtN) trough the activity of PmrC or L-4-aminoarabinose (L-Ara4N) through the action of Ugd and PbgP. P: phosphorylated (from Yu *et al.*, 2015).

One of the primary roles of PmrAB activation is LPS modification. These modifications mask phosphate groups with positively charged moieties, affecting the electrostatic interaction of certain cationic APs (e.g. polymyxin) with the bacterial cell surface dramatically (Gunn, 1998). It should be noted that the pI of lysozyme (N-acetylmuramoylhydrolase) from chicken EW is unusually high at 11.35 (Wetter & Deutsch, 1951) with an optimal activity at pH 9.2 (matching the pH of egg white after laying) (Davies *et al.*, 1969). Thus, lysozyme is highly cationic and as such the PmrAB response might be expected to lessen lysozyme association with the outer membrane.

Polymyxins are a type of non-ribosomal cyclic, lipopeptide, cationic, antibiotic produced by certain Gram-positive bacteria. They were originally discovered in 1947 and since 1959 polymyxin E has been used for the treatment of Gram-negative bacterial infection (Yu *et al.*, 2015). They bind to the to the outer membrane LPS of Gram-negative bacteria disrupting both inner and outer membranes, probably via a 'detergent-like' action. There are three antibacterial pathways for polymyxin activity (Fig 6.3); membrane lysis causing death, vesicle-vesicle contact and hydroxyl radical death (Yu *et al.*, 2015).

The first pathway involves the selective binding of polymyxin to LPS causing loss of integrity of the phospholipid bilayer of the cytoplasmic membrane (CM) through membrane thinning, by straddling the interface of the hydrophilic head-groups and fatty-acyl chains, leading to CM lysis and cell death (Yu *et al.*, 2015). The alternative 'vesicle-vesicle' pathway is believed to occur when polymyxin binds to <u>both</u> the anionic phospholipid vesicles, namely the inner phospholipid leaflets of the OM and CM, promoting the exchange of phospholipids between vesicles causing the loss of specificity of phospholipid composition (Yu *et al.*, 2015). The third pathway is 'hydroxyl radical death' through the accumulation of hydroxyl radicals causing oxidative stress due to polymyxin induced formation of reactive oxygen species (Yu *et al.*, 2015).



Figure 6.3: Antibacterial mechanisms of polymyxin: (a) classic mechanism of membrane lysis; (b) alternative mechanism of vesicle-vesicle contact (Yu *et al.*, 2015). The polymyxin is coloured as magenta. LPS: lipopolysaccharide.

According to a random transposon mutagenesis study (Tamayo *et al.*, 2002), there are three different phenotypic classes of genes regulated by PmrA-PmrB and/or PMB: those necessary for PMB resistance and regulated by PmrA; those necessary for PMB resistance and not regulated by PmrA; and PmrA-regulated genes not required for PMB resistance. PmrA-regulated loci so far identified include *dgoA* (a *hex* gene) and *yibD* (or *waaH*, encoding a LPS (HepIII)-glucuronic acid glycosyl transferase; Klein *et al.*, 2013), which demonstrated a 500-and 2,500-fold activation by PmrA, respectively (Tamayo *et al.*, 2002). However, according to Tamayo *et al.* (2002), both *dgoA* and *yibD* showed no effect on PM resistance, and no effect on resistance to high iron concentrations or virulence in the mouse model. *dgoA* showed no role in PmrA-regulated resistance to high iron concentrations, PMB and or in virulence in mice. For further characterization of the PmrA-regulated gene mutants, the promoter region of *dgoA* was analysed for the presence of a putative PmrA-binding site, but no consensus PmrA-binding sequence was identified for *dgoA*, either within the putative promoter region upstream of *dgoA* nor within the putative promoter upstream of *dgoA* and *vib*.

suggested that regulation of *dgoA* by PmrA may be indirect. Note that none of the other '*hex* genes' were shown to be PMB/PmrAB regulated (Tamayo *et al.*, 2002).

The above observations thus suggest that the *hex* genes induced in *S*E during exposure to EW in response to lysozyme might be under PmrAB control, which could in turn be PhoPQ dependent (Fig. 6.3).

6.1.1 Aims of this chapter

In this chapter, the impact of *pmrAB* and *phoPQ* mutation on *hex* gene EW/lysozyme induction in *Salmonella* was determined.

6.2 Transformation hex-lacZ fusions into Salmonella phoP and pmrA mutant strains

The first objective was to transform the transcriptional fusions created in chapter 3 into the *S*. Typhimurium (*S*T) *phoP* and *pmrA* mutant strains (see Table 2.2 for strain details). Six fusions were selected (SEN2977-, SEN2978-, SEN1432-, SEN1435-, SEN1436- and *dgoR-lacZ*). Previous results (chapter 4) showed that these genes are strongly (SEN1436, 1250 U; SEN2977, 1360 U), weakly (SEN2978, 350 U; *dgoR*,190 U, SEN1435, 200 U) and moderately (SEN1432, 740 U) expressed, and that SEN1436, SEN1432, SEN2977, and *dgoR* possess promoters that are EW induced about 60-, 21-, 13- and 21-fold, respectively. In addition, the microarray results indicated that four are induced by EW (not SEN2978 or SEN1432).

These fusions were electrotransformed (as described in section 2.2.13-14) into three strains:

wild-type ST ATCC 14028s (JSG210);

ST pmrA::Tn10d-Tc^R (JSG421); and

ST phoP::Tn10d-Tc^R (JSG425).

The wildtype and two mutants were prepared as a competent cells for electroporation as described in section 2.2.13.

Note that as the *pmrA* and *phoP* genes are upstream of the co-operonic *pmrB* and *phoQ* genes, respectively, it is expected that the corresponding downstream genes would not be well

Chapter 6

expressed. The plasmid DNA was extracted from transformants (section 2.2.2.1) to confirm their identity by agarose gel electrophoresis; DNA of the expected mobility was observed as shown in representative samples from each strain (Fig. 6.4).



Figure 6.4: Electrophoretic analysis of pRS1274 *hex* **gene fusions in** *S***T.** Lane 1, Fermentas GeneRuler[™] 1kb ladder; lanes 2-4, SEN1436-lacZ fusions in wild-type; lanes 5-7, SEN2977-lacZ fusions in JSG421; lanes 8-10, dgoR-lacZ fusions in JSG425. Electrophoresis was performed using a 1% agarose gel and at 60 V for 70 min.

6.3 Phenotypic analysis of ST transformants

The expression levels of the *lacZ* fusions were tested in both the wildtype and mutant strains to investigate the effect of *pmrA* and *phoP* deletion. Expression was measured with and without 0.1 mg/ml lysozyme (from chicken EW) and 50 μ g/ μ l polymyxin B.

As observed from Fig. 6.5, SEN1435 showed little response to the *pmrA* and *phoP* mutations, and was only weakly induced by PMB and lysozyme (2 and 3 fold, respectively). Indeed, none of the fusions showed any notable response to the *pmrA* and *phoP* mutations in the absence of PMB or lysozyme. However, five of the six (not SEN1435) showed strong PMB as well as lysozyme induction in the wildtype: SEN1432, SEN1436, SEN2977, SEN2978 and *dgoR* by 2.5-, 18-, 23-, 5- and 9-fold, respectively (Fig. 6.5). However, in all cases induction by lysozyme was greater than that by PMB in the wildtype, by an average of ~twofold.

Absence of PmrA resulted in loss of PMB induction for SEN1436, SEN2977 and SEN2978, and reduced PMB induction for SEN1432 and *dgoR* by ~twofold. Loss of PhoP also resulted in loss of PMB induction for SEN1436 and 2977 (but, unlike PmrA had no effect on SEN2978), a complete loss of PMB induction for SEN1432 (where lack of PmrA gave a partial loss), and a partial loss of PMB induction for *dgoR*. These results are thus consistent with roles for the PmrAB and PhoPQ systems in inducing expression of the *hex* genes in response to PMB. Given that the effects of the *pmrA* and *phoP* mutations on PMB induction are largely similar, this indicates that the induction effect is mediated through direct regulation by PmrAB in response to PhoPQ (acting as the primary sensor for PMB-induced membrane disturbance).

For lysozyme-dependent induction of the *hex* genes, the loss of PmrA or PhoP had a less dramatic effect, with on average only an ~twofold reduction in induction seen. SEN1432 showed a slightly lower lysozyme induction (2 fold) in the phoQ mutant, but no effect in the pmrA mutant, suggesting a partial dependence on PhoPQ, but not PmrAB. SEN1436 showed a major reduction in lysozyme induction (by 6.4-fold) in the absence of PmrA, but little effect when PhoP was absent. This indicates direct control of SEN1436 expression by PmrAB in response to lysozyme with little contribution by PhoPQ. This is in contrast to the response to PMB which appears to be directly PmrAB dependent, but also requires PhoPQ (presumably as the initial PMB sensor). For SEN2977, the response to lysozyme resembles its response to PMB in that loss of either PmrA or PhoP resulted in a similarly-diminished lysozyme induction (reduced by ~twofold); this indicates that PmrAB is acting as the direct regulator, as for the PMB effect, with PhopPQ likely acting as the direct sensor. SEN2978 showed a slightly reduced lysozyme induction (reduced by ~twofold) in the *phoP* mutant, but there was no effect on the lysozyme-induction response caused by absence of PmrA. For dgoR, loss of PmrA and PhoP gave similar reductions is lysozyme induction (3 and 2 fold reductions, respectively), which resembles the effect observed for PMB where both PhoPQ and PmrAB appeared to contribute to PMB induction. Thus, both PhoPQ and PmrAB appear to contribute to lysozyme induction of dgoR, but at least one other factor must also contribute. Testing both mutant together contribute in further confirmation for results.

In summary, the *hex* gene fusions are clearly subject to strong induction by PMB, which is dependent on the PhoPQ-PmrAB system. However, in general the response to lysozyme is only moderately controlled by these factors (around twofold) and thus the lysozyme response of the *hex* genes appears to be largely controlled by an additional, unknown regulatory pathway.



Figure 6.5: Graph showing expression of *hex* **genes in** *phoP* **and** *pmrA* **mutants in response to lysozyme and polymyxin.** Expression was measured as before (2.2.12) after 45 min incubation at 42 °C in ST. 210, wild-type; 421, *pmrA* mutant; 425, *phoP* mutant.

6.4 Is the effect of the *pmrA* mutation on *hex* gene expression reversed by *pmrAB* complementation?

In this section, the impact of *pmrA* mutation on *hex* gene expression was confrimed through construction abd utilisation of a complementing plasmid carrying the deleted *pmrAB* genes in the medium-copy vector pSU18.

6.4.1 Primer design and the amplification of *pmrAB*.

To confirm the role of PmrA on *hex* gene induction by PMB (and by lysozyme), the *pmrAB* locus was cloned in order to enable complementation of the *pmrA* mutant. Specific primers were used to amplify an appropriate fragment to incorporate the *pmrAB* coding regions as well as the upstream promoter (section 2.8). Primers were designed to anneal at least 150 bp upstream of the start codon of the upstream *yjdB/eptA* gene, and 150 bp downstream of *pmrB* (*basS*) so that the *eptA-pmrAB* operon would be present within the amplified region (Fig. 6.6) (see Appendix 9 for details) together with the promoter. PCR was as described in section 2.2.4. The genomic DNA of wild-type *ST* (JSG210) was used as a template.



Figure 6.6: Schematic representation of the *pmrAB* (*basRS*) genes of *ST* JSG210. *basR* is referred to as *pmrA*, and *basS* is referred to as *pmrB*. The target region for amplification is indicated inside the purple rectangle. https://www.ncbi.nlm.nih.gov/gene/1255818.

As shown by agarose gel electrophoresis (Fig. 6.7), the target sequences were amplified successfully with a band at approximately ~ 3712 bp for the *eptA-pmrAB* operon apparent, which corresponded to the sizes of the target fragment. This was purified using Thermo Scientific GeneJETTM PCR purification kits to remove any contaminants (section 2.2.7).



Figure 6.7: Gel electrophoresis of PCR amplification products of the *eptA-pmrAB* genes. Lane 1, GeneRulerTM 1kb ladder (250-10,000 bp); lane 2 & 3, *eptA-pmrAB* PCR product (expected size 3712 bp). Electrophoresis was performed in a 1% agarose gel and at 60 V for 70 min.

6.4.2 Cloning of eptA-pmrAB into pSU18

In order to create sticky ends for cloning the PCR fragment, the purified PCR product was ligated with the intermediate vector (pJET1.2) as described in section 2.2.10, and the ligations products were transformed into *E. coli* TOP10 (section 2.2.1). Five resulting colonies from the transformation plate were selected for plasmid DNA extraction (section 2.2.2.1). As shown in Fig. 6.8, bands were observed at ~3 kb; to confirm the presence of the inserted fragment, one isolate was digested with single restriction enzymes (section 2.2.8) which resulted in a linear fragment of the expected size (~6 kb). This plasmid was designated pJET*-pmrAB*. Its identity was further confirmed by sequencing (with T7-F and pJET1.2 reverse primer) and no errors were observed within the sequenced regions.



Figure 6.8: Gel electrophoresis of pJET-*pmrAB*. Lanes 1, GeneRulerTM 1kb ladder (250-10,000 bp); lanes 2-6, pJET-*pmrAB* isolates; lane 7, pJET-*pmrAB* single digest with *Bam*HI; lane 8, pJET-*pmrAB* single digests with *Eco*RI. Electrophoresis was performed as above.

The product insert was released from pJET-*pmrAB* by double digestion with *Bam*HI and *Eco*RI as described in section 2.2.8, purified by gel extraction (section 2.2.9) and then introduced into the medium-copy vector pSU18 (Cm^R), which was also double digested with same enzymes (*Bam*HI and *Eco*RI) (see section 2.2.10). The resulting ligations reactions were used to transform competent cells and a selection of the Cm^R colonies thus obtained were subjected to plasmid DNA isolation (section 2.2.2.1). These were analysed by agarose gel electrophoresis which indicated a mobility consistent with the presence of the ~3.7 kb insert (Fig. 6.9). The plasmids were then analysed by double restriction digestion using *Bam*HI and *Eco*RI, as described in section 2.2.8. As shown in Fig. 6.10, all plasmids showed bands of the expected size: a 3.7 kb insert and a 2.3 kb vector fragment. Subsequent nucleotide sequencing confirmed their identity. The plasmid was designated pSU18-*eptA-pmrAB*. Two step cloning shows efficiency higher than one step.



Figure 6.9: Electrophoretic analysis of putative pSU18-*eptA-pmrAB* clones. Lane 1, Fermentas GeneRulerTM 1kb ladder; lanes 2-6, undigested putative pSU18-*eptA-pmrAB* DNA. Electrophoresis was performed in a 1% agarose gel and at 60 V for 70 min.



Figure 6.10: Electrophoretic analysis of putative pSU18-*eptA-pmrAB* clones double digested with *Bam*HI and *Eco*RI. Lane 1, Fermentas GeneRulerTM 1kb ladder; lanes 2, 4 and 6 undigested plasmid DNA from isolates 1, 2 & 3, respectively; lanes 3, 5 and 7, double digested plasmid DNA from isolates 1, 2 & 3, respectively. Electrophoresis was performed on 1% agarose gel and at 60 V for 70 min.

6.4.3 Effect of complementation of the *ST pmrA* and *phoP* mutant, with pSU18-*eptA*-*pmrAB*, on *hex* gene induction by lysozyme and PMB

The pSU-*eptA-pmrAB* plasmid or vector control (pSU18) were transformed into the *S*T wildtype, *pmrA* and *phoP* strains carrying the SEN2977-, SEN1436- and *dgoR-lacZ* fusions (as described in section 2.2.14). The transformants were then tested for the effect of lysozyme and PMB on the expression of the three *hex* gene fusions, to determine whether the pSU18-encoded

Chapter 6



pmrAB (pSU18::*pmrAB*) could reverse the impact of the *pmrA* (or *phoP*) mutations on *hex* gene induction.

Figure 6.11: Graph showing effect of *pmrAB* complementation on the expression of three *hex* genes in *phoP* and *pmrA* mutant backgrounds, in response to lysozyme and polymyxin. The strains carry the corresponding *lacZ* fusion plasmids and either pSU-*eptA*-*pmrAB* or pSU18, as indicated. See Fig. 6.5 for details.

As observed in Fig. 6.11, the presence or absence of the complementing plasmid had very little impact on *hex* gene expression in the wildtype, with expression levels remaining similar in the vector control and *pmrAB*-complemented wildtype in all cases, with and without PMB or lysozyme. For SEN1436 expression in the *pmrA* mutant (using pSU-SEN1436-421), there was a clear reduction in the degree of PMB and lysozyme induction (by 6 and 12 fold, respectively) in comparison to levels seen in the wildtype. However, the provision of *pmrAB* in multicopy largely reversed this reduced PMB and lysozyme induction, such that induction levels were only ~twofold lower than seen in the wildtype. Thus, the *pmrAB* plasmid successfully

complemented the reduced expression seen in the *pmrA* mutant for this gene. In the *phoP* mutant, levels of SEN1436 induction by lysozyme were little affected. However, there was a ~9 fold reduction in induction caused by PMB (as seen previously; Fig. 6.5) cf, the wildtype vector control. This low PMB induction, caused by lack of *phoP* (425), was largely reversed by multicopy *pmrAB*, such that expression was only ~1.8 fold lower than seen in the wildtype control. This strongly suggests that the weak PMB induction of SEN1436 observed in the *phoP* mutant is caused by weak PmrAB activity. Thus, PmrAB is likely to be the direct regulator for SEN1436 in response to PMB. Given the modest impact of the *phoP* mutation of the lysozyme induction of SEN1436, in comparison to the relatively strong effect seen for the *pmrA* mutation, it is likely that PmrAB responds directly the lysozyme signal, but indirectly to the PMB signal (through PhoPQ).

For SEN2977, the effect of the *pmrAB* complementation was relatively modest (Fig. 6.11). The *pmrA* mutation resulted in a reduced induction by PMB and lysozyme by 4.5- and 3-fold, respectively, but, surprisingly, complementation by *pmrAB* little affected this reduced induction, with the only notable effect being an ~twofold increase in induction by PMB. The reason for this failure of the mutlicopy *pmrBA* to reverse the effect of the *pmrA* mutation is unclear, but could be related to artefacts caused by the multicopy nature of the complementation that might result in, for instance alter expression of other regulatory factors influencing SEN2977 expression. In the case of the *phoP* mutant, there was a ~twofold lowered PMB and lysozyme induction and this reduction was unaffected by multicopy *pmrAB*, thus suggesting that the effect observed is directly mediated by PhoPQ.

For dgoR, there was a 3-4 fold reduction in lysozyme induction caused by the pmrA mutation that was largely reversed by multicopy *pmrAB* (Fig. 6.11). The *pmrA* mutation little impacted PMB induction, as did addition of multicopy *pmrAB* to the *pmrA* mutant. In the *phoA* mutant, the degree of dgoR induction by PMS and lysozyme was reduced by ~twofold (Fig. 6.11).

This reduced lysozyme induction in the *phoP* mutant was largely reversed by multicopy *pmrAB*, but the lowered PMB induction was unaffected by multicopy *pmrAB*. The relatively low expression level of *dgoR* in the *phoA* strain in the presence of PMB, and the failure of multicopy *pmrAB* to reverse this effect suggests that PhoPQ directly induced *dgoR* in response to PMB.

It is again notable, that the lysozyme induction effect is only partly mediated by PmrAB and/or PhoPQ. The major regulatory effect described above are summarised in the model below. A greater degree of understanding the relatively complicated regulatory influences revealed here will require further experimentation, which should include identification of PhoP and PmrA binding interactions with *hex* gene targets.



Figure 6.12. Summary of the major regulatory effect exerted by PMB and lysozyme through PmrAB and PhoPQ. Model derived from data in Fig. 6.11. Arrows indicate activation/stimulation effects, blue arrows indicate an environmental signal, solid orange arrows indicate a direct transcriptional induction, and broken orange arrow indicates an enhanced activity effect.

6.5 Discussion

In this chapter, the polymyxin B and lysozyme induction of *hex* gene fusions was investigated in the presence and absence of two 2-component transcriptional regulators (PmrA and PhoP) that might have a role in mediating the *hex* gene response in EW. The PmrAB and PhoPQ systems coordinate the expression of genes that enhance survival under conditions where membrane integrity is threatened, through inducing modifications of the LPS that decrease AMP (anti-microbial peptide) binding and bacterial cell entry (Gunn, 2008). PmrAB was shown to induce dgoA by 500 fold (Tamayo et al., 2002); this finding led to the studies described in this chapter. The PmrAB system responds (directly or indirectly) to high extracellular levels of ferric iron or Al^{3+} and acidity, as well as AMPs like PMB (Zhou, 1999; Ryan et al., 2015; Tamayo et al., 2002). Interestingly, the yibD (waaH) gene is also greatly (2,500-fold) induced by PmrAB in response to PMB (Tamayo et al., 2002) and this gene specifies an enzyme (LPS(HepIII)-glucuronic acid glycosyltransferase) that incorporates glucoronate (a hexuronate) into LPS to improve resistance to SDS and other factors (e.g. PMB) that damage the outer membrane (Klein et al., 2013). This suggests the possibility that the reason for the induction of the *hex* genes by lysozyme (and PMB) might be to generate hexonate/hexuronates for addition to LPS in order to raise membrane resistance.

In this chapter *Salmonella* serovar Typhimurium (JSG210) and two isogenic mutants, *pmrA* (JSG421) and *phoP* (JSG425) were utilised, with six *hex* gene *lacZ* fusions in lysozyme and PMB assay. In general, the six *hex* genes studied showed strong induction by PMB, and this induction was (generally) markedly reduced or eliminated by loss of either PmrA or PhoP (Fig. 6.5). This strongly indicates that the *hex* genes have a role in protection against outer-membrane damage as elicited by exposure to PMB, and that their PMB induction depends on both PhoPQ and PmrAB. As indicated in Fig. 6.12, it is likely that PhoPQ acts as the direct sensor for PMB activity, and then activates PmrA through PmrD, resulting in PMB-induction

of the *hex* genes. The loss of either PmrA or PhoP, in general, only slightly reduced hex gene induction by lysozyme (by around twofold). Thus, it is clear that the response to lysozyme is distinct to that for PMB. In addition, the strong residual *hex*-gene response to lysozyme in the absence of PmrA or PhoP indicates that some other factor is mainly responsible to lysozyme induction. Candidates include RpoE and CpxAR since these regulators respond to envelope damage and there was strong evidence of their role in the regulatory response of *S*E to EW (Baron *et al.*, 2017). Another possibility is that the lysozyme response is controlled directly by both PhoPQ and PmrAB, such that absence of one or other system only weakly affects the induction observed. Thus, *pmrA phoP* double mutants should be employed in any future work to test for this possibility. It is interesting to note that PmrB has no extensive, recognisable sensory domain (just a short 30 residue segment in the periplasm) whereas PhoQ has a large periplasmic domain (~130 residues) that is presumed to respond to the various extracellular (periplasmic) signals that induce the PhoPQ regulon.

To confirm that the *pmrA* deletion was indeed responsible for the corresponding *hex-gene* regulatory effects, pSU-*eptA-pmrAB* complementing plasmid was generated and deployed. Subsequent expression analysis showed that provision of *eptA-pmrAB* in *trans* reversed the lack of induction by lysozyme and PMB of SEN1436 in the *pmrA* mutant, clearly supporting the direct control of this gene by PmrAB (Fig. 6.11 & 6.12). The results are also consistent with a direct response of the PmrAB system to lysozyme, and an indirect response to PMB via PhoPQ-mediated control (Fig. 6.12). For *dgoR*, the results suggested direct regulatory control by PhoPQ in response to PMB, and a degree of direct regulatory control by PmrAB in response to lysozyme. Thus, the control of the hex genes by the PmrAB and PhoPQ systems is complex, and involves additional factors. Such additional factors previously identified include Crp and the various GntR-related repressors responding to Hex compounds (Robert-Baudouy & Stoeber, 1973; Portalier *et al.*, 1980; Blanco *et al.*, 1986; Zeng *et al.*, 2001; see Fig. 4.5).

In summary, the results in this chapter clearly show that the *hex* genes are subject to PMB induction and that this is largely controlled by PmrAB-PhoPQ. However, the response to lysozyme is only partly controlled by these factors indicating the involvement of another regulator. The results are consistent with a role for the observed *hex* gene induction by lysozyme in preserving the integrity of the cell envelope. Further work is required using a mixtures of lysozyme and PMB to determine whether these factors induce gene expression in an additive fashion, which would confirm the use of distinct regulatory pathways for the response to these factors. Further, a double *pmrAB-phoPQ* mutant should be used to test the possibility that in the absence of one system, the other provides a compensatory activity for lysozyme-dependent induction. In addition, the possible role of CpxAR and RpoE in the observed lysozyme induction should be tested, particularly as both these systems were predicted to be activated in response to EW exposure (Baron *et al.*, 2017) and a potential CpxR site was identified in the promoter region of one *hex* gene by BPROM. It would be particularly interesting to perform a transcriptomic analysis of the effect of lysozyme on global expression in *SE*.

Chapter 7: Attempted overexpression of SEN1432 and dgoD

Two of the *hex* genes (SEN1432 and *dgoD*, encoding a predicted transcription factor and Dgalactonate dehydratase, respectively) were targeted for overexpression to enable isolation of the corresponding proteins for generation of antibodies to allow western blot analysis of expression, as a second approach for monitoring expression effects. Also, the isolated proteins could be studied for their biochemical/regulatory activities, encode a regulator and an enzyme respectively. The vector pET21a was used to overexpress the native and His-tagged version of the proteins from *E. coli* BL21/ λ DE3. Next, the His-tagged proteins was to be purified using Ni-affinity chromatography for further work (e.g. raising antisera), as indicated above. The native proteins were also to be purified using alternative chromatographic approaches.

7.1 Amplification of SE genes of interest.

One forward and two reverse primers were designed to amplify each gene (Table 2.9). The primers were designed to add restriction sites, *NdeI* and *Hin*dIII, which allows subsequent cloning into the overexpression vector pET21a (Appendix 10). SEN1432 uses the start codon TTG which is rarely used by *E. coli* (~1%) and has a translation efficiency 2-3x lower than ATG (Makrides, 1996). Therefore, the start codon was changed in the forwardprimer to ATG using as part of the generation of an *NdeI* restriction site to avoid this problem. The first reverse (21R) is used to produce the native protein His-tagged protein. The stop codon has been removed from the end of the gene such that to allowed translation is allowed to continue into pET21a during overexpression, which contains the 6x His (CAC) codons containing region of the vector, such that so a C-terminal His-tagged protein is produced.

While the second primer (28R), two stop codons (TAA) were are added to the end of the gene to terminate translation and produce the native protein during overexpression. Figure Table 7.1 shows the primerss names and combination used to amplify the target genes.

Gene	Primer combination used in PCR	PCR product	The protein produced during overexpression
SEN1432	SEN1432-FOR / SEN1432-21R	SEN1432-21R	His-tagged SEN1432
	SEN1432-FOR / SEN1432-28R	SEN1432-28R	Native SEN1432
dgoD	dgoD-FOR / dgoD-21R	dgoD-21R	His-tagged dgoD
	dgoD-FOR / dgoD-28R	dgoD-28R	Native dgoD

Table 7.1: The primers	combination used t	o amplify each	gene and the gen	e product during	overexpression.
		1 0	0 0	•	.

PCR was carried out to amplify the genes of interest (section 2.2.4) using genomic DNA of SE as a template. As shown in figure Fig. 7.1, successful amplification was observed by gel electrophoresis of where the PCR products of the expected sizes (were ~720 bp and ~1149 bp for SEN1432 and dgoD, respectively).



Figure 7.1: Gel electrophoresis of PCR amplification products of SEN1432 and *dgoD* genes. Lanes 1, 4 and 8,: GeneRulerTM 1kb ladder (250-10,000 bp); Lane lane 2: , SEN1432-21R (expected size 720 bp); lane 3: SEN1432-28R (expected size 720 bp); lanes 5-8:, *dgoD*-21R (expected size 1149 bp); lanes 9-11, *dgoD*-28R (expected size 1149 bp). Electrophoresis was performed on 1% agarose gel and at 60 V for 70 min.

All PCR products (dgoD-21R, dgoD-28R, SEN1432-21R & SEN1432R-28R) were purified to

remove contaminants (section 2.2.7).

7.2 Cloning amplified genes into the intermediate vector pJET1.2./blunt.

In order to create sticky ends for easy cloning of the PCR fragments, the purified PCR products were cloned cloning into an intermediate vector (pJET1.2) as described in section 2.2.10 and transformed into *E. coli* TOP10 (section 2.2.1). This step is also useful for confirming the fragments sequence through nucleotide sequencing service.

Three colonies from each transformation plate were selected for plasmid DNA extraction (section 2.2.2.1). As shown in figure Fig. 7.2, bands were observed at ~3 kb, corresponding to the expected size of pJET were present as expected size carrying the inserted fragments.



Figure 7.2: Gel electrophoresis of purified pJET-*dgoD***21R & 28R and pJET**-**SEN1432-21R & 28R.** Lanes 1,: GeneRulerTM 1kb ladder (250-10,000 bp); lanes 2-4, pJET-*dgoD***21R**; lanes 6-8: , pJET-*dgoD***28R**. Lanes lanes 8-10:, pJET-SEN1432-21R; lanes 11-13: , pJET-SEN1432-28R. Electrophoresis was performed on 1% agarose gel and at 60 V for 70 min.

To confirm the presence of the inserted fragments, 12 isolates were then digested with restriction enzymes *Hin*dIII and *Nde*I (section 2.2.8) and electrophoresed in order to confirm the presence of the inserts. Three bands were expected for each sample: pJET has a *Hin*dIII restriction site 253bp downstream of the MCS, so bands of ~3 kb and ~250bp were expected, plus a band at either ~720bp (SEN1432) or ~1200bp (*dgoD*), depending on the insert.

In figure Fig 8.3, the electrophoresis of the double digested of the constructed pJET1.2 plasmid DNA shows the expected sizes of bands confirming that SEN1432-21R & 28R had been successfully cloned into all pJET isolates. While the figure Fig. 8.4 shows bands of the expected size for pJET fragments, plus two unexpected bands (A) ~800bp and (B) ~500bp. The bands in lanes 8 & 9 are faint, probably due to low sample concentration. There were no bands at ~1200 bp, the expected size of *dgoD*. This is due to the presence of additional internal restriction sites of *Hind*III & and *Nde*I in the *dgoD* sequence. These sites were not recognised during *in-silico* analysis which may be a technical error in the software program, and this caused problems extracting the insert from pJET for subcloning to pET21a. Partial digestion was attempted in order to release the insert without cutting it, but due to time constraints, this was not achieved.

Isolate #1 of each construct SEN1432-21R and SEN143-28R were also sent for sequencing in order to confirm the identity of the inserts. Mutations can occur during PCR amplification which may prevent the protein being translated correctly during overexpression, so it was important to confirm the identity of the inserts before proceeding. The sequences of our constructs were aligned with the sequences of the genes of interest showing a 100% match (no conflicts) between the insert sequence and the expected sequence of target genes (see Appendix 11).

Bands at ~720 bp corresponding to the expected size of SEN1432 (Fig. 7.3, indicates inside rectangles) were extracted from the gel as described in section 2.2.9 for cloning into pET21a overexpression vector.



Figure 7.3: Gel electrophoresis of pJET-SEN1432-21R & 28R digested with *Hind***III &** *Nde***I. Lanes 1 & 5,: GeneRulerTM 1kb ladder (250-10,000 bp); lanes 2-4, pJET-SEN143-21R; lane 9 undigested pJET, lanes 6-8, pJET-SEN1432-28R. Electrophoresis was performed on 1% agarose gel and at 60 V for 70 min.**



Figure 7.4: Gel electrophoresis of pJET- *dgoD*-21R & 28R digested with *Hin*dIII & *Nde*I. Lanes 1 & 6,: GeneRulerTM 1kb ladder (250-10,000 bp); lanes 2-4, pJET- *dgoD* -21R; lane 5 undigested pJET, lanes 7-9, pJET- *dgoD* -28R. Electrophoresis was performed on 1% agarose gel and at 60 V for 70 min.

7.3 Cloning amplified genes into overexpression vector pET21a

In order to construct the overexpression plasmids vector, the purified fragments of SEN1432-21R and & 28R, which possess *Hind*III & *Nde*I ends, were cloned ligated with into digested pET21a (amp^R) overexpression vector as described in section 2.2.10. Then, the constructed plasmids were transformed into competent *E. coli* TOP10 as described in section 2.2.1. Before transformation use in cloning, the identity of the prepared pET21a DNA (5443 bp) was confirmed using by single digest using restriction mapping with *Xho*I digestion and by double digest using *Hin*dIII & *Nde*I (Fig. 7.5).



Figure 7.5: Gel electrophoresis of pET21a digested with *XhoI*, or *HindIII & NdeI*. Lane 1, GeneRulerTM 1kb ladder (250-10,000 bp); lane 2, undigested pET21a; lane 3, single digested pET21a with *XhoI*; lane 4, double digested pET21a with *HindIII & NdeI*. Electrophoresis was performed on 1% agarose gel and at 60 V for 70 min.

As shown in figure Fig. 8.6, the expected size of pET21a bands in different treatments was obtained. The double digested form was extracted from the gel, as described in section 2.2.9, in preparation forming to use it in cloning. The inserts isolated above were then cloned ligated into with the purified and digested pET21a DNA (section 2.2.10), and the reaction products were then transformed into competent *E. coli* XL1-blue, as described in section 2.2.1). XL1-blue
was chosen in order to propagate the plasmids and it is *endA* deficient so it should provide high quality plasmid DNA (Stratagene, 2004).

Five transformants isolates for each form cloning of vector 21R and 28R were selected for double digestion with *Hin*dIII & and *Nde*I (section 2.2.8), followed by analysis by DNA electrophoresis (figure Fig. 7.6) to confirm that they contained the insert. Bands of ~5.5 kb and 0.72 kb0bp were expected, corresponding to pET21a and SEN1432 respectively. However, not all showed transformants show the insert which might be due to the additional restriction site and this could also explain why there were no bands of ~750 bp in lanes 3, 6, 9, 11, 12 & 13 12.(lanes 2, 4, 5 and 10).



Figure 7.6: Gel electrophoresis of potential pET21a-SEN1432-21R & and pET21a-SEN1432-28R isolates digested with *HindIII & NdeI*. Lanes 1 & 8,: GeneRuler[™] 1kb ladder (250-10,000 bp); lanes 2-6, digested pET21a-SEN1432-21R; lane 7, undigested pET21a-SEN1432-21R; lanes 9-13, digested pET21a-SEN1432-28R; lane 14, undigested pET21a-SEN1432-28R. Electrophoresis was performed on 1% agarose gel and at 60 V for 70 min.

The plasmids constructs those showing the presence of the insert were sent submitted to Eurofins Genomics for sequencing. Sequence alignment was carried out (see Appendix 11), which. The identity of the insert was confirmed the correct sequence for 28R#10 to give a predicted as the sequence matched the template. It's found to be missing the double stop codon

at the end of the gene which would result in the His-tagged protein being produced during upon overexpression. Two of 21R constructs (#2 & #4) showed unexpected mutations in sequences, which are suspected to be errors during sequencing, as the concentration of the samples was lower than recommended (avg. conc. $23ng/\mu$ l), but because the identity could not be confirmed, these samples were not used. However, for #5 was confirmed the sequence matched that of the template (see Appendix 11). Thus, the pET21a-SEN1432-28R#10 and -21R #5 constructs were used to the next step through to transforming them into *E. coli* BL21 (λ DE3) as described in section 2.2.1. The transformants colonies obtained were confirmed to contain pET21a using by plasmid DNA extraction (2.2.2.1).

7.4 Small-scale overexpression of SEN1432.

Small-scale overexpression of SEN1432 was carried out using overexpression strain in *E. coli* BL21(λ DE3). *E. coli* BL21 (λ DE3) has a T7 RNA polymerase gene under control of a *lac* promoter which is induced by 1 mM IPTG, allowing. When BL21(DE3) is transformed with a recombinant pET21a vector and grown in media with added IPTG, induction of the T7 polymerase promoter drives expression of the target gene from pET21a.

The pET21a-SEN1432-28R#10 and -21R #5 transformants of *E. coli* BL21 (λ DE3) were streaked out on ampicillin LA plates and grown overnight at 37 °C. Single colonies were selected, and following propagated in a small-scale overexpression experiment (section 2.2.17.8), and SDS-PAGE was used to analyse the samples (0.5 OD units) obtained for protein overexpression content which 0.5 ODs of lysed cells taken (section 2.2.17.6). The bands between 25.0 and 35.0 kDa were expected, as the size of the His-tagged SEN1432 polypeptide was calculated expected to be 28.7 kDa.

As shown in Figs. 7.7 and 7.8 for the 28R#10 and 21R #5 transformants respectively, the expected band does was not appear apparent indicating that overexpression did not work was unsuccessful. The protocol results obtained were compared to those achieved with a negative

was verified by using negative (empty vector, BL21(DE3) with- pET21a) and positive control (BL21(DE3) with- pET21a-N-term-*mbfA*) which showed good production of MbfA at the expected mass, and no notable difference between the negative control and the pET21a-SEN1432 samples (data not shown). In an attempt to enable overexpression, two recipient other host strains were used. Rosetta® (Novagen, 2011DE3) (Novagen, 2010) which provides tRNAs for 6 rare codons: (AUA, AGG, AGA, CUA, CCC, and GGA) and BL21(DE3)Star® (Invitrogen, 2010) which increases mRNA stability due to being deficiency in RNAseE. However, negative and no bands were no corresponding to the expected size of SEN1432use of these strains failed to improve expression (data not shown).



Figure 7.7: SDS-PAGE analysis SEN1432 overexpression from of BL21/DE3-(pET21a-SEN1432-21R) (#5) **following overexpression.** Lane 1, Fermentas unstained protein ladder; lane 2, before adding IPTG induction; lane 3-8 - 1, 2, 3, 4, 6 & 16 h post induction. Electrophoresis was carried out in a 15% polyacrylamide gel at 60 mA for 80 min.



Figure 7.8: SDS-PAGE analysis SEN1432 overexpression from of BL21/DE3 (pET21a-SEN1432-28R). Lane 1, Fermentas unstained protein ladder; lane 2, before induction; lanes 3-8 - 1, 2, 3, 4, 6 & 16 h post induction. Electrophoresis was carried out in 15 % polyacrylamide gel at 60mA for 80 min. See Fig. 8.7 for details.

Further analysis of SEN1432 shows that codon usage bias could be behind explains the failed overexpression. It contains of the 37 rarely used (<10%) codons that are used <10% of the time by *E. coli*, of which only seven supported with by that RNAs by Rosetta strain. The presence of rare codons in mRNA can cause transcription to terminate prematurely, which negatively affects protein expression. There is a correlation between gene expression levels and codon bias (Gouy and Gautier, 1982). Codon usage analysis using GenScript confirmed that SEN1432 contains a relatively high number of rare codons, compared to a gene that is highly expressed by *E. coli* (*ompC*) (figure Fig. 7.9). The SEN1432 encodes as is a regulator which is likely to target only a few operators, its normal expression level is predicted to be expected to be low, and it was therefore suspected to contain rare codons. Fig. 8.9A & C show the relative codon usage frequency along the gene for SEN1432 and *ompC*, respectively. Fig. 7.9B & D show the distribution of codons in the 'codon quality groups', where codons with values lower than 30 are likely to negatively affect protein expression. Thus, the failure to achieve overexpression of SEN1432 likely relates to its poor codon usage; this problem could be

Chapter 7

corrected by codon optimisation in any future work. In addition to the GC content of the gene as <30% or >70% will negatively affect transcription and translation efficiency.



Figure 7.9: Codon usage analysis of SEN1432 and *ompC* **generated by GenScript.** Figures A & C show the relative codon usage frequency along the gene. Figures B & D show the % distribution of codons in the codon quality groups.

7.5 Discussion.

The aim was to overexpress two SE hex genes, SEN1432 and dgoD, in order to discover more about their function. The genes chosen were: dgoD, which encodes D-galactonate dehydratase, and SEN1432, which encodes a suspected regulator of the SEN1432-6 cluster. In order to create constructs for overexpression of native and His-tagged versions of the proteins, the genes of interest were amplified and cloned into an intermediate vector, pJET1.2. The presence of *Hin*dIII & *Nde*I restriction sites within *dgoD* meant that the gene was digested during release of the insert from pJET1.2 for subcloning into pET21a, and so it was not possible to clone the gene into pET21a. SEN1432 was subcloned into pET21a successfully, however, the overexpression of SEN1432 from E. coli BL21(DE3) in both its native and His-tagged from failed. Sequencing showed that the correct gene sequence was present in the constructs, in the correct cloning context, it is thus assumed that the transcript was produced but that translation was poor. Several problems could affect translation. During initiation of translation the Shine-Delgarno (SD) sequence recruits the ribosome to the mRNA and aligns it with the start codon – in E. coli the consensus sequence is UAAGGAGGUGA and it is spaced -2 to -15 bp from the start codon AUG (optimum -8) (McCarthy and Tuite, 2013). Translation can be negatively affected if the SD sequence is spaced too close or too far from the start codon, however, sequencing of our constructs shows that the SD sequence of pET21a is located at -9 bp from the start codon of the SEN1432 gene, so this should not be a problem. During elongation, the formation of secondary structures (stem-loop structures) can also affect translation, as it interrupts the activity of the ribosome (Hall et al., 1982). Higher translation efficiency has been linked to high A/T content downstream of the start codon, as high G/C content is associated with secondary structures (Qing et al., 2003). The SEN1432 gene has a relatively low G/C content at the start of the gene (see Appendix 12) so secondary structures should not be causing a problem with translation. Likewise, the second codon has been shown to affect translation in *E. coli*, but the second codon of SEN1432 (AGC) is not associated with poor translation (Looman *et al.*, 1987). Therefore, as suggested above, it is probable that translation is impaired by suboptimal codon usage. Codon optimization should correct this problem, so this should be considered for future work. This can be achieved by gene synthesis, which can also be used to eliminate undesired restriction sites. Further analysis in future work, the availability of the purified, over-expressed SEN1432 protein would allow direct DNA binding studies to progress along with effector interaction investigations, using techniques such as gel retardation and DNase I foot printing.

SEN1432 belongs to the GntR/FadR family (Haydon and Guest, 1991), with GntR first recognised as a gluconate operon repressor from *Bacillus subtilis* (Rigali et al., 2002; Suvorova et al., 2015). According to the Pfam database (PF00392), there are 49,014 sequences of proteins belonging to this family nearly entirely within the Bacteria, mostly the Proteobacteria, Actinobacteria and Firmicute phyla. This type of transcriptional regulator controls transcription through allosteric structural alteration upon binding to metabolite effector molecule. The members of the GntR family consist of a DNA binding N-terminal helix-turn-helix domain (a winged helix structure consisting of a three-helix bundle and a small β-sheet) and a varied Cterminal effector-binding/oligomerization domain designated as a 'FadR C-terminal Domain' (FCD), which is α-helical (<u>http://pfam.xfam.org/family/PF00392</u>; Haydon and Guest, 1991; Van Aalten et al., 2000). These proteins interact with DNA as homodimers, where they act as repressors. Binding of an inducer (usually the substrate of the metabolic pathways that the transcription factor regulates; Jain, 2015) appears to trigger a change in conformation which releases the transcription factor from the DNA (Resch et al., 2010). SEN1432 is suspected to utilise a gluconate-like metabolite as its coeffector, and evidence provided here suggests that it represses the SEN1432-6 gene cluster, presumably through direct interaction with the corresponding promoter regions.

Chapter 8: General discussion

8.1. Introduction

Salmonella Enteritidis is one of the most common issues threating humans worldwide due its responsibility for about 90% of foodborne infections via consumption of poultry products (EFSA BIOHAZ Panel, 2014). This servor presents a particular capacity to survive encounter with the extreme conditions of EW (Gantois et al., 2008). The most common vehicle associated with SE outbreaks are animal origin products such as egg, poultry, pork, beef and raw dairy products (Peris et al., 2010). Thus, SE is often used as a model pathogen to analyse the mechanisms by pathogens survival within EW (Cogan et al., 2004). Avian albumen provides efficient physical and chemical barriers for protecting the embryo from contamination (Van Dijk *et al.*, 2008). Molecular genetic studies provide various explanations for the survival of SE upon exposure to the antimicrobial effects of EW (Gantois et al., 2008). In order to further understand SE behaviour towards the bactericidal mechanisms of the EW, Baron et al. (2017) conducted a global transcriptional response study of the effect of SE during exposure to EW for 7-45 min at 45 °C using microarrays. This study revealed a large-scale global shift in transcription (18.7% of genes affected) in involving many genes related to stress-response, energy metabolism and micronutrient provision. Of particular interest, was the high degree of induction of hexonate/hexuronate (Hex) utilization genes: the dgoRKADT operon (13.6- to 31.1-fold), the uxuAB-uxaC operon (10.7- to 28.2-fold) and the SEN1433-6 genes (5.17- to 33.4-fold). Yet, there was no previous evidence indicating the presence of hexonates/hexurnates in EW.

Therefore, this study aimed to determine the role of the hexonate/hexuronate utilisation genes in EW survival, and to discover whether these genes are subject to induction by a common regulatory pathway within egg white and if so to characterise the regulatory mechanism and identify the environmental inducing signal within EW. It was anticipated that such information would contribute new overall understanding of the mechanisms applied by pathogenic bacteria to encounter host defence, particularly with regard to EW.

8.2 Exploring the potential promoter regions of *hex* genes.

First objective was to determine the patterns of transcriptional regulation of the genes of interest that located in the three distinct hex gene clusters: the dgoRKADT operon, the uxuABuxaC (SEN2978-SEN2980) operon and the SEN1432-6 locus. To achieve this, transcriptional fusions were generated using the promoterless pRS1274 lacZYA transcriptional fusion vector. The transcriptional fusion data in E. coli TOP10 indicated that seven of the fusions had activity markedly above that of the vector control, but two (dgoT-lacZ; SEN2979-lacZ) had weak activity only slightly higher than the vector suggesting no promoter is present, although the insilico analysis for dgoT showed a strongly predicted promoter. The other fusions could be divided into three groups on the basis of their relative expression levels during the rapid growth phase in L-broth. The group with strong activity (SEN1436-lacZ, 1250 U; SEN2977-lacZ, 1350 U) consisted of genes that are divergently arranged with respect to adjacent operons (SEN2978-80 and SEN1435-33). The moderate activity group (SEN1432-lacZ, 740 U), is consisted of a single gene encoding a putative transcriptional regulator that likely has an independent proximal promoter. The encoded regulator could control the SEN1435 and SEN1435-33 genes through interaction with divergent putative promoters at the SEN1436-35 intergenic region. The weakly expressed group (SEN1435-, SEN2978-, dgoR- and ybhC-lacZ) gave relatively weak log-phase activities (maximum of 180-350 U) suggesting that these genes are repressed under the conditions employed in this work (LB medium, with E. coli as host). Two fusions (dgoT- and SEN2979-lacZ) gave very weak activity that was only slightly above that of the vector control (Fig. 3.22), as indicated above.

A previous study showed that several genes are up regulated (2.5-3.5 fold) in operons involved in the transport and metabolism of D-galactonate (dgo), D-gluconate (gntU, kdgT, and kduD), and L-idonate (idn) in SE that are indicative of its metabolism in macerated leaf tissue in cilantro and lettuce soft rot lesions (Goudeau et al., 2013). However, the precise environmental factor inducing their expression was not clear. Interestingly, genes involved in the utilisation of gluconate and related hexonates (gntT, STM3134, dgoT, dgoK and dgoA) were up-regulated in S. Typhimurium upon macrophage colonisation, as were Entner–Douderoff pathway genes involved in the interconversion of these sugars to pyruvate and glyceraldehyde-3-phosphate (Eriksson et al. 2003). The reason for this is unclear but it was suggested that hexonates may be an important source of carbon for intracellular bacteria (Eriksson et al. 2003). The promoter finder program, BPROM, was used to recognize promoters for the hex genes as well as potential transcription factor binding sites (section 3.4); these remain to be proven as valid. To summarise the effect of hexonates on control of the *hex* genes, a range of relatively-moderate regulatory responses was observed suggesting no common mode of control with respect to hexonate availability. This suggests that the induction of the *hex* genes by EW is unlikely to be Hex mediated.

8.3 EW factors influencing expression of SE hex genes

The inducers responsible for the up-regulation of the *hex* genes in EW have not been identified. Further, there is no evidence available for the presence of hexonates or hexuronates within EW (Guérin-Dubiard *et al.*, 2010). Thus, an important aim was to confirm the proposed induction of the *hex* genes in EW and to investigate of the relevant environmental factors affecting this increase in expression. Initially, the role of Hex compounds was explored, and subsequently the role of WE components was investigated.

8.3.1 Utilisation the hexonates as energy/carbon sources and the effect of hexonates on expression

The ability of SE to grow using different hexonates as carbon sources was tested. In addition, the effect of these substrates on hex gene expression was examined, using the lacZ fusions generated in chapter 3. Firstly, two factors were tested (standard carbon sources and temperature) to establish control conditions for comparison with growth tests with the Hex compounds. Thus, glucose was selected as it is present in EW at 0.4-0.5% (Guérin-Dubiard et al., 2010), and glycerol was used as an example of a non-fermentable carbon source that does not induce catabolite repression. Mammal and hen body temperatures (37 and 42 °C; Raspoet et al., 2014; Baron et al., 2017) were tested. SE grew well at 42 °C with glycerol or glucose at 0.4%, and increasing glucose concentration showed a quantitative increase in growth, although growth was weaker than that obtained at 37 °C. The ability of SE to utilise four available hexonates (D-galactonic acid; D-mannono-1,4-Lactone; L-(+)-gulonic acid y-lactone and gluconate) was tested at 0.1-1.6% w/v. All Hex compounds acted as good sole carbon/energy sources and supported the growth of SE at both 37 and 42 °C. However, some differences in the degree of growth supported were apparent. Best growth was achieved with galactonate, followed by gluconate, then mannonate, and finally gulonate (maximum ODs of 1.1, 0.9, 0.8, 0.7, respectively, at 42 °C). This finding is supported by previous work showing the ability of Salmonella to grow on galactonate and gulonate (Cooper, 1980).

SE carries the genes of the GntI system (Parkhill *et al.*, 2008). These genes are likely to be subject to catabolite repression (Rodionov *et al.*, 2000) and to be induced by gluconate through GntR transcriptional control. The gulonate degradation pathway in SE is unclear. However, this pathway is likely to involve one or more of the following: the GntI system, the Dgo pathway or the SEN1433-6 pathway. Galactonate is expected to be catabolised via the Dgo pathway and feed end products into the glycolytic pathway. The Dgo pathway showed induction by D-

galactonate in *E. coli* and is subject to catabolite repression (Deacon & Cooper, 1977; Cooper, 1978). For mannonate, the utilisation pathway in *SE* is expected to be that operated by SEN2977-90 (UxuAB/UxaA) system (Robert-Baudouy & Stoeber, 1973; Portalier *et al.*, 1980; Blanco *et al.*, 1986; Zeng *et al.*, 2001). However, this needs to be proven in *SE*.

To test the effect of hexonates on *hex* gene expression in M9 medium at 42 °C, all four Hex substrates were provided at 0.4% and expression effects were compared with that for 0.4% glycerol. For D-galactonate, the greatest induction (sixfold) effect was seen for dgoR and a repression effect was seen for sen1436, sen1432 and sen2977 by 6-, 3.5- and 20-fold, respectively. This suggested that likely DgoR acts as a repressor for the dgo genes. dgoR is the first gene in the *dgoRKDA-T* cluster of SE. This response was previously reported in E. coli, as the dgoR gene showed induction by D-galactonate (Neidhardt, 2005). However, the induction level observed does not match that seen in EWMM (up to 28.7 fold; Baron et al., 2017) suggesting with that D-galactonate is either not the relevant inducer in EWMM (the medium used by Baron et al., 2017) or that the conditions used here are not sufficiently similar to those used by Baron et al. (2017) to enable the same level of induction to be achieved. Mannonate showed significant induction of SEN2977 (14-fold). However, there was little effect on the expression of the other hex genes suggesting a role for the sen2977-uxuAB-uxaA genes in utilisation of mannonate and/or related compounds. In contrast, this fusion showed repression by galactonate and gluconate, again consistent with a role in mannonate utilisation. Gluconate showed a repression effect on most of the fusions tested with greatest effect was seen for sen1436 (17-fold). It is likely that this response is mediated by GlnR (the gluconate-responsive repressor) and reflects the need to repress the *hex* genes whilst inducing the GntI system (gluconate catabolism) when gluconate is provided (see Fig, 4. 5). Gulonate showed a significant induction on three fusions (sen1435, sen1432, dgoR) at twofold or more. No previous data on gulonate-dependent gene control in SE or E. coli could be found in the

literature so the regulator responsible is not clear. However, three corresponding fusions were induced (*sen1432*, x2.1; *sen1435*, x4; *sen1436*, x1.6) which might suggest a role for these genes in gulonate utilisation with a potential role for the GntR-like *sen1432* product in mediating this regulatory control. To summarise, it is clear that the *hex* genes are indeed subject to regulatory control by hexonates, and that different hexonates show distinct regulatory responses suggestive of multiple regulatory pathways. Arguments for roles of DgoR, GntR and SEN1432 in mediating many of the hexonate-dependent responses observed have been provided. However, these possibilities need confirmation through further investigation with relevant regulatory mutants.

8.3.2 Effect of EW on growth and hex gene expression in SE

The changes in expression level of several of the *hex* genes due to the presence of hexonates indicate that any hexonates released during exposure of *SE* to EW could induce the change in expression levels of the *hex* genes as observed by Baron *et al.* (2017). However, there is no clear explanation about how Hex compounds could be generated by *SE* exposure to EW, and the degrees of induction observed does not match that seen in EW. Therefore, further investigation was performed to explore the effect of EW on the growth of *SE* and its impact on *hex* gene expression. EW was tested at 0.05-10%, at hen body temperature (42 °C), to confirm the growth inhibitory effect and to determine appropriate EW levels to employ in subsequent EW expression experiments. Impaired growth was observed at 37 °C at all EW concentrations tested. On the other hand, at 30 and 42 °C, growth was totally inhibited at relatively low EW levels (2.5%). Such an impact of temperature on the antimicrobial activity of EW has been reported previously by Baron *et al.* (2011). This effect reflects the well observed antimicrobial activity of EW for many bacterial species (Sahin *et al.*, 2003; Wellman-Labadie *et al.*, 2009). Adding 20 μ M ferric citrate was found to restore the growth of *SE* in 0.1% EW to levels similar to those achieved without EW.

SE carrying pRS-SEN1436-lacZ (encoding a predicted D-galactonate dehydratase) was selected for further study as a representative *hex* gene that showed good expression in the previous experiments and was the most greatly induced gene in response to EWMM in the previous work of Baron et al. (2017). Initially, the effect of EW on hex gene expression was tested in M9 medium at 42 °C using different levels of EW (0.0001-10%; prepared at lab) in M9 medium at 42 °C. In contrast, the EW exposure experiments of Baron et al. (2017) employed an 'EW model medium' (EWMM) composed of EW filtrate with 10% EW protein to mimic EW medium as far possible. The results showed that *sen1436* expression is induced by 22-61 fold with 0.01-10% EW. The induction observed with 10% EW (61-fold) is higher than (33-fold) that reported by Baron *et al.* (2017), but is similar in scale, and is far greater than that seen above with hexonates, where a maximum 7 fold induction was observed. The experiment was repeated with three other hex gene fusions (sen1432, dgoR and sen2977) and the results showed that expression of all three fusions is induced by EW by 21-, 21- and 13fold for *sen1432*, *dgoR* and *sen2977*, respectively, using concentrations of EW at 10%. These findings thus support the presence a *hex* gene inducer within EW. The experimental conditions applied by Baron et al. (2017) showed hex gene induction depended on the presence of EW proteins since EW (10 kDa cutoff) filtrate without addition of EW failed to induce the hex genes. Therefore, EW filtrate (10 kDa cutoff) was tested in place of EW to confirm that the EW proteins of >10 kDa are indeed responsible for the induction observed for sen1436. The results showed the EW filtrate gave only a very weak induction of *sen1436* expression, of just under twofold compared, compared with the expression level in the M9 medium. This results narrow the suspected inducer to be an EW proteins of >10 kDa.

Further experiments showed that the induction of SEN1436 by EW is not influenced by provision of iron. Thus, it can be assumed that the EW response of the *hex* genes is unrelated to the low iron availability in EW.

8.3.2.1 The role of lysozyme in inducing the *hex* genes in EW

The above finding led to experiments testing the impact of individual EW proteins on SEN1436-*lacZ* induction. Four EW proteins (albumin, conalbumin, ovomucoid, and lysozyme) were tested at three different concentrations (0.01, 0.1 and 1 mg/ml). The results showed lysozyme gave a very strong induction effect for SEN1436 expression. The greatest induction (48 fold), with lysozyme, was seen at 7 min with 0.1 mg/ml lysozyme suggesting this protein is primarily responsible for the EWMM-induction of the *hex* genes. The other three *hex* gene *lacZ* fusions (SEN1432, *dgoR* and SEN2977) tested are also subject to induction by lysozyme. The results showed that the expression of all three fusions was increased, by 19-, 13- and 14-fold (respectively). The roles of lysozyme was confirmed in several ways: different combinations of EW proteins, different sources of lysozyme, heat inactivation, and examination at different pH values, iron levels and temperatures). Thus, the absence of lysozyme clearly lead to lack of any induction with EW supports the suggestion that lysozyme is the key factor in EW induction of *hex* gene expression which is novel found. However, the mechanism of by which lysozyme induces the *hex* genes remained unclear.

In EW, lysozyme is considered to be more effective against bacteria due to the synergistic activity of other EW components. Such synergistic components potentially include the chelating activity of ovotransferrin which removes metals associated with the LPS moieties of the outer membrane of Gram-negative bacteria which could in turn disrupt this membrane and allow lysozyme access to the peptidoglycan layer (Baron *et al.*, 2015). The bactericidal activity of lysozyme is reported to involve three main mechanisms (Baron *et al.*, 2015). The membrane disruption is reported as one of lysozyme's activities against Gram negative bacteria (Masschalck *et al.*, 2003). In addition, induction of pore formation in the outer membrane of *E. coli* has been recognized as another lysozyme activity (Derdre *et al.*, 2013). Moreover, lysozyme has a high affinity (presumably due to its very high pI) for the LPS and is able to

insert into the latter as causing reorganization of the LPS monolayer (Derdre *et al.* 2014). Although, there is possibility that SE resists the peptidoglycan lytic activity of lysozyme due to the protection provided by the outer membrane and the periplasmic lysozyme inhibitor (PliC), Baron *et al.* (2015) indicated that the particular conditions provide by EW (e.g. high pH, metalion limitation) might increase SE sensitivity to lysozyme. Various studies have indicated the potential role of genes involved in LPS biosynthesis in EW survival. The *rfal* mutant, in which an enzyme that catalyzes the early step in LPS biosynthesis is absent, was unable to survive in EW at 42 °C (Raspoet *et al.* 2014). A *murA* gene, encoding an enzyme involved in the synthesis of peptidoglycan, showed an induction in SE during hen oviduct colonization and in contaminated eggs suggesting a response to the permeabilization of the peptidoglycan by lysozyme (Gantois *et al.* 2008). The recognition of lysozyme as the main *hex* gene inducer in EW allows a hypothesis to be proposed whereby the induction observed is caused by the release of an endogenous inducer from SE in response to cell envelope damage elicited by lysozyme. Characterizing the mechanisms might contribute to improvements in food product preservation against foodborne pathogens infection e.g. by enhancing the impact of lysozyme.

8.4 Role of the *hex* gene regulators, SEN1432 and DgoR

Two *hex* genes (SEN1432 and *dgoR*) were selected for further analysis through knock out to investigate whether deletion of these genes has any obvious phenotypic effect. This technology has been used extensively in *E. coli* and *S.* Typhimurium (Murphy and Campellone, 2003). DgoR and SEN1432 were selected as likely being involved in mediating many of the hexonate-dependent responses identified in this study. The promoter activity measurement showed moderate activity (~740 U) for SEN1432-*lacZ* with 21- and 19-fold of induction towards 10% EW and 0.1 mg/ml lysozyme, respectively, suggests that it has an independent proximal promoter. However, the report by Baron *et al.* (2017) suggests that it is not subject to EW

induction. The *sen1432* gene is separated from the rest of the cluster (SEN1435-33) by ~90 bp and specifies a putative GntR-family regulatory protein. Although the microarray data reported high induction of *dgoR* (27 fold; Baron *et al.*, 2017), the promoter activity assay showed relatively low expression (~180-190 U), suggesting that the corresponding operon is repressed under the conditions studied, or that its promoter is weak. However, it showed strong induction of 21- and 13-fold by 10% EW and 0.1 mg/ml lysozyme, respectively. *dgoR* is the first gene in the apparent *dgoRK*-SEN3645-*dgoT* operon, indicating that such control of expression (as exhibited by *dgoR*) would extend to the entire operon.

The deletion of *dgoR* caused a moderate increase in the expression of *dgoR-lacZ*, and likewise, deletion of SEN1432 caused a moderate induction of the SEN1432- and SEN1436-*lacZ* fusions. These results indicated a role for the GntR-like products of DgoR and SEN1432 in repression of the corresponding genes. This effect was reversed when complementary plasmid-borne versions of the SEN1432 and *dgoR* genes were introduced to the mutants. The regulatory mutations did not affect induction by EW or lysozyme, indicating that neither DgoR nor SEN1432 are involved in the induction of the *hex* genes by EW lysozyme. Both SEN1432 and DgoR are GntR-like transcriptional repressors with common structural organisations (Jain, 2015). Previous work showed that the deletion of the entire SEN1432–SEN1436 locus decreased of the bacterial load in the spleen of chickens at 14 days post-infection suggesting a minor role for this system in systemic colonization (Coward *et al.* 2012).

8.5 The two-component sensor-regulators, PmrAB and PhoPQ, mediate the response of

the *hex* genes to polymyxin B, and have a minor role in the response to lysozyme/EW The most likely reason for the lysozyme-dependent induction observed in chapter 4 would appear to be the release of an endogenous inducer/signal generated by *S*E in response to cellenvelope damage. Alternatively, any lysozyme-mediated alteration in the structure/integrity of the envelope might trigger a protective response leading to *hex* gene induction. The expression data of Baron *et al.* (2017) were consistent with a considerable membrane-stress response imposed by EW on SE, a stress that is likely to be caused by lysozyme, to some degree at least. The genes thus up-regulated, that are related to membrane-stress, include degP (a periplasmic/membrane-associated serine endoprotease that degrades abnormal proteins), Tol-Pal system genes (involved in the maintenance of cell-envelope integrity) and *ompC* (encoding an outer-membrane porin). Gantois *et al.* (2008) suggest that maintenance of cell-envelope integrity is a significant feature of resistance to EW, with cell-wall disruption and progressive cell lysis reported as the major mechanisms of EW-mediated bactericidal action at 45 °C for *E. coli* (Jan *et al.*, 2013; Baron *et al.*, 2017); a similar effect can be anticipated for SE.

The polymyxin B and lysozyme induction of the *hex* gene fusions was investigated in the presence and absence of two 2-component transcriptional regulators (PmrA and PhoP) that might have a role in mediating the *hex* gene response in EW. The PmrAB and PhoPQ systems coordinate the expression of genes that enhance survival under conditions where membrane integrity is threatened, through inducing modifications of the LPS that decrease AMP (antimicrobial peptide) binding and bacterial-cell entry (Gunn, 2008). PmrAB was shown to induce one of the hex genes, dgoR, by 500 fold (Tamayo et al., 2002); this finding thus leads to the suggestion that all of the hex genes might be subject to major regulatory induction by PmrAB in response to membrane damage exerted by lysozyme in EW. The PmrAB system responds (directly or indirectly) to high extracellular levels of ferric iron or Al³⁺, external acidity and AMPs such as PMB (Zhou, 1999; Ryan et al., 2015; Tamayo et al., 2002). Interestingly, the yibD (waaH) gene is also greatly (2,500-fold) induced by PmrAB in response to PMB (Tamayo et al., 2002) and this gene specifies an enzyme (LPS(HepIII)-glucuronic acid glycosyltransferase) that incorporates glucoronate (a hexuronate) into LPS to improve resistance to SDS and other factors (e.g. PMB) that damage the outer membrane (Klein et al., 2013). This suggests the possibility that the reason for the induction of the *hex* genes by

lysozyme (and PMB) might be to generate hexonate/hexuronates for addition to LPS in order to raise membrane resistance. However, the manner in which such modification might result in lysozyme resistance is unclear. Addition of hexonates would be expected to raise the negative charge of the outer membrane and thus would be expected to promote binding of lysozyme due to the strong positive charge of this enzyme. A possibility to consider is that such modification might trap lysozyme at the surface of the OM since, thus providing resistance to lysozyme damage to peptidoglycan. This suggestion requires further investigation.

To study the role of the PmrAB and PhoPQ systems in *hex* gene expression, *Salmonella* serovar Typhimurium (JSG210) and two isogenic mutants, pmrA (JSG421) and phoP (JSG425), were utilised with six *hex* gene *lacZ* fusions. In general, the six *hex* genes studied showed strong induction by PMB, and this induction was (generally) markedly reduced or eliminated by loss of either PmrA or PhoP. This strongly indicates that the hex genes have a role in protection against outer-membrane damage as elicited by exposure to PMB, and that their PMB induction depends on both PhoPQ and PmrAB. It is likely that PhoPQ acts as the direct sensor for PMB activity, and then activates PmrA through PmrD (the *pmrD* gene is induced by PhoP, and PmrD activates PmrA by inhibiting its dephosphorylation; (Kato et al., 2007), resulting in PMB-induction of the hex genes. The loss of either PmrA or PhoP, in general, only slightly reduced *hex* gene induction by lysozyme (by around twofold). Thus, it is clear that the response to lysozyme is distinct to that for PMB. In addition, the strong residual *hex*-gene response to lysozyme in the absence of PmrA or PhoP shows that some other factor is mainly responsible to lysozyme induction. Candidates include RpoE and CpxAR since these regulators respond to envelope damage and there was strong evidence of their role in the regulatory response of SE to EW (Baron et al., 2017). Another possibility is that the lysozyme response is controlled directly by both PhoPQ and PmrAB, such that absence of one or other system only weakly affects the induction observed. Thus, pmrA phoP double mutants should be employed in any future work to test for this possibility. It is interesting to note that PmrB has no extensive, recognisable sensory domain (just a short 30 residue segment in the periplasm) whereas PhoQ has a large periplasmic domain (~130 residues) that is presumed to respond to the various extracellular (periplasmic) signals that induce the PhoPQ regulon.

The pSU-*eptA-pmrAB* complementing plasmid was generated and showed that provision of *eptA-pmrAB* in *trans* reversed the lack of induction by lysozyme and PMB of SEN1436 in the *pmrA* mutant, clearly supporting the direct control of this gene by PmrAB (Fig. 6.11 & 6.12). The results are also consistent with a direct response of the PmrAB system to lysozyme, and an indirect response to PMB via PhoPQ-mediated control. For *dgoR*, the results suggested direct regulatory control by PhoPQ in response to PMB, and a degree of direct regulatory control by PmrAB in response to lysozyme. Thus, the control of the *hex* genes by the PmrAB and PhoPQ systems in response to lysozyme and PMB is complex, and involves additional unidentified factor(s). Such additional regulator factors previously identified include Crp (responding to glucose) and the various GntR-related repressors (e.g. GntR) responding to Hex compounds (Robert-Baudouy & Stoeber, 1973; Portalier *et al.*, 1980; Blanco *et al.*, 1986; Zeng *et al.*, 2001; see Fig. 4.5).

In summary, the results clearly show that the *hex* genes are subject to PMB induction and that this is largely controlled by PmrAB-PhoPQ. However, the response to lysozyme is only partly controlled by these factors indicating the involvement of another regulator. The results are consistent with a role for the observed *hex* gene induction by lysozyme in preserving the integrity of the cell envelope.

8.6 Attempted overexpression of SEN1432 and dgoD

Two of the *hex* genes (SEN1432 and *dgoD*, encoding a predicted transcription factor and D-galactonate dehydratase, respectively) were targeted for overexpression and purification, partly to enable antibody production for the monitoring of expression effects by western blotting (providing

a second method to monitor *hex* gene expression). Overexpression vectors were generated for SEN1432 but not for *dgoD* due to the presence of *Hin*dIII and *Nde*I restriction sites within *dgoD* meaning that the gene was digested during release of the insert from pJET1.2 for subcloning into pET21a, and so it was not possible to clone the gene into pET21a. In future work, this problem could be overcome through a gene-synthesis approach or by using cloning by Gibson assembly (Gibson *et al.*, 2009).

Subcloning of SEN1432 into pET21a was successful, however the overexpression of SEN1432 from *E. coli* BL21(DE3) in both its native and His-tagged form failed. Sequencing showed that the correct gene sequence was present in the constructs, in the correct cloning context; it is thus assumed that the transcript was produced but that translation was poor. A few reasons for this effect include the formation of a secondary structure (stem-loop structures) interrupting the activity of the ribosome and thus negatively affecting translation (Hall *et al.*, 1982; Qing *et al.*, 2003). The SEN1432 gene has a relatively low G/C content at the start of the gene so secondary structures should not be causing a problem with translation. Likewise, the second codon has been shown to affect translation in *E. coli*, but the second codon of SEN1432 (AGC) is not associated with poor translation (Looman *et al.*, 1987). Therefore, it is probable that translation is impaired by suboptimal codon usage as indicated by the relatively high level of suboptimal codons carried by this gene.

8.7 Suggested future work

In any future work following from that described in this thesis, there are several priorities that should be considered. The effects of the dgoR and SEN1432 mutations on the remaining *hex* gene *lacZ* fusions should be investigated such that a more complete indication of the regulatory influences of the corresponding GntR-like regulators can be deduced. Any such additional experiments should also include mutations in relevant regulatory genes (e.g. *gntR*, *idnR*). In addition, the effects of the various hexonates on *hex* gene expression with each regulatory

mutant should be performed to investigate how the regulatory-impact of hexonates on hex gene expression is affected by absence of these regulators. In this way, it should be possible to define the effectors that each regulator responds to. It would also be beneficial to include generate multiple deletions in the genes encoding the regulators of relevance and then study the effects of such mutations on hex gene regulation by hexonates. This would further clarify the regulatory processes governing the expression of the hex genes. Purification of the DgoR and SEN1432 proteins would enable direct DNA and ligand binding experiments to proceed which would extend and support the work with the *lacZ* fusions. Further work is required on the lysozyme and PMB induction effects observed, using a mixtures of lysozyme and PMB to determine whether these factors induce gene expression in an additive fashion; this would confirm that these two factors induce hex gene expression by different pathways. Further, a double *pmrAB-phoPQ* mutant should be used to test the possibility that in the absence of one system, the other provides a compensatory activity for lysozyme-dependent induction. The possible role of CpxAR and RpoE in the observed lysozyme induction should be tested, particularly as both these systems were predicted to be activated in response to EW exposure (Baron et al., 2017) and a potential CpxR site was identified upstream of dgoT. It would be particularly interesting to perform a global expression analysis of the effect of lysozyme on in SE. Codon optimization using programmes such as 'GeneOptimizer' (ThermoFisher) should assist in correcting the overexpression problem, so this should be considered for future work. This can be achieved by gene synthesis, which can also be used to eliminate undesired restriction sites. The availability of the purified, over-expressed SEN1432 and DgoR proteins would allow direct DNA-binding studies to progress along with effector interaction investigations, using techniques such as gel retardation and DNase I foot printing.

9. References

Ahlborn, G. and B. W. Sheldon. 2005. Enzymatic and microbiological inhibitory activity in Egg shell membranes as influenced by layer strains and age and storage variables. *Poultry Science*. **84:**1935-41.

Akiba, M., M. Kusumoto, and I. Iwata. 2011. Rapid identification of *Salmonella* enterica serovars, Typhimurium, Choleraesuis, Infantis, Hadar, Enteritidis, Dublin and Gallinarum, by multiplex PCR. *Journal of Microbiological Methods*. **85:**9-15.

Alabdeh, M., V. Lechevalier, F. Nau, M. Gautier, MF. Cochet, F. Gonnet, S. Jan and F. Baron. 2011. Role of incubation conditions and protein fraction on the antimicrobial activity of egg white against *Salmonella* Enteritidis and *Escherichia coli*. *Journal of Food Protection*. **74:** 24-31.

Alabdeh, M., V. Nau and A. Françoise. 2010. Role of Incubation Conditions and Protein Fraction on the Antimicrobial Activity of Egg White against *Salmonella* Enteritidis and *Escherichia coli*. *Journal of Food Protection*. **74:** 24-31.

Alison W., G. Dolan. M. Philip, A. C. Jenkins. T. Dallman. 2018. Epidemiological analysis of *Salmonella* clusters identified by whole genome sequencing, England and Wales 2014. *Food Microbiology*.**71**:39-45.

Amavisit, P., D. Lightfoot, G, Browning and P. Markham. 2003. Variation between Pathogenic Serovars within *Salmonella* Pathogenicity Islands. *Journal of Bacteriology*. **185**: 3624-3635.

Andrews, S. 1998. Iron storage in bacteria. Advanced Microbial Physiology. 40: 281-351.

Andrews, S., A. Robinson and F. Rodriguez. 2003. Bacterial iron homeostasis. *FEMS Microbiology Research*. 27:215–237.

Bang I., A. Frye, G. Jonathan, A. McClelland, A. Michael, J. Velayudhan and F.Ferric C.2005. Alternative sigma factor interactions in *Salmonella*: σE and σH promote antioxidant defences by enhancing σS levels. *Molecular Microbiology*. **56**: 1365-2958.

Baron, F, M. Gautier and Brule G. 1997. Factors Involved in the inhibition of growth of *Salmonella* Enteritidis in liquid egg white. *Journal of Food Protection*. **60**: 1318–1323.

Baron, F. 1998. Etude du comportement de *Salmonella* Enteridis dans le blanc d'oeuf (Thèse de doctorat). France.

Baron, F. and JAN, S. 2011. Egg and egg product microbiology. Inra Prod.Anim. 23: 193-204.

Baron, F. and S. Jan. 2010. Microbiologie de l'œuf. Productions Animales. 23:193-204.

Baron, F., F. Nau, C. Guérin-Dubiard, M. Gautier, S. Bonnassie, M. Gautier, S. Andrews, S. Jan. 2015. Egg white versus Salmonella Enteritidis! A harsh medium meets a resilient pathogen. *Food microbiology*. **53**:82-3.

Baron, F., M. Cochet, W. Ablain, N. Grosset, S. Jan, Sophie and M GAUTIER. 2006. Rapid and cost-effective method for micro-organism enumeration based on miniaturization of the conventional plate-counting technique. *INRA*, *EDP Sciences*. **86**: 251-257.

Baron, F., S. Bonnassie, M. Alabdeh, F. Cochet, F. Nau, C. Guérin-Dubiard, M. Gautier, S. Andrews, S. Jan. 2017. Global Gene-expression Analysis of the Response of Salmonella Enteritidis to Egg White Exposure Reveals Multiple Egg White-imposed Stress Responses. *Frontiers in Microbiology*. **8**:829.

Bartolomé, B., Y. Jubete, E. Martínez, and F. de la Cruz. 1991. Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. Gene. 102:75-78.

Bausch, C., N. Peekhaus, C. Utz, T. Blais, E. Murray, T. Lowary and T. Conway. 1998. Sequence analysis of the GntII (subsidiary) system for gluconate metabolism reveals a novel pathway for L-idonic acid catabolism in *Escherichia coli*. *Journal of Bacteriology*. **180**: 3704-10.

Beckett, D. 2007. Biotin sensing: Universal influence of biotin status on transcription. *Annual Review of Genetics*. **41**: 443–464.

Bedrani, L., E. Helloin, N. Guyot, S. Réhault-Godbert and Y. Nys. 2013. Passive maternal exposure to environmental microbes selectively modulates the innate defences of chicken egg white by increasing some of its antibacterial activities. *BMC Microbiology*. **13**:128-131.

Belitz, H. D., W. Grosch and P. Schieberle. 2009. Eggs. Food Chemistry. 2:546-561.

Bera, A., Herbert, S., Jakob, A., Vollmer, W. and F. Gotz. 2005. Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major determinant for lysozyme resistance of *Staphylococcus aureus*. *Molecular Microbiology*. **55**: 778–787.

Berge, A., J. Adaska and W. Sischo. 2004. Use of Antimicrobiotic Susceptibility Patterns and Pulsed-Field Gel Electrophoresis to compare Historic and Contemporary Isolates of Multi-Drug-Resistant *Salmonella* enterica subsp. Enterica Serovar Newport. *Applied and Environmental Microbiology*.**70**: 318-323.

Betancora, L, L. Yima, A., Martíneza, M., Fookesc, S., Sasiasa, F., Schelottob, N., Thomsonc, D., Maskelld and J. Chabalgoitya. 2012. Access Genomic Comparison of the Closely Related *Salmonella enterica* Serovars Enteritidis and Dublin. *The Open Microbiology Journal.* **6:** 5-13.

Bogs, J. and K. Geider .2000. Molecular analysis of sucrose metabolism of Erwinia amylovora and Influence on bacterial virulence. *Journal of Bacteriology*. **182:**5351–5358.

Bouchrif, B., B. Paglietti, M. Murgia, A. Piana, N. Cohen, Ennaji, M., S. Rubino and M. Timinouni. 2009. Prevalence and antibiotic-resistance of *Salmonella* isolated from food in Morocco. *Journal of Infection in Developing Countries*. **28**: 35-40.

Bowden, S., G. Rowley, J. Hinton and A. Thompson. 2009. Glucose and Glycolysis Are Required for the Successful Infection of Macrophages and Mice by Salmonella enterica Serovar Typhimurium. *Infection and Immunity*. 77: 3117–3126.

Boyen, F., F. Haesebrouck, D. Maes, F. Van Immerseel, R. Ducatelle, and F. Pasmans. 2008. Non-typhoidal *Salmonella* infections in pigs: A closer look at epidemiology, pathogenesis and control. *Veterinary Microbiology*. **130**: 1-19.

Brenner, F., R. Villar, F. Angulo, R. Tauxe, and B. Swaminathan. 2000. *Salmonella* Nomenclature. *Journal of Clinical Microbiology*. **38:** 2465–2467.

Carroll, P., R.M. La Ragione, A.R. Sayers and M. J. Woodward, M. J. 2004. The O-antigen of *Salmonella* enterica serotype Enteritidis PT4: a significant factor in gastrointestinal colonisation of young but not newly hatched chicks. *Veterinary Microbiology*. **102**: 73-85.

Callewaert, L., A. Aertsen, D. Deckers, K. Vanoirbeek, J. Van Herreweghe, B. Masschalck, D., Nakimbugwe, J., Robben, and C. Michiels 2008. A new family of lysozyme inhibitors contributing to lysozyme tolerance in gram-negative bacteria. *PLOS Pathogens*. **4**: e1000019.

Castilla, K.S., Moreno, A.M., Nunes, I.A., and A.J.P. Ferreira. 2006. Distribution of virulence genes *sefC*, *pefA* and *spvC* in *Salmonella* Enteritidis Phage Type 4 strains isolated in Brazil. *Brazilian Journal of Microbiology*. **37**:135-139.

Chang, D., D. Smalley, D. Tucker, M. Leatham, W. Norris, S. Stevenson, A. Anderson, J. Grissom, D. Laux, P. Cohen and T. Conway. 2004. Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proceedings of the National Academy of Sciences*.**19:**7427-7432.

Charalampopoulos, D., S. Pandiella and C. Webb. 2002. Growth studies of potentially probiotic lactic acid bacteria in cereal-based substrates Journal of Applied Microbiology. **92:** 851–859.

Chart, H. and B. Rowe. 1993. Iron restriction and the growth of *Salmonella* enteritidis. *Epidemiology and Infection*. **110**: 41-47.

Chen, H. and E. Groisman. 2013. The Biology of the PmrA/PmrB Two-Component System: The Major Regulator of Lipopolysaccharide Modifications. *Annual Review of Microbiology*. **67:**83-112.

Chen, J., H. T. Shallo and W.L. Kerr. 2005. Outgrowth of Salmonellae and the physical property of albumen and vitelline membrane as influenced by egg storage conditions. *Journal of Food Protection*. **68**: 2553–2558.

Chimalizeni, Y., K. Kawaza and E. Molyneux, E. 2010. The epidemiology and management of non-typhoidal *salmonella* infections. *Advances in Experimental Medicine and Biology*.**659**: 33-46.

Clavijo, R. I., C. Loui, G. L. Andersen, L. W. Riley, and L. Sangwei. 2006. Identification of genes associated with survival of *Salmonella enterica* serovar Enteritidis in chicken egg albumen. *Applied and Environmental Microbiology*.**72**:1055–1064.

Clay, C. E., and R. G. Board. 1991. Growth of *Salmonella* Enteritidis in artificially contaminated hen's eggs. *Epidemiology and Infection*. **106**:271–281.

Clifton-Hadley, F. A., M. Breslin, L. M. Venables, K. A. Sprigings, S. W. Cooles, S. Houghton and M. J. Woodward . 2002. A laboratory study of an inactivated bivalent iron restricted *Salmonella* enterica serovars Enteritidis and Typhimurium dual vaccine against Typhimurium challenge in chickens. *Veterinary Microbiology*. **89**: 167-179.

Cogan, T. A., F. Jorgensen, H. M. Lappin-Scott, C. E. Benson, M. J. Woodward, and T. J. Humphrey. 2004. Flagella and curli fimbriae are important for the growth of *Salmonella enterica* serovars in hen eggs. *Microbiology*. **150**:1063–1071.

Cogan, TA. and TJ. Humphrey. 2003. The rise and fall of Salmonella Enteritidis in the UK. *Journal of Applied Microbiology*. **94:** 114S-119S.

Collighan, R. J. and M. J. Woodward. 2001. The SEF14 fimbrial antigen of Salmonella enterica serovar Enteritidis is encoded within a pathogenicity islet. *Veterinary Microbiology*. **80**:235-245.

Connor, B. A. and E. Schwartz, E. 2005. Typhoid and paratyphoid fever in travellers. *The Lancet Infectious Diseases.* **5**: 623-628.

Cooper R. 1980. The pathway for L-gulonate catabolism in Escherichia coli K-12 and Salmonella typhimurium LT-2. *Federation of European Biochemical Societies*. 115:63-7

Cooper, R. A. 1978. The utilisation of D-galactonate and D-2-oxo-3- deoxygalactonate by Escherichia coli K-12. Biochemical and genetical studies. *Archives of Microbiology*. **118**:199–206.

Coward, C., L. Sait, L. Williams, T. J. Humphrey, T. Cogan and D. J. Maskell. 2012. Investigation into the role of five Salmonella enterica serovar Enteritidis genomic islands in colonization of the chicken reproductive tract and other organs following oral challenge. *FEMS Microbiology Letters*. **336**: 73–78.

Cox, J. M. and A. Pavic. 2010. Advances in entero pathogen control in poultry production. *Journal of Applied Microbiology*. **108**: 745-755.

Crouch, M. V., M. Castor, J. E. Karlinsey, T. Kalhorn and F. C. Fang. 2008. Biosynthesis and IroC-dependent export of the siderophore salmochelin are essential for virulence of *Salmonella* enterica serovar Typhimurium. *Molecular Microbiology*. **67**: 971–983.

Danese P.N., B. Snyder, L. Cosma, J. Davis and J. Silhavy. 1995 The Cpx two-component signal transduction pathway of Escherichia coli regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. *Genes*. **9:**387-398.

Datsenko, K. and B. Wanner. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*. **97:** 6640–6645.

Davies, R., A. Neuberger and B. Wilson.1969. The dependence of lysozyme activity on pH and ionic strength. *Biochim Biophys Acta*. **178**: 294-305.

De Reu K, K. Grijspeerdt, W. Messens, A. Heyndrickx, M. Uyttendaele, J. Debevere and L. Herman. 2006. Eggshell factors influencing eggshell penetration and whole egg contamination by different bacteria, including *Salmonella* enteritidis. *International Journal of Food Microbiology*. **112:**253–260.

De Vylder, J., R. Raspoet, J. Dewulf, F. Haesebrouck, R. Ducatelle, and F. Van Immerseel. 2013. *Salmonella* Enteritidis is superior in egg white survival compared with other *Salmonella* serotypes. *Poultry Science*. **92**:842–845.

Deacon, J. and R. Cooper. 1977. D-galactonate utilisation by enteric bacteria The catabolic pathway in Escherichia coli. *FEBS letters Federation of European Biochemical Societies*. **77**:201-2015.

Deacon, J., RA. Cooper. 1977. D-Galactonate utilisation by enteric bacteria. The catabolic pathway in *Escherichia coli*. *Federation of European Biochemical Societies Letters*. **15**:201-5.

Derde, M., F. Nau, C. Guérin-Dubiard, V. Lechevalier, G. Paboeuf, S. Jan, F. Baron, M. Gautier and V. Vié. 2015. Native and dry-heated lysozyme interactions with membrane lipidmonolayers:

Lipid packing modifications of a phospholipid mixture, model of the *Escherichia coli* cytoplasmic membrane. *Biochimica et Biophysica Acta*. **1848**:1065-1073.

Derde, M., V. Lechevalier, C. Guérin-Dubiard, M. Cochet, S. Jan, F. Baron, M. Gautier, V. Vié, and F. Nau. 2013. Hen egg white lysozyme permeabilizes *Escherichia coli* outer and inner membranes. *Journal of Agricultural and Food Chemistry*. **61**: 9922-9929.

Desin, T.S., P. Lam, B. Koch, C. Claudia Mickael, E. Emil Berberov, A. Wisner, G.H., Townsend, A.A. Potter, and W. Koster. 2009. *Salmonella enterica* Serovar Enteritidis Pathogenicity Island 1 Is Not Essential for but Facilitates Rapid Systemic Spread in Chickens. *Infection and Immunity*.**77**: 2866-2875.

EFSA BIOHAZ Panel (European Food Safety Autority Panel on Biological Hazards) 2014. Scientific Opinion on the public health risks of table eggs due to deterioration and development of pathogens. *EFSA J European food safety authority*. 3782, 1–147.

Eisenberg, R. and W. Dobrogosz. 1967. Gluconate metabolism in Escherichia coli. *Journal of Bacteriology*. **93:** 941-9.

Eriksson, S., S. Lucchini, A. Thompson, M. Rhen and J. C. Hinton. 2003. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Molecular Microbiology*. **47**: 103–118.

Escolar, L., J. Perez-Martin and V. de Lorenzo. 1999. Opening the iron box: transcriptional metallo regulation by the Fur protein. *Journal of Bacteriology*. **181**: 6223–6229.

European Food Safety Authority (EFSA). 2012. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2010. *European Food Safety Authority*. **10**:1- 442.

Flint, JA., YT. Van Duynhoven and FJ. Angulo. 2005. Estimating the burden of acute gastroenteritis, foodborne disease, and pathogens commonly transmitted by food: an international review. *Clinical Infectious Diseases*. **41**:698-704.

Fong, L., A. Lerman, I. Lam, B. Palsson and P. Charusanti. 2013. Reconciling a Salmonella enterica metabolic model with experimental data confirms that overexpression of the glyoxylate shunt can rescue a lethal ppc deletion mutant. *FEMS Microbiology Letters*. 342: 62–69.

Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck and F. Van Immerseel. 2008. *Salmonella enterica* Serovar Enteritidis Genes Induced during Oviduct Colonization and Egg Contamination in Laying Hens. *Applied and Environmental Microbiology*.**74**: 6616–6622.

Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, F. Van Immerseel. 2009. The *Salmonella* Enteritidis lipopolysaccharide biosynthesis gene rfbH is required for survival in egg albumen. *Zoonoses Public Health.* **56**: 145-149.

Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, R. Gast, TJ. Humphrey and F. Van Immerseel. 2009. Mechanisms of egg contamination by *salmonella* enteritidis. *FEMS Microbiology Reviews*. **33**: 718-738.

Gast, RK. and PS. Holt. 2000. Influence of the level and location of contamination on the multiplication of *Salmonella enteritidis* at different storage temperatures in experimentally inoculated eggs. *Poultry Science*.**79:**559–563.

Gaul, S. B., S. Wedel, M. M. Erdman, D. L. Harris, I. T. Harris, K. E. Ferris, and L. Hoffman. 2007. Use of pulsed-field gel electrophoresis of conserved XbaI fragments for identification of swine *Salmonella* serotypes. *Journal of Clinical Microbiology*. **45**:472-476.

Geng, S, X, Jiao, Z. Pan, Q. Fang, Z. Wen and X. Chen. 2011. A simpler method for the efficient and precise deletion of genes in Salmonella sp. *Journal of Animal and Veterinary Advances*. **10**: 2090-2094.

Geng, S., X. Jiao, Z. Pan, X. Chen, X. Zhang, and X. Chen. 2009. Methodology Report An Improved Method to Knock Out the asd Gene of Salmonella enterica Serovar Pullorum. *Journal of Biomedicine and Biotechnology*. 8.

George, S., I. Salas González, L. Lorca, and M. Teplitskia. 2016. Contribution of the *Salmonella enterica* KdgR Regulon to Persistence of the Pathogen in Vegetable Soft Rots. *Applied and environmental Microbiology*. **82**: 1353–1360.

Gillespie, I. and R. Elson. 2005. Successful reduction of human *Salmonella* Enteritidis infection in England and Wales. *Eurosurveillance*. **10**: 2834.

Gillespie, IA., SJ. O'Brien, GK. Adak, LR. Ward and HR. Smith. 2005. Foodborne general outbreaks of Salmonella Enteritidis phage type 4 infection, England and Wales, 1992-2002: where are the risks? *Epidemiology and Infection*. **133**: 795-801.

Gibson, D., L. Young, R. Chuang, J. Venter, C. Hutchison and H. Smith. 2009. "Enzymatic assembly of DNA molecules up to several hundred kilobases". *Nature Methods*. **6**:343-345

Gómez, K., A. Rodríguez, Y. Rodriguez and A. Ramírez . 2011. The subsidiary GntII system for gluconate metabolism in Escherichia coli: alternative induction of the gntV gene. *Biological Research*. **44**:269-275.

Gormley, FJ., CL. Little, N. Rawal, IA. Gillespie, S. Lebaigue and GK.Adak. 2011. A17-year review of foodborne outbreaks: describing the continuing decline in England and Wales (1992-2008). *Epidemiology and Infection*. **139**:688-699.

Goto, R., T. Miki, N. Nakamura, M. Fujimoto and N. Okada. 2017. *Salmonella* Typhimurium PagP- and UgtL-dependent resistance to antimicrobial peptides contributes to the gut colonization. *PLoS ONE*. **12**: e0190095.

Goudeaua, D. M., C. T. Parker, Y. Zhou, S. Sela, Y. Kroupitski and M. T. Brandl. 2013. The Salmonella Transcriptome in Lettuce and Cilantro Soft Rot Reveals a Niche Overlap with the Animal Host Intestine. Applied *and Environmental Microbiology*.**79**: 250-262.

Gouy, M. & C. Gautier. 1982. Codon usage in bacteria: correlated with gene expressivity. *Nucleic acids research*. **10**: 7055–7074.

Greger, M., 2007. The human/animal interface: emergence and resurgence of zoonotic infectious diseases. *Critical Reviews in Microbiology*.**33**: 243-299.

Greig J. and A. Ravel .2009. Analysis of foodborne outbreak data reported internationally for source attribution. *International Journal of Food Microbiology*. 1**30**:77-87.

Guan, J., C. Grenier, and B. W. Brooks. 2006. In Vitro Study of *Salmonella* enteritidis and Salmonella typhimurium Definitive Type 104: Survival in Egg Albumen and Penetration through the Vitelline Membrane. *Poultry science*. **85:**1678-1681.

Guard-Petter, J. 2001. The chicken, the egg and *Salmonella* enteritidis. *Environmental Microbiology*. **3**: 421-430.

Guard-Petter, J.1998. Variants of smooth *Salmonella enterica* serovars Enteritidis that grow to higher cell density than the wild type are more virulent. *Applied and Environmental Microbiology*. **64:** 2166–2172.

Guérin-Dubiard C., M. Pasco, A. Hietanen, A. Quiros Del Bosque, F. Nau and T. Croguennec.2005. Hen egg white fractionation by ion-exchanger chromatography. *Journal of Chromatography*. **1090:** 58-67.

Gunn J. 2008. The *Salmonella* PmrAB regulon: lipopolysaccharide modifications, antimicrobial peptide resistance and more. *Trends in Microbiology*. **16**: 284-293.

Gunn, J., K. Lim, J. Krueger, K. Kim, L. Guo, M. Hackett and I. Samuel. 1998. PmrA–PmrBregulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Molecular Microbiology*. **27**: 1171–1182. Gunn, J.S. and S.I. Miller.1996. PhoP/PhoQ activates transcription of pmrA/B, encoding a twocomponent system involved in *Salmonella typhimurium* antimicrobial peptide resistance. *Journal of Bacteriology*. **178:** 6857–6864.

Guo, C., RM. Hoekstra and CM. Schroeder. 2011. Application of Bayesian techniques to model the burden of human salmonellosis attributable to U.S. food commodities at the point of processing: adaptation of a Danish model. *Foodborne Pathogens and Disease*. **8**:509-516.

H.S. Johnson and S. F. Ridlen. 2015. Structure of the Egg. University of Illinois Extension, University of Illinois at Urbana-Champaign, College of ACES: http://extension.illinois.edu/eggs/res16-egg.html.

Hahn, H and T. Stachelhaus. 2004. Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proceedings of the National Academy of Sciences*. 101: 44 15585–15590.

Halbrooks, P.J., A.M. Giannetti, J.S. Klein, P. J. Björkman, J.R. Larouche, V.C. Smith, R.T.A. MacGillivray, S.J. Everse and A.B. Mason, A.B. 2005. Composition of pH-sensitive triad in C-lobe of human serum transferring comparison to sequences of ovotransferrin and lactoferrin provides insight into functional differences in iron release. *Biochemistry*. **44** : 15451-15460.

Hall, M, J. Gabay, M. Débarbouillé and M. Schwartz.1982. A role for mRNA secondary structure in the control of translation initiation. *Nature*. **295:** 616-618.

Hanning, IB., JD. Nutt and SC. Ricke. 2009. Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention measures. *Foodborne Pathogens and Disease*. **6**: 635-648.

Hansen-Wester, I. and M. Hensel. 2002. Genome-based identification of chromosomal regions specific for *Salmonella* spp. *Infection and Immunity*.**70**: 2351–2360.

Hantke, K., G. Nicholson, W. Rabsch and G. Winkelmann. 2003. Salmochelins, siderophores of *Salmonella* enterica and uropathogenic *Escherichia coli* strains, are recognized by the outer membrane receptor IroN. *Proceedings of the National Academy of Sciences*. **100**: 3677-3682.

Haydon, D., and J Guest. 1991. A new family of bacterial regulatory proteins. *Microbiology Letters*. 63: 291-295.

Hendriksen, R.S., R.V. Antonio, S. Karlsmose, M.A. Danilo, L. F. Wong, A. B. Jensen, H.C. Wegener and F.M. Aarestrup. 2011. Global monitoring of *Salmonella* Serovar distribution from the World Health Organisation global foodborne infections network country data bank: results of quality assured laboratories from 2001 to 2007. *Foodborne Pathogens and Disease*. **8**: 887-900.

Hermans, D., K. Van Deun, A. Martel, F.Van Immerseel, W. Messens, M. Heyndrickx, F. Haesebrouck and F.Pasmans. 2011. Colonization factors of *Campylobacter jejuni* in the chicken gut. *Veterinary Research*. **42**:82-90.

Hiidenhovi J. 2007. Ovomucin. In: Huopalahti R., López-Fandiño R., Anton M., Schade R. (eds) Bioactive Egg Compounds. Springer, Berlin, Heidelberg. Chapter 9: 61-68.

Hope B. K., A. R. Baker, E.D. Edel, A. T. Hogue, Schlosser, R. Whiting, R. M. McDowell, R. A. Morales and B.K. Hope. 2002. An overview of the *Salmonella* Enteritidis risk assessment for shell eggs and egg products. *Risk Analysis*. **22**:203-18.

Hugouvieux-Cotte-Pattat, N. and J. Robert-Baudouy.1981. Isolation of fusions between the lac genes and several genes of the exu regulon: analysis of their regulation, determination of the transcription direction of the uxaC-uxaA operon, in Escherichia coli K-12. *Molecular Genetics and Genomics.* **182**: 279-87.

Hurley, D., M. McCusker, S. Fanning and M. Martins. 2014. *Salmonella*–Host Interactions – Modulation of the Host Innate Immune System. *Frontiers in Immunology*. **5**: 481-492.

Ibrahim, H., Y. Sugimoto and T. Aoki, 2000. Ovotransferrin antimicrobial peptide (OTAP-92) kills bacteria through a membrane damage mechanism. *Biochim Biophys Acta*. **18**: 196-205.

Jain D. 2015. Allosteric control of transcription in GntR family of transcription regulators: A structural overview. *International Union of Biochemistry and Molecular Biology Life*. **67:** 556–563.

Janmohamed, K., D. Zenner, C. Little, C. Lane, J. Wain, A. Charlett, and D. Morgan. 2011.National outbreak of *Salmonella* Enteritidis phage type 14b in England, September to December 2009: case–control study. *Eurosurveillance*. **16**: 19840.

Jurtshuk, Peter, and Jr. "Bacterial Metabolism. "Medical Microbiology. 4th edition. U.S. National Library of Medicine, 01 Jan. 1996.

Kamelia M. Osman, M. S. Zeinab, G. Amin, M. A. Bakry and Elsafty. M. M. 2011. Identification of the *Salmonella* Genes *Yafd* and *Xtha*: A Step into the Control of *Salmonella* Food Poisoning Associated with Consumption of Eggs in Egypt. *Global Veterinaria*. **7**: 625-631.

Kammler, M., C.Schon and K.Hantke.1993. Characterization of the ferrous iron uptake system of *Escherichia coli. Journal of Bacteriology*. **175:** 6212–6219.

Kammler, M., C. Schön and K. Hantke, K.1993. Characterization of the ferrous iron uptake system of Escherichia coli. *Journal of Bacteriology*. **175**: 6212–6219.

Kang, H., C. Loui, R. I. Clavijo, L. W. Riley and S. Lu. 2006. Survival characteristics of *Salmonella* enterica serovars Enteritidis in chicken egg albumen. *Epidemiology and Infection*. **134:** 967–976.

Kaoru Ichimura, K. H.Kihara, T. Yamamura," and K.Satake. 1989. Negative Cooperativity of Chicken Ovotransferrin on Al (III)-Binding. *The Journal of Biochemistry*. **106**: 50-54.

Kato, A. and E. Groisman. 2004. Connecting two-component regulatory systems by a protein that protects a response regulator from dephosphorylation by its cognate sensor. *Genes & Development*. **18**: 2302-2313.

Kato, A, A. Mitrophanov and E. Groisman.2007. A connector of two-component regulatory systems promotes signal amplification and persistence of expression. *Proceedings of the National Academy of Sciences*. **104:**12063-8.

Keller LH. et al. 1997. Invasion of chicken reproductive tissues and forming eggs is not unique to *Salmonella enteritidis*. *Avian Diseases*. **41:**535-539.

Keller, L. H., C. E. Benson, K. Krotec, and R. J. Eckroade. 1995. *Salmonella* Enteritidis colonization of the reproductive tract and forming and freshly laid eggs. *Infection and Immunity*. **63**:2443-2449.

Kenyon, W. J.; S. M. Thomas, E. Johnson, M. J. Pallenand M. P. Spector. 2006. Shifts from glucose to certain secondary carbon-sources result in activation of the extra cytoplasmic function sigma factor σE in *Salmonella enterica* serovarTyphimurium. *Microbiology*. **151**: 2373-2383.

Klein, G., S. Müller-Loennies, B. Lindner, N. Kobylak, H. Brade and S. Raina. 2013. Molecular And Structural Basis Of Inner Core Lipopolysaccharide Alterations In *Escherichia Coli*: Incorporation Of Glucuronic Acid And Phosphoethanolamine In The Heptose Region. *The Journal of Biological Chemistry*. **288**: 8111-8127.

Ko, K. Y., A. F. Mendonca, and D. U. Ahn. 2009. Effect of ethylene diamine tetraacetate and lysozyme on the antimicrobial activity of ovotransferrin against *Listeria monocytogenes*. *Poultry Science*.**87:**1649-1658.

Ko, M., and C. Park .2000. Two novel flagellar components and H-NS are involved in the motor function of Escherichia coli. *Journal of Molecular Biology*. **303:** 371-382.

Kobayashi, H., Hall, G. S., Tuohy, M. J., Knothe, U., G. W. Procop and Bauer, T. W. 2009. Bilateral periprosthetic joint infection caused by *Salmonella* enterica serotype Entertitidis, and identification of *Salmonella* sp. using molecular techniques. *International Journal of Infectious Diseases.***13**: e463-e466.

Kottwitz, L. B. M., De Oliveira, T. C. R. M., Alcocer, I., Farah, S. M., Abraha, W. S. M. and Rodrigues, D. P. 2010. Epidemiological data of salmonellosis outbreaks occurred between 1999 and 2008 in Parana State, Brazil. *ActaSci Health Sci.* **32**: 9-15.

Kottwitz, L.B.M., Scheffer, M.C. Dalla-Costa, L.M., Farah, S.M., Moscalewski, W.S., M. Magnani and De Oliveira, T.C. 2011. Molecular characterization and resistance profile of *Salmonella* Enteritidis PT4 and PT9 strains isolated in Brazil. *Journal of Medical Microbiology*. **60**:1026-1031.

Krewulak, K., K. Vogel, J. Hans. 2008. Structural biology of bacterial iron uptake. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1778**: 1781-1804.

Lamble, HJ., CC. Milburn, GL. Taylor, DW. Hough and MJ. Danson. 2004. Gluconate dehydratase from the promiscuous Entner-Doudoroff pathway in Sulfolobus solfataricus. *FEBS Letters*. **8**: 133-6.

Lane, C. R., S. LeBaigue, O. B. Esan, A. A. Awofisyo, N. L. Adams, I. S.T. Fisher, K. A. Grant, T. M. Peters, L. Larkin, R. H. Davies and G. K. Adak. 2014. *Salmonella* enterica Serovar Enteritidis, England and Wales, 1945–2011. *Emerging Infectious Diseases*. **20**: 1908-1194.

Lang, E.R. and C. Rha. 1982. Apparent shear viscosity of native egg white. *International Journal of Food Science & Technology*. **17:** 595–606.

Latendresse, M., J.P. Malerich, M. Travers and P.D. Karp. 2012. Accurate atom-mapping computation for biochemical reactions. *Journal of Chemical Information and Modeling*. **26:** 2970-82.

Le Minor, L. 1984. Genus III. *Salmonella*, p. 427-458. *In* N. R. Krieg, D. H. Bergey, and J. G. Holt (eds.), Bergey's manual of systematic bacteriology. Baltimore, USA.

Le Minor, L., and M. Popoff. 1987. Request for an opinion. Designation of Salmonella enterica sp. nov., nom. rev., as the type and only species of the genus Salmonella. *International Journal of Systematic Bacteriology*. **37:**465–468.

Lee, S.Y., K.S. Yim, H.N. Chang and Y.K. Chang. 1994. Construction of plasmids, estimation of plasmid stability, and use of stable plasmids for the production of poly(3-hydroxybutyric acid) by recombinant *Escherichia coli*. *Journal of Biotechnology*. **14**: 203-211.

Lesic, B. and L. Rahme. 2008. Use of the lambda Red recombinase system to rapidly generate mutants in Pseudomonas aeruginosa. *BMC Molecular Biology*. **9:** 20-27.

Li Y., W. Liang, D. Wang, G. Qi, B. Kan, S. Gao and Y. Liu. 2005. Functional analysis of promoters of Vibrio cholerea typing phage VP1 with reporter system. *Wei Sheng Wu Xue Bao*. **45**:846-50.

Li-Chan, E. and S. Nakai. 1989. Biochemical basis for the properties of egg white. *Critical reviews in poultry biology* (USA).

Lijestroem P., I. Laamanen and T. Palva .1988. Structure and expression of the ompB operon, the regulatory locus for the outer membrane porin regulon in *Salmonella typhimurium* LT-2. *Journal of Molecular Biology*. **201**:663-673

Looman, A., J. Bodlaender, L. Comstock, D. Eaton, P. Jhurani, H. de Boer and P. van Knippenberg .1987. Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in *Escherichia coli*. *European Molecular Biology Organization Journal*. **68**: 2489–2492.

Lu, S., P. B. Killoran and L. W. Riley. 2003. Association of *Salmonella enterica* serovar Enteritidis *yafD* with resistance to chicken egg albumen. *Infection and Immunity*. **71**:6734–6741.

Lu, S., P. Killoran, F. Fang and L. Riley. 2002. The Global Regulator ArcA Controls Resistance to Reactive Nitrogen and Oxygen Intermediates in *Salmonella enterica* Serovar Enteritidis. *Infect Immun*.70: 451-461.

Lucisano, M., A. Hidalgo, E.M. Comelli, M. Rossi, M. 1996. Evolution of Chemical and Physical Albumen Characteristics during the Storage of Shell Eggs. *Journal of Agricultural and Food Chemistry*. **44**: 1235-1240.

Lüders, S., Fallet, C., and Franco-Lara, E. 2009. Proteome analysis of the *Escherichia coli* heat shock response under steady-state conditions. *Proteome Sci.* **7**: 36.

Majowicz, SE., J.Musto and E. Scallan E. 2010. The global burden of non typhoidal *Salmonella* gastroenteritis. *Clinical Infectious Diseases*. **50**:882-889.

Makrides, SC. 1996. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological Reviews*. 60: 512–538.

Marchal, K. 2004. In silico identification and experimental validation of PmrAB targets in *Salmonella* typhimurium by regulatory motif detection. *Genome Biology*. **5**.

Masschalck, B. and C.W. Michiels. 2003. Antimicrobial properties of lysozyme in relation to foodborne vegetative bacteria. Critical Reviews in Microbiology. **29:** 191–214.

Maurer, L.M., E. Yohannes, S.S. Bondurant, M. Radmacher and J.L. Slonczewski. 2005. pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. *Journal of Bacteriology*. **187**: 304–319.

McCarthy JEG & Tuite MF. (2013) Post-transcriptional control of gene expression. Berlin: Springer.

Miethke, M., and M. Marahiel. 2007. Siderophore-based iron acquisition and pathogen control. *Microbiology and Molecular Biology Reviews*. **71:**413–451.

Miller, JH. 1972. Experiments in Molecular Genetics, p 217-352. Cold Springs Harbor Laboratory Press. Cold Springs Harbor, New York.

Mine, Y. 1995. Recent advances in the understanding of egg-white protein functionality. *Trends in Food Science & Technology*. **6:** 225-232.

Mine, Y., and M. Yang. 2008. Recent advances in the understanding of egg allergens:Basic, industrial, and clinical perspectives. *Journal of Agricultural and FoodChemistry*.**56:** 4874-4900.

Morales, C. A., S. Porwollik, J. G. Frye, H. Kinde, M. McClelland and J. Guard-Bouldin. 2005. Correlation of phenotype with the genotype of egg contaminating *Salmonella enterica* serovar Enteritidis. *Applied and Environmental Microbiology*. **71**:4388–4399.

Müller, SI., M. Valdebenito and K. Hantke. 2009. Salmochelin, the long-overlooked catecholate siderophore of *Salmonella.Biometals*.**22**: 691-695.

Murphy, K. and K. Campellone. 2003. Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic E. coli. *BMC Molecular Biology*. **4:** 11.

Murray L. and T. Conway. 2005. Multiple regulators control expression of the Entner-Doudoroff aldolase (Eda) of Escherichia coli. *Journal of Bacteriology* .**187**: 991–1000.

Nasser W, S. Reverchon and J. Robert-Baudouy. 1992. Purification and functional characterization of the KdgR protein, a major repressor of pectinolysis genes of *Erwinia chrysanthemi*. *Molecular Microbiology*. **6**:257–265.

Nasser W, S. Reverchon, G. Condemine and J. Robert-Baudouy. 1994. Specific interactions of *Erwinia chrysanthemi* KdgR repressor with different operators of genes involved in pectinolysis. *Journal of Molecular Biology*. **236**:427–440.

Neidhardt, C.F. 2005. Escherichia coli and Salmonella: Cellular and Molecular Biology. p:321.

Nemoz, G, J Robert-Baudouy, F Stoeber .1976. Physiological and genetic regulation of the aldohexuronate transport system in E. coli. *J. Bacteriol.* **127:**706-18.
Newell, D. G., M. Koopmans, L. Verhoef, E. Duizer, A. Aidara-Kane, H. Sprong, J. V. D. Giessen and H. Kruse. 2010. Food-borne diseases-the challenges of 20 years ago still persist while new ones continue to emerge. *International Journal of Food Microbiology*. **139:** S3-S15.

Nikaido, H. 2003. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiology and Molecular Biology Reviews*. **67:** 593–656.

Nisbet, AD., RH. Saundry, AJ. Moir, LA. Fothergill and JF. Fothergill. 1981. The complete amino-acid sequence of hen ovalbumin, *European Journal of Biochemistry/FEBS*. **115**: 335.

Nolle, N., A. Felsl, R. Heermann and T. Fuchsa. 2017. Genetic Characterization of the Galactitol Utilization Pathway of Salmonella enterica Serovar Typhimurium. *Journal of Bacteriology*.**199**: e00595-16.

Nonaka, G., B. Herman, A. Gross and A. Rhodius. 2006. Regulon and promoter analysis of the E. coli heat-shock factor, s32, reveals a multifaceted cellular response to heat stress. *Genes & Development*. **20**: 1776–1789.

Okamura, M., et al. 2001. Differences among six *Salmonella* serovars in abilities to colonize reproductive organs and to contaminate eggs in laying hens. *Avian Diseases*.**45**: 61–69.

Parker, C. T., B. Harmon and J. Guard-Petter. 2002. Mitigation of avian reproductive tract function by *Salmonella* Entertiidis producing high-molecular- mass lipopolysaccharide. *Environmental Microbiologyl.* **4**:538–545.

Patrick M. E., P. Adcock, T. Gomez, S. Altekruse, B. Holland, R. Tauxe and D. Swerdlow .2004. *Salmonella enteritidis* infections, United States, 1985–1999. *Emerging Infectious Diseases journal*. **10**: 1-7.

Peekhaus N. and T. Conway.1998. What are for dinner? Entner-Doudoroff metabolism in Escherichia coli. *Journal of Bacteriology*. **180**:3495–3502.

Pellegrini A, U. Thomas, N. Bramaz, S. Klauser, P. Hunziker and R. Von Fellenberg. 1997. Identification and isolation of a bactericidal domain in chicken egg white lysozyme. *Journal of Applied Microbiology*. **82:**372-378.

Pellegrini, A., U. Thomas, P. Wild, E. Schraner and R. von Fellenberg. 2000. Effect of lysozyme or modified lysozyme fragments on DNA and RNA synthesis and membrane permeability of Escherichia coli. *Microbiol. Res.* **155**: 69-77.

Pezzotti F. and M. Therisod. 2006. Enzymatic synthesis of aldonic acids. *Carbohydrate Research*. **25:** 2290–2292.

Stancik, L., D. Stancik, B. Schmidt, D. Barnhart, Y. Yoncheva and J. Slonczewski. 2002. pH-dependent expression of periplasmic proteins and amino acid catabolism in *Escherichia coli*. *Journal of Bacteriology*. **184**: 4246–4258.

Pires, SM., H.Vigre, P. Makela and T.Hald. 2010. Using outbreak data for source attribution of human salmonellosis and campylobacteriosis in Europe. *Foodborne Pathogens and Disease*.**7:**1351-1361.

Pires, SM., L. de Knegt and T. Hald. 2011. Estimation of the relative contribution of different food and animal sources to human *Salmonella* infections in the European Union: *European Food Safety Authority*. P: 1-80.

Portalier, R, J. Robert-Baudouy and F. Stoeber .1980. Regulation of Escherichia coli K-12 hexuronate system genes: exu regulon. *J Bacteriol*.43:1095-107.

Porwollik, S. et al. 2005. Differences in Gene Content Between Salmonella enterica Serovar Enteritidis Isolates and Comparison to Closely Related Serovars Gallinarum and Dublin. Journal of Bacteriology. **187:** 6545-6555.

Postle, K. and RG. Kadner. 2003. Touch and go: tying TonB to transport. *Molecular Microbiology*. **49:**869–882.

Pui, C. F., Wong, W. C., Chai, L. C., R. Tunung, P. Jeyaletchumi, M. S. Noor Hidayah, A. Ubong, M. G. Farinazleen, Y. K. Cheah and R. Son. 2011. Review Article *Salmonella*: A foodborne pathogen. *International Food Research Journal*. **18**: 465-473.

Rabsch, W., U. Methner, W. Voigt, H. Tscha[•]pe, R. Reissbrodt and P. H. Williams. 2003. Role of Receptor Proteins for Enterobactin and 2,3-Dihydroxybenzoylserine in Virulence of *Salmonella enteric*. *Infection and Immunity*.**71**: 6953–6961.

Radziejewska, R. C., G. Le-nierowski, and J. Kijowski. 2008. Properties and application of egg white lysozyme and its modified preparations. *Polish Journal of Food and Nutrition Sciences*. **58:**5–10.

Ramachandran S., P. Fontanille, A. Pandey and C. Larroche1. 2006. Gluconic Acid: Properties, Applications and Microbial Production. *A Review, Food Technol. Biotechnol.***44**: 185–195.

Ran, N., K. Draths and J. Frost.2004. "Creation of a shikimate pathway variant. *J Am Chem Soc.* **126:** 6856-7.

Raspoet, R., C. Appia-Ayme, N. Shearer, A. Martel, F., Haesebrouck. 2014. Microarray-based detection of Salmonella enterica serovar Enteritidis genes involved in chicken reproductive tract colonization. *Applied and Environmental Microbiology*. **80**: 7710-7716.

Raspoet, R., I. Gantois, R. Devloo, A. Martel, F. Haesebrouck, F. Pasmans, R. Ducatelleand F. Van Immerseel .2011. *Salmonella* Enteritidis universal stress protein (usp) gene expression is stimulated by egg white and supports oviduct colonization and egg contamination in laying hens. *Veterinary Microbiology*. **153**:186–190.

Raspoet, R., N. Shearer, C. Appia-Ayme, F. Haesebrouck, R. Ducatelle, A. Thompson, F. Van Immerseel. 2014. A genome-wide screen identifies *Salmonella* Enteritidis lipopolysaccharide biosynthesis and the HtrA heat shock protein as crucial factors involved in egg white persistence at chicken body temperature. *Poultry Science*. **93**: 1263–1269.

Ratledge, C., and L. G. Dover. 2000. Iron metabolism in pathogenic bacteria. *Annual Review of Microbiology*. **54**:881-941.

Rawlings, N.D. et al. 2004. Evolutionary families of peptidase inhibitors. *Biochemical Journal*. **378**: 705-716.

Resch, M., E. Schiltz, F. Titgemeyer and Y. Muller. 2010. Insight into the induction mechanism of the GntR/HutC bacterial transcription regulator YvoA. *Nucleic Acids Res.* **38**:2485–2497.

Rigali, S., A. Derouaux, F. Giannotta and J. Dusart. 2002. Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem.* **277**:12507–15.

Ritzenthaler P and M. Mata-Gilsinger. 1982. Use of In Vitro Gene Fusions to Study the uxuR Regulatory Gene in *Escherichia coli* K-12: Direction of Transcription and Regulation of Its Expression. *Journal of Bacteriology*.**150**:1040-1047.

Robert-Baudouy, J. M., R. C. Portalier, and F. R. Stoeber. 1974. Regulation of hexuronate metabolism in *Escherichia coli* K-12. Kinetics of enzyme induction of the hexuronate system. *European Journal of Biochemistry*. **43:**1-15.

Robyt, J.F. 1997. Essentials of Carbohydrate Chemistry, Springer-Verlag, New York.13-14.

Rodionov A., M. Gelfand, N. Hugouvieux-Cotte-Pattat. 2004. Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria. *Microbiology*.**150**:3571-90.

Rodionov, A., A. Mironov, B. Rakhmaninova, A. B. and M. Gelfand. 2000. Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria. *Molecular Microbiology*. **38**: 673-683.

Roland, K.L. 1993. Spontaneous pmrA mutants of *Salmonella* typhimurium LT2 define a new two-component regulatory system with a possible role in virulence. *Journal of Bacteriology*. 175, 4154–4164.

Růzicková, V. 1994. Growth and survival of Salmonella enteritidis in selected egg foods. Vet Med (Praha). **39:** 187–195.

Ryan, D., N. Pati, U. Ojha, C. Padhi, S. Ray, S. Jaiswal, G. Singh, G. Mannala, T. Schultze, T. Chakraborty and M. Suar .2015. Global Transcriptome and Mutagenic Analyses of the Acid Tolerance Response of *Salmonella enterica* Serovar Typhimurium. *Applied and Environmental Microbiology*. **81**: 23 8054-8065.

Sahin, O., P. Kobalka and Q. Zhang.2003. Detection and survival of Campylobacter in chicken eggs. *Journal of Applied Microbiology*. **95:** 1070–1079.

Salehi, T.Z., O. Madadgar, S. Naserli, B.N. Fasaei, M.M. Ghafari and I. A. Tamai. 2011. Detection of sef14, sef17 and sef21 Fimbrial Virulence Genes of *Salmonella* enteritid is by Multiplex PCR. *Journal of Animal and Veterinary Advances*. **10**: 1421-1426.

Schaffer, S., K. Hantke and V. Braun. 1985. Nucleotide sequence of the iron regulatory gene fur. *Molecular Genetics and Genomics.* **200**: 110–113.

Sellier, N., M. L. Vidal, F. Baron, J. Michel, J. Gautron, M. Protais, C. Beaumont, M. Gautier and Y. Nys. 2007. Estimations of repeatability and heritability of egg albumen antimicrobial activity and of lysozyme and ovotransferrin concentrations. *British Poultry Science*. **48**: 559-566.

Shachar, D. and S. Yaron. 2006. Heat tolerance of *Salmonella enterica* serovars Agona, Enteritidis, and Typhimurium in peanut butter. *Journal of Food Protection*.**69**:2687-91.

Shah D.H. 2014. RNA Sequencing Reveals Differences between the Global Transcriptomes of *Salmonella enterica* Serovar Enteritidis Strains with High and Low Pathogenicities. *Applied and Environmental Microbiology*. **80:** 896–906.

Shah D. H., C. Casavant, Q. Hawley, T. Addwebi, D.R. Call and J. Guard. 2012. *Salmonella* Enteritidis strains from poultry exhibit differential responses to acid stress, oxidative stress, and survival in the egg albumen. *Foodborne Pathogens and Disease*. **9:** 258-264.

Shah, D.H., X. Zhou, T. Addwebi, M. A. Davis, L. Orfe, R. Douglas Call, J. Guard and T.B. Besse. 2011. Cell invasion of poultry-associated *Salmonella* enterica serovar Enteritidis isolates is associated with pathogenicity, motility and proteins secreted by the type III secretion system. *Microbiology*. **157**: 1428–1445.

Shan, Y., M. Ma, Z. Yuan, H. Zhu, X. Huang and S. Zuo. 2013. Antibacterial Activity of Ovomucin and Its Stability to pH and Temperature. *Food Science*. **34**: 1-4.

Siguret, V., A.S. Ribba, G. Chérel, D. Meyer, and G. Piétu. 1994. Effect of Plasmid Size on Transformation Efficiency by Electroporation of *Escherichia coli* DH5 Alpha. *BioTechniques*. **16**: 422-426.

Silva, C. A., C. J. Blondel, C.P. Quezada and C. A. Santiviago. 2012. Infection of Mice by *Salmonella enterica* Serovar Enteritidis Involves Additional Genes That Are Absent in the Genome of Serovar Typhimurium. *Infection and Immunity*. **80**: p. 839–849.

Simons, R. W., R. Houman, and N. Kleckner. 1987. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene*. **53**:85–96.

Sood, S., P. Rishi, H. Vohra, S. Sharma and N. K. Ganguly. 2005. Cellular immune response induced by *Salmonella enterica serotype Typhi* iron-regulated outer-membrane proteins at peripheral and mucosal levels. *Journal of Medical Microbiology*. **54**:815–821.

Stratagene .2004 XL-1 Blue Competent Cells. Available at: http://www.chem-agilent.com/pdf/strata/200249.pdf.]. XL1-Blue Competent Cells Catalog #200249.

Suvorova, I., M. Tutukina, D. Ravcheev, O. and M. Gelfand. 2011. Comparative Genomic Analysis of the Hexuronate Metabolism Genes and Their Regulation in Gammaproteobacteria. *Journal of Bacteriology*.**193**: 3956-3963.

Szumiło, T. 1981. Pathway for D-galactonate catabolism in nonpathogenic mycobacteria. *Journal of Bacteriology*. **148:** 368-70.

Tamayo, R., A. Prouty and J. Gunn. 2005. Identification and functional analysis of *Salmonella enterica* serovar Typhimurium PmrA-regulated genes.FEMS Immunology & Medical Microbiology. **43**: 249-258.

Tamayo, R., S. Ryan, A. McCoy and J. Gunn. 2002. Identification and Genetic Characterization of PmrA-Regulated Genes and Genes Involved in Polymyxin B Resistance in *Salmonella enterica* Serovar Typhimurium. *Infection and Immunity*. **70**: 6770-6778.

Thapon, J.-L. and V. Audiot. 1994. Composition de l'oeuf de poule, in: Thapon, J.-L., Bourgeois, C.M. (Eds.), L'oeuf et Les Ovoproduits, Tec and Doc. Paris, pp. 6–26.

Thiagarajan, D., Saeed, M., Turek, and J., Asem, E.K. 1996. In vitro attachment and invasion of chicken ovarian granulosa cell by *Salmonella* Enteritidis phage type 8. *Infection and Immunity*. **64:** 5015-5021.

Thomson, NR, DJ. Clayton and D. Windhorst. 2008. Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Research.* **18**: 1624-1637.

Van Aalten D., C. DiRusso, J. Knudsen J, R. Wierenga. 2000. Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. *European Molecular Biology Organization Journal*. **19:**5167-5177.

Van Dijk, A., E. Veldhuizen and H. Haagsman. 2008. Avian defensins. *Veterinary Immunology and Immunopathology*. **124:**1-18.

Velišek J. (2014): The chemistry of food. John Wiley & Sons, Inc., New Jersey: 235-236.

Venkataramani S.; J. Truntzer and D. R. Coleman. 2013. Thermal stability of high concentration lysozyme across varying pH: A Fourier Transform Infrared study. *Journal of Pharmacy And Bioallied Sciences*. **5**: 148-153.

Walters, M. V. Srikannathasan, A. McEwan, J.Naismith, C. Fierke and E.Toone. 2008. Characterization and Crystal Structure of *Escherichia coli* KDPGal Aldolase. *Bioorg Med Chem.* **16:** 710–720.

Wellman-Labadie, O., J. Picman and M.T. Hincke. 2008. Comparative antibacterial activity of avian egg white protein extracts. *British Poultry Science*.**49**: 125-132.

Western Kentucky University. 2013. "WKU BIO 113 Carbohydrates". wku.edu.

Wetter, L. And H. Deutsch. 1951. Immunological studies on egg white proteins: iv. immunochemical and physical studies of lysozyme. *Journal of Biological Chemistry*. **192**: 237-242.

WHO, 2005. The World Health Report 2005 - making every mother and child count. Geneva, Switzerland WHO.

Wisner, A. L. S., T. S. Desin, B. Koch, P. S. Lam, E. M. Berberov, C. S. Mickael, A. A. Potter and W. Koster. 2010. *Salmonella* enterica subspecies enterica serovars Enteritidis *Salmonella* pathogenicity island 2 type III secretion system: role in intestinal colonization of chickens and systemic spread. *Microbiology*. **156**: 2770-2781.

Wolfe, A. J., N. Parikh, P. Lima and B. Zemaitaitis. 2008. Signal integration by the twocomponent signal transduction response regulator CpxR. *Journal of Bacteriology*. **190**: 2314-2322. Wong KK, et al.1998. Identification and sequence analysis of a 27- kilobase chromosomal fragment containing a *Salmonella* pathogenicity island located at 92 minutes on the chromosome map of *Salmonella enteric* serovar Typhimurium LT2. *Infection and Immunity*. **66**:3365–3371.

Woods, D. F., F.J. Reen, D. Gilroy, J. Buckley, J.G Frye and E.F. Boyd. 2008. Rapid multiplex PCR and real-time TaqMan PCR assays for detection of *Salmonella enterica* and the highly virulent serovars Choleraesuis and Paratyphi C. *Journal of Clinical Microbiology*. **46**: 4018-4022.

Wooldridge, K. G. and P.H Williams. 1993. Iron uptake mechanisms of pathogenic bacteria. *FEMS Microbiology Reviews*. **12**: 325-348.

Wosten, M. M., L. F. Kox, S. Chamnongpol, F. C. Soncini, and E. A. Groisman. 2000. A signal transduction system that responds to extracellular iron. *Cell.* **103**:113-125.

Wu, J., and A. Acer-Lopez. 2012. Ovotransferrin: Structure, bio activities and preparation. *Food Research International*. **46**:480-487.

WU, J., and A. Acer-Lopez. 2012. Ovotransferrin: Structure, bioactivities, and preparation. *Food Research International*. **46**: 480-487.

Yang, X.-W., He, Y., Xu, J., Xiao, X., and F. Wang. 2013. The regulatory role of ferric uptake regulator (Fur) during anaerobic respiration of Shewanella piezotoleransWP3. *PLoS ONE*. **8**:1371.

Yim.L., Betancor, L., Martinez, A., Giossa, G., Bryant, C., Maskell, D. & Chabalgoity, J. A. 2010. Differential phenotypic diversity among epidemic-spanning *Salmonella* enterica serovar Enteritidis isolates from humans or animals. *Applied and Environmental Microbiology*.**76**: 6812-6820.

Yohannes, E., M. Barnhart L. and Slonczewski. 2004. pH-dependent catabolic protein expression during anaerobic growth of Escherichia coli K-12. *Journal of Bacteriology*. **186:** 192-199.

Yoke-Kqueen, C., L. Learn-Han, A.S. Noorzaleha, R.Son, S. Sabrina, S. Jiun-Horng and K. Chai-Hoon. 2008. Characterization of multiple-antimicrobial resistant *Salmonella enterica* subsp. Enterica isolated from indigenous vegetables and poultry in Malaysia. *Letters in Applied Microbiology*. **46**: 318-324.

Yu, Z., W. Qin, J. Lin, S. Fang, and J. Qiu. 2015. Antibacterial Mechanisms of Polymyxin and Bacterial Resistance. *BioMed Research International*. ID 679109.

Zarate-Bonilla, L.J., P. del Portillo, H. Saenz-Suarez, J. Gonzales-Santos, G. E. Barreto-Sampaio, R.A. Poutou-Pinales, A. F. Rey and J. G. Rey. 2014. Computational modeling and preliminary

*iro*N, *fep*A, and *cir*A gene expressionin *Salmonella* Enteritidis under iron-deficiency-induced conditions1. *Poultry Science*. **93:**221-230.

Zhao K., M. Liu and R. R. Burgess. 2007. Adaptation in bacterial flagellar and motility systems: From regulon members to "foraging"-like behavior in *E. coli. Nucleic Acids Research*. **35:** 4441-4452.

Zhao YW, W. Liang, D. Wang, G. Qi, B. Kan, S. Gao and Y. Liu. 2005. Sequencing and analysis of *Vibrio cholerae* typing phage VP3 genome. *Wei Sheng Wu Xue Bao*. **45**:851-5.

Zhensheng, P. et al. 2009. Identification of genetic and phenotypic differences associated with prevalent and non-prevalent *Salmonella* Enteritidis phage types: analysis of variation in amino acid transport. *Microbiology*. **155**:3200-3213.

Zhou J. and E. Rudd. 2013. EcoGene 3.0. Nucleic Acids Research. 41: D613–D624.

Zhou, D., W. Hardt and J. E. Galan. 1999. *Salmonella typhimurium* Encodes a Putative Iron Transport System within the Centisome 63 Pathogenicity Island. *Infection and Immunity*. **67**:1974–1981.

Zhou, QM., DJ. Fan and JB., Xie, 2010. Method for generating precise gene deletions and insertions in Escherichia coli. World Journal of Microbiology and Biotechnology. **26:** 1323–1329.

Zhou, Z. 1999. Lipid A modifications characteristic of Salmonella typhimurium are induced by NH4VO3 in Escherichia coli K12. *The Journal of Biological Chemistry*. **274:** 18503–18514.

Zhu, M., M. Valdebenito, G. Winkelmann and K. Hantke. 2005. Functions of the siderophore esterases IroD and IroE in iron-salmochelin utilization.*Microbiology*. **151**: 2363–2372.

Zou, W., W.J. Lin, S.L. Foley, C. H. Chen, R. Nayak and J. J. Chen. 2010. Evaluation of pulsed-field gel electrophoresis profiles for identification of *Salmonella* serotypes. *Journal of Clinical Microbiology*. **48:** 3122–3126.

Appendixs

Appendix 1: Translated sequence of induced genes (using Vector NTI program)

	1	TTCTCATCGG	ATAAGGTTAT	TTTAAAATAT	AAGTTTTTGT	TGATTTTTAC	TGATTAAGAA	CCAATCCCAT	TAGGCTCTAA	GTGATGATGT	CCARAGTAGC
-	101	AATACACATT	CTTTTAGAAA	AACGGTATTT	TCTGTTTGAT	TAATGAGGTT	ATCTAATACA	TCAGGAGGGA	GAATACGGCC	TGAAAAGGGG	AATCCCGTTT
-	+3	TTATGTGTAA	GAAAATCTTT	TTGCCATAAA	AGACAAACTA	ATTACTCCAA	TAGATTATGT	AGTCCTCCCT	CTTATGCCGG	ACTTTTCCCC	M A I
	201	GTGAATCTTA	TCCCCCTTTC	TCTCGGATTC	TTTGCTGTTT	TTTGAGCGGA	ATCGCGTTAG	CATGAGTCAG	GACTCATTTA	ATGCGGGGGAT	TTCATGGCTA
yidA	+3	I K L I	A I D	M D G	T L L L	P D H	T I S	P A V F	N A I	A A A	R E K G
229 aa-690 bp.	301	TCAAACTTAT AGTTTGAATA	TGCTATCGAC ACGATAGCTG	ATGGATGGCA TACCTACCGT	CCCTTCTGCT	GCCCGATCAC	ACCATTTCTC	CGGCGGTTAA	AAACGCGATT	GCCGCTGCGC CGGCGACGCG	GTGAAAAAGG CACTTTTTCC
Fr; 3,907,912	+3	G V N V	V L T	T G R P	Y A G	V H S	Y L K F	L H M	E Q P	G D Y C	I T Y
To; 3,908,601	401	GGTAAATGTG CCATTTACAC	GTGCTGACCA CACGACTGGT	CAGGCCGTCC GTCCGGCAGG	GTATGCGGGT CATACGCCCA	GTGCACAGTT CACGTGTCAA	ACCTGAAAGA TGGACTTTCT	ACTTCACATG TGAAGTGTAC	GAACAGCCCG CTTGTCGGGC	GCGATTATTG CGCTAATAAC	CATCACCTAT GTAGTGGATA
B5OUP2	+3	N G A		A G D	G S T	V A Q T	ALS	Y D D	Y R Y L	EKL	SREV
Galactonate	301	TTGCCCCGCG	ACCACGTCTT	TCGTCCCCTG	CCGTCATGCC	AACGCGTTTG	CCGCGAGTCG	ATACTACTGA	TGGCAATGGA	CCTTTTTGAC	AGGGCACTCC
operon	+3 601	TGGGTTCTCA	E H A	TTAGACCGAA	ATACGCTTTA	TACCECTAAC	CGCGATATCA	GCTACTACAC	GGTGCATGAG	TCGTATGTGG	CGACCATTCC
transcriptional -	12	PLVF	C E A	AATCTGGCTT	P N T	ATEGCEATTE	GCGCTATAGT	CGATGATGTG	P A V	AGCATACACC	GCTGGTAAGG
repressor	701	GCTGGTATTT	TGTGAAGCGG	AGAAGATGGA	CCCGAACACC	CAGTTCCTGA	AAGTTATGAT	GATCGATGAG	CCTGCCGTTC	TCGACCGGGC	GATTGCGCGT
	+3	I P A	E V K E	K Y T	V L K	GTCAAGGACT	TTCAATACTA	I L D	K R V N	K G T	G V KS
	801	ATACCESCAS TATESCCETC	AGGTGAAGGA TCCACTTCCT	AAAGTACACC TTTCATGTGG	GTGCTGAAAA Cacgactttt	GCGCGCCGTA CGCGCGGCAT	CTTCCTTGAA GAAGGAACTT	ATCCTCGATA TAGGAGCTAT	AACGGGTTAA TTGCCCAATT	TAAAGGCACC ATTTCCGTGG	GGCGTAAAAT CCGCATTTTA
-	+3	S L A E	A L G	IKP	EEVN	í A I G	DQE	N D I A	MIE	Y A G	MGVA
	901	CACTESCCEA STEACCESCT	GGCGCTGGGT CCGCGACCCA	ATTAAGCCAG TAATTCGGTC	ACCACCACTA TCCTCCACTA	GGCGATTGGC CCGCTAACCG	GATCAGGAAA CTAGTCCTTT	ACGACATTGC TGCTGTAACG	GATGATCGAA CTACTAGCTT	TACGCCGGTA ATGCGGCCAT	TGGGCGTGGC ACCCGCACCG
	+3	A M D N	A I P GCCATTCCGT	S V K E	V A N GGTGGCTAAC	TTTGTGACTA	AATCGAACCT	TGAAGATGGT	U A W	A I E K	ATTIGTGCTG
-		TTACCTGTTG	CGGTAAGGCA	GCCAGTTTCT	CCACCGATTG	AAACACTGAT	TTAGCTTGGA	ACTTCTACCA	CAACGGACCC	GCTAACTTTT	TAAACACGAC
	+3 1101	AACCCCGATC	ACTCATCCGG	CCATTTCCCC	GCCCGATAAG	GCATAGCCGC	CATCGGGCAA	ATACGCGCTT	AACGACCCGC	ACTTGCTGCG	GGTTTTTTTA
-	1201	TTGGGGGCTAG TGTCTTTCGT	TGAGTAGGCC TTACGTCTTA	GGTAAAGGGG TAACGTTCCC	CGGGCTATTC ATAACCAATT	CGTATCGGCG GTTGTTTTTG	GTAGCCCGTT TGATCTAAAT	TATGCGCGAA TGTAGTACAA	TTGCTGGGCG CATAATTATG	TGAACGACGC TTGTACTACA	CCAAAAAAAT TTAATGGCAT
-		ACAGAAAGCA	AATGCAGAAT	ATTGCAAGGG	TATTGGTTAA	CAACAAAAAC	ACTAGATTTA	ACATCATGTT	GTATTAATAC	AACATGATGT	AATTACCGTA
	1301	GATAACGACG	GTTGATATCA	CGCTAGTACT	ACAAAATTGC	GGCGTAATTC	AGCTATCGCG	GTAAAGTAAG	AGAGTTCACA	TCGAGCACAA	GGACTCTCTA
-	+2	M T L N	K T D	R I V	I T L G	K Q I	V S G	K Y V F	G S A	L P A	E A D L
	1401	TGACTCTCAA ACTGAGAGTT	TAAAACCGAT ATTTTGGCTA	CGCATCGTTA GCGTAGCAAT	TCACGCTGGG AGTGCGACCC	CARACAGATT GTTTGTCTAR	GTCAGCGGTA	AATACGTACC TTATGCATGG	<u>CGGTTCGGCG</u> GCCAAGCCGC	CTGCCAGCGG GACGGTCGCC	AAGCGGATCT TTCGCCTAGA
dgoR	+2	LCEE	F E T	S R N I	IRE	V F R	S L M A	KRL	IEM	K R Y R	GAF
281 aa-846 bp.	1501	GTGCGAGGAG CACGCTCCTC	TTTGAAACGT AAACTTTGCA	CGCGCAACAT GCGCGTTGTA	CATTCGCGAA GTAAGCGCTT	GTGTTTCGTT CACAAAGCAA	CGCTTATGGC GCGAATACCG	GAAGCGGCTA CTTCGCCGAT	ATTGAAATGA TAACTTTACT	AGCGCTATCG TCGCGATAGC	GCCGCGCGTTT
294-1139 Er: 3 908 862	+2 1601	I A P	R N Q W	GAATTATCTC	D T D GATACCGACG	TGCTGCAATG	GGTGCTGGAA	AATGACTACG	D P R L	I S A	M S E I
To; 3,909,707		TAGCGTGGCG	CATTGGTCAC	CTTAATAGAG	CTATGGCTGC	ACGACGTTAC	CCACGACCTT	TTACTGATGC	TGGGCTCCGA	ATAGTCGCGC	TACTCGCTTT
D5OLID2	1701	TACGAAACCT	GGTGGAGCCA	GCAATAGCAC	GCTGGGCGGC	GGAACGGGCA	ACATCAAGCG	ATCTGGCTGA	AATTGAGTCG	GCGCTAAACG	ACATGATTGC
DoQUP5 Uncharacterized	+2	ATGCTTTGGA	D R E	CGTTATCGTG A F N E	CGACCCGCCG	R Y H	E A V I	TAGACCGACT	H N P	CGCGATTTGC	L N V
protein	1801	CAACAACCAG GTTGTTGGTC	GACCGGGAAG CTGGCCCTTC	CGTTTAACGA GCAAATTGCT	GGCGGATATT	CGCTATCACG GCGATAGTGC	AAGCAGTGTT TTCGTCACAA	GCAGTCGGTG	CATAACCCGG GTATTGGGCC	TACTGCAACA ATGACGTTGT	GTTAAATGTA CAATTTACAT
	+2	A I S	S L Q R	A V F	ERT	W M G E	A A N	M P K	T L Q F	H K A	
_	1901	CGCTAGTCGA	GCGATGTCGC	AGCGGTATTT TCGCCATAAA	CTTGCCTGGA	CCTACCCGCT	ACGCCGGTTA	TACGGCTTTT	GCGAGGTCCT	TGTATTCCGC	GATAAGCTAC
	+2	AIR H	Q D G		EQAA		I A S	S T R R			ARY
	2001	CGATACGGCA	TCAGGATGGC	GATGCGGCAG	AGCAGGCGGC	ATTAACCATG	ATCGCCAGCT	CGACACGAAG	GTTAAAGGAA	ATCACATGAC	AGCTCGCTAC
-	+1	I A I	D W G S	T N L	R A W	L Y Q G	D K C	L E S	R Q S E	A G V	TRLN
	2101	ATCGCAATTG TAGCGTTAAC	ACTGGGGATC TGACCCCTAG	GACCAATCTG CTGGTTAGAC	CGCGCCTGGC GCGCGGACCG	TTTACCAGGG AAATGGTCCC	CGACAAATGC GCTGTTTACG	CTGGAGAGCA GACCTCTCGT	GGCAATCAGA CCGTTAGTCT	AGCAGGCGTT TCGTCCGCAA	ACACGCCTGA TGTGCGGACT
	+1	NGK S	P D A	V L A	E V T T	H W R	D S A	T P V V	MAG	MIG	S N V G
dgoK	2201	TGCCATTTAG	AGGACTGCGC	CACAATCGTC	TTCAGTGTTG	TGTGACCGCG	CTGTCGCGGT	GGGGTCACCA	TTACCGCCCG	TACTAGCCGT	CATTGCATCC
292 aa- 879 bp. 2086-2964	+1 2301	CTGGCAAAAT	GCGCCTTATC	TGCCGGTTCC	CGCCCTGTTC	TCCGCTATTG	GCGAACAGTT	AACCGCCGTT	GGCGACAACA	TCTGGATCAT	TCCCGGATTG
Fr; 3,906,436	+1	GACCGTTTTA C V S	CGCGGAATAG	ACGGCCAAGG	GCGGGACAAG M R G	AGGCGATAAC	CGCTTGTCAA	A R E	L S P S	AGACCTAGTA S V Y	AGGGCCTAAC V M P G
To; 3,907,053	2401	TGTGTCTCAC	GCGAGGATAA	CCACAACGTG	ATGCGCGGTG	AAGAGACGCA	ACTGCTTGGC	GCCCGCGAAC	TTTCTCCTTC	TTCTGTCTAT	GTCATGCCCG
B5QUP1	+1	G T H C	K W V	QTD	Τ _ Q _ Q _ I	HDF	R T V	M T G E	LHH	LLL	R H S L
2-dehydro-3-	2501	GCACGCATTG CGTGCGTAAC	CAAATGGGTA GTTTACCCAT	CAGACTGATA GTCTGACTAT	CGCAACAAAT GCGTTGTTTA	TCATGATTTT AGTACTAAAA	CGTACTGTGA GCATGACACT	TGACAGGCGA ACTGTCCGCT	ACTCCATCAC TGAGGTAGTG	TTGTTGCTGC AACAACGACG	GTCACTCGCT CAGTGAGCGA
galactonokinase	+1	L V G A	G L P	E Q E V	S G D	A Y A	A G L E	R G L	N S P	A V L P	TTCTCTTTTT
		CCAGCCCCGA	CCAAACGGCC	TTGTCCTTCA	AAGACCGCTG	CGGATACGGC	GCCCCGACCT	CGCGCCAGAA	TTAAGAGGAC	GGCAGGACGG	AAGAGAAAAA
	+1 2701	GAGGTTCGCG	CCTCGCACGT	GTTGGGACAC	сттесесете	AGCAGGTCAG	CGACTTCCTC	тесевсствт	TGATTGGCGC	GGAAGTCGCC	AGCATGAGCG
-	+1	E S F A	A Q Q	A I T	GAACGCGCAC L V A G	P A L	GCTGAAGGAG I S R	AGGCCGGACA Y Q Q A	ACTAACCGCG F S A	CCTTCAGCGG I G R	TCGTACTCGC D V S T
	2801	AATCCTTCGC	GGCGCAACAG	GCTATCACTC	TCGTCGCTGG	ACCCGCGCTG	ATCTCACGTT	ACCAACAGGC	GTTTAGTGCT	ATTGGGCGTG	ATGTTTCAAC
-	+2	TROOMOCO		COATAOTOAO	2002000200	M		T N L F	LIA	I L R	G I T P
	+1 2901	CGTGGATGGC	D M A GATATGGCAT	F Q A G	I R S AATAAGGAGC	I A H ATCGCTCATG	A V A N	CTAATCTCCC	TCTCATCGCT	ATCTTACGCG	GTATTACGCC
dgoD		GCACCTACCG	CTATACCGTA	AAGTCCGACC	TTATTCCTCG	TAGCGAGTAC	GTCACCGTTT	GATTAGAGGG	AGAGTAGCGA	TAGAATGCGC	CATAATGCGG T S F
205 aa-618 bp.	3001	CGATGATGCC	CTGGCGCACG	TTGGCGCGGT	GGTGGATGCG	GGATTTGACG	CTATAGAAAT	TCCGCTTAAC	TCCCCACAGT	GGGAAAAAG	CATTTCTTTC
Fr;3,906,436	+2	V V K	A Y G G	R A L	I G A	G T V I	GATATCTTTA K P E	AGGCGAATTG	AGGGGTGTCA	MGC	K L I V
To; 3,907,053	3101	GTGGTGAAGG CACCACTTCC	CGTATGGCGG GCATACCGCC	CAGGGCGCTT GTCCCGCGAA	ATTGGCGCTG TAACCGCGAC	GTACCGTACT CATGGCATGA	GAAACCGGAA CTTTGGCCTT	CAGGTAGACC GTCCATCTGG	AGCTTGCCGG TCGAACGGCC	GATGGGCTGC CTACCCGACG	AAGCTGATCG TTCGACTAGC
B5QUP0 2-debydro-3	+2	V T P N	I Q P	EVI	R R A V	S Y G	M T V	C P G C	ATA	TEA	F S A L
deoxy-6-	3201	AGTGCGGCTT	ATAGGTTGGC	GAGGTGATCC CTCCACTAGG	CGGCCCGCCA	GAGCTATGGC CTCGATACCG	ATGACCGTGT TACTGGCACA	CAGGCCCGAC	GCGGTGCCGT	ACGGAAGCCT TGCCTTCGGA	AAAGACGCGA
phospho	+2 3301	L D A G	A Q A GCACAGGCGT	L K I F	P S S CCCGTCGTCG	A F G GCGTTTGGTC	P G Y I CGGGCTACAT	CAGCGCGCTG	K A V	L P P D TTCCGCCGGA	TGTTCCGCTA
galactonate		CCTACGTCCG	CGTGTCCGCA G G V T	ATTTTTAAAA	GGGCAGCAGC	CGCAAACCAG	GCCCGATGTA G C V	GTCGCGCGAC	TTTCGCCATG	AAGGCGGCCT L Y R	ACAAGGCGAT A G O S
andonase	3401	TTTGCCGTCG	GCGGCGTGAC	GCCGGAAAAC	CTGGCGCAAT	GGATTAAAGC	COCCUTOTOTO	GGCGCGGGAT	TGGGTAGCGA	TCTCTATCGC	GCCGGGCAAT
			JULUCAUIG	2000011110	JACCOUCTIA.	- DOLMALICO	JOGGAGAGAG	ALOODOCCE.	1000410001		I / / / / / / / / / / / / / / / /

Appendix

	12	SVERTAQ QAAAFVNAYR EAVK *
	+4	
	+1	
	3501	CONTRACTS CREECECERS CREECING CATTURE IS CONTROL RECONTRACT AND AND A CONTROL OF
-		BEARCHIGE GIGGGEGEE GIEGAEGEE GIEAACAATI ACCEATACT LECEGEACT INCLINIC INACIDIATI GEIGEATGE AAAGAAGI
	+1	алия и али али али али али али али али али а
	3601	COTTGGATGT TCCTGAAAAT CGAAACGGAT GAAGGCGTGG TTGGCTGGGG ACAGCCGGTC ATTGAAGGTC GGCCACGTAC TGTACAGGCG GCACTACATG
-		GEARCETRER REGRETTITA GETTIGEETA ETTECEGACE ARCEGRECECE TETEGGECAG TARCITECAG CECEGIGERIG ACATETECEGE CETATATAC
	+1	EFADYLIGKDPARINDLWQVMYRAGFYRGGPIMM
dao A	3701	AGTTTGCCGA CTACCTGATA GGGAAAGATC CGGCGCGTAT CAACGACCTA TGGCAGGTAA TGTACCGGGC CGGTTTTTAT CGCGGCGGCC CGATTATGAT
202 1140 h-		TCANACGGCT GATGGACTAT CCCTTTCTAG GCCGCGCATA GTTGCTGGAT ACCGTCCATT ACATGGCCCG GCCAMAATA GCGCCGCCGG GCTAATACTA
382 aa-1149 bp.	+1	MS À I À G I D Q À L W D I K G K V L N À P V W Q L M G G L V R D K
3562-4710	3801	GAGCGCCATC GCCGGTATTG ACCAGGCATT GTGGGATATC AAAGGCAAGG TGTTGAATGC GCCGGTCTGG CAGCTCATGG GCGGCCTAGT GCGCGACAAA
Fr;3,905,291	_	CTCGCGGTAG CGGCCATAAC TGGTCCGTAA CACCCTATAG TTTCCGTTCC ACAACTTACG CGGCCAGACC GTCGAGTACC CGCCGGATCA CGCGCTGTTT
To:3.906.439	+1	I KAYSW V G G D R PAD V I D G I E KL R G I G F D T F KL NG
B5OUN9	3901	ATCAAGGCCT ATAGCTGGGT GGGTGGCGAT CGTCCGGCAG ACGTCATTGA CGGTATTGAA AAATTGCGCG GTATTGGTTT TGACACCTTC AAGCTGAACG
BIQUNI		TAGTTCCGGA TATCGACCCA CCCACCGCTA GCAGGCCGTC TGCAGTAACT GCCATAACTT TTTAACGGGC CATAACCAAA ACTGTGGAAG TTCGACTTGC
D-galactonate	+1	G C E E M G V I D N S R À V D À À V N T V À Q I R E À F G S E I E F
dehvdratase	4001	GCTGTGAAGA GATGGGCGTG ATTGATAACT CCCGTGCGGT GGATGCGGCG GTCAATACCG TGGCGCAAAT CCGCGAAGCT TTCGGCAGTG AAATTGAGTT
		CGACACTTCT CTACCCGCAC TAACTATTGA GGGCACGCCA CCTACGCCGC CAGTTATGGC ACCGCGTTTA GGCGCTTCGA AAGCCGTCAC TTTAACTCAA
	+1	FGLDFHGRVSÀPMÀ KVLIKELEPY RPLFIE EPVL
	4101	тбобстебле ттеслеботе бебттлоебе бесблтобеб алботбетба тталаблает болассетат ебесебетот ттаттолаба бесботбетб 👘
_		ACCCGAGCTG AAGGTGCCAG CGCAATCGCG CGGCTACCGC TTCCACGACT AATTTCTTGA CCTTGGGATA GCGGGCGACA AATAACTTCT CGGCCACGAC
	+1	A E Q A E Y Y P R L A A Q T H I P I A A G E R M F S R F E F K R V L
	4201	geggaaleage eggaaltatta teegegeete geagegeaal egeatattee gattgeegea geegaalegta tettetegee ttttgaalttt aaalegegtge
_		CGCCTTGTCC GCCTTATAAT AGGCGCGGAC CGTCGCGTTT GCGTATAAGG CTAACGGCGT CCGCTTGCAT ACAAGAGCGC AAAACTTAAA TTTGCGCACG
	+1	L D À G G L À I L Q P D L S H À G G I T E C Y K I À G M À E À Y D V
	4301	тобаловодо своеттовод аттотасабо обратттато соловодоводо вобаттарода алтостатала алтовосовала сататоятот 👘
_		ACCTGCGCCC GCCCAACCGC TAAGATGTCG GCCTAAATAG GGTGCGCCCG CCGTAATGGC TTACGATATT TTAGCGGCCT TACCGCCTTC GTATACTACA
	+1	VÀLÀ PHCPLGPIÀLÀ CLHIDFVS RNÀ VFQEQSM
	4401	GGGGCTGGGG CCGCATTGCC CGCTGGGTCC AATCGCCCTG GCTGCCTGCC TGCATATCGA TTTTGTTTCG CGCAACGCGG TATTCCAGGA GCAGAGCATG
_		CCGCGACCGC GGCGTAACGG GCGACCCAGG TTAGCGGGAC CGACGGACGG ACGTATAGCT AAAACAAAGC GCGTTGCGCC ATTAGGTCCT CGTCTCGTAC
	+1	GIHYNKGAEL LDFVKNKEDFSMDGGFFKKLTKPG
	4501	GGCATTQACT ATAACAAGGG CGCGGAGCTG CTGACTTTG TGAAAAACAA AGAAGACTTC AGCATGGACG GCGCCTTCTT TAAACCCCTTA ACCAAACCGG
-		CUGIARGIAR IATIGUTUCE GEOCUTIGAE GAGETGARAE ACTITUCIA TOTICIDAE IEGRACUTO CUCUGARGAR ATTIGUGARTI IGGITUGUE
	+1	
	4601	GIGINGEGA TERESTANCE GREGELAGGE IGANIGARETA INCLARAGE GEGEGERTI GEGENALETE GININGEGE EREGENERE
-		LAR W *
	4701	CALETERTAL TEACCHERT STARSTEN, CHARGETER, COTCERSED, ACCOUNTER, ANDERSAN, CACHERTER, STARTARA, SECONDERT,
	1101	
		GCTCACCACT AGCGGTGCGA CATCCGAGTT GTTTGCAGCG GGAGGCCCGT TGGGTTAAAT TTATATTTTT GTGTGGGAGA CATTAAATGT CCCGTACCAC
_	+1	CTCACCACT AGCGGTGCGA CATCCGAGTT GTTTGCAGCG GGAGGCCCCGT TGGGTTAAAT TTATATTTTT GTGTGGGAGA CATTAAATGT CCCGTACCAC
-	+1 4801	GCTCACCACT ACCEPTEGEAGA CATCCGAGTT GTTTGCACCG GGGGCCCCG TGGGTAAAT TTATATTTT GTGTGGGAGA CATTAAATGT CCCGTACCAC M D I S V T A A Q P G R R R Y L T L V M I ACCEGCCCG CTATGCCCAG AATCTGGAGA CAGATGACGA TGGATATTTC ACTTACAGCA GGACACCGG GCGTGCCGC CTATGTGACG CTGGTGATGA
-	+1 4801	CONCECTOR ACCEPTION ACCEPTION AND A CONCERNMENT AND A CONCERNMENTAL AND A CO
-	+1 4801 +1	GCTCACCACT ACCGGIGCGA CATCCGAGIT GTTTCCACGG GGAGGCCCGI TGGTTAAAT TTATATTTT GTGGGAGA CATTAAATGT CCCGTACCAC M D I S V T A A P G R R Y T L V MI AGCGGCCCCG CTATGCCCAG AATCTGGAGA CAGATGACGA TGGATATTTC AGTALAGCA GGACACGCCG GGACGCCGG CTATCTGACG CTACGTGTGAGA TCGCCGGAGC GATACGGGTC TTAGACTTCT GTCATGTCGT ACTATATAG TCAATGTCGC CCGGTCGCCG GATAGACTGC GACACACTACT I F I V V I C Y D R N L V A S M H Q K E G I T K A M M G K <t< th=""></t<>
_	+1 4801 +1 4901	GCTCACCACT ACCGGIGCGA CATCCGACTIT GTTTCCCACG GGAGGCCCCT TIGGTTAAAT TIATATITIT GTTGGGAGA CATTAAATGT CCCGTACGCCAC M D I S V T A A P G R R Y L T L V MI AGCGGCCCCG GTATGCCCAG AATCTGGAGA CAGATGACGA TGGATATTTC AGTACAGCA GGACACGCCGG GGACACGCGG CTATCTGACG CTATCTGACG CTATCTGACG CTATCTGACG CTATCTGACG CTATCTGACG CTATCTGACGC CGCGCAGCGGC GATAGACGGC CGCGCAGCGGC GATAGACGGC CGCGCAGCGGC GATAGACGGC CGCGCAGCGGC GATAGACGGC CGCGCACCGGC GATAGACGGC CGCGCACCGCGC GATAGACGGC GATAGACGGC GACACCACTACT TCTTTATTAC V V I C Y V A N A N A S M H Q K E F G I T K A M GATAGCGGC ATTACAAAG CGATAGCGGC CGCACACTACT T T K A M M </th
-	+1 4801 +1 4901	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$
-	+1 4801 +1 4901 +1	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$
-	+1 4801 +1 4901 +1 5001	$\begin{array}{c} \hline \textbf{GCTCATCCACT} \textbf{AGCGTGCGA} \textbf{CATCCGAGTT} \textbf{GTTTGCACGC} \textbf{GGAGGCCCGT} \textbf{TGGTTAAAT TTATATTTTT} \textbf{GTTGGGAGA} \textbf{CATTAAATGT} \textbf{CCTTATGCCCAC} \\ \hline \textbf{AGCGGCCTCG} \textbf{CTATGCCCAG} \textbf{AATCTGGAGA} \textbf{CAGATGACGA} \textbf{TGGTTATTC} \textbf{ATTACGGCG} \textbf{A} \textbf{Q} \textbf{P} \textbf{G} \textbf{R} \textbf{R} \textbf{Y} \textbf{L} \textbf{T} \textbf{L} \textbf{V} \textbf{M} \textbf{I} \\ \hline \textbf{TGCTCGCGGAGC} \textbf{GATACGGGTC} \textbf{TTAGACTCT} \textbf{GTTACTGCAC} \textbf{GATACGGCC} \textbf{GGATATTTC} \textbf{ACTTACAGCA} \textbf{GGACAGCCGG} \textbf{GGCGTCGCCG} \textbf{CTATTGCACGCG} \textbf{CTGGTTATA} \\ \hline \textbf{TGCTCGCGGAGC} \textbf{GATACGGGTC} \textbf{TAGACTCCT} \textbf{GTTACTGCT} \textbf{ACCTATTAAAGC} \textbf{CCATGTCGTCG} \textbf{CGTGTGGGCC} \textbf{CGGCGGCGCG} \textbf{GATAGACTGC} \textbf{GACCACTACT} \\ \hline \textbf{I} \textbf{F} \textbf{I} \textbf{V} \textbf{V} \textbf{I} \textbf{C} \textbf{Y} \textbf{V} \textbf{D} \textbf{R} \textbf{R} \textbf{R} \textbf{L} \textbf{V} \textbf{R} \textbf{R} $
- - dgoT	+1 4801 +1 4901 +1 5001	$\begin{array}{c c c c c c c c c c c c c c c c c c c $
<i>dgoT</i> 430 aa-1293 bp	+1 4801 +1 4901 +1 5001 +1	$\begin{array}{c} \hline \textbf{GCTCATCCACT} \textbf{AGCGTGCGA} \textbf{CATCCGAGTT} \textbf{GTTTGCAGCG} \textbf{GGAGGCCCGT} \textbf{TGGTTAAAT TTATATTTTT} \textbf{GTTGGGAGA} \textbf{CATTAAATGT} \textbf{CCCGTACCAC} \\ \hline \textbf{M} \textbf{D} \textbf{I} \textbf{S} \textbf{V} \textbf{T} \textbf{\lambda} \textbf{\lambda} \textbf{Q} \textbf{F} \textbf{G} \textbf{R} \textbf{R} \textbf{R} \textbf{Y} \textbf{L} \textbf{T} \textbf{L} \textbf{V} \textbf{M} \textbf{I} \\ \hline \textbf{AGCGGCCTCG} \textbf{CTATGGCCCAG} \textbf{AATCTGGAGA} \textbf{CAGATGACGA} \textbf{TGGTTATTC} \textbf{AGTTACAGCA} \textbf{GGAAGCCCGG} \textbf{GGCTGCCGC} \textbf{CTATGTGCAGCG} \textbf{CTGGTGGAGA} \textbf{CTGGTGAGAGA} \\ \hline \textbf{TCGCCGGGAGC} \textbf{GATACGGCTC} \textbf{TTAGACTCCT} \textbf{GCTATTTC} \textbf{AGTTACAGCA} \textbf{GGAAGCCGG} \textbf{GGATGACGGG} \textbf{GGATGACGGGG} \textbf{GGATGACGGGGG} \textbf{GGATGACGGGGGGGGG } \textbf{GGATGACGGGGG} \textbf{GGATGACGGGGC} \textbf{GGATGACGGGGGGGGGGT } \textbf{GGATGACGGGGGGGGT} \textbf{GGATGACGGGGGGGGA \\ \textbf{GGATGACGGGGGAATGGGG \\ \textbf{GGATGACGGGGAAATGGG \\ \textbf{GGACGATATAAGGGGGGGGG & TAGGGGGGGGGT \textbf{GGGACGGGGAC } TATAGGTTT T TCTAAGGGG AATGGTTTG GGCTTAAGGGGA \\ GAAATAATG GGACGATTAA AGGATGGGGG & TAGGGCGGGT \textbf{GGGGGGGGGTGT T TCTCGAGGGT ATTGGTTCG \\ \textbf{GGGGGGAAAGG \\ \textbf{GGGGGAAAAG \\ \textbf{GGGGGAAAAG \\ \textbf{GGGGGGAAAAG \\ \textbf{GGGCGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGGCGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGGCGGAAAAC \\ \textbf{GGGCGGAAAAC \\ \textbf{GGGCGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGGGGGAAAAC \\ \textbf{GGCGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf{GGCGGGGAAAAC \\ \textbf{GGCGGGAAAAC \\ \textbf{GGCGGGGAAAAC \\ \textbf{GGCGGGGAAAAC \\ \textbf{GGCGGGAAAAC \\ \textbf{GGGGGGGAAAAC \\ \textbf$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132	+1 4801 +1 4901 +1 5001 +1 5101	$ \begin{array}{c} \hline \textbf{GCTCACCACT} \textbf{AGCGGGCGGA} \textbf{CATCCGAGTT} \textbf{GTTTGCACGC} \textbf{GGAGGCCCGT} \textbf{TGGTTAAAT TTATATTTT} \textbf{GTTGGGAGA} \textbf{CATTAAATGT} \textbf{CCCGTACCAC} \\ \hline \textbf{M} \textbf{D} \textbf{I} \textbf{S} \textbf{V} \textbf{T} \textbf{\lambda} \textbf{\lambda} \textbf{Q} \textbf{F} \textbf{G} \textbf{R} \textbf{R} \textbf{R} \textbf{Y} \textbf{L} \textbf{T} \textbf{L} \textbf{V} \textbf{M} \\ \hline \textbf{AGCGGCCTCG} \textbf{CTATGCCCAG} \textbf{AATCTGGAGA} \textbf{CAGATGACGA} \textbf{TGGATATTTC} \textbf{AGTACAGCA} \textbf{GCACACCGCG} \textbf{GGCGTCGCCG} \textbf{CTTGTGGGCG} \textbf{CTGTTGGGCG} \textbf{CCTGTGGGCG} \textbf{CCGCGACGGC} \textbf{GATGACTGC} \textbf{CTGGTGAGAGA} \textbf{TGGATATTTC} \textbf{AGTACAGCG} \textbf{CCGGCGGCGC} \textbf{CGCCGACGGC} \textbf{CATATGCTGC} \textbf{CTGTGGGGCC} \textbf{CCGCGACGGC} \textbf{GATGACTGC} \textbf{CTGGTGAGAGA} \textbf{TGGATATTTC} \textbf{AGTACAGCGC} \textbf{CCGGCGCGCC} \textbf{CGCCGACGGC} \textbf{GATGACTGC} \textbf{CTGTGTGGCGC} \textbf{CCGCGACGGC} \textbf{GATGACTGC} \textbf{GACCACTACT} \\ \hline \textbf{I} \textbf{F} \textbf{I} \textbf{V} \textbf{V} \textbf{I} \textbf{C} \textbf{V} \textbf{V} \textbf{L} \textbf{V} \textbf{L} \textbf{V} \textbf{V} \textbf{A} \textbf{M} \textbf{H} \textbf{I} \textbf{Q} \textbf{K} \textbf{E} \textbf{F} \textbf{G} \textbf{I} \textbf{T} \textbf{K} \textbf{A} \textbf{E} \textbf{M} \textbf{G} \\ \textbf{GCAAATGAGGCC} \textbf{CCTGGCGGTAT} \textbf{GCGCCGACGC} \textbf{TATGCAATGCGC} \textbf{CCTGGCGCGT} \textbf{GCACCACCTGC} \textbf{ATTATCCAAAA} \textbf{AGAATTCGGCC} \textbf{ATTGCCATGCC} \textbf{TATAGGTTTT} \textbf{CCGGGTGACGC} \textbf{TATGAGTTTC} \textbf{CCGGGAAATGGG} \\ \textbf{GCAAATGATTGCCG} \textbf{CCGCGCATTATA} \textbf{ACGATTCGACGC} \textbf{CCACGCGGCT} \textbf{CCACGGCGCT} \textbf{ATTGGTTTC} \textbf{CCGGCTTATGCC} \textbf{GGACGCGCAA} \textbf{CCACGCGCGT} \textbf{TTTAAGCCGC} \textbf{TATGGTTTC} \textbf{CCGGCTTATGCC} \\ \textbf{GTATGCTTTC} \textbf{CCGGCTTTTG} \textbf{CCGGCTTATATCCC} \textbf{GGGCGGGGGTT} \textbf{TTGCAATACGG} \textbf{CCCCCGCCAA} \textbf{AGGACTGCGC} \textbf{ATTGGGTCCC} \textbf{GGCTGGCCA} \textbf{ATTGCAATCGCG} \\ \textbf{CAGAATCGGCAAAAC \\ \textbf{GGCCGAAAAC \\ GGGCCGAAAAC \\ GGGCCGAAAAC GGGCCGACGACA TATGCAATACG G GTCTAGGGAC CCCCCGCCCAA & AGAGCTGCCCA & ATTGGGGCCCCAGGCCAA \\ \textbf{CAGAATCGCGC} \textbf{CGCCCACGGC} \textbf{ATGCAATATCGCC} \\ CAGAATCGCCCTG & CCGCGCCCCAA & ACGACGCCCAC & AGAATCGCCTG & CCGCCCCCCACA & AGACGCCCCA & ATTGCAATGCCCC & GCGCCCACAC & ATTGCAATGCCCC & GCGCCCCCCAA & ACGACGCCCACA & ATTGCAATGCCCCCC & GCGCCCCCACA & AGAGCTGCCCC & ATTCCCCGGGCCCCAC$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 E-r 3 903 869	+1 4801 +1 4901 +1 5001 +1 5101	$ \begin{array}{c} \hline \textbf{GCTCACCACT} \textbf{ACCGGTGCGA} \textbf{CATCCGACTT} \textbf{GTTTCCACCG} \textbf{GGAGGCCCGT} \textbf{TGGTTAAAT TTATATTTT} \textbf{GTTGGGAGA} \textbf{CATTAAATGT} \textbf{CCCGTACCAC} \\ \hline \textbf{M} \textbf{D} \textbf{I} \textbf{S} \textbf{V} \textbf{T} \textbf{\lambda} \textbf{\lambda} \textbf{Q} \textbf{F} \textbf{G} \textbf{R} \textbf{R} \textbf{R} \textbf{Y} \textbf{L} \textbf{T} \textbf{L} \textbf{V} \textbf{M} \\ \hline \textbf{TGGCGGGAGC} \textbf{GGATGCCGGA} \textbf{TGGTTAATTTC} \textbf{AGTTACAGCA} \textbf{GGCACCGCGG} \textbf{GGCGTCGCCG} \textbf{GGCTGCGCG} \textbf{GGCGTCGCCG} \textbf{GGCGTCGCCG} \textbf{CTGTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTGTGGCGC} \textbf{CTGTGTGGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGGCC} \textbf{CTGTTGGGGCC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGTGGCGC} \textbf{CTGTTGGGTCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTGTGTGCGC} \textbf{CTGTTGGGCGC} \textbf{CTGTGTGTGCGC} \textbf{CTGTGGTGTT} \textbf{CTGGGTGTT} \textbf{CTGGGTGGTT} \textbf{CTGGGCGCTA} \textbf{CGGCGGGGTTT} \textbf{CTGGGCGCT} \textbf{CGGCGGGCGGTGGTT} \textbf{TTTAACCGCG} \textbf{CGCTTACGC} \textbf{CGGCGCGGCT} \textbf{CGGCGCGGCT} \textbf{CGGCGCGGCCCAA} \textbf{CGGCGCGGCCCAA} \textbf{CGGCGCGCGCCAA} \textbf{CGGCGCGGCCCAA} \textbf{CGGCGCGGCCCAA} \textbf{CGGCGCGGCCCAA} \textbf{CGGCGCGGCCCAA} \textbf{CGGCGCGGCCCAA} \textbf{CGGCGCGGCCCAA} \textbf{CGGCGCGCGCCCAA} \textbf{CGGGCGGCGCCCAA} \textbf{CGGCGCGGCCCCAA} \textbf{CGGCGCGCGCCCAA} \textbf{CGGCGCGCCCAA} \textbf{CGGCGCGCGCCCAA} \textbf{CGGCGCGCGCCCAA} \textbf{CGGCGCGCGCCCAA} \textbf{CGGCGCGCGCCCAA} \textbf{CGGCGCGCGCCCAA} \textbf{CGGCGCGCGCCCAA} \textbf{CGGCGCGCCCCAA} \textbf{CGGCGCGCGCCCCAA} CGGCGCGCGCCCCAA \\ CGCGCGCGCGCCCCAGGCCCAGGCCCCAGGCCCCCAACGCGGCCCCCAACGGCGCGCCCCCAACGCGGCCCCCC$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869	+1 4801 +1 4901 +1 5001 +1 5101 +1	$ \begin{array}{c} \begin{array}{c} \begin{array}{c} \mbox{GCTCACCT} & \mbox{ACCGGAGAGA} & \mbox{CATTACAGAT} & \mbox{GCTCACCT} & \mbox{ACCGGAGAGAC} & \mbox{CATTACAGAGA} & \mbox{CACACTCC} & \mbox{CATTACAGAGA} & \mbox{CACACTCC} & \mbox{CATTACAGAGA} & \mbox{CACATACTCCC} & \mbox{CATTACAGAGA} & \mbox{CACATACACAAA} & \mbox{CATTACAGAAG} & \mbox{CACATACACAAA} & \mbox{CACATACACACAC} & CACACACACACACACACACACACACACACACACACACA$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201	$ \begin{array}{c} \hline \textbf{GCTCACCACT} \textbf{ACCGGTGCGA} \textbf{CATCCGACTT} \textbf{GTTTCCACCG} \textbf{GGAGGCCCGT} \textbf{TGGTTAAAT TTATATTTTT} \textbf{GTTGGGAGAC CATTAAATGT CCCGTGACCAC} \\ \hline \textbf{M} \textbf{D} \textbf{I} \textbf{S} \textbf{V} \textbf{T} \textbf{A} \textbf{A} \textbf{Q} \textbf{F} \textbf{G} \textbf{R} \textbf{R} \textbf{R} \textbf{V} \textbf{L} \textbf{T} \textbf{L} \textbf{V} \textbf{M} \\ \hline \textbf{AGCGGCCTCG} \textbf{CTATGCCCAG} \textbf{AATCTGGAGA} \textbf{CAGATGACGA} \textbf{TGGTTATTTC} \textbf{ACTTACAGCA} \textbf{CGCACACCGC} \textbf{CGCCACCGCCG} \textbf{CTATTCGACGC} \textbf{CTGGTTATAC} \\ \hline \textbf{TGCTCGCGGAGC} \textbf{GATACGGGTC} \textbf{TTAGACTCTC} \textbf{GTCTACTGCT} \textbf{ACCTATAAAG} \textbf{TCAATGTCGT} \textbf{CGCGACGCCGC} \textbf{CGCCACGCGC} \textbf{CTATCTGACGC} \textbf{CGCACACTCC} \\ \hline \textbf{I} \textbf{F} \textbf{I} \textbf{V} \textbf{V} \textbf{I} \textbf{C} \textbf{V} \textbf{V} \textbf{L} \textbf{V} \textbf{A} \textbf{N} \textbf{H} \textbf{I} \textbf{Q} \textbf{K} \textbf{E} \textbf{F} \textbf{G} \textbf{I} \textbf{T} \textbf{K} \textbf{A} \textbf{E} \textbf{M} \textbf{G} \\ \hline \textbf{TCTTTATTAC} \textbf{CCTGGTGATT} \textbf{TGCTACTGCT} \textbf{ACCTATAAAG} \textbf{CCTATGCCGT} \textbf{CGTGTCGGCC} \textbf{CGCGACGCGCC} \textbf{CGCGAACTCGC} \\ \hline \textbf{AGAAATAATG GCACCACTAAA ACGATGCAGC TAGGCGCCCAA CCTTGCCGTG \textbf{GGTTCCATGC} \textbf{ATATTCAAAAA} ACGATTCGGC \textbf{ATATCCAAAG} \textbf{CGGAAATGGGG} \\ \hline \textbf{AGAAATAATG GCACCACTAAA ACGATGCAGC TAGGCGCCCAA CCTTGCCGTG \textbf{CTAAGGGTACG } TATAAGTTTT TCTTAAGCGC ATATGGTTTC GCGCTTTACCC \\ \hline \textbf{GY V F} \textbf{S} \textbf{A} \textbf{F} \textbf{A} \textbf{W} \textbf{L} \textbf{Y} \textbf{T} \textbf{L} \textbf{C} \textbf{Q} \textbf{K} \textbf{K} \textbf{L} \textbf{T} \textbf{Y} \textbf{F} \textbf{A} \\ \hline \textbf{AGAACGGGCAAAAC GGACGGAGAT ATGCTTATGC \\ \textbf{GAGCGGAAAAC } & GGACGGAAAA GGACCGAGAT ATGCTAATGG \\ \hline \textbf{GCTTTGCGGCCA A AGAGCTGGCA \\ \textbf{AGAGCTGGCA } AGCGCGAAAAC \\ \textbf{GGCCGAAAAAC \\ GGACCGAGATT & TTCAAATGCG \\ \textbf{GAGCGGACCAG TTAGCGGCCAAAAC \\ GAGCGGAAAAC \\ GGCCGAAAAAC \\ GGCCGCAAAAC \\ GGCCCCAAAAC \\ GGCCGCAAAAC \\ GGCCCAAAAC \\ GGCCGCAAAAC \\ GGCCGCAAAAC \\ GGCCCAAAAAC \\ GGCCCCAAA \\ AGAGCTGGCCA \\ AAAAAAAACGGG \\ CCGCCAAAAAC \\ GGCCCCAAAAC \\ GGCCGCAAAAC \\ GGCCGCAAAAC \\ GGCCCCAAAC \\ GGCCCAAACCCCG \\ GGCCCAAAAAC \\ GGCCCAAAAAC \\ GGCC$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201	$ \begin{array}{c} \hline \textbf{GCTCACCACT} \textbf{ACCGTGCGA} \textbf{CATCCGACTT} \textbf{GTTTCCACCG} \textbf{GGAGGCCCGT} \textbf{TGGTTAAAT TTATATTTT } \textbf{GTTGGGAGA} \textbf{CATTAAATGT} \textbf{CCCGTACCAC} \\ \hline \textbf{M} \textbf{D} \textbf{I} \textbf{S} \textbf{V} \textbf{T} \textbf{A} \textbf{A} \textbf{Q} \textbf{F} \textbf{G} \textbf{R} \textbf{R} \textbf{R} \textbf{V} \textbf{L} \textbf{T} \textbf{L} \textbf{V} \textbf{M} \\ \hline \textbf{AGCGGCCTCG} \textbf{CTATGGCCAG} \textbf{AATCTGGAGA} \textbf{CAGATGACGA} \textbf{TGGTTATTC} \textbf{ACTTACAGCA} \textbf{CGCACACCGC} \textbf{CGCCACCGCG} \textbf{GGCGTCGCCG} \textbf{CTATTGGCGCAG} \textbf{CGCATTACTGC} \\ \hline \textbf{TGCTCGCGGAGC} \textbf{TGGTTATTC} \textbf{ACTTACAGCA} \textbf{CGATGTCGCC} \textbf{CGTGCGGCC} \textbf{CGCCACGCCG} \textbf{CGCCACCACCTACT} \\ \hline \textbf{I} \textbf{F} \textbf{I} \textbf{V} \textbf{V} \textbf{I} \textbf{C} \textbf{V} \textbf{V} \textbf{A} \textbf{N} \textbf{L} \textbf{A} \textbf{V} \textbf{A} \textbf{S} \textbf{M} \textbf{H} \textbf{I} \textbf{Q} \textbf{K} \textbf{E} \textbf{F} \textbf{G} \textbf{I} \textbf{T} \textbf{K} \textbf{A} \textbf{E} \textbf{M} \textbf{G} \\ \hline \textbf{TCTTTATTAC} \textbf{CCTGGTGATT} \textbf{TGCTACTGCC} \textbf{ACCTATTAAAGC} \textbf{CCTTACTGCC} \textbf{CGTGTCGGCC} \textbf{CGCGACGCGCC} \textbf{CGCGACGCGCC} \\ \hline \textbf{AGAAATAATG GCACCACTAA ACGATGCACG TAGGCGCCCA CCTTGCCGTG \textbf{GGTTCCATGCC} \\ \hline \textbf{GCTATGTCTTC} \textbf{CCGGTTATA} \textbf{ACGATGCACG} \textbf{TGCGTCATGC} \textbf{CGAGGGCACC} \textbf{CGAAGGGCCC} \textbf{TAAGGTTTT} \textbf{TCTTAAGGCC} \textbf{AATGGTTTC} \textbf{GCGTACACTAT} \\ \hline \textbf{CTATGTCTTC} \textbf{TCCGCTTTTG} \textbf{CCTGGCTTA TACGTTATGC} \textbf{CAGATCCCTG} \textbf{GGGGCTGGTT} \textbf{TCTCGACGGC} \textbf{AATGGTTTC} \textbf{GCTTACCCC} \\ \hline \textbf{GATACAGAAG} AGGCGAAAAC \\ \textbf{GGACGGAAAAC \\ \textbf{GGACGGAAAAC \\ \textbf{GGACGGAAAAC \\ GGACCGGAAA ACGGCGAAAAC \\ \textbf{GGACCGGAAAAC \\ GGACCGGAAAAC \\ \textbf{GGCCGAAAAC \\ GGACCGGAAAAC \\ GGACCGGAAAAC \\ GGACCGGGAAAAC \\ GGACCGGAAAAC \\ GGACCGGAAAAC \\ GGACCGGAAAAC \\ CCCGCCAAAAC \\ CCCGTTCCGGCC \\ CAGCGTAAAGCCG \\ CTCCCAAAAC \\ CTTCCGGCCTAAGCG \\ CTCCCAAAAG \\ CTTCCGGGCTAAGCG \\ TTATGGCCC \\ TACCAGGAA \\ TTTCCGGCC \\ TTATGGCCC \\ TTACAACCGA \\ TTTCTCGGCC \\ TTATGCCCC \\ TTACAGCGGACAAGCCG \\ TTTATGCCC \\ TTCCGGGACTACTA \\ GCTGCCAAAGC \\ CTCCCAAAGC \\ TTTCCGGCC \\ TTAAGACCGA \\ TTTCCGGCC \\ TTACAACCGG \\ TTTCCGGCC \\ TTACAACCGG \\ TTTCCGGCC \\ TTACAACGCGA \\ TTTCCGCGCAAAGCC \\ TTCCGCGAAAGCCG \\ TTTCCGGCCAAAGC \\ TTCCGCGAAAGCCG \\ TTTCCGGCCAAAGC \\ TTCC$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201 +1 5201	$ \begin{array}{c} \hline \textbf{GCTCATCT} \textbf{ACCGTGCGA} \textbf{CATCCGACTT} \textbf{GTTTCCACCG} \textbf{GGAGGCCCGT} \textbf{TGGTTAAAT} \textbf{TTATATTTT} \textbf{GTTGGGAGA} \textbf{CATTAAATGT} \textbf{CCTTATGCCCAC} \\ \hline \textbf{M} \textbf{D} \textbf{I} \textbf{S} \textbf{V} \textbf{T} \textbf{A} \textbf{A} \textbf{Q} \textbf{F} \textbf{G} \textbf{R} \textbf{R} \textbf{R} \textbf{V} \textbf{L} \textbf{T} \textbf{L} \textbf{V} \textbf{M} \\ \hline \textbf{AGCGGCCTCG} \textbf{CTATGCCCAG} \textbf{AATCTGGAGA} \textbf{CAGATTATTC} \textbf{AGTTACTGCT} \textbf{GCALAGCCGG} \textbf{GGCATGCGCG} \textbf{CTATTCGACG} \textbf{CTGGTTAATG} \\ \hline \textbf{TCGCCGGAGCG} \textbf{CATATGGCGCC} \textbf{TAGACTCTC} \textbf{TTATATTTC} \textbf{ACTTATAAGCA} \textbf{CCATATGACGA} \textbf{CGATGCTGCC} \textbf{CGCTCCATGC} \textbf{CCTATCGGCGC} \textbf{CTGGTGATGAC} \\ \hline \textbf{TGGTTATTAC} \textbf{CCTGCGTGATT} \textbf{TCCTACTGCT} \textbf{ACCTATAAAG} \textbf{TCAATGTCGT} \textbf{CTTATGCCC} \textbf{CTTATCGACG} \textbf{CGGATATGGG} \\ \hline \textbf{AGAAATAATG} \textbf{GCACCACTAA} \textbf{ACGATGCAGC} \textbf{TGGGTTAAGA} \textbf{CCTTACTGCT} \textbf{CTTATGCCCAC} \\ \hline \textbf{AGAAATAATG} \textbf{GCACCACTAA} \textbf{ACGATGCAGC} \textbf{TAGGCGGCGCA} \textbf{CCTTGCCGTG} \textbf{GCTTCCATGC} \textbf{ATATTCAAAAA} \textbf{AGAATTCGGC} \textbf{ATATGCAAAG} \textbf{CGGAAATTGGG} \\ \hline \textbf{AGAAATAATG} \textbf{GCACCACTAA} \textbf{ACGATGCAGC} \textbf{TAGGCGGCGAC} \textbf{CGAAGGCGCAC} \textbf{CGAAGGCTGC} \textbf{TATAAGTTTT TCTTAAGCGC} \textbf{ATATGCAAAG} \textbf{CGCGAAACCGGCAC} \\ \hline \textbf{AGAAATAATG} \textbf{GCACCACTAA} \textbf{ACGATGCAGC} \textbf{TAGGGTTATGC} \textbf{CTTGGGTAGAC} \textbf{ATAGGTTTC} \textbf{CCTGGTGAATC} \\ \hline \textbf{GATACAGAAA} \textbf{AGGGGGAAAAC} \textbf{GGACGCAGAGA} \textbf{ATGCGATGCGGGCA} \textbf{CGAGGGGTACG} \textbf{TATAAGTTTT} \textbf{TTTAAGGCCC} \\ \hline \textbf{AATGGTCCCC} \textbf{GGCCGAAAAC} \textbf{GGCCGAAAAC} \\ \hline \textbf{GGCGGAAAAC} \textbf{GGCCGAAAAC} \textbf{GGCCGAAAAC} \textbf{GGCGGCAAAC} \textbf{CCTTCGGGCCAC} \\ \hline \textbf{ATGGGCCGGCA} \textbf{AAGCCCAGGGC} \textbf{AAATAAGGGGC} \\ \hline \textbf{AAAATAGCGG} \textbf{AAAAAAACGGG} \textbf{AAGCCGGAAAC} \textbf{AGCGGGGAAC} \textbf{AGGCGGACCG} \textbf{TAAGGCTTGC} \textbf{GGGCTAAACG} \\ \hline \textbf{CCTTTCCGGCC} \textbf{AAAAAACCGG} \textbf{ATGGTCAAAG} \textbf{GCGCGCAAAC} \textbf{AGCGGGAACGCACG} \textbf{AAAAAAACGGG} \\ \hline \textbf{ATGCTAGCAGCGG } \textbf{AAAAAAACGGG} \textbf{ATGGTCAAAG} \textbf{GCGCGCAAGA} AGGGGGCGCCCGCCGAAAAAAAAAAACGGG A \ TAGGCCGGAAA A AAAAAAACGGG A \ TAGGCCGGAAA A AAAAAACGGG A \ TAGGCCGGAAA A & GCGGGAAAGGCCG CCGGCAAAAA & CGGCGGAAAAC & G$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201 +1 5301	$ \begin{array}{c} \hline \textbf{GCTCATCT} \textbf{ACCGTGCGA} & \textbf{CATCCGACTT} & \textbf{GTTTCCACCG} & \textbf{GGAGGCCCGT} & \textbf{TGGTTAAAT} & \textbf{TTATATTTT} & \textbf{GTGTGGAGA} & \textbf{CATTAAATGT} & \textbf{CCCGTACCAC} \\ \hline \textbf{M} & \textbf{D} & \textbf{I} & \textbf{S} & \textbf{V} & \textbf{A} & \textbf{A} & \textbf{Q} & \textbf{F} & \textbf{G} & \textbf{R} & \textbf{R} & \textbf{Y} & \textbf{L} & \textbf{T} & \textbf{L} & \textbf{V} & \textbf{M} \\ \hline \textbf{AGCGGCCTCG} & \textbf{CTATGGCCAG} & \textbf{AATCTGGAGA} & \textbf{CAGATGACGA} & \textbf{TGGTTAATGC} & \textbf{GCALAGCCGG} & \textbf{GCGTGCGCG} & \textbf{CTTATGACGC} & \textbf{CL} & \textbf{V} & \textbf{M} \\ \hline \textbf{T} & \textbf{C} & \textbf{Y} & \textbf{V} & \textbf{L} & \textbf{X} & \textbf{Q} & \textbf{F} & \textbf{G} & \textbf{R} & \textbf{R} & \textbf{Y} & \textbf{L} & \textbf{T} & \textbf{L} & \textbf{V} & \textbf{M} \\ \hline \textbf{T} & \textbf{C} & \textbf{Y} & \textbf{V} & \textbf{L} & \textbf{C} & \textbf{Y} & \textbf{V} & \textbf{L} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{H} & \textbf{I} & \textbf{Q} & \textbf{K} & \textbf{E} & \textbf{F} & \textbf{G} \\ \hline \textbf{T} & \textbf{T} & \textbf{V} & \textbf{V} & \textbf{L} & \textbf{C} & \textbf{Y} & \textbf{V} & \textbf{L} & \textbf{X} & \textbf{X} & \textbf{X} & \textbf{M} & \textbf{H} & \textbf{I} & \textbf{Q} & \textbf{K} & \textbf{E} & \textbf{F} & \textbf{G} & \textbf{I} & \textbf{T} & \textbf{K} & K$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5201 +1 5301	GCTCACCACT ACCGGTGCGA CATCCGACTT GTTTGCACCG GGAGGCCCGT TGGTTAAAT TTATATTTT GTTGGGAGA CATTAAATGT CCCGTACCACC AGCGGGCCTG CTATGGCCAG AATCTGGAGA CAGATGACGA TGGATATTTC AGTACGGGC GCACAGCCGG GGAGGCCGG CTATGGCGG CTATGGCGGC CTATGGCGCG CTATGGCGGC CTATGGCGGC CTATGGCGGC CTATGGCGGC CTATGGCGGC CACACACCCG GGAGGCCGGC CACAGTGGGC CTATGGCGCG CTATGGCGCG CTATGGCACGG CTATGGCGCG CTATGGCGCG CTATGGCGCG CTATGGCGCG CACACTCA CCTATGGCGGC CACACTCA CCTATGGCGCG CACACTCAC CCTATGGCGCG CACACTCAC CCTATGGCGCG CACACTCAC CCTATGGCGC CACACTCAC CCTATGGCGCA CCTATGGCCAC CACACTCAC CACATCAC A S M H L Q K F G R R L N T K A E M G A S A CCTATGCGCC
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5101 +1 5301 +1 5401	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5201 +1 5301 +1 5401	GCTCACCACT ACCGGTGCGA CATCCGACTT GTTTCCACCG GGAGGCCCGT TGGTTAATAT TTATATTTT GTTGGGAGAA CATTACATGT CCCGTACCAC AGCGGCCTCG CTTATGCCCAG AATCTGGAGA CAGATGACGA TGGTTATTTC ACTTACAGCG GGCGCTGCCG CTTATGACCAC CTGGTGATAT TCGCCGGAGC GATACGGGTC TTAGACCTCT GTCTACTGCT ACCATTACAGCA CCGTTACTGCGT CAGATGACGA TGGTTACTGCT CTGTTGTGATGA I F I V V I C Y V R N L A N H Q K E F G I T K A E M G GCACACACCGG GCGCGGCGC CTTTCTGAGC ACCATTCAGCA CCGGGAGCG CGAGGCGCCG CGGGGGCGCGC CGGGGGCGCGC ACGATTGGGC ACCATTCAGCGC ACCATTCAGCGC ACCATTCAGCGC ACCATTCAGCGC ACGATTGGGC ACCATTCAGCGC ACGATTGGGC ACGATTGGCG ATTCCCAAGC ACGATTGGCGCAA CCTTTAGCCCC GCGGCGCGCGCG ATTATCAAAA ACGATTGGCG ATTATCAAAA ACGATTGGCG ATTATCAAAA ACGATGCCAA ACGATGCGGCAA ACGAGATGAG
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 5001 +1 5101 +1 5201 +1 5301 +1 5401	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5201 +1 5301 +1 5401 +1 5501	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201 +1 5401 +1 5501	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201 +1 5401 +1 5501 +1	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 5001 +1 5101 +1 5201 +1 5301 +1 5501 +1 5501 +1 5601	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 5001 +1 5101 +1 5201 +1 5301 +1 5501 +1 5501 +1 5501	$ \begin{array}{c} \text{CTCACCACT} \text{ ACCCGFTCCG} CTATCGCAGTT GTTTGCAGCG GAGGCCCCT TGGGTTAAT TTATATTTTT GTGTGGGGACA CATTAAATGT CCCGFTACCAC CATCGGGCCT ACCCGFTACTAGT CTGCGTATAA TGGTTCCT GTCATGCA CATTGGTACG CATCGGAGCA CATTAAATGT CCCGFTACCAC CATCGGAGCA CATTGGGAGCA CATTAAATGT CCCGFTACCAC CATCGGAGCA CATTGGGAGCA CATTGGGAGCA CATTGGAGCA CATTGGGAGCA CATTGGAGCA CATTGGGAGCA CATTGGGAGCA CATTGGGAGCA CATTGGGAGCA CATTGGGAGCA CATAGGGGAGA CATAGGGGAGA CAGATGGGAGA CAGATGGGAGA CAGATGGGAGA CATAGGGAGA CAGATGGGAGA CAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5201 +1 5201 +1 5301 +1 5501 +1 5601 +1	$ \begin{array}{c} \text{CTTCACCACT } \text{K} \text{GGGFGCGG} & \text{CHTCGAFG} & \text{GCTCCGAFG} & \text{GGGFGCGG} & \text{TGGGGGAG} & \text{CHTCARGG} & \text{CCTTCACCAC} \\ \text{M} D I S & Y T \lambda & \lambda P G R R R Y L T \\ \text{CGGCGGAGC } & \text{CHTCGGAGA} & \text{CAGATGAGGA} & \text{CAGATGAGGA} & \text{CGGTGAGGC} & \text{GCCCACCGGG} & \text{GCGCGTGCGC} & \text{CTTCTCATGAC} & \text{CCTGTGATGC} & \text{CCTGCGGGC} & \text{CGCCGCGGGC} & \text{CTATCGAGA} & \text{CCTGCGGGT} & \text{CCTGCGGC} & \text{CCCCACCGGG} & \text{GCGCGCGGC} & \text{CTATCGAGAA} & \text{CCTGCGGGT} & \text{CCTGCGGC} & \text{CCCCACCGGG} & \text{CCACCGGGC} & \text{CCTGCGGC} & \text{CCCCACCGG} & \text{CCACGGCG} & \text{CCTGCGGC} & \text{CCCCGGCGGC} & \text{CCCCGGCGGC} & \text{CCCCGGCGGC} & \text{CCTCCCGGC} & \text{CCTGCGGC} & \text{CCTCCGGGC} & \text{CCCCGGCGGGC} & \text{CCCCGGGCGGGC} & \text{CCCCGGGCGGGC} & \text{CCCCGGGCGGGC} & \text{CCCCGGGCGGGC} & \text{CCCCGGGCGGGC} & \text{CCCCGGGCGGGGC} & \text{CCCCGGGGCG} & CCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 5001 +1 5101 +1 5201 +1 5301 +1 5501 +1 5501 +1 5701	$ \begin{array}{c} \text{CTTACCACT } \text{ KeGGGFGCG } \text{ CTTCGAFGG } \text{ CTTCGAFGG } \text{ CTTCGAFGG } \text{ CTTTCGAGG } \text{ CTTTCGGAGG } \text{ CTTTCGAGG } \text{ CTTTCGGAG } \text{ CTTTCGGAG } \text{ CTTTCGGAGG } \text{ CTTTCGGAG } \text{ CTTTCGGG } CTTTCGGG$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201 +1 5401 +1 5501 +1 5601 +1 5701	$ \begin{array}{c} CTATCACLET RECERTEGATEGA CATTACATE COCCURTE GARGECCOT TOGOTTALT TITTE TITTETTTTT GUETGGARGA CATTALATE COCCURATE M D I S V T A A Q P G R R R V L T L V M I ACCOGGACC CATAGECCAG CATTACACE COCCURATE TOGATACAGET TATACTEGA CAGATGARGA CAGATATTC AGTATATC COUNT AND CONTRACTOR CATTACATAGE CACCULATE I F I V V I C Y V R R N L V A S M H I Q R F G R R R V I T K A E M G TETTATTAC CONSTRAINT TOCTACOGG ATGEGECCAA CCTTOCOTE GENEROGCE COCCUCACEGEC GATAGACTEG GACACATCT TOTTATTAC CONSTRAINT TOCTACOGG ATGEGECCAA CCTTOCOTE GENEROGCE CAGACAGEGE CATAGACTEG GACACATCT AGAAATATATE CONSTRAINT TOCTACOGG ATGEGECCAA CCTTOCOTE GENEROGCE TATATACTTT TOCTAAGEGE ATTACCAAGE GOAAATGEG AGAAATATATE CONSTRAINT TOCTACOGG ATACCAGECTA AGGEGECT GACACACCC CAGAGGACCA TATATAGATTT TOCTAAGEGE TATAGETTTE CONSTRAINT TOCTAAGEGE TATAGETTATE CAAGAACTAA ACGATGAAC TACCATAA CGATCAACE TAGEGEGETT GACACGACCA CAGAGGACA CATTAGEGEC TAATGETTE COCCUTACECCE AGAAATATAG CACACATAA ACGATGAAC TACCATACE TOCTACAGE CONSTRACT TATATAGTTT TOCTAAGEGE CAACACACE AGAAATATAG GAACGAGAA ATGACATAGE COCCUCACAC CAAGAGGACA CATTAGEGEC TAATGETTE CACGGATTTE CAGAGGACA CATTAGEGEC TAATGETTA TACCAAGE GACCACATAGE CACACAGE AGAAATACCGATAA ACGATGAA ATGACGAGAA TAGEGEGECA ACGATGACCATAA ACGATGAG ATGEGECCAA COTOCCUCACA CATAGEGEC TATAGACTTT TOCTAAGEGE ATAGEGEC TAATGEGECT AAATTCGACC GACACAGAA ATGACGAGAA ATGACATAGE GTCTAGGGAC TATGGCCTCA ATAGEGECA TATACCGAGE CAGGAGGACA TAGEGECAACATCG ATTOTAGAATCGE ACCTCAGE CAGGEGEAT AGGEGETTE CCAGEGGACA TAGEGECAA TAGEGEGACA CAGEGEGAC CAGACAGEGE CAGGAGGAC CAGGCAGACE CAGGAGGAC CAGGCAGGE CAGGCAGGE CCGACAGACE CAGGAGGE CCGACAGACE CAGGGAGE CCGACAGACE CAGGGAGE CCGACAGAA TAGEGEGA CAGGCAGAATCGE CAAGGGAGE CAGGCAGGAC CAGGCAGAA TAGEGEGA CAGGCAGAACE CAGGCAGGE CAGGCAGGE CAGGCAGGE CAGGCAGGE CAAGGCAGA TAGEGCCAAA ATAGEGE TTAGAATACEGE TAGAATACEGE CACGCAGGAA TAGEGEGAATAGE CAAGGCAGA TAGEGCCAAA TAGEGEGA CATTAGAAATACE CCTTTCCGCC GAAAAACCEGE CTGGAATCAA GGAAATCET A GCAGGGACATCAE GACGCCCAAA ATAGCCCAA ATAGCAGAATACE GAAAAACCAE AAAACCAE AAACCCAAA CCCACACTCAE COCCCACACA CATTACCAACACE CAA$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201 +1 5301 +1 5501 +1 5601 +1 5701 +1	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 5001 +1 5101 +1 5201 +1 5301 +1 5501 +1 5501 +1 5501 +1 5501 +1 5501	$ \begin{array}{c} CTRCCACT ACCORTCGA CATCCGAGTT GITTGCACCG GAGGCCCCT [GGGCTALL] INSTITUT GOTGGGGG CATCCGGGGGG GAGGCGG CATTACTOR CCCGGGCCCAC CATTALLITTLE CONTROL CATCCACCG GGGCCG CATTACTACG CATTACCALCG CATCCACCG GGCCGGC CATCCACG CAGGCGGC GALAGACGG GGCCALCCT IT CGCACGGCG TTAGCGCCCA CATTACCALGA ATTACCALGA CATTACCALGA ATTACCALGA CATTACCALGA ATTACCALGA CATTACCALGA ATTACCALGA CATTACCALGA ATTACCALGA CATTACCALGA ATTACCALGA ATTACCALGA$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 5001 +1 5101 +1 5201 +1 5201 +1 5501 +1 5501 +1 5501 +1 5501 +1 5501	$ \begin{array}{c} \text{CCLCCLCL} \\ \textbf{ACCGRCTGC} \\ \textbf{ACCGRCTGC} \\ \textbf{ATCTGGCCAGGGC} \\ \textbf{ACGTAGCGGCC} \\ \textbf{ACGTAGCGGCC} \\ \textbf{ACGTGGCCAG} \\ \textbf{ACGTGGCCAG} \\ \textbf{ACGTGGCCAG} \\ \textbf{ACGTGGCCAGGGCC} \\ \textbf{ACGTGGCCAG} \\ \textbf{ACGTGGCCAGGC} \\ \textbf{ACGTGGCCAGGC} \\ \textbf{ACGTGGCCAGGC} \\ \textbf{ACGTGGCCAGGC} \\ \textbf{ATATATGGCC} \\ \textbf{ACGTGGCCAGGC} \\ \textbf{ACGTGGCCAGGC} \\ \textbf{ATATATGGCCC} \\ \textbf{ATATATGGCCCAGGCCAGGCCGC} \\ \textbf{ATATATGGCCCC} \\ \textbf{ATTATGGCCCC} \\ \textbf{ATTATGGCCCCG} \\ \textbf{ATATATGGCCCCG} \\ \textbf{ATATAGGGCCCAG } \\ \textbf{ATATAGGGCCCAG } \\ \textbf{ATATAGGGCCCAG } \\ \textbf{ACGTGGCCAG } \\ \textbf{ATATAGGGCCCCG } \\ \textbf{ATAGGGCCCAG } \\ \textbf{ATAGGGCCCCG } \\ \textbf{ATAGGGCCCCG } \\ \textbf{ATAGGGCCCAG } \\ \textbf{ATAGGGCCCCCGCGCCCCCC } \\ \textbf{ATGGCCCAGGCCG } \\ \textbf{ATGGCCCAGGCCG } \\ \textbf{ATGGCCCAGGC } \\ \textbf{ATGGCCCAGGC } \\ \textbf{ATGGCCCAGGC } \\ \textbf{ATGGCCCAGGC } \\ \textbf{ATGGCCGAGGC } \\ \textbf{ATGGCCGAGGC } \\ \textbf{ATGGCCGAGGC } \\ \textbf{ATGGCCGAGGC } \\ \textbf{AGGCGCAAAAGGC } \\ \textbf{ATGGCCGAGGC } \\ \textbf{ATGGCCGAGC } \\ \textbf{ATGGCGCGC } \\ \textbf{ATGGCGGAGC } \\ \textbf{ATGGCGAGGC } \\ \textbf{ATGGCGAGGCG } \\ \textbf{ATGGCGAGGCG } \\ \textbf{ATGGCGAGGCG } \\ \textbf{ATGGCGAGGCG } \\ \textbf{ATGGCGAGGC } \\ \textbf{ATGGCGAGGCG } \\ \textbf{ATGGCGAGGC } \\ \textbf{ATGGCGGAGGC } \\ \textbf{ATGGCGGAGGC } \\ \textbf{ATGGCGGCGC } \\ \textbf{ATGGCGGCGC } \\ \textbf{ATGGCGGCGCGC } \\ \textbf{ATGGCGGCGCGC } \\ \textbf{ATGGCGGCGC } \\ \textbf{ATGGCGGCGC } \\ ATGGCGGCGC \\ \textbf{ATGGCGGCGCG$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5201 +1 5201 +1 5301 +1 5501 +1 5501 +1 5501 +1 5501 +1 5501 +1 5501	$ \begin{array}{c} \text{CCCCCACT} \text{ACCORTGEGA} \text{CATCRARTGE} \text{CHARTGEGARA} \text{CATTARTGET} \text{CCCCARTGEGARA} \text{CCCCARTGEGARA} \text{CCCCARTGEGARA} \text{CCCCARTGEGARA} \text{CCCCARTGEGARA} \text{CCCCARTGEGARA} \text{CCCCARTGEGARA} \text{CCCCARTGEGARA} \text{CCCCARTGEARA} \text{CCCCARTGEARA} \text{CCCCARTGEARA} \text{CCCCCARTGEARA} \text{CCCCARTGEARA} \text{CCCCCARTGEARA} \text{CCCCARTGEARA} \text{CCCCCARTGEARA} \text{CCCCARTGEARA} \text{CCCCCARTGEARA} \text{CCCCCCCCCCCCCCCCCCCARA} CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5201 +1 5201 +1 5501 +1 5501 +1 5701 +1 5701 +1 5701 +1 5701	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5201 +1 5201 +1 5401 +1 5501 +1 5701 +1 5701 +1 5701 +1 5701	$ \begin{array}{c} \text{Generative} \\ Gene$
<i>dgoT</i> 430 aa-1293 bp. 4840-6132 Fr; 3,903,869 To; 3,905,161 B5QUN8 D-galactonate transporter	+1 4801 +1 4901 +1 5001 +1 5101 +1 5201 +1 5401 +1 5501 +1 5601 +1 5601 +1 5601 +1 5601 +1 50	$ \begin{array}{c} \text{Generation} \\ Gene$

				R V G	×						
	6101	TGTTAGTGGG	CGATGTGAAA	CGCGTCGGAT	AAGTTGGCCG	TGGCGAGCCG	GAAGCTA TTA	TGGCGAAGGA	TGATCGGCTA	TCCATGCGTC	GAGCGCCGTC
		ACAATCACCC	GCTACACTTT	GCGCAGCCTA	TTCAACCGGC	ACCGCTCGGC	CTTCGAT AAT	ACCECTTCCT	ACTAGCCGAT	AGGTACGCAG	CTCGCGGCAG
	-2						* '	PSP	H D A I	WAD	LATV
	6201	ACCCCCTGCG	TCACGATAGT	TTTAATATCT DDDTTDTDCD	ATATCGCTTA	GCGGTTGCAA	TTCCAGCTCC	CGACAGGCTT	CGGTTGCCTG	CGTCATCCCC	AGGCTGGCGC
	-2	V G O T	V I T	K I D	I D S L	POL	E L E	R C A E	TAO	T M G	LSAC
	6301	AGCCGCTTTT	TAATTTATGC	GCCAGCCGCT	TAATGTTAAC	GTCGTCATTC	ATGGCGCGCG	CCGCCTCGAT	CTCCTCGACT	AAAGGCAGCG	CGCTGTCTTT
		TCGGCGAAAA	ATTAAATACG	CGGTCGGCGA	ATTACAATTG	CAGCAGTAAG	TACCGCGCGC	GGCGGAGCTA	GAGGAGCTGA	TTTCCGTCGC	GCGACAGAAA
	-2	CGSK	_ К Н	A L R F	C I N V	N	MARA	AEI	EEV	L P L A	's'd'k'
	6401	GAATAACGCG	ATCCATTGCC	GGAGTTTTTC	CGGGCCGACA	GACGCCATAT	GCGGCCGGCAAG	GACTACAAGT	TTCAGCATCG AAGTCGTAGC	GEGCECECETT	GTGAGATTTG
torS	-2	F L A	I W O R	LKE	PGV	S A M D	G A L	O H E	N L M A	R A N	H S K G
6158-8893	6501	CCCTGTAAAT	AGTGAGCTAT	CATGCGGTAA	AGCTCCTCGC	GGGGAACGGG	CTTTTGAATA	ATCCCGCAGA	AAAGACCGGC	GGTACGCTGA	CGAAGATTAT
FR; 3,901,108		GGGACATTTA	TCACTCGATA	GTACGCCATT	TCGAGGAGCG	CCCCTTGCCC	GAAAACTTAT	TAGGGCGTCT	TTTCTGGCCG	CCATGCGACT	GCTTCTAATA
To; 3,903,843	-2	G Q L Y	H A I	M R Y	LEER	P V P	K Q I	I G C F	L G A	T R Q	R L N D
BSQUN7	0001	GCAGATAGTG	TACGCGCGCTA	TTTGGTTACG	CAAAGTACCG	CCCTATGACT	GAGTAGTTGA	CGACGCGGGTC	CCAATTCGGC	AGTATTAGGC	CATTTAGTTT
rwo-	-2	DDIV	HAS	FGIR	K M A	P Y Q	S M L Q	QAL	TLG	D Y D P	LDF
sensor protein	6701	ATCCACCAGC	GCCACATCAA	AGGACTCGCC	TTCGGCTAAA	CAGCGTAGCG	CGTCATTCGC	GCTTTCAGCT	ACTGAGACCT	TCACCCCTTT	ACCCGTGAGC
histidine		TAGGTGGTCG	CGGTGTAGTT	TCCTGAGCGG	AAGCCGATTT	GTCGCATCGC	GCAGTAAGCG	CGAAAGTCGA	TGACTCTGGA	AGTGGGGAAA	TGGGCACTCG
	-2	D V L	A V D F	CGTCAGCATA	E A L	C R L A	CCTAATCCC	S E A	V S V K	GAACGCACTT	G T L M
	0001	TAAAGTCGCC	ATTATGCGAC	GCAGTCGTAT	AATAGAAGTT	AGTCGTTGTC	CGCATTAGGC	AATTCTAAAT	AGCCAAAGGA	CTTGCGTGAA	AACGAACGCC
	-2	MEAT	IRQ	TLM	NDEI	LLL	RLG	N L N I	PKR	FAS	K S A P
	6901	GTTTGGGGTG	GCGAACCGGC	AGTTGTAAAC	GAAAACAACT	GCCAACATGA	AGCGTACTGG	TGACGGTTAG	CGTCCCGCCC	ATCGCCTCTG	CCAGGCTGGC
	_2	CAAACCCCCAC	CECTTEECCE	TCAACATTTG	CTTTTGTTGA	CEETTETACT	TCUCATUACC	ACTUCCAATC		TAUCUUAUAC	GETCCGACCE
	7001	GCTGATAGCC	AGCCCCAGGC	CCGTCCCGCC	CCGGCGTCCT	GTCGCCTGCA	CAAACGGTTT	GAAAATAGCC	GTCAGCTTCG	CTTCAGGGAT	ACCGCATCCG
		CGACTATCGG	TCGGGGTCCG	GGCAGGGCGG	GGCCGCAGGA	CAGCGGACGT	GTTTGCCAAA	CTTTTATCGG	CAGTCGAAGC	GAAGTCCCTA	TGGCGTAGGC
	-2	SIA	L G L G	TGG	RRG	T A Q V	FPK	FIA	TLKA	EPI	GCGT
	7101	GTATCTTCAA	CCTCAATAAA GGDGTTDTTT	CCAGCTCTGA	TCGTCGCAAA	ACGTGCGCAG	CACTATECTE	CCGCGATCGG	TAAATTTCGC	CGCATTACTT	AAAAGATTAA TTTTCTDDTT
	-2	T D E V	EIF	W S Q	D D C F	TRL	V I S	G R D T	FKA	A N S	LLNI
	7201	TCACAATTTG	ACGTATCCGG	CGCGGATCGC	CTTGTAGCGT	GGAGGGGAGT	TGCTCGCTAA	AGTCGGCGAT	GAGCGCCACC	TGCACGCGGC	TGTGCATTAA
		AGTGTTAAAC	TGCATAGGCC	GCGCCTAGCG	GAACATCGCA	CCTCCCCTCA	ACGAGCGATT	TCAGCCGCTA	CTCGCGGTGG	ACGTGCGCCG	ACACGTAATT
	-2	I V I Q	R I R	ATTOTOCOCO	TTCAAACGGC	S P L	Q E S F	D A I	L A V	Q V R S	H M L
	7301	TACGTCGCGC	GATAAGTTGT	TAACAGCGCC	AAGTTTGCCG	AGGAGCGATT	ATCTTTGCAA	CCACGGCGGA	TGAAGTTAGC	GGCTCATTAG	TTCTTATAGC
	-2	HLA	S N L L	QRP	EFP	EESI	SVN	TGG	VEIA	SYD	LIDN
	7401	TTAAGAATCG	CCAGTAGCGA	TTCGCCTGAA	TCATTAATGG	CCTGTAAGTC	GTCGCGGTAG	TTTGCCATGA	GCGGTTTATC	GGCCAGCAAT	TGTACCGTAC
	-2	AATTCTTAGC	GGTCATCGCT	AAGCGGACTT	AGTAATTACC	GGACATTCAG	CAUCECCATC	AAACGGTACT	CUCCAAATAU	CCGGTCGTTA	ACATEGCATE
	7501	0030338000	GTAAAGCGGC	GTGCGDDTTT			D IC I		00007777000		Q V 1 G
	7301	COAGAAIGCC	0 Innno COOC	010000001111	CATGGUTCAT	GRCGRCAARE	AACGTCGATT	TCGCCTCGTT	Geccrittec	GCTTCGGCGC	BCBCCTBCCB
	/501	GCTCTTACGG	CATTTCGCCG	CACGCTTAAA	GTACCGAGTA	CCGCCGTTCC	TTGCAGCTAA	AGCGGAGCAA	CCGGAAAAGG	CGAAGCCGCG	CGCGGACGGC
_	-2	G L I G	CATTICGCCG	CACGCTTAAA T R I E	GTACCGAGTA	CCGCCGTTCC	TTGCAGCTAA F T S K	AGCGGAGCAA	CCGGAAAAGG	GCTTCGGCGC CGAAGCCGCG A E A R TECCCCCCTT	CGCCGGACGGC
_	-2 7601	GTGTTCCAGC GACAAGGTCG	CATTTCGCCG Y L P ACCAGCGCAT TGGTCGCGTA	CACGCTTAAA T R I B GAAGCTCCGC CTTCGAGGCG	CGTTTGCGAA GCAAACGCTT	CGCCGCGTTCC A A L CGCACCTGTT GCGTGGACAA	TTGCAGCTAA F T S K CCGCCAAATC GGCGGTTTAG	AGCGGAGCAA A E N CTCACGATGG GAGTGCTACC	CCGGAAAAGG	CGAAGCCGCG A E A R TGCGCACGTT ACGCGTGCAA	GCGCGGACGGC A Q R GGCGCGAAAC CCGCGCTTTG
_	-2 7601 -2	GTGTTCCAGG GTGTTCCAGC CACAAGGTCG H E L	CATTTCGCCG Y L P ACCAGCGCGTA TGGTCGCGTA V L A H	CACGCTTAAA T R I E GAAGCTCCGC CTTCGAGGCG L E A	CATEGOTAT GTACCGAGTA 2 H S M CGTTTGCGAA GCAAACGCTT T Q S	CCCCCCTTCC A A L CCCCCTGTT CCCCCCTGTT GCGTGGACAA R V Q E	F T S K CCGCCAAATC GGCGGTTTAG C A L D	A E N CTCACGATGG GAGTGCTACC E R H	CGGTTCAATT GCCAAGTTAA R N L K	CGAAGCCGCG A E A R TGCGCACGTT ACGCGTGCAA R V N	GCGCCTECCG CGCGGACGGC A Q R GGCGCGAAAC CCGCGCTTG A R F A
_	-2 7601 -2 7701	GL I G GTGTTCCAGC CACAAGGTCG H E L GCTTCCATCA GCTTCCATCA	CATTTCGCCG Y L P ACCAGCGCGAT TGGTCGCGGA V L A H GACGGCTGAT CTGCCGCGCGA	CACGCTTAAA T R I E GAAGCTCCGC CTTCGAGGCG GGTATCCAGT GGTATCCAGT	GTACCGAGTA GTACCGAGTA CGTTTGCGAA GCAAACGCTT T Q S TCGCTGACGC ACCGCTGCGC	CGCCCGTTCC A A L CGCACCTGTT GCGTGGACAA R V Q E CAGCCGCCTCC CCGCCGCCTC	TTGCAGCTAA F T S K CCGCCAAATC GGCGGTTTAG CGGAAACGGT CGGAAACGGT	AGCGGAGCAA A E N CTCACGATGG GAGTGCTACC E R H GAGTCAATAT CTCACGATATA	CCGGAAAAGG A K E CCGGTTCAATT GCCAAGTTAA R N L K CGCCCTCCAG GCGCGEGEGTC	CGAAGCCGCG A E A R TGCGCACGTT ACGCGTGCAA TAGCCGCTGT TAGCCGCTGT TAGCCGCTGT	CGCGGACGGC A Q R GGCGGCGAAAC CCGCGCTTTG A F F A AAGGCCTGCG TTCGCGGCGCGC
_	-2 7601 -2 7701	$\begin{array}{c} \textbf{G} \textbf{CTACGG} \\ \textbf{G} \textbf{CTCTACGG} \\ \textbf{G} \textbf{L} \textbf{I} \textbf{G} \\ \textbf{G} \textbf{G} \textbf{G} \textbf{TTCCAGC} \\ \textbf{CACAAGGTCG} \\ \textbf{H} \textbf{E} \textbf{L} \\ \textbf{H} \textbf{E} \textbf{CCACCA} \\ \textbf{G} \textbf{CGACGGTAGT} \\ \textbf{CGAGGGTAGT} \\ \textbf{A} \textbf{E} \textbf{M} \textbf{L} \end{array}$	CATTCGCCG Y L P ACCAGCGCAT TGGTCGCGTA V L A H GACGGCTGAT CTGCCGGCTA	CACGCTTAAA T R I E GAAGCTCCGC CTTCGAGGCG GTATCCAGT CCATAGGTCA CCATAGGTCA	CATEGECTCAT GTACCGAGTA CGTTGCGAGA GCAAACGCTT T Q S TCGCTGACGC AGCGACTGCG CGCCGACGCG CGCCGACGCG CGCCGACGCG CGCCGACGCG CGCCGACGCG CGCCGACGCG CGCCGACGCG CGCCGACGCG CGCCGACGCG CGCCGACGCG CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGC CGCCGACGCC CGCCGCC CGCCGCC CGCCGCC CGCCGCC CGCCGCC CGCCGCC CGCCGCC CGCCGCC CGCCGCCCC CGCCGCCCC CGCCGCCC CGCCGCCCC CGCCGCCCC CGCCGCCCCC CGCCGCCCCC CGCCGCCCCC CGCCGCCCCC CGCCGCCCCC CGCCGCCCCC CGCCGCCCCC CGCCGCCCCCC CGCCCCCCCC	CCECECAAGE CCECCGTTCC A A L CCCACCTGTT GCGTGGACAA R V Q E CAGCCGCTC GTCGECGGAG GTCGECGGAG	TTGCAGCTAA F T S K CCGCCAAATC GGCGGTTAG CGGAACGGT GCCTTGCCA P F P	AGCGGAGCAA A E N CTCACGATGG GAGTGCTACC E R H GAGTCAATAT CTCAGGTTATA S D I D	GCCGGAAAAGG A K E GCGAGTCAATT GCCAAGTTAA R N L K GCCCCCCAG GCGCGCCCAG GCGCGAGGTCC	GAAGCCGCG CGAAGCCGCG A E A R TGCGCACGT ACGCGTGCAA R V N TAGCCGCGCGACA ATCGGCGACA	GCGCCTECCG CGCGGACGGC GCGCGAAAC CCGCGCTTG A R F A AAGGCCTGCG TTCCGGACGC L A O T
_	-2 7601 -2 7701 -2 7801	$\begin{array}{c} \textbf{G} \textbf{CTGTTACGG}\\ \textbf{G} \textbf{L} \textbf{I} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{TTCCAGC}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{TTCCAGC}\\ \textbf{G} \textbf{CACAAGGTCG}\\ \textbf{H} \textbf{E} \textbf{L}\\ \textbf{G} \textbf{CTTCCATCA}\\ \textbf{CGAAGGTAGT}\\ \textbf{A} \textbf{E} \textbf{M} \textbf{L}\\ \textbf{TTTGTTGCGC}\\ \end{array}$	CATTCGCCCG Y L P ACCAGCGCAT TGGTCGCGTA V L A H GACCGCGCTGAT CTGCCGACTA R S I CAGCGGCGGGCGG	$ \begin{array}{c} \textbf{CACGCTTAAA} \\ \textbf{CACGCTTAAA} \\ \textbf{T} \textbf{R} \textbf{I} \textbf{E} \\ \hline \textbf{GAAGCTCCGC} \\ \textbf{CTTCGAGGCG} \\ \textbf{CTTCGAGGCG} \\ \textbf{GGTATCCAGT} \\ \textbf{CCATAGGTCA} \\ \textbf{T} \textbf{D} \textbf{L} \\ \hline \textbf{GATACGGAAC} \\ \end{array} $	$\begin{array}{c} \textbf{GTACCGAGTA}\\ \textbf{GTACCGAGTA}\\ \textbf{CGTTTGCGAA}\\ \textbf{GCAAACGCTT}\\ \textbf{T} \textbf{Q} \textbf{S}\\ \textbf{TCGCTGACGCC}\\ \textbf{AGCGACTGCCG}\\ \textbf{E} \textbf{S} \textbf{V} \textbf{G}\\ \textbf{GGTAAACCCC}\\ \end{array}$	CCGCCGTCCC CCGCCGTTCC CGCACCTGTT GCGTGGACAA R V Q E CAGCCGCCTCC GTCGGCGAGA CAGCCGCCTCC GTCGGCGAAGA	$\begin{array}{c} \textbf{ARCETCGATT}\\ \textbf{TTGCAGCTAA}\\ \textbf{F} \textbf{T} \textbf{S} \textbf{K}\\ \textbf{CCGCCAAATC}\\ \textbf{GGCGGTTAG}\\ \textbf{CGCAACGGT}\\ \textbf{GCCTTTGCCA}\\ \textbf{F} \textbf{F} \textbf{F} \\ \textbf{ATAAAACTTA} \end{array}$	$\begin{array}{c} \textbf{TCGCCTCGTT}\\ \textbf{AGCGGAGCAA}\\ \hline \textbf{A} & \textbf{E} & \textbf{N} \\ \hline \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \hline \textbf{E} & \textbf{R} & \textbf{H} \\ \hline \textbf{GAGTCAATAT}\\ \textbf{CTCAGTTATA}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{ATGAGCACAG} \end{array}$	CGCCGGAAAAGG A K E CGGTCAATT GCCAAGGTTAA R N L K CGCCCCCCAG GCGGGAGGTC G E L CGCCAGGATA	$\begin{array}{c} \textbf{GTATCGGCGC}\\ \textbf{GAAGCCGCG}\\ \textbf{A} & \textbf{E} & \textbf{A} & \textbf{R} \\ \textbf{TGCGCACGTT}\\ \textbf{ACGCGTGCAA}\\ \textbf{CGCGCGCGAC}\\ \textbf{A} & \textbf{V} & \textbf{N} \\ \textbf{TAGCCGGCGACA}\\ \textbf{L} & \textbf{R} & \textbf{Q} \\ \textbf{CCTAAGATCA} \end{array}$	$\begin{array}{c} \textbf{GCGCCTECCE}\\ \textbf{GCGCGACGGC}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GGCGCGGCTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} \\ \textbf{A} & \textbf{R} & \textbf{F} \\ \textbf{A} & \textbf{AGGCCTGCG} \\ \hline \textbf{TCCGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTGCCCTAT} \\ \hline \end{array}$
	-2 7601 -2 7701 -2 7801	GTGTTCAGG GTGTTCCAGC GL I G GTGTTCCAGC CACAAGGTCG H E L GCTTCCATCA GCGAGGTAGT A E M L TTTGTTGCGCC AAACAACGCG	CATTCGCCG Y L P ACCAGCGCAT TGGTCGCGTA V L A H GACGGCGCAT CTGCCGACTA CTGCCGGCCGA CTGCCGGCCGG GTCGCCCGCC	$\begin{array}{c} \textbf{GACGCTTAAA}\\ \textbf{T} \textbf{R} \textbf{I} \textbf{E}\\ \textbf{GAAGCTCCGC}\\ \textbf{CTTCGAGGCC}\\ \textbf{CTTCGAGGCCA}\\ \textbf{CTTCGAGGCCA}\\ \textbf{GGTATCCAGT}\\ \textbf{GGTATCCAGT}\\ \textbf{GATACGGAAC}\\ \textbf{CTATGCCTTG}\\ \end{array}$	GTACCGAGTA GTACCGAGTA CGTTTGCGAA GCAAACGCTT T Q S TCGCTGACGC AGCGACGCCGC GGTAAACCAC CCATTTGGTG	CCGCCGTCC A A L CGCACCTGTT GCGTGGACAA R V Q CAGCCGCTCC GTCGGCGGAG GTCGGCGGAG TCGCCAAAGA AGCGGTTTCT	TIGCAGCTAA F T S K CCGCCAAATC GCCGGTTAG GCCGGTTAG CGGAAACGGT GCCTTGCCA P F P ATAAAACTTA TATTTGAAT	TCGCCTCGTT AGCGGAGGCAA $\downarrow E N$ CTCACGATGG GAGTGCTACC $\downarrow E R H$ GAGTCAATAT CTCAGTTATA S D I D D ATGAGCACAG TATCAGTGTC	CCGGAAAAGG A K E CGGTTCAATT GCCAAGTTAA R N L K CGCCCCCCAG GCGGGAGGTC GCGCAGGATA GCGCAGGATA GCGCCCCTCTA	$\begin{array}{c} \textbf{GCARGCCGCG}\\ \textbf{CARGCCGCGG}\\ \textbf{A} & \textbf{E} & \textbf{A} & \textbf{R}\\ \hline \textbf{TGCGCACGTT}\\ \textbf{ACGCGTGCAA}\\ \textbf{ACGCGTGCAA}\\ \textbf{CCGTGCAACAT}\\ \textbf{ATCGCGCACA}\\ \textbf{L} & \textbf{R} & \textbf{Q}\\ \hline \textbf{CCTAAGATCA}\\ \textbf{GGATTCTAGT}\\ \end{array}$	$\begin{array}{c} \textbf{GCGCCTGCCG}\\ \textbf{GCGCGACGGC}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GGCGCGAAAC}\\ \textbf{CCGCGCTTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \textbf{AAGGCCTGCG}\\ \textbf{TTCCGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAACCCTAT}\\ \textbf{GATTGGGATA} \\ \hline \end{array}$
	-2 7601 -2 7701 -2 7801 -2 7801	$\begin{array}{c} \textbf{Catabat bec}\\ \textbf{GTCTTAC6G}\\ \textbf{GTGTTCCAFG}\\ \textbf{GTGTTCCAGC}\\ \textbf{GTGTTCCAFCG}\\ \textbf{H} & \textbf{E} & \textbf{L} \\ \textbf{GCTTCCATCA}\\ \textbf{GCTTCCATCA}\\ \textbf{GCTTCCATCA}\\ \textbf{GCTAGTGCGC}\\ \textbf{AA} & \textbf{E} & \textbf{M} & \textbf{L} \\ \textbf{TTGTTGTGCGC}\\ \textbf{AACAACGCG}\\ \textbf{T} & \textbf{Q} & \textbf{Q} \\ \textbf{A} \\ \textbf{TTGCGCAACGCG} \\ \textbf{T} \\ \textbf{GCGCACGCG} \\ \textbf{GCGCACGC} \\ \textbf{GCGCACGCG} \\ \textbf{GCGCACGC} \\ \textbf{GCGCCACGC} \\ \textbf{GCGCACGC} \\ \textbf{GCGCCACGC} \\ \textbf{GCGCCCACGC} \\ \textbf{GCGCCCACGC} \\ \textbf{GCGCCCACGC} \\ \textbf{GCGCCCACGC} \\ \textbf{GCGCCCACGC} \\ \textbf{GCGCCCACGC} \\ \textbf{GCGCCCCACGC} \\ \textbf{GCGCCCCACGC} \\ \textbf{GCGCCCCACGC} \\ \textbf{GCGCCCCACGC} \\ GCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATTICGCCG T L ACCAGCGCAT T6GTCGCGTA GACGGCGCAT CTGCCGACTA R S CAGCGGCGGG GTCGCCGCCGC L P CAGCGGCCGC TGCCCGCCCCC L P CAGCGGGCGCG GCCGCCCCCC L P CAGCCGCCCCC L P	TO CONTINA CACECITAAA T R I E GAAGCTCCGC CTTCGAGGCG L E A GGTATCCAGT CCATAGGTCA T D L GATACGGAAC CTATGCCTTG S V S R	$\begin{array}{c} \textbf{Categorian}\\ \textbf{Categorian}\\$	CCGCCCETTCC CCGCCCGTCCC CCGCCCGTCCC CCGCCCGTCCC CCGCCCGTCCC CCGCCCGC	$\begin{array}{c} \textbf{AACGFICATT}\\ \textbf{TIGCAGCTAA}\\ \textbf{F} & \textbf{T} & \textbf{S} & \textbf{K}\\ \textbf{CCGCCAAATC}\\ \textbf{GGCGGGTTAG}\\ \textbf{GCCTTAGCA}\\ \textbf{CGGAAACGGT}\\ \textbf{GCCTTGCCA}\\ \textbf{P} & \textbf{F} & \textbf{P}\\ \textbf{ATAAAACTTA}\\ \textbf{TATTTGCAA}\\ \textbf{I} & \textbf{F} & \textbf{S} & \textbf{L}\\ \textbf{CTTTCCCA}\\ \textbf{CTTCCCCA}\\ \textbf{CTTTCCCA}\\ \textbf{CTTCCCCA}\\ \textbf{CTTCCCCA}\\ \textbf{CTTCCCCA}\\ \textbf{CTTCCCCCA}\\ \textbf{CTTCCCCA}\\ CTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \textbf{A} \in \textbf{CrCCTCHT}\\ \textbf{A} \in \textbf{CGACAA}\\ \textbf{A} = \textbf{E} & \textbf{N}\\ \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} & \textbf{R} & \textbf{H}\\ \textbf{GAGTGCAATAT}\\ \textbf{CTCAGTATAT}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D}\\ \textbf{ATGAGCACAG}\\ \textbf{TACTGGTGCT}\\ \textbf{S} & \textbf{C} & \textbf{L}\\ \textbf{S} & \textbf{C} & \textbf{C}\\ \textbf{S} & \textbf{C} & \textbf{C} & \textbf{C}\\ \textbf{S} & \textbf{C} & \textbf{C}\\ \textbf{S} & \textbf{C} & \textbf{C} & \textbf{C} & \textbf{C} \\ \textbf{S} & \textbf{C} & \textbf{C} & \textbf{C} & \textbf{C} & \textbf{C} \\ \textbf{S} & \textbf{C} \\ \textbf{S} & \textbf{C} & \textbf{C}$	CCGCGAAAAGG A K CGGTTCAATT GCCAAGTTAA R N L K CGCCCCAGG GCGGGAGGTC H GCGCGAGGGTC H CGCCCCCCGG GCGGGAGGTC H L CGCCCCCACG AL AL ALCTCCCCCCCCCCCCCC	CTCACGACCCCCC CGAAGCCCCCC A E A R TGCGCACGTT ACGCGCGCGCA R TGCGCCGCTG ACCGCCGCT CCTAAGATCA GGATCTAGG CCTAAGATCA GGATCTAGG CTCCCCTAC	$\begin{array}{c} \textbf{GCGCCTGCCG}\\ \textbf{GCGCGACGGC}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GGCGCGAAAC}\\ \textbf{CCGCGCTTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \textbf{AAGGCCTGCG}\\ \textbf{TTCCGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAACCCTAT}\\ \textbf{GATTGGGATA}\\ \textbf{L} & \textbf{G} & \textbf{I} \\ \hline \textbf{CACCCTCD} \\ \hline \end{array}$
	-2 7601 -2 7701 -2 7801 -2 7801 -2 7901	$\begin{array}{c} \textbf{Catchartec} \\ \textbf{G} \\ \textbf{Catchartec} \\ \textbf{G} \\ \textbf{G} \\ \textbf{T} \\ \textbf{T} \\ \textbf{Catcharger} \\ \textbf{G} \\ \textbf{G} \\ \textbf{T} \\ \textbf{Catcharger} \\ \textbf{G} \\ \textbf{G} \\ \textbf{Catcharger} \\ \textbf{G} \\$	$\begin{array}{c} \textbf{CATTICGCCG}\\ \textbf{Y} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$\begin{array}{c c} \textbf{CACGCTTAAA}\\ \textbf{T} \textbf{R} \textbf{I} \textbf{E}\\ \textbf{GAGGCTCGCC}\\ \textbf{CTTCGAGGCC}\\ \textbf{CTTCGAGGCC}\\ \textbf{GGTATCCAGT}\\ \textbf{GGTATCCAGT}\\ \textbf{GGTATCGGCATG}\\ \textbf{S} \textbf{V} \textbf{S} \textbf{R}\\ \textbf{GTGTAAGACG}\\ \textbf{GTGTAAGACGC}\\ \textbf{GACATTCFGC}\\ \textbf{CACATTCTGC}\\ \textbf{CACATTCTCTGC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACTTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCC}\\ \textbf{CACATTCTCC}\\ \textbf{CACATTCTCTCC}\\ \textbf{CACATTCTCC}\\ \textbf{CACATTCTCC}\\ \textbf{CACATTCTCC}\\ \textbf{CACATTCTCC}\\ \textbf{CACATTCCC}\\ \textbf{CACATTCCC}\\ \textbf{CACATTCCC}\\ \textbf{CACATTCCC}\\ \textbf{CACATTCCC}\\ \textbf{CACATTCCC}\\ \textbf{CACATTCCC}\\ CACTTCC$	CATECECTAT GTACCGAGTA CETTTGCGAA GCAAACGCTT T $ Q$ S TCGCTGACGC ABCCGACTGCG E S V G GGTAAACCAC CCATTTGGTG QCATACCCA	CGCCCGGTCC CGCCCGTCCC CGCCCGTCCC CGCACCTGT CGCACCGCTCC CAGCCGCCCC CAGCCGCGCAC CGCCCGGAAGA AGCGGTTCC CGCGCGAAGA	$\begin{array}{c c} \textbf{AACGFICGATM}\\ \hline \textbf{TIGCAGCTAA}\\ \hline \textbf{TIGCAGCGAAATC}\\ \textbf{GGCGGTTAG}\\ \hline \textbf{GGCGGTTAG}\\ \hline \textbf{GGCGTTGCG}\\ \hline \textbf{GCCTTGCGA}\\ \hline \textbf{CGGAAACGGT}\\ \hline \textbf{GCCTTGCGA}\\ \hline \textbf{TIGCAGAT}\\ \hline \textbf{TIGCAGAT}\\ \hline \textbf{TIGCAGAT}\\ \hline \textbf{TIGCAGAT}\\ \hline \textbf{GTTTTCGAT}\\ \hline \textbf{CAAAAAGCTA}\\ \hline \textbf{CAAAAACCTA}\\ \hline \textbf{CAAAAACCTA}\\ \hline \textbf{CAAAAAACCTA}\\ \hline \textbf{CAAAAAACCTA}\\ \hline \textbf{CAAAAAACCTA}\\ \hline \textbf{CAAAAAACCTA}\\ \hline \textbf{CAAAAAACCTA}\\ \hline \textbf{CAACACCTA}\\ \hline \textbf{CAACACCTA}\\ \hline \textbf{CACACCTA}\\ \hline \textbf{CACACCTA}\\ \hline \textbf{CACACCTA}\\ \hline \textbf{CCCCACCTA}\\ \hline \textbf{CCCCACCTA}\\ \hline \textbf{CCCCACCTA}\\ \hline \textbf{CCCCACCTA}\\ \hline \textbf{CCCCCACTTC}\\ \hline \textbf{CCCCCACTCACTA}\\ \hline \textbf{CCCCCCACTCACTC}\\ \hline \textbf{CCCCCCACTCACTC}\\ \hline \textbf{CCCCCCACTCACTC}\\ \hline \textbf{CCCCCCCACTCC}\\ \hline CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \text{TCGCCTCFT}\\ \text{A} \subseteq \text{CGCAA}\\ \text{A} \equiv M \\ \text{CTCACGATGG}\\ \text{GAGTGCTACC}\\ \text{E} R \\ \text{H}\\ \text{GAGTCAATAT}\\ \text{CTCAGTGATAT}\\ \text{CTCAGTATAT}\\ \text{S} D \\ \text{D} I \\ \text{D}\\ \text{ATGAGCACAG}\\ \text{TACTGGTGTC}\\ \text{S} C \\ \text{CGCCAATGGGCACAG}\\ \text{GGCGTTCACC}\\ GGCGTT$	$\begin{array}{c c} \textbf{GCCCTTTCCC}\\ \hline \textbf{CCGCAAAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CCGGTTCAATT}\\ \hline \textbf{GCCAAGTTAA}\\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{CGCCTCCAG}\\ \hline \textbf{G} & \textbf{G} & \textbf{E} \\ \hline \textbf{G} & \textbf{G} & \textbf{C} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} \hline \textbf{G} \hline \hline \vec$	$\begin{array}{c} \textbf{GATCGCCCC}\\ \textbf{CGAAGCCGCG}\\ \textbf{A} \hspace{0.5mm} \sqsubseteq \hspace{0.5mm} A \hspace{0.5mm} \end{matrix} \\ \textbf{A} \hspace{0.5mm} \rule{0.5mm} \blacksquare \hspace{0.5mm} A \hspace{0.5mm} \end{matrix} \\ \textbf{A} \hspace{0.5mm} \rule{0.5mm} \blacksquare \hspace{0.5mm} A \hspace{0.5mm} \end{matrix} \\ \textbf{A} \hspace{0.5mm} \rule{0.5mm} \blacksquare \hspace{0.5mm} A \hspace{0.5mm} \end{matrix} \\ \textbf{A} \hspace{0.5mm} \rule{0.5mm} \blacksquare \hspace{0.5mm} A \hspace{0.5mm} \end{matrix} \\ \textbf{A} \hspace{0.5mm} \rule{0.5mm} \blacksquare \hspace{0.5mm} A \hspace{0.5mm} \rule{0.5mm} \blacksquare \hspace{0.5mm} A \hspace{0.5mm} \rule{0.5mm} \blacksquare 0.5mm$	$\begin{array}{c} \textbf{GCGCCTGCCG}\\ \textbf{GCGCGACGGC}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GGCGCGAAAC}\\ \textbf{CCGCGCTTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \textbf{AGGCCTGCG}\\ \textbf{TTCCGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAACCCTAT}\\ \textbf{GATTGGGATA}\\ \textbf{L} & \textbf{G} & \textbf{I} \\ \hline \textbf{GAAGCGGGTA}\\ \textbf{CTTCGCCCAT} \end{array}$
	-2 7601 -2 7701 -2 7801 -2 7801 -2 7901 -2	$\begin{array}{c} \textbf{Constant occ}\\ \textbf{G} \textbf{Catchatacog}\\ \textbf{G} \textbf{Catchatacog}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{T} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} G$	$\begin{array}{c} \textbf{CATTICGCCG}\\ \textbf{Y} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$\begin{array}{c c} \textbf{CACCGUARTIAA}\\ \textbf{T} \textbf{R} \textbf{I} \textbf{E}\\ \textbf{GAGEGTCGGC}\\ \textbf{CTTCGAGGCG}\\ \textbf{CTTCGAGGCG}\\ \textbf{CTTCGAGGCG}\\ \textbf{CTTCGAGGCGGAC}\\ \textbf{CTATGGCCTTG}\\ \textbf{S} \textbf{V} \textbf{S} \textbf{R}\\ \textbf{GTGTAGGCGAC}\\ \textbf{CTATGGCCTTG}\\ \textbf{CACATTCFGC}\\ \textbf{T} \textbf{L} \textbf{R} \end{array}$	CATECECTAT GTACCGAGTA CETTTGCEAA GCAAACGCTT TOSCTAACGCTT TOSCTAACGCT CGCTGACGCC ABCCGACTGCG GGTAAACCAC CCATTGGGT CGCTAACCCA CCATTGGGT CGCATGGGT A L G	CCGCCGGTCC CCGCCGTTCC CCGCCGCTCC CCGCGCGCCC CCGCGCCCC CCGCCCCCC CCGCCCCCC	$\begin{array}{c c} \textbf{ARCOFICATIV}\\ \textbf{TIGCAGCTAA}\\ \textbf{F} & \textbf{TIGCAGCTAA}\\ \textbf{F} & \textbf{TIGCAGCCAAATC}\\ \textbf{GGCGGTTAG}\\ \textbf{GGCGGTTAG}\\ \textbf{GGCGTTAGCGGT}\\ \textbf{GGCGTTGCCA}\\ \textbf{F} & \textbf{F} & \textbf{F} \\ \textbf{GCTTTGCAT}\\ \textbf{TATAAACTTA}\\ \textbf{ATAAAACTTA}\\ \textbf{ATATAACTTA}\\ \textbf{ATATTTCGAT}\\ \textbf{I} & \textbf{F} & \textbf{S} & \textbf{L} \\ \textbf{GTTTTCGAT}\\ \textbf{CAAAAGCTA}\\ \textbf{K} & \textbf{E} & \textbf{I} \\ \end{array}$	$\begin{array}{c} \textbf{TCGCCTGTT}\\ \textbf{A} & \textbf{E} & \textbf{N} \\ \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} & \textbf{R} & \textbf{H} \\ \textbf{GAGTCAATAT}\\ \textbf{CTCAGTTATA}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{A} & \textbf{GAGTCAATAT}\\ \textbf{GGCGTCACCGTGTC}\\ \textbf{GGCGTTCACCC}\\ \textbf{GGCGTTCACCC}\\ \textbf{A} & \textbf{N} & \textbf{V} \end{array}$	CCGCAAAAGG CCGCAAAAGG CCGCAAGTTAA R N L K CGCCTCCAG CGCCACGCACA CGCCCTCCAG CCCCACGAA CCCCCCCCAG CCCCCCCCAG CCCCCCCCACA CCCCCCCCACA CCCCCCCCACA CCCCCCCC	CTTCGCCGC GAAGCCGCCGT A E A R TGCGCGCGCT ACGCGGCGCAC CGCACGGCGCAC L R Q CCTAAGGATCA GGATCCAGT G L I V CTCCCGTACT GAAGCCATGA A CTCCGTACT	$\begin{array}{c} \textbf{GCCGCGACGGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGAAGC}\\ \textbf{GCCGCGAAAC}\\ \textbf{CCGCGCTTGG}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{AAGGCCTGCG}\\ \textbf{TTCCGCGGACGC}\\ \hline \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{GATGGGATA}\\ \textbf{GATGGGATA}\\ \hline \textbf{GATGGGGGTA}\\ \hline \textbf{GAGCGGGTA}\\ \hline \textbf{CTTCGCCCCAT}\\ \hline \hline \textbf{F} & \textbf{R} & \textbf{T} & \textbf{F} \end{array}$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001	$\begin{array}{c} \textbf{Constructor}\\ \textbf{G} \textbf{C} \textbf{C} \textbf{T} \textbf{C} \textbf{T} \textbf{A} \textbf{C} \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{G} \textbf{T} \textbf{C} \textbf{A} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} G$	$\begin{array}{c} \textbf{CATTICGCCG}\\ \textbf{Y} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$\begin{array}{c c} \textbf{GACGCTTAAA}\\ \textbf{T} \textbf{R} \textbf{I} \textbf{E}\\ \hline \textbf{GAGGCTCGGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCG}\\ \textbf{CTTCGAGGCG}\\ \textbf{GGTATCGGGAC}\\ \textbf{CTATGGCTTG}\\ \textbf{S} \textbf{V} \textbf{S} \textbf{R}\\ \hline \textbf{GTGTAGGCGAC}\\ \textbf{CTATGGCTTG}\\ \textbf{CACATGCGCTT}\\ \textbf{CACATGCGCT}\\ \textbf{CACATGCGCT}\\ \textbf{CACATGCGCT}\\ \textbf{CACATGCGCT}\\ \textbf{CACAGGCCT}\\ \textbf{CACATGCGT}\\ \textbf{CACAGGCCT}\\ \textbf$	CATEGORIA GTACCGAGATA CETTTGCEAA GCAAACGCTT T COTTGCCAA GCAAACGCT T CGCTGACGCA AGCGACTGCG GGAAACCAC CCATTGGTG CCATGTG CCATGGTG CCATTGGTG CCATGTG CCATTGGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGGTG CCATGGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGGTG CCATGGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGTG CCATGG CCATGGGT CCATGGTG CCATGG CCATGGTG CCATG CCATG CCATGC CCATG	$\begin{array}{c} \textbf{GCCCCCCC}\\ \textbf{CCGCCGTTCC}\\ \textbf{CCGCCGTTCC}\\ \textbf{CCGCCGCTCC}\\ \textbf{GCCGGCCCCC}\\ \textbf{GCCGGCCCCC}\\ \textbf{GCCGGCCCCC}\\ \textbf{GCCGGCCCCC}\\ \textbf{GCCGGCCGCTCC}\\ \textbf{GCCGGCCGCTCC}\\ \textbf{GCCGCCTCCTTCC}\\ \textbf{CCCGCAAGA}\\ \textbf{A} \\ \textbf{C} \\ \textbf{K} \\ $	$\begin{array}{c} \textbf{AACCFICGATT}\\ \textbf{TIGCAGCTAA}\\ \textbf{F} & \textbf{T} & \textbf{S} & \textbf{K}\\ \textbf{CCGCCGAATC}\\ \textbf{GGCGGTTAG}\\ \textbf{GGCGTTAGCGGT}\\ \textbf{GCCTTIGCCA}\\ \textbf{GCCTTIGCCA}\\ \textbf{F} & \textbf{F} & \textbf{F}\\ \textbf{ATAAACCTA}\\ \textbf{I} & \textbf{F} & \textbf{S} & \textbf{L}\\ \textbf{GTTITCGAT}\\ \textbf{I} & \textbf{F} & \textbf{S} & \textbf{L}\\ \textbf{GTTITCCATA}\\ \textbf{GTTITCCTA}\\ \textbf{GTTTCCTA}\\ \textbf{GTTTCCTA}\\ \textbf{GTTTCCTA}\\ \end{array}$	$\begin{array}{c} \textbf{A} \subset \textbf{CacCreating}\\ \textbf{A} \subset \textbf{CacCatGatGG}\\ \textbf{CacCGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} \textbf{R} \textbf{H} \\ \textbf{GAGTCAATAT}\\ \textbf{CTCAGTTATA}\\ \textbf{S} \textbf{D} \textbf{I} \textbf{D} \\ \textbf{ATGAGCACAGT}\\ \textbf{ACCGTGTC}\\ \textbf{GGCTTCACCC}\\ \textbf{GGCCTTCACCC}\\ \textbf{CCCCCAAGTGG}\\ \textbf{A} \textbf{N} \textbf{V} \\ \textbf{CCARACAGGGACGAGTGG} \\ \textbf{CCCCCAAGTGGGC} \\ \textbf{A} \textbf{N} \textbf{V} \\ \textbf{CCARACAGGGAGTGGC} \\ \textbf{CCCCCAAGTGGC} \\ \textbf{CCCCCAAGTGGC} \\ \textbf{CCCCCCAAGTGGC} \\ \textbf{CCCCCCAAGTGGC} \\ \textbf{CCCCCCACGTGGC} \\ CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} \textbf{GCCCTTTTCCC}\\ \hline \textbf{CCCGCAAAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CCGCTCAATT}\\ \hline \textbf{GCCAAGTTAA}\\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{CGCCTCCAG}\\ \hline \textbf{G} & \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{CGCCTCCAG}\\ \hline \textbf{G} & \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{G} & \textbf{CGCCTCCAG} \\ \hline \textbf{G} & \textbf{G} & \textbf{L} & \textbf{L} \\ \hline \textbf{G} & \textbf{G} & \textbf{L} & \textbf{L} \\ \hline \textbf{G} & \textbf{G} & \textbf{C} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} $	$\begin{array}{c c} \textbf{GetaGccGcG}\\ \textbf{CGAAGCCGCCG}\\ \textbf{A} & \textbf{E} & \textbf{A} & \textbf{R} \\ \textbf{TGCGCGCCGTT}\\ \textbf{ACGCGGCGCAGT}\\ \textbf{ACGCGGCGCAGT}\\ \textbf{ACGCGGCGCAGT}\\ \textbf{ATCGGCGCAGT}\\ \textbf{ATCGGCGCAGT}\\ \textbf{CTAGGCGCAGT}\\ \textbf{G} & \textbf{L} & \textbf{T} & \textbf{V} \\ \textbf{CTTCGGTGCT}\\ \textbf{G} & \textbf{L} & \textbf{T} & \textbf{V} \\ \textbf{GAGGCCATGA}\\ \textbf{GAAGCGCATGA}\\ \textbf{E} & \textbf{T} & \textbf{S} \\ \textbf{GGTGTATCGG}\\ \textbf{G} \\ \textbf{GTGTGTATCGG} \\ \textbf{G} \\ \textbf{GTGTGTATCGG} \\ \textbf{G} \\ \textbf{GTGTGTATCGG} \\ \textbf{G} \\ \textbf{GTGTGTCGGG} \\ \textbf{G} \\$	$\begin{array}{c} \textbf{GCCGCCGCCGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGAAGC}\\ \hline \textbf{GCCGCGAAACCCGCCGAAACCCCGCGCTTG}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{AGGCCTGCG}\\ \hline \textbf{TCCGCGGACGC}\\ \hline \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAACCCTAT}\\ \hline \textbf{GAATGCGGATA}\\ \hline \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \textbf{GAACCGGGTA}\\ \hline \textbf{CTTCGCCCAT}\\ \hline \hline \textbf{F} & \textbf{R} & \textbf{T} & \textbf{F} \\ \hline \textbf{TTAATTTC}\\ \hline \textbf{TTATTTC}\\ \hline \textbf{CTCGCCCAT} \\ \hline \textbf{GAATGCGGCACACCCAT} \\ \hline \textbf{GAATCGGGCACCACCAT} \\ \hline \textbf{GAATCGGGCACCACCACCACCACCCCCAT} \\ \hline \textbf{GAATCGGGCACCACCACCACCACCCCCAT} \\ \hline \textbf{GAATCGGGCACCACCACCACCACCACCCCCAT} \\ \hline \textbf{GAATCGGGCACCACCACCACCACCCCCAT} \\ \hline GAATCGCGGCACCACCACCACCCACCCCCACCCCCCCCCACCCCCC$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2	$\begin{array}{c} \textbf{Cachager}_{\textbf{C}} \textbf{Cachager}_{\textbf{C}} \textbf{C}\\ \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{C} \textbf{C} \textbf{A} \textbf{C} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{T} \textbf{G} \textbf{T} \textbf{C} \textbf{A} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{C} \textbf{G} \textbf{C} \textbf{A} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{H} \textbf{E} \textbf{L} \textbf{L}\\ \textbf{G} \textbf{C} \textbf{T} \textbf{C} \textbf{C} \textbf{A} \textbf{A} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{H} \textbf{E} \textbf{L} \textbf{L}\\ \textbf{G} \textbf{C} \textbf{T} \textbf{C} \textbf{C} \textbf{A} \textbf{A} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{H} \textbf{E} \textbf{L} \textbf{L}\\ \textbf{G} \textbf{G} \textbf{T} \textbf{C} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{H} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}$	$\begin{array}{c} \text{CATTICGCCG}\\ \begin{array}{c} Y \\ - L \end{array} \\ \begin{array}{c} P \\ \end{array} \\ \begin{array}{c} \text{ACCABCGCAT} \\ \hline \\ \text{TGGTCGCGAT} \\ \hline \\ \text{GACGGCTGAT} \\ \hline \\ \text{CTGCCGACTA} \\ \hline \\ \hline \\ \begin{array}{c} R \\ \end{array} \\ \begin{array}{c} S \\ \end{array} \\ \begin{array}{c} \text{CACCGGCGAT} \\ \hline \\ \hline \\ \hline \\ \begin{array}{c} R \\ \end{array} \\ \begin{array}{c} \text{CACCGGCGAT} \\ \hline \\ $	$\begin{array}{c c} \textbf{Caccertraa}\\ \textbf{T} & \textbf{R} & \textbf{I} & \textbf{E}\\ \hline \textbf{GAGEGTCGGC}\\ \textbf{CTTCGAGGCG}\\ \textbf{CTTCGAGGCG}\\ \textbf{CTTCGAGGCG}\\ \hline \textbf{L} & \textbf{E} & \textbf{A}\\ \hline \textbf{GGTATCGGCA}\\ \hline \textbf{CATAGGCTTG}\\ \hline \textbf{S} & \textbf{V} & \textbf{S} & \textbf{R}\\ \hline \textbf{GTGTAGGCGAC}\\ \hline \textbf{CATAGGCTTG}\\ \hline \textbf{CACATGCTTGGC}\\ \hline \textbf{T} & \textbf{L} & \textbf{R}\\ \hline \textbf{CATAGGCTTGGCAA}\\ \hline \textbf{GTGTAAGGCGTA}\\ \hline \textbf{GTGTAAGGCGTA}\\ \hline \textbf{GTGTAAGCGTT}\\ \hline \textbf{GTATTCGCAA}\\ \hline \textbf{M} & \textbf{L} & \textbf{T} \\ \hline \end{array}$	CATEGECTAAT GTACCGAGATA CGTTTGCGAA GCAAACGCTT T	$\begin{array}{c} \textbf{GCCGCGTTCC}\\ \hline \textbf{CGCCGTTCC}\\ \hline \textbf{A} & \textbf{A} & \textbf{L} \\ \hline \textbf{CGCGCGTTCC}\\ \hline \textbf{GCGTGGACAA}\\ \hline \textbf{GCGTGGCCTC}\\ \hline \textbf{GCGCGCCCC}\\ \hline \textbf{GCGGCGCCC}\\ \hline \textbf{GCGGCGCTC}\\ \hline \textbf{GCGCGGTTCC}\\ \hline \textbf{R} & \textbf{W} & \textbf{L} \\ \hline \textbf{GCCGGTTCC}\\ \hline \textbf{CGCGAAGA}\\ \hline \textbf{A} & \textbf{E} & \textbf{N} & \textbf{F} \\ \hline \textbf{GCCGGGATAGC}\\ \hline \textbf{GCGCCTATCG}\\ \hline \textbf{GCGCCTATCG}\\ \hline \textbf{GCCCCTATCG}\\ \hline \textbf{GCCCCTATCG}\\ \hline \textbf{GCCCCATCG}\\ \hline \textbf{GCCCCTATCG}\\ \hline \textbf{GCCCCATCG}\\ \hline \ \textbf{GCCCCATCG}\\ \hline \hline \hline \textbf{GCCCCATCG}\\ \hline \hline \hline \textbf{GCCCCATCG}\\ \hline \hline \hline \textbf{GCCCCCATCG}\\ \hline \hline \hline \hline \textbf{GCCCCCATCG}\\ \hline \hline \hline \textbf{GCCCCCATCG}\\ \hline \hline \hline \hline \textbf{GCCCCCATCC}\\ \hline \hline \hline \hline \hline \hline \textbf{GCCCCCATCC}\\ \hline \hline$	$\begin{array}{c} \textbf{AACGFICGATT}\\ \textbf{TIGCAGGCTAA}\\ \textbf{F} & \textbf{T} & \textbf{S} & \textbf{K}\\ \textbf{CCGCCGAATC}\\ \textbf{GGCGGTTAG}\\ \textbf{GGCGGTTAG}\\ \textbf{GGCGTTGCCAA}\\ \textbf{GCGTTGCCA}\\ \textbf{GCGTTGCCA}\\ \textbf{GCGTTGCCA}\\ \textbf{GCGTTGCCA}\\ \textbf{GTTTGCAT}\\ \textbf{GTTTTCCAT}\\ \textbf{GTTTTCCAT}\\ \textbf{GTTTTCCATA}\\ \textbf{GTTTTCCATA}\\ \textbf{GTAGGGAT}\\ \textbf{CAAAGGGAT}\\ \textbf{GAAGGGAT}\\ \textbf{GAAGGGAT}\\ \textbf{GAAGGGAT}\\ \textbf{GAAGGGAT}\\ \textbf{GTTCCTA}\\ \textbf{GTTCCTA}\\$	$\begin{array}{c} \textbf{TCGCCTCFT}\\ \textbf{A} & \textbf{E} & \textbf{N} \\ \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} & \textbf{R} & \textbf{H} \\ \textbf{GAGTCATAT}\\ \textbf{CTCAGTTATA}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{A} & \textbf{GAGTCATAT}\\ \textbf{CTCAGTTATA}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{A} & \textbf{A} & \textbf{C} \\ \textbf{CGCGATACGTGTC}\\ \textbf{GGCGTTCACCC}\\ \textbf{GGCGTTCACCC}\\ \textbf{GCCATAGTGGC}\\ \textbf{A} & \textbf{N} & \nabla \\ \textbf{CCGAAACAGGGGCTTGTCCCC}\\ \textbf{R} & \textbf{F} & \textbf{I} & \textbf{L} \\ \end{array}$	CCGCAAAAGG CCCGCAAAAGG CCGCAAGTTAA R N L K CGCCTCCAG GCGCGAGGTC G E L G E L AACTGCGCACA AACTGCGCACA AACTGCGCACA ATTGGCCACG ATTGGCCACG ATTGGCCACG ATTGGCCACG	$\begin{array}{c c} \textbf{Getafcccccc}\\ \textbf{Caracccccc}\\ \textbf{A} & \textbf{E} & \textbf{A} & \textbf{R} \\ \textbf{TGCCCCCCT}\\ \textbf{ACCCCCCCT}\\ \textbf{ACCCCCCCT}\\ \textbf{ACCCCCCCCT}\\ \textbf{ACCCCCCCCT}\\ \textbf{ACCCCCCCCCC}\\ \textbf{ACCCCCCCCCCC}\\ \textbf{ACCCCCCCCCCCC}\\ ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \textbf{GCCGCCGCCGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGAAGC}\\ \hline \textbf{GCCGCGAAAC}\\ \hline \textbf{GCCGCGCTTG}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{AAGGCTGCG}\\ \hline \textbf{TTCGGGATG}\\ \hline \textbf{TCGGGATG}\\ \hline \textbf{CTAGCCGGT}\\ \hline \textbf{GATGGGATA}\\ \hline \textbf{GATGGGATA}\\ \hline \textbf{GATGGGATA}\\ \hline \textbf{F} & \textbf{R} & \textbf{T} & \textbf{F} \\ \hline \textbf{TTATTTCT}\\ \hline \textbf{AAGAAGGA}\\ \hline \textbf{AAGAAAGGA}\\ \hline \textbf{A} & \textbf{T} & \textbf{K} & \textbf{T} \\ \hline \textbf{GATAAAAGGA}\\ \hline \textbf{GATGGAAAGGA}\\ \hline \textbf{GATGGAA}\\ \hline \textbf{GATGGGATA}\\ \hline \textbf{GAAAGAAGGA}\\ \hline \textbf{GAAAAAGGA}\\ \hline \textbf{GAAAAAGGA}\\ \hline \textbf{GAAAAAGGA}\\ \hline \textbf{GAAAAAAGGA}\\ \hline \textbf{GAAAAAAAGGA}\\ \hline \textbf{GAAAAAAAAGGA}\\ \hline GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8001 -2 8101	$\begin{array}{c} \textbf{Cacacada focus } \textbf{Cocacada focus } \textbf{Cocacada focus } \textbf{Cocacada focus } \textbf{Cocacada focus } \textbf{Cacada focus } Caca$	$\begin{array}{c} \textbf{CATTICGCCG}\\ \textbf{Y} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$\begin{array}{c c} \textbf{CACCGUARTIAL} \\ \hline \textbf{CACCGUTAAA} \\ \hline \textbf{T} & \textbf{R} & \textbf{I} & \textbf{E} \\ \hline \textbf{GABAGCUCGCC} \\ \textbf{CTTCGAGGCC} \\ \hline \textbf{CTTCGAGGCC} \\ \hline \textbf{CTTCGAGGCC} \\ \hline \textbf{CATACGGAAC} \\ \hline \textbf{CATACGGAAC} \\ \hline \textbf{CATACGGAAC} \\ \hline \textbf{CACATCGCCTTG} \\ \hline \textbf{CATACGCCTTG} \\ \hline \textbf{CATACGCCTTG} \\ \hline \textbf{CATACGCCTT} \\ \hline \textbf{GTATTCGCAA} \\ \hline \textbf{M} & \textbf{L} & \textbf{T} \\ \hline \textbf{CGCCGAACGGACGGCT} \\ \hline \textbf{GCCCGAACGGACGCG} \\ \hline CGCCCAACGCCCAACGCCCCCCCCCCCCCCCCCCCCCC$	CATEGECTAAT GTACCGABCTA CGTTGCGAA GCAAACGCTT T S _ TCGCTGACGC AGCGACTGCG GGTAAACCAC CCATTGGTG CCATTGGTG CCATTGGTG CGCTAACCCA CGCTAACCCA CGCTAACCGA CCATTGGTG CGCTAACCGA CCATTGGTG CCATTGGCG CGCTACCCA CCATTGGCG CGCTACCCA CGCTACCCA CCATTGGCG CGCTACCCA CGCTCCCA CGCTACCA CGCTACCCA CGCTACCCA CGCTACCCA CGCTACCCA CGCTACCCA CGCTACCCA CGCTACCCA CGCTAC	CCGCCGCTCC CTCCCCCTTT CCGCCGCCTC CTCCCCCAAGA CCGCCCCCC CTCCGCCAACA CCGCCCCCCTCC CCGCGCCAACG CCGCGCCCACCG CCGCGCCCACCG CCGCGCCCACCG CCGCGCCCACCG CCGCGCCCCCCCC	$\begin{array}{c} \textbf{AACGFICGATM}\\ \textbf{F} \textbf{TIGCAGCTAA}\\ \textbf{F} \textbf{TIGCAGCTAATC}\\ \textbf{GGCGGTTAGC}\\ \textbf{GGCGGTTAGCAT}\\ \textbf{GGCGGTTAGCAT}\\ \textbf{GCGGAAACGGT}\\ \textbf{GCTTTGCCA}\\ \textbf{F} \textbf{F} \textbf{F} \\ \textbf{ATAAAACTTA}\\ \textbf{TATTTGGAAT}\\ \textbf{TATTTGGAAT}\\ \textbf{TATTTGGAAT}\\ \textbf{CAAAAAGGAAT}\\ \textbf{K} \textbf{E} \textbf{I} \\ \textbf{GTTTTCCTTA}\\ \textbf{CAAAAGGAAT}\\ \textbf{N} \textbf{E} \textbf{K} \\ \textbf{TGGCGGCGGGG} \end{array}$	$\begin{array}{c} \textbf{A} \subset \textbf{CGACTGUT}\\ \textbf{A} \subset \textbf{CGACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{C} \leftarrow \textbf{R} \textbf{H} \textbf{H} \textbf{GAGTGCTACC}\\ \textbf{GAGTCATAT}\\ \textbf{CTCAGTTATA}\\ \textbf{S} \textbf{D} \textbf{I} \textbf{D} \\ \textbf{ATGAGCACAG}\\ \textbf{GCTTCACCGTGTC}\\ \textbf{CCGCAAGTGGT}\\ \textbf{CCGCAAGTGGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} \textbf{F} \textbf{L} \textbf{T} \\ \textbf{ATGAACAGGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} \textbf{R} \textbf{L} \textbf{L} \textbf{T} \\ \textbf{ATGAACATTT} \\ \textbf{TATAAATTTT} \end{array}$	$\begin{array}{c c} \textbf{GCCTTTTCCC}\\ \textbf{CCGGAAAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CCGGTCAATT}\\ \textbf{GCCAAGTTAA}\\ \hline \textbf{GCCAAGTTAA}\\ \hline \textbf{GCCATCCAG}\\ \hline \textbf{GCCCTCCAG}\\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGGCTCCTAT}\\ \hline \textbf{A} & \textbf{L} & \textbf{L} \\ \hline \textbf{ACGGGCCTT}\\ \hline \textbf{TGACGGCTGT}\\ \hline \textbf{ATGGCGCTGT}\\ \hline \textbf{ATGGCCAGTGT}\\ \hline \textbf{ATGGCCACTGT}\\ \hline \textbf{ATGGCAACTGA}\\ \hline \textbf{ACCAACTGA}\\ \hline \ \ \textbf{ACCAACTGA}\\ \hline \ \textbf{ACCAACTGA}\\ \hline \ \ \ \textbf{ACCAACTGA}\\ \hline \ \ \ \textbf{ACCAACTGA}\\ \hline \ \ \ \ \ \$	$\begin{array}{c} \text{GenarccGcG}\\ \textbf{A} & \textbf{E} & \textbf{A} & \textbf{R}\\ \textbf{TFCGGCCGCT}\\ \textbf{ACGCGTGCAA}\\ \textbf{ACGCGTGCAA}\\ \textbf{ACGCGTGCAA}\\ \textbf{ACGCGTGCAA}\\ \textbf{ACGCGTGCAA}\\ \textbf{ACGCGTGCAAC}\\ \textbf{CCTAAGATCA}\\ \textbf{G} & \textbf{L} & \textbf{T} & \textbf{V}\\ \textbf{CTTCGGTACTAGT}\\ \textbf{G} & \textbf{L} & \textbf{T} & \textbf{V}\\ \textbf{G}\\ \textbf{GGTGTACTGAG}\\ \textbf{G}\\ \textbf{G}\\ \textbf{CCAAGAACATGA}\\ \textbf{CACCGTGACTGA}\\ \textbf{CCACATGACT}\\ \textbf{T} & \textbf{Y} & \textbf{Q}\\ CTTCGTACTTTTTTAGTTTTTTAGTTTTTTAGTTTTTTTT$	$\begin{array}{c} \textbf{GCCGCGACGGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGAAGC}\\ \hline \textbf{GCCGCGAAAC}\\ \hline \textbf{CCGCGCTTGG}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{AGGCCTGCG}\\ \hline \textbf{TCCGGGACGC}\\ \hline \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAACCCTAT}\\ \hline \textbf{GAATGGGGATA}\\ \hline \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \textbf{GAACCGGGTA}\\ \hline \textbf{CTTGCCCAT}\\ \hline \textbf{GAACCGGGTA}\\ \hline \textbf{CTTGCCCAT}\\ \hline \textbf{TTAATTTCT}\\ \hline \textbf{AAAAAAAA}\\ \hline \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \textbf{CACCGFETC} \\ \hline \textbf{CACGGFETC} \\ \hline \textbf{CACGFETCC} \\ \hline \textbf{GACGGFETC} \\ \hline \end{array}$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8001 -2 8101	$\begin{array}{c} \textbf{Ccaccadaticce}\\ \textbf{G} \textbf{CtcTtaCc6G}\\ \textbf{G} \textbf{CaccadGtc6}\\ \textbf{G} \textbf{CaccadGtc6}\\ \textbf{G} \textbf{CaccadGtc6}\\ \textbf{H} \textbf{H} \textbf{E} \textbf{L}\\ \textbf{G} \textbf{CtTcCatCa}\\ \textbf{CGAaGGtaGtaGt}\\ \textbf{A} \textbf{E} \textbf{M} \textbf{L}\\ \textbf{G} \textbf{CtTcCatCa}\\ \textbf{G} \textbf{CaccaGGtaGt}\\ \textbf{A} \textbf{L}\\ \textbf{G} \textbf{CttCcatCa}\\ \textbf{G} \textbf{CaccaGGtaGt}\\ \textbf{G} \textbf{G} \textbf{R}\\ \textbf{A} \textbf{CGGGTGCG}\\ \textbf{G} \textbf{G} \textbf{R}\\ \textbf{A} \textbf{CGGGAGGTGG}\\ \textbf{G} \textbf{G} \textbf{R}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G}\\ \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}$	$\begin{array}{c} \text{CATTICGCCG}\\ \underline{\vee} & \underline{\perp} & \underline{P} \\ \text{ACCABCGCAT} \\ \textbf{TGGTCGCGAT} \\ \textbf{TGGTCGCGAT} \\ \hline \textbf{GACGGCTGAT} \\ \textbf{CTGCCGACTA} \\ \underline{\vee} & \underline{R} & \underline{S} & \underline{\bot} \\ \textbf{GACGGCTGAT} \\ \textbf{CACCGGCGGCG} \\ \hline \underline{L} & \underline{P} & \underline{R} \\ \textbf{TGACTGGCGT} \\ \textbf{ACTGACCGCA} \\ \textbf{GGTATTGGC} \\ \textbf{CCATAACCG} \\ \hline \underline{N} & \underline{N} & \underline{A} \\ \textbf{GGCGATTGGC} \\ \textbf{CCGCTBAACC} \\ \end{array}$	$\begin{array}{c c} \textbf{GACGCTTAAA}\\ \hline \textbf{CACGCTTAAA}\\ \hline \textbf{T} & \textbf{R} & \textbf{I} & \textbf{E}\\ \hline \textbf{GAGAGCTCGGC}\\ \hline \textbf{CTTCGAGGCC}\\ \hline \textbf{CTTCGAGGCC}\\ \hline \textbf{CTTCGAGGCC}\\ \hline \textbf{CATACGGAC}\\ \hline \textbf{CTTGCTTG}\\ \hline \textbf{CATACGGAC}\\ \hline \textbf{CATACGGAC}\\ \hline \textbf{CACATCTGC}\\ \hline \textbf{CACATCTGC}\\ \hline \textbf{CACATCTGC}\\ \hline \textbf{CACATCGCCTT}\\ \hline \textbf{GATACGGAC}\\ \hline \textbf{CACATCTGC}\\ \hline \textbf{CACATCTGC}\\ \hline \textbf{CACATCGCCT}\\ \hline \textbf{GATACGGAC}\\ \hline \textbf{GCGCGAACGGC}\\ \hline \textbf{GCGCGAACGGC}\\ \hline \textbf{GCGCCGAACGGC}\\ \hline \textbf{GCGCCAACGGCCCCCCC}\\ \hline \ \textbf{CCCCCACACGCCCCCC}\\ \hline \ \textbf{CCCCCACACGCCCCCC}\\ \hline \textbf{CCCCCACACGCCCCCC}\\ \hline \textbf{CCCCCCACACGCCCCCC}\\ \hline \textbf{CCCCCACACGCCCCCC}\\ \hline \textbf{CCCCCACACGCCCCCCC}\\ \hline \textbf{CCCCCCACACGCCCCCC}\\ \hline \textbf{CCCCCCACACGCCCCCC}\\ \hline \textbf{CCCCCCACACGCCCCCC}\\ \hline \ \textbf{CCCCCCACACGCCCCCC}\\ \hline \ \textbf{CCCCCCACACCGCCCCCC}\\ \hline \ \textbf{CCCCCCACACCGCCCCCC}\\ \hline \ \textbf{CCCCCCACACCCCCCC}\\ \hline \ \textbf{CCCCCCACACCGCCCCCCCCC}\\ \hline \ \textbf{CCCCCCACACCGCCCCCCCCCC}\\ \hline \ \textbf{CCCCCCACACCCCCCCCCC}\\ \hline \ \textbf{CCCCCCCCCCCCCCCCCCCCCCCC}\\ \hline \ CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATECECTAAT GTACCEABGTA COTTTGCEAA GCAAACGCTT T S TCGCTGACGC AGCGACTECG E	$\begin{array}{c} \textbf{LCCECCAAGE}\\ \textbf{CCGCCGTTCC}\\ \textbf{CCGCCGTTCC}\\ \textbf{CCGCCGTCC}\\ \textbf{CCGCCGCTC}\\ \textbf{CCGCCGCCTC}\\ \textbf{CCGCGCCAACG}\\ \textbf{CCGCCAACGC}\\ \textbf{CCGGCCAACGC}\\ \textbf{CCGGACCAACGC}\\ \textbf{CCGGACCAACGC}\\ \textbf{CCGCGCTACCGCCTC}\\ \textbf{CCGCGGATAGC}\\ CGCGGCTACCGCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} \textbf{AACGFICGATT}\\ \textbf{TTGCAGCTAA}\\ \textbf{F} & \textbf{T} & \textbf{S} & \textbf{K}\\ \textbf{CCGCCAAATC}\\ \textbf{GGCGGTTAGC}\\ \textbf{GGCGTTAGCAT}\\ \textbf{GCGGAAACGGT}\\ \textbf{GCCTTGCCA}\\ \textbf{GCTTTGCCAT}\\ \textbf{ATAAAACTTA}\\ \textbf{TATTTGGAT}\\ \textbf{TATTTGCAT}\\ \textbf{CAAAAAGGAT}\\ \textbf{K} & \textbf{E} & \textbf{I} \\ \textbf{GTTTTCCTTA}\\ \textbf{CAAAAAGGAAT}\\ \textbf{N} & \textbf{E} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{M} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} \\ \textbf{K} & $	$\begin{array}{c} \textbf{A} \textbf{C} \textbf{C} \textbf{A} \textbf{C} \textbf{C} \textbf{A} \textbf{C} \textbf{A} \textbf{C} \textbf{A} \textbf{C} $	CCGCARAAGG CCGCARAAGG CCGCARATT GCCAGGTCAATT GCCAGGTCAATT GCCAGGTCA CGCCCTCCAG GCGCCTCCAG GCGGCCCCAG GCGGCCCCAT CGCCCAGGATA CGCCCTCCAG CGCCCCCCAG GCGGCCCCAT TGCCGCCGT TGCCCCGCCAG TGCCCCGCCAG TGCCCCCAG TGCCCCCAG TGCCCCCAG TGCCCCCAG TGCCCCCAG TGCCCCCAG TGCCCCCCAG TGCCCCCCAG TGCCCCCCAG TGCCCCCCAG TGCCCCCCAG TGCCCCCCCAG TGCCCCCCCAG TGCCCCCCCCAG TGCCCCCCCAG TGCCCCCCCAG TGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{c c} \textbf{GCTTCGCCGC}\\ \textbf{G} & \textbf{E} & \textbf{A} & \textbf{R} \\ \textbf{TGCGCACGTT}\\ \textbf{ACGCGTCGAA}\\ \textbf{R} & \textbf{V} & \textbf{N} \\ \textbf{TGCCGCTGT}\\ \textbf{ACGCGTGCAA}\\ \textbf{ACGCTGCAA}\\ \textbf{CCTAAGGATCA}\\ \textbf{GGATCTCAGT}\\ \textbf{G} & \textbf{L} & \textbf{I} & \textbf{V} \\ \textbf{CTTCCGTACT}\\ \textbf{GAAGGCATGA}\\ \textbf{GAAGGCATGAC}\\ \textbf{GACGCATGACT}\\ \textbf{GACGCATGACT}\\ \textbf{T} & \textbf{Y} & \textbf{Q} \\ \textbf{CCAAGAAGATTAT}\\ \textbf{ATTCAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAA} \\ \textbf{ATTCAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ \textbf{ATTCAAAAAAA} \\ ATTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA$	$\begin{array}{c} \textbf{GCCGCCGCCGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGAAAC}\\ \textbf{CCGCGCGAAAC}\\ \textbf{CCGCGCTTGC}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{AGGCCTGCG}\\ \textbf{TTCGGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAACCCTAT}\\ \textbf{GATCGGGATA}\\ \hline \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \textbf{CTACCCTAT}\\ \textbf{GAACCGGGATA}\\ \hline \textbf{R} & \textbf{T} & \textbf{F} \\ \hline \textbf{TTATTTCT}\\ \textbf{AARAAAGA}\\ \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \textbf{CACCGFGTC}\\ \textbf{GATGGCCACGGGTA} \\ \hline \textbf{GAACCACGGGTA} \\ \hline \textbf{CTCGCCCT} \\ \hline \textbf{TTATTTCT}\\ \textbf{AARAAAAGA}\\ \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \textbf{CACCGFGTC}\\ \textbf{GAAGCCACGGGTA} \\ \hline \textbf{GAAGCCACGGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline \textbf{GAAGGCGGGTA} \\ \hline \textbf{GAAGGCGGTA} \\ \hline \textbf{GAAGGCGGTA} \\ \hline \textbf{GAAGGCGGTA} \\ \hline \textbf{GAAGGCGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline \textbf{GAAGCGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline \textbf{GAAGCGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline \textbf{GAAGCGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline \textbf{GAAGCGGTA} \\ \hline \textbf{GAAGCGGGTA} \\ \hline $
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8001 -2 8101 -2	$\begin{array}{c} \textbf{Cchoral focus }\\ \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{C} \textbf{C} \textbf{A} \textbf{C} \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{G} \textbf{C} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} G$	CATTICECCG Y L P ACCASCEGCAT TGGTCECCGAT TGGTCECCGAT CTGCCGGCTA CTGCCGCGAT CTGCCGCGAT CACCGGCCGAT CACCGGCCGAC L P R TGACGGCCGCC L P R TGACGGCCGCC CCASTACCGCA ACCGGCCGAT CCGCTAAACCG CCGCTAACCG CCGCTACCG CCGCTAACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCTACCG CCGCCG CCGCTACCG CCGCCGCCG CCGCCG CCGCCG CCGCGCCG CCGCG	$\begin{array}{c c} \textbf{GACGCTTAAA}\\ \textbf{T} \textbf{R} \textbf{I} \textbf{E}\\ \hline \textbf{GAGGCTCGGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCCGGC}\\ \textbf{CTTCGAGGCTG}\\ \textbf{GATACGGAAC}\\ \textbf{CATACGGAAC}\\ \textbf{CATACGCTG}\\ \textbf{S} \textbf{V} \textbf{S} \textbf{R}\\ \hline \textbf{GTGTAAGACG}\\ \textbf{GATACGGAC}\\ \textbf{CACATCTGC}\\ \textbf{CACATTCGCA}\\ \textbf{M} \textbf{L} \textbf{T}\\ \hline \textbf{GCGCGAACGGC}\\ \textbf{GCGCCAACGGC}\\ \textbf{GCGCCAACGGC}\\ \textbf{S} \textbf{R} \textbf{V} \textbf{T} \\ \hline \textbf{T} \textbf{T} \\ \hline \textbf{CACATCGCA}\\ \textbf{S} \textbf{R} \textbf{V} \textbf{T} \\ \hline \textbf{S} \textbf{CCCCCACCGCC}\\ \textbf{S} \textbf{R} \textbf{V} \textbf{T} \\ \hline \textbf{S} \textbf{CCCCCACCGCC}\\ \textbf{S} \textbf{CCCCCACCGCCCCCC}\\ \textbf{S} \textbf{R} \textbf{V} \textbf{T} \\ \hline \textbf{S} \textbf{CCCCCCACCCCCCCCCCC}\\ \textbf{S} \textbf{R} \textbf{V} \textbf{V} \textbf{T} \\ \hline \textbf{S} CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATEGECTAAT TH S M COTTECEAA GCAAACGCTT T	CCGCCGCTCC CAGCCGCTCC CAGCCGCTCC CAGCCGCCTC CTCGGCGCGC CAGCCGCCCC CTCGGCGCGC CCGCGCCAAGG A A A E TCGCCCAAGG A C T CGCGGCGATCC CGGAGCAAGC GCGCTGTTCC CGGAGCAAGC GCGCTGTTCC TCGCGGATAGC GCGCCTACCG CGGACCTGC TTATACGCTGC ATATGCGACG ATATGCGACG	$\begin{array}{c c} \textbf{AACGFICGATT}\\ \textbf{TTGCAGCTAA}\\ \textbf{F} & \textbf{T} & \textbf{S} & \textbf{K}\\ \textbf{CCGCCGAAATC}\\ \textbf{GGCGGTTAGC}\\ \textbf{GGCGTTAGCAT}\\ \textbf{GGCAACGGT}\\ \textbf{GCCTTGCCA}\\ \textbf{GCCTTGCCA}\\ \textbf{GCCTTGCCAT}\\ \textbf{ATAAAACTAA}\\ \textbf{TATTTGGAT}\\ \textbf{TATTTGGAT}\\ \textbf{CAAAAAGGAA}\\ \textbf{GTTTTCCTA}\\ \textbf{CAAAAAGGAA}\\ \textbf{K} & \textbf{k} & \textbf{k} \\ \textbf{GTTTTCCTA}\\ \textbf{CAAAAAGGAAT}\\ \textbf{N} & \textbf{k} & \textbf{K} \\ \textbf{K} & \textbf{K} & \textbf{K} \\ \textbf{GTCTCCTA}\\ \textbf{CGAAAGGGAT}\\ \textbf{N} & \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} \\ \textbf{K} & \textbf{K} \\ \textbf{K} \\ \textbf{K} & \textbf{K} \\ \textbf$	A E N A E N CTCACGATGG GAGTGCTACG E R H GAGTGCTACG GAGTCAATAT CTCACGATAT CTCACGGTATAT S D I ATGAGCACAG C GGCGTTCACC S C C CGCGAAGGGG GCGTTGTCCC R F C CGAACAGGG GCTTTGTCCC R R F TATAAAATTT TATATAAAATTT	$\begin{array}{c} \textbf{GCCTTTTCCC}\\ \textbf{CCGGAAAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CCGTTCAATT}\\ \textbf{GCCAAGTTAA}\\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{GCCATCCAG}\\ \hline \textbf{GCCCTCCAG}\\ \hline \textbf{GCCGTCCTCAG}\\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGCCTCAT}\\ \hline \textbf{A} & \textbf{L} & \textbf{I} \\ \hline \textbf{ACCAACGGTGT}\\ \hline \textbf{ATTGGCCAGA}\\ \hline \textbf{TTGACCGTGT}\\ \hline \textbf{ATTGGTCACAC}\\ \hline \textbf{TGGTGACTGA}\\ \hline \textbf{TGGTGACACTGA}\\ \hline \textbf{TGGTGACTGAC}\\ \hline \textbf{V} & \textbf{L} & \textbf{U} \\ \hline \textbf{CGCCTCACTGAC}\\ \hline \textbf{GCGACTGACTGAC}\\ \hline \textbf{GCGACACTGACTGAC}\\ \hline \textbf{GCGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGACTGAC}\\ \hline \textbf{GCGACTGACTGACTGACTGACTGAC}\\ \hline GCGACTGACTGACTGACTGACTGCCCCCGCCCCCCCCCC$	$\begin{array}{c c} \textbf{GCTTCGCCGC}\\ \textbf{G} & \textbf{E} & \textbf{A} & \textbf{R} \\ \textbf{TGCGCACGTT}\\ \textbf{ACGCGTCGAA}\\ \textbf{R} & \textbf{V} & \textbf{N} \\ \textbf{TGGCGGCGTGA}\\ \textbf{ACGCGTGCAA}\\ \textbf{ACGCGTGCAA}\\ \textbf{CCTAAGGTCAA}\\ \textbf{CCTAAGGTCAGGTGAC}\\ \textbf{G} & \textbf{L} & \textbf{I} & \textbf{V} \\ \textbf{CTTCGGTACTAGT}\\ \textbf{G} & \textbf{L} & \textbf{I} & \textbf{V} \\ \textbf{GGTGTACTGAGT}\\ \textbf{GGTGTACTGAGT}\\ \textbf{GGTGTACTGAA}\\ \textbf{CCACATGACT}\\ \textbf{ACCCCAATGACT}\\ \textbf{T} & \textbf{V} \\ \textbf{CTTCGGTACTGA}\\ \textbf{CCACATGACT}\\ \textbf{R} \\ \textbf{N} & \textbf{L} & \textbf{K} & \textbf{E} \\ \textbf{K} \\$	$\begin{array}{c} \textbf{GCCGCFGCCGGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GGCGCGAAAC}\\ \hline \textbf{CCCGCGTTG}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{AGGCCFGCG}\\ \hline \textbf{TCCGGGACGC}\\ \hline \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAACCCTAT}\\ \hline \textbf{GATCGGGATA}\\ \hline \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \textbf{CTACCCCTAT}\\ \hline \textbf{GAACCGGGTA}\\ \hline \textbf{R} & \textbf{T} & \textbf{T} \\ \hline \textbf{TTATTTTCT}\\ \hline \textbf{AAATAAAAGA}\\ \hline \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \textbf{CATCGGCCACAG}\\ \hline \textbf{GTGGCCACAG}\\ \hline \textbf{G} & \textbf{T} & \textbf{D} \\ \hline \textbf{CATCGCCCACGGTTC} \\ \hline \textbf{GATCGGCCACGGTTC}\\ \hline \textbf{GATCGGCCCCCT} \\ \hline \textbf{GATCGGCCCCCT} \\ \hline \textbf{GATCGGCCCCCCT} \\ \hline \textbf{GATCGGCCCCCCT} \\ \hline GATCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8001 -2 8101 -2 8101	$\begin{array}{c} \textbf{Caccadatice}\\ \textbf{G} \textbf{CatcTacGG}\\ \textbf{G} \textbf{CatcTacGG}\\ \textbf{G} \textbf{CacadGtcG}\\ \textbf{H} \textbf{L} \textbf{L} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{T} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} G$	$\begin{array}{c} \text{CATTICGCCG}\\ \hline \mathbf{Y} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$\begin{array}{c c} \textbf{GACGCTTAAA}\\ \textbf{T} \textbf{R} \textbf{I} \textbf{E}\\ \hline \textbf{GAGGCTCGGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCCGGC}\\ \textbf{CTTCGAGGCTGG}\\ \textbf{GATACGGAAC}\\ \textbf{CATACGGAAC}\\ \textbf{CATACGCTGG}\\ \textbf{S} \textbf{V} \textbf{S} \textbf{R}\\ \hline \textbf{GTGTAAGACG}\\ \textbf{GATACGGAC}\\ \textbf{CACATCTGC}\\ \textbf{CACATTCGCA}\\ \hline \textbf{M} \textbf{L} \textbf{T}\\ \hline \textbf{CATAAGCGTT}\\ \textbf{GATACGCACGGT}\\ \textbf{GATACGCACGGCGCGCTGCC}\\ \textbf{A} \textbf{R} \textbf{V} \textbf{T}\\ \hline \textbf{GCGCGAACGGCCGTATA}\\ \hline \textbf{GCGCGATATA}\\ \hline \textbf{GCGCGATAACGCGACACGCGCGCGCGTAAA}\\ \hline \textbf{GCGCGATAACGCGACACGCGCGCGCGTAAA}\\ \hline GCGCGATAACGCGAACGGCGCGCGCATACACGCGCGCGCG$	CATEGECTAAT TH S M COTTECEAA GCAAACGCTT TGCTGACGC AGCCACTECCA GGTAAACGAC CATTGETG GGTAAACCA CCATTGETG CCATGETG CCATGETG CCATGETG CCATG	CCGCCGCTCC CAGCCGCTCC CAGCCGCTGC CAGCCGCCTC CTCGGCGCGC CAGCCGCCCC CTCGGCGCGC CCGCGGCGCGC CCGCGCGCG	THECAGECTEANT TIGCAGECTAA F T S K CCGCCCAAATC GGCGGTTAG GCGGTTAGCA F F L D CGGAAACGGT GCCTTTGCCA F F F ATAAAACTA TATTITGAAT I F S L GTTTTCCAT CAAAAGGAA K E K CGAAAGGAAT N E K TGCCGCCGCC Q R S S TAACCCCAC	ACCGAAGGCAA A E M CTCACGATGG GAGTGCTACC E R H GAGTCAATAT CTCACGATAT CTCACGATATA CTCACGATATA CTCACGATATA CTCACGTATAT CTCACGTATAT CTCACGTAC CCGCAAGTGG CCCCCAAGTGG CCCCCAAGTGG CCCTTGTCCC R F L T ATAAAATTTT ATATTAAAA L T K GGCATTGACCC	$\begin{array}{c} \textbf{GCCCTTTCCCC}\\ \textbf{CCCGGAAAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CCGTCAATT}\\ \textbf{GCCAAGTTAA}\\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{GCCCTCCAG}\\ \hline \textbf{GCCCTCCAG}\\ \hline \textbf{GCCGCCCCCAG}\\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGCCTCCAT}\\ \hline \textbf{A} & \textbf{L} & \textbf{I} \\ \hline \textbf{ACCAGGGTG}\\ \hline \textbf{ATGGCCGCTGT}\\ \hline \textbf{ATGGCCAGGTG}\\ \hline \textbf{ATGGCCACGTGT}\\ \hline \textbf{ATGGCCACTGA}\\ \hline \textbf{ACCAACTGA}\\ \hline \textbf{AGCCAACTGA}\\ \hline \textbf{GCGTTGACC}\\ \hline \textbf{V} & \textbf{L} & \textbf{Q} \\ \hline \textbf{GCGTGACGTG}\\ \hline \textbf{GCGTGACGTG}\\ \hline \textbf{GCGTGACGTG}\\ \hline \textbf{GCGTGACGTG}\\ \hline \end{array}$	$\begin{array}{c c} \textbf{GCTTCGCCGC}\\ \textbf{G} & \textbf{E} & \textbf{A} & \textbf{R} \\ \textbf{TGCGCCGCGT}\\ \textbf{ACCGCTGCAA}\\ \textbf{R} & \textbf{V} & \textbf{N} \\ \textbf{TGCCGCGCGTGT}\\ \textbf{ACCGCTGCAA}\\ \textbf{ACCGTGCAA}\\ \textbf{CCTAAGGCGCGCACA}\\ \textbf{L} & \textbf{R} & \textbf{Q} \\ \textbf{CCTAAGGCATGA}\\ \textbf{G} & \textbf{L} & \textbf{I} & \textbf{V} \\ \textbf{GGTGTCGGTACTGGTACTGAGTCGAAGGCATGAACA}\\ \textbf{GGTGTACTGATTTTAAGTTGAAGAACA}\\ \textbf{N} & \textbf{L} & \textbf{K} & \textbf{E} \\ \textbf{TCATCTGATTTT}\\ \textbf{AGTCAGACTAAA}\\ \textbf{N} & \textbf{L} & \textbf{K} & \textbf{E} \\ \textbf{TCATCTGATT}\\ GGTGTACTGATTAGTCGATGAACAACAACAACAACAACAACAACAACAACAACAACAAC$	$\begin{array}{c} \textbf{GCGCFGCCGGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GGCGCGAAAC}\\ \hline \textbf{CCGCGCTTG}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} \\ \hline \textbf{A} & \textbf{R} & \textbf{F} \\ \hline \textbf{A} & \textbf{R} & \textbf{F} \\ \hline \textbf{A} & \textbf{GCCTGCG} \\ \hline \textbf{TCCGGGACGC}\\ \hline \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAACCCTAT} \\ \hline \textbf{GAACCGGGATA} \\ \hline \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \textbf{CTAGCCGGATA} \\ \hline \textbf{R} & \textbf{R} & \textbf{T} \\ \hline \textbf{TTTATTTTCT} \\ \hline \textbf{AAATAAAAGA} \\ \hline \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \textbf{CACGGGTTC} \\ \hline \textbf{CACGGGTTC} \\ \hline \textbf{GTAGCCACAG} \\ \hline \textbf{D} & \textbf{T} & \textbf{D} \\ \hline \textbf{CACATACTCA} \\ \hline \textbf{GTGTATGAGT} \\ \hline \textbf{CACATACTCA} \\ \hline \textbf{CACATACTACTCA} \\ \hline \textbf{CACATACTACTCA} \\ \hline \textbf{CACATACTCA} \\ \hline CACATACTAC$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8001 -2 8101 -2 8201 -2	$\begin{array}{c} \textbf{Cacchartec} \\ \textbf{G} \\ \textbf{Catchartec} \\ \textbf{G} \\ \textbf{G} \\ \textbf{Cacchartec} \\ \textbf{G} \\ G$	$\begin{array}{c} \text{CATTICGCCG}\\ \begin{array}{c} \textbf{Y} & \textbf{L} & \textbf{P} \\ \text{ACCAGCGCAT} \\ \text{TGGTCGCGGAT} \\ \text{TGGTCGCGGAT} \\ \text{CTGCCGGCTA} \\ \text{CTGCCGGCTA} \\ \text{CACCGGGCGGC} \\ \begin{array}{c} \textbf{L} & \textbf{P} \\ \text{CACGGCGCGC} \\ \text{CCGCTATACCG} \\ \text{CCGCTATATGGC} \\ \text{CCGCTATATGGC} \\ \text{CCGCTATACCG} \\ \text{CCGCTATACCG} \\ \text{CCGCTATACCG} \\ \begin{array}{c} \textbf{R} \\ \textbf{N} \\ \text{N} \\ \text{N} \\ \text{CGCGCTATACCG} \\ \text{CCGCTAAACCG} \\ \text{CCGCCACCGCA} \\ \text{CCGCCACCGCA} \\ \text{CCGCCACCGCA} \\ \text{CCGCCACCGCA} \\ \text{CCGCCGCAACCG} \\ \text{CCGCCACCGCCA} \\ \text{CCGCCGCAACCG} \\ \text{CCGCCGCCACCGCCACCG} \\ \text{CCGCCGCCACCGCCACCG} \\ \text{CCGCCGCCACCGCCACCG} \\ \text{CCGCCCACCGCCCACCG} \\ \text{CCGCCCACCGCCCACCG} \\ \text{CCGCCCACCGCCCACCG} \\ CCGCCCACCGCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} \textbf{CACCGUTTAAA}\\ T & R & I & E\\ \hline \textbf{GAGEGUTCGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCCGGC}\\ \textbf{CTTCGAGGCTC}\\ \hline \textbf{T} & D & L\\ \hline \textbf{GATACGGAAC}\\ \textbf{CATACGCAAC}\\ \hline \textbf{CACCTTG}\\ \hline \textbf{CACACGUTCTGC}\\ \hline \textbf{CACATCGCCTG}\\ \hline \textbf{CACATCGCCTG}\\ \hline \textbf{CACATCGCC}\\ \textbf{CACATCGCCTGC}\\ \hline \textbf{CACATCGCC}\\ \hline \textbf{CACATCGCC}\\ \hline \textbf{CACATCGCC}\\ \hline \textbf{CCCCCACACGGCCCTGCC}\\ \hline \textbf{A} & R & V & T\\ \hline \textbf{GCCCCGACGCTTAAC}\\ \hline \textbf{CCCCCGCCTTACCCCTGC}\\ \hline \textbf{A} & R & V & T\\ \hline GCCCCCACACGGCCTTACCCCCCCCCCCCCCCCCCCCCC$	CATEGECTAAT T H S M CETTECEAA GCAAACGCTT T Q S TCGCTGACGC AGCGACTGCG E S V G GGTAAACCA GGTAAACCA CCATTGGTG CCATTGC CCATTGC CCATTGGT CCATTGC CCATTGGT CCATTGGT CCATTGC CCATTGGT CCATTGGT CCATTGC CCATTGGT CCATTGGT CCATTGGT CCATTGC CCATTGGT CCATTGGT CCATTGC CCATTGGT CCATTGC CCATTGGT CCATTGC CCATTGGT CCATTGC CCATTGC CCATTGC CCATTGGT CCATTGGT CCATTGGT CCATTGGT CCATTGGT CCATTGGT CCATTGC CCATTGGT CCATTGC CCATTGGT CCATTG	CCGCCGTTCC CAGCCGCTGCC CAGCCGCTGC CAGCCGCCTC CTCGGCGCGC CAGCCGCCCC CTCGGCGCGC CAGCCGCCAAGG A A A E CCGCGCAAGG AGCGGTTCC CGGAGCGAAGC CGGAGCAAGC CGGACAGC CGGACAGC CGGACAGC CGGACAGC CGGACAGC CGGACAGC CGGACAGC CGGACAGC CGGACAGC CGGACAGC CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCACC CGGACACC CGGCGCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCACC CGGCGCAACG CGGCCAACG CGGCCAACG CGGCCAACG CGGCCACGC CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGCGCAACG CGGCCACC CGGGCGAACG CGGCCACC CGGGCGAACG CGGCGCAACG CGGCCACC CGGGCGAACG CGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCACC CGGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCAACG CGGCGCCACC CGGGGCGCACG CGCGCCACC CGGGCGCCACC CGGGCGCCACC CGGGCGCCACC CGGGGCCACC CGGGCCCCC CGGGCCCCC CGGGCCCCC CGGGCCCCC CGGGCCCCC CGGGCCCCC CGGGCCCCCC	$\begin{array}{c c} \textbf{AACCEFICEATT}\\ \textbf{TTGCAGCTAA}\\ \textbf{F} & \textbf{T} & \textbf{S} & \textbf{K}\\ \textbf{CCGCCAAATC}\\ \textbf{GCCGTTAGC}\\ \textbf{GCCGTTAGCA}\\ \textbf{GCGGAAACGGT}\\ \textbf{GCCTTTGCCA}\\ \textbf{GCTTTGCCA}\\ \textbf{TATATAGAACTA}\\ \textbf{T} & \textbf{F} & \textbf{S} & \textbf{L}\\ \textbf{GTTTTCCAT}\\ \textbf{CAAAAAGGAAT}\\ \textbf{K} & \textbf{K} & \textbf{K} & \textbf{I}\\ \textbf{GTTTTCCTA}\\ \textbf{CAAAAAGGAAT}\\ \textbf{N} & \textbf{K} & \textbf{K} & \textbf{K}\\ \textbf{TGGCGGCGCGC}\\ \textbf{Q} & \textbf{R} & \textbf{R} & \textbf{S}\\ \textbf{TGACCGCGCC}\\ \textbf{T} & \textbf{R} & \textbf{L} \\ \end{array}$	$\begin{array}{c} \textbf{A} \subset \textbf{GacGrade} \\ \textbf{A} \subset \textbf{GacGrade} \\ \textbf{A} \subset \textbf{GacGrade} \\ $	$\begin{array}{c} \textbf{GCCTATTCCC}\\ \textbf{CCGCARAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CGGTCAATT}\\ \textbf{GCCATCAAT}\\ \textbf{GCCATCAAGTTAA}\\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{GCCCTCCAG}\\ \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGCCTCAT}\\ \hline \textbf{A} & \textbf{L} & \textbf{I} \\ \hline \textbf{A} & \textbf{CGCCTGT}\\ \hline \textbf{ATGGCCAGCTGT}\\ \hline \textbf{ATGGCCAGTGT}\\ \hline \textbf{ATGGCACACTGA}\\ \hline \textbf{ACCAACTGA}\\ \hline \textbf{GCGACACTGAT}\\ \hline \textbf{GCGCAACTGATA}\\ \hline \textbf{GCGCTGATA}\\ \hline \textbf{GCGTGACTA}\\ \hline \textbf{R} & \textbf{L} & \textbf{L} & \textbf{N} \\ \hline \textbf{K} & \textbf{L} & \textbf{K} \\ \end{array}$	$\begin{array}{c c} GCTTGGCGCC\\ G & E & A & R \\ \hline TGCGCACGTG\\ TGCGCACGTG\\ A & C & A & R \\ R & V & N \\ \hline TGCGCGCGCG\\ A & CCTAGGCGACA \\ \hline A & CCTAGGCGACA \\ \hline L & R & Q \\ \hline CCTAGGCGACA \\ \hline CCTAGGCACGA \\ \hline CCTAGGCATGA \\ \hline CGTGTGTACTGG\\ \hline CCCCATGGCATGA \\ \hline CCCCATGGCATGA \\ \hline CCCCATGGCATGA \\ \hline CCCCATGGCATGA \\ \hline CCCCCCCCCCCCCCCCCC \\ \hline T & V & Q \\ \hline CTTCGTACT \\ \hline TTAGGTTGTTT \\ \hline AATTCGATAT \\ \hline AATTCGATAA \\ N & L & K & E \\ \hline CCTCCCTA \\ \hline CCCCCA \\ \hline CCCCCCCCC \\ \hline CCCCCCCCCCCCC \\ \hline CCCCCCCCCCCCCCCCCC \\ \hline CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \textbf{GCCGCCFGCCG}\\ \textbf{A} \bigcirc & \textbf{R} \\ \hline \textbf{GGCGCGAAAC}\\ \textbf{CCGCGCTTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} \\ \hline \textbf{A} & \textbf{GCCTGCG}\\ \textbf{TCCGGACTTG}\\ \textbf{TCCGGACGGCTT}\\ \textbf{CTAACCCTAT}\\ \hline \textbf{CTAACCCTAT}\\ \textbf{GATCGGGATA}\\ \textbf{L} & \textbf{G} \\ \hline \textbf{T} \\ \textbf{CTCGCCCAT}\\ \hline \textbf{F} & \textbf{R} & \textbf{T} \\ \hline \textbf{TTTATTTCT}\\ \textbf{AAATAAAAGA}\\ \textbf{N} \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \textbf{CATCGGGTGTC}\\ \textbf{GTACCACACAG}\\ \hline & \textbf{D} & \textbf{T} \\ \hline \textbf{CACAGGTGTCG}\\ \hline \textbf{GTGTATCACACAG}\\ \hline \textbf{H} \\ \textbf{V} \textbf{Y} & \textbf{E} \\ \hline \end{array}$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8101 -2 8101 -2 8201 -2 8301	$\begin{array}{c} \textbf{CtCharles} \\ \textbf{G} \\ \textbf{CtCtTACGG} \\ \textbf{G} \\ $	CATTICECCG CATTICECCG T L P ACCASCEGCAT TGGTCECGTA TGGTCECGGAT CTGCCGGCTA CTGCCGGCTA CCGCCGAC L P R TGACGGCGCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCTAAACCG CCGCCGCA S L G Q CAACCGCCAC	$\begin{array}{c c} \textbf{GACGCTTAAA}\\ \hline \textbf{CACGCTTAAA}\\ \hline \textbf{CACGCTTAAA}\\ \hline \textbf{CACGCTAAGGTCCGC}\\ \hline \textbf{CTTCGAGGCGC}\\ \hline \textbf{CTTCGAGGCGC}\\ \hline \textbf{CTTCGAGGCAC}\\ \hline \textbf{CATACGGAAC}\\ \hline \textbf{CATACGGAAC}\\ \hline \textbf{CATACGGAAC}\\ \hline \textbf{CACACGGAACCGCTTG}\\ \hline \textbf{GATACGGAACGGTT}\\ \hline \textbf{GATACGGAACGGCGT}\\ \hline \textbf{GCACAGCGTTGCC}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGCTTTCCCCAA}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGCGCTT}\\ \hline \textbf{CCGCGCCTTGCC}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGCCTTGCC}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGCCTTTCCCCAA}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGCCTTTCCCCAA}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGCCCTTGCC}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGCCCTTTCCCCAA}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGCCCTTTCCCCCAA}\\ \hline \textbf{A} & \textbf{V} & \textbf{V} & \textbf{T}\\ \hline GCGCGCCCCTTTCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATEGECTAAT T H S M COTTECCAA GCAAACGCTT T \rightarrow S M COTTECCAA GCAAACGCT T \rightarrow S S M CGCTAACCCA GCAACGCTGAC GGTAAACCAC CCATTGGTG CCATTGGC CCATTGGTG CCATGGT	CCGCCGTTCC CAGCCGCTGCC CAGCCGCTGC CAGCCGCCTC CGCGCCCTC CGCGCCCAAGG CAGCCGCCAAGG CGCGCCAAGG CGCGCCAAGG CGGGCGAAGC CGGGGCAAGC CGGGCGCACG CGGGCGCACG CGGGCGCACG CGGGCGCACG CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGGCGCGC CGGCGCGC CGGGCGCC CGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGCGCGC CGGGGCGC CGGGCGC CGGGGCGC CGGGGCGC CGGGCGC CGGGGCGC CGGGCGC CGGGCGC CGGGCGC CGGGCGC CGGGCGC CGGGGCGC CGGGCGC CGGGGCGC CGGGGCGC CGGGGCGC CGGGGCGC CGGGCGC CGGGCGCC CGGGGCGC CGGGGCC CGGGGCGC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGCC CGGGCC CGGGCC CGGGGCC CGGGCC CGGGCC CGGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGCC CGGGGCC CGGGGCC CGGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGCC CGGGC CGGCC CGGGCC CC	$\begin{array}{c c} \textbf{AACCECTICATT}\\ \textbf{TTGCAGCTAA}\\ \textbf{F} & \textbf{T} & \textbf{S} & \textbf{K}\\ \textbf{CCGCCAAATC}\\ \textbf{GCCGTTAGCA}\\ \textbf{GCCGTTAGCA}\\ \textbf{GCCTTTGCCA}\\ \textbf{GCCTTTGCCA}\\ \textbf{TATATAGAACTTA}\\ \textbf{T} & \textbf{F} & \textbf{S} & \textbf{L}\\ \textbf{GTTTTCCAT}\\ \textbf{CAAAAAGGAA}\\ \textbf{GTTTTCCTA}\\ \textbf{CAAAAAGGAA}\\ \textbf{K} & \textbf{E} & \textbf{K}\\ \textbf{TGGCGGCGCGC}\\ \textbf{Q} & \textbf{R} & \textbf{K} & \textbf{S}\\ \textbf{TGGCGGCGCGCGC}\\ \textbf{ATTGGCCTCGA}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTTCCG}\\ \textbf{ACCGCCTCCGCCCC}\\ \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \textbf{GCCGCCCCCCCCC}\\ \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \textbf{GCCGCCCCCCCCCCC}\\ \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ \textbf{GCCGCCCCCCCCCCC}\\ \textbf{G} & \textbf{G} & \textbf{G} & \textbf{G} \\ GCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \textbf{AccGaaGCAA}\\ \textbf{A} & \textbf{E} & \textbf{M} \\ \hline \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACCG}\\ \textbf{GAGTGCTACCG}\\ \textbf{GAGTGCTACCATAT}\\ \textbf{CTCAGTTATA}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{ATGAGCACAGG}\\ \textbf{GCTTGTCCC}\\ \textbf{GCGTTCACCGTGTCC}\\ \textbf{CCGCAAGTGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{ATAAAATTTT}\\ \textbf{ATAAAGATCGT}\\ \textbf{GCGTTAAACCG}\\ \textbf{GCGTTGACCGGCAACGGG}\\ \textbf{GCTTGTCCCC}\\ \textbf{CCGCAAACGGG}\\ \textbf{GCTTGTCCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{ATAAGATCGT}\\ \textbf{GGCATTGACCG}\\ \textbf{GCGTTAACCGGCAACGGG}\\ \textbf{GCTTGTCCCC}\\ \textbf{GCGTAACTGGCCCC}\\ \textbf{GCGTAACTGGCCCCCC}\\ GCGTACCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} GCCTATTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} GCTTCGCCGC\\ G & E & A & R \\ TGCGCACGTC\\ TGCGCACGTT\\ ACGCGTGCAA\\ R & \forall & N \\ TAGCCGCGCGAC\\ A & E & N & N \\ TAGCCGGCGACA\\ L & R & Q \\ CTAAGATCA\\ G & L & I & V \\ CTTCGGTACT\\ G \\ G & L & I & N \\ CTTCGGTACT\\ G \\ GGATCTAGT\\ G \\ GGATCACATGACT\\ T & Y & Q \\ TTAAGTTTT \\ AATTCAAAAA\\ N & L & K & E \\ TCATCGATA\\ N & L & K & E \\ T \\ GATGTACTGAT\\ AGTCAAAAA\\ N & L & K & E \\ T \\ CATCGATA\\ N & Q & N \\ TTGGGTGGCG\\ \end{array}$	$ \begin{array}{c} \textbf{GCGCFGCCGGC}\\ \textbf{GCGCGACGGC}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \\ \textbf{GGCGCGAAAC}\\ \textbf{CCGCGCTTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} \\ \textbf{A} & \textbf{GCCTGCG}\\ \textbf{TCCGGACGC}\\ \textbf{TCCGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \\ \textbf{CTACCCTAT}\\ \textbf{GATCGGGATA}\\ \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \\ \textbf{GATCGGCGAT}\\ \textbf{GATCGGCGATA}\\ \textbf{T} & \textbf{K} & \textbf{E} \\ \hline \\ \textbf{GATCGGCGATA}\\ \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \\ \textbf{CATCGGGTGTC}\\ \textbf{GTATCACCACAG}\\ \textbf{H} & \textbf{H} & \textbf{T} \\ \hline \\ \textbf{GATCGGTGTC}\\ \textbf{GTATCACCACAG}\\ \textbf{GTGTATGAGT}\\ \textbf{V} & \textbf{Y} & \textbf{E} \\ \hline \\ \textbf{CCCGCACATG}\\ \hline \\ \textbf{GTGTATGAGT}\\ \hline \\ \textbf{V} & \textbf{Y} & \textbf{E} \\ \hline \\ \textbf{CCCGCAGATG} \\ \hline \end{array} $
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8101 -2 8201 -2 8201 -2 8301	$\begin{array}{c} \textbf{CtCabArbec}\\ \textbf{G} \textbf{CtCTTACGG}\\ \textbf{G} \textbf{CTCTACGG}\\ \textbf{G} \textbf{G} \textbf{CTTCCAGC}\\ \textbf{CACAAGGTCG}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{T} \textbf{CAAGGTCG}\\ \textbf{H} \textbf{G} \textbf{E} \textbf{L} \textbf{L} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{CTTCCATCA}\\ \textbf{CGAAGGTAGT}\\ \textbf{A} \textbf{E} \textbf{M} \textbf{L}\\ \textbf{G} \textbf{CTTGTGCGC}\\ \textbf{AACGACGCG}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{R} \textbf{G}\\ \textbf{AACGACGCG}\\ \textbf{AACGGCGGTGCG}\\ \textbf{AACGGCGGTGCG}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C}\\ \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C}\\ \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{G} \textbf{C} \textbf{C}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{A}\\ \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{A}\\ \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{A}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{A}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{A}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{G} \textbf{A} \textbf{G} \textbf{G} \textbf{A} \textbf{G} \textbf{G} \textbf{A} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{C} \textbf{C} \textbf{C} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} G$	CATTICECCG CATTICECCG TGTCECCGAT TGGTCECGGAT TGGTCECGGAT CTGCCGGCGAT CTGCCGGCGA CTGCCGGCGAT CTGCCGGCGA CAGCGGGGGGG CAGCGGGGGGG CCGCTATACCG CCGCTATACCG CCGCTATACCG CCGCTAACC CCGCTACC CCGCACC CCGCTACC CCGCC CCGCTACC CCGCC CCGCTACC CCGCC CC	$\begin{array}{c c} \textbf{GacGetTTAAA}\\ \hline & \textbf{R} & \textbf{I} & \textbf{E}\\ \hline \textbf{GAGEGTCGGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCGC}\\ \hline \textbf{L} & \textbf{E} & \textbf{A}\\ \hline \textbf{GGTATCAGT}\\ \textbf{GGTATCAGT}\\ \hline \textbf{GATACGGAAC}\\ \hline \textbf{CTATGCTTG}\\ \hline \textbf{S} & \textbf{V} & \textbf{S} & \textbf{R}\\ \hline \textbf{GTGTAAGACG}\\ \hline \textbf{GTATCGGAAC}\\ \hline \textbf{GTATCGCAA}\\ \hline \textbf{GTATCGCAA}\\ \hline \textbf{GTATCGCAA}\\ \hline \textbf{GTATCGCAA}\\ \hline \textbf{M} & \textbf{L} & \textbf{T}\\ \hline \textbf{GCGCGGACGGC}\\ \hline \textbf{GCGCGTTTCC}\\ \hline \textbf{GCGCGCTTTC}\\ \hline \textbf{GCGCGCGTTTC}\\ \hline \textbf{GCGCGCGTTTC}\\ \hline \textbf{GCGCGCGTTTC}\\ \hline \textbf{GTCAAGCCCG}\\ \hline \textbf{A} & \textbf{D} & \textbf{K}\\ \hline \textbf{GTCAAGCCCC}\\ \hline \textbf{GTCAAGCCCC}\\ \hline \textbf{GCGCGCCCCCC}\\ \hline \textbf{GTCAAGCCCCC}\\ \hline \textbf{GTCAAGCCCCC}\\ \hline \textbf{GTCAAGCCCCC}\\ \hline \textbf{GTCAAGCCCCC}\\ \hline \textbf{GTCAAGCCCCC}\\ \hline \textbf{GTCAAGCCCCC}\\ \hline \textbf{GTCCAGCCCCC}\\ \hline \textbf{GTCCAGCCCCC}\\ \hline \textbf{GTCCAGCCCCC}\\ \hline \textbf{GTCCAGCCCCC}\\ \hline \textbf{GTCCAGCCCCC}\\ \hline \textbf{GTCCAGCCCCC}\\ \hline \textbf{GTCCAGCCCCCC}\\ \hline GTCCAGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATEGECTANT T + G + S M CETTTECEAA GCAAACGCTT T + G + S M CETTTECEAA GCAAACGCTT T + G + S M TCGCTGACGCC AGCGACTGCG E + S V - G GGTAAACCAC CCATTGGTG CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGCCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGGCC CCATTGCC CCATTGCC CCATTGGCC CCATTGGCC CCATTGCC CCATTGGCC CCATTGCC CCATTGCC CCATTGCC CCATTGGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTGCC CCATTCGCC CCATTGCC CCATTGCC CCATTGCC CCATTCC CCACTCC CCATTCC CCATTGCC CCATTCC CCATTGCC CCATTGCC CCATGCC CCATTGCC CCATGC	CCGCCGTTCC CAGCCGCTGT CCGCCGTGT CCGCACCTGTT CCGCGCCGTCC CGCCGCCGCGC CGCCGCCGCC CGCCGCCGCCGC CGCGCCGAAGA ACCGGTTCT CCGCGACAAGA ACCGGTTCT CCGCGACAAGA ACCGCTTC CCGCGACAAGA CGCGCGATAGC CGCGGATAGC CGCCGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGCTACC CGCGCCTTC CGCGGCTACC CGCGCCTC CGCGCCACC CGCGCCTC CGCGCCTC CGCGCCTC CGCGCCCCC CGCGCCCCCCCC	$\begin{array}{c c} \textbf{AACCECTCAAT}\\ \hline \textbf{T} & \textbf{T} & \textbf{GCGCCAAATC}\\ \hline \textbf{GCCGCAAATC}\\ \hline \textbf{GCCGAAACGGT}\\ \hline \textbf{GCCGAAACGGT}\\ \hline \textbf{GCCTTTGCCA}\\ \hline \textbf{GCCTTTGCCA}\\ \hline \textbf{TATATTGAAT}\\ \hline \textbf{I} & \textbf{F} & \textbf{S} & \textbf{L}\\ \hline \textbf{GTTTTCCAT}\\ \hline \textbf{CAAAAGGAAT}\\ \hline \textbf{ATAAAAGGAA}\\ \hline \textbf{K} & \textbf{E} & \textbf{L}\\ \hline \textbf{GTTTTCCTA}\\ \hline \textbf{CAAAAGGGAAT}\\ \hline \textbf{N} & \textbf{E} & \textbf{K}\\ \hline \textbf{TGGCGGCGCGGC}\\ \hline \textbf{Q} & \textbf{R} & \textbf{K} & \textbf{S}\\ \hline \textbf{TGGCGGCGCGGC}\\ \hline \textbf{ATATGGCGTCG}\\ \hline \textbf{ACCGCCTTTCG}\\ \hline \textbf{TGGCCGCAAAGC}\\ \hline \textbf{ACCGCCTTCG}\\ \hline \textbf{GGCCGCAAAGCC}\\ \hline \textbf{GCCCTTCG}\\ \hline \textbf{GGCCGCAAAGCC}\\ \hline \textbf{GCCCTTCC}\\ \hline \textbf{GGCCGCAAAGCC}\\ \hline \textbf{GCCCCTTCC}\\ \hline \textbf{GGCCGCCAAAGCC}\\ \hline \textbf{GCCCCTTCC}\\ \hline \textbf{GGCCGCAAAGCC}\\ \hline \textbf{GCCCTTCC}\\ \hline \textbf{GCCCCTTCC}\\ \hline \hline \textbf{GCCCCTTCC}\\ \hline \hline \textbf{GCCCCTTCC}\\ \hline \hline \textbf{GCCCCTTCCC}\\ \hline \hline \textbf{GCCCCTTCC}\\ \hline \hline \textbf{GCCCCCCAAACCC}\\ \hline \hline \textbf{GCCCCCCAAACCC}\\ \hline \hline \textbf{GCCCCCCCAAACCC}\\ \hline \hline \textbf{GCCCCCTTCCC}\\ \hline \hline \textbf{GCCCCCTCCCACCCC}\\ \hline \hline \textbf{GCCCCCTCCCCACCCCC}\\ \hline \hline \textbf{GCCCCCCCCCCCCCCCC}\\ \hline \hline GCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \textbf{ACCGARGCAA}\\ \textbf{A} & \textbf{E} & \textbf{M} \\ \hline \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACCC}\\ \textbf{E} & \textbf{R} & \textbf{H} \\ \hline \textbf{CTCACGATAT}\\ \textbf{CTCACGATATA}\\ \textbf{CCGGTATAT}\\ \textbf{CTCACGTATAT}\\ \textbf{CCGGTATAT}\\ \textbf{CCGGTATAT}\\ \textbf{CCGGTACCC}\\ \textbf{GCGTCACCC}\\ \textbf{CCGCAAGTGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \hline \textbf{ATAAGATCTT}\\ \textbf{ATAAGAACGGC}\\ \textbf{GCGTTGACCC}\\ \textbf{CCGCAAACGGG}\\ \textbf{GCTTGACCC}\\ \textbf{CCGCAAACGGG}\\ \textbf{GCTTGTCCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \hline \textbf{ATAAGATCTT}\\ \textbf{ATAAGATCGT}\\ \textbf{ATAAGACGTCGT}\\ \textbf{ATAAGACGGT}\\ \textbf{ATAAGACGGT}\\ \textbf{ATAAGACGGT}\\ \textbf{ATAAGACGT}\\ \textbf{CGCGTACGCC}\\ \textbf{CCGTAACGGC}\\ \textbf{A} & \textbf{N} & \textbf{V} \\ \hline \textbf{ATAAGATCGT}\\ \textbf{TATTCTACCACCGT}\\ \textbf{ATAAGACGTCGT}\\ \textbf{ATAAGACCGT}\\ \textbf{ATAAGACCGT}\\ \textbf{A} & \textbf{N} & \textbf{V} \\ \hline \textbf{ATAAGATCGT}\\ \textbf{A} & \textbf{N} & \textbf{V} \\ \hline \textbf{ATAAGATCGT}\\ \textbf{A} & \textbf{N} & \textbf{V} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{V} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{V} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{V} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} & \textbf{A} \\ \hline \textbf{A} & \textbf{A} & \textbf{A}$	$\begin{array}{c} \textbf{GCCTTTTCCC}\\ \textbf{CCGCAAAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CGCTCAATT}\\ \textbf{GCCAAGTTAA}\\ \hline \textbf{GCCAAGTTAA}\\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{GCCCTCCAG}\\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGGCTCCAT}\\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{GCGCTCCAG}\\ \hline \textbf{GCGGCCCTAT}\\ \hline \textbf{A} & \textbf{L} & \textbf{I} \\ \hline \textbf{AACCAGCGTGT}\\ \hline \textbf{ATGGCCAGTGT}\\ \hline \textbf{ATGGCACAGTGT}\\ \hline \textbf{AACCAACTGA}\\ \hline \textbf{GGCCACAGTGT}\\ \hline \textbf{GCGATCGACACTGT}\\ \hline \textbf{GCGCACACTGT}\\ \hline \textbf{GCGCACACTGT}\\ \hline \textbf{GCGCACACTGT}\\ \hline \textbf{GCGCTGAGTA}\\ \hline \textbf{GCGCTGAGTA}\\ \hline \textbf{GCGCTGCGCGCTT}\\ \hline \textbf{GCGCCGCGCGCTTATGGCCGCGCTTATGGCCGCGCTT}\\ \hline GCGCCCGCGCCGCTTATGGCCGCGCTTTAGGCCGCGCTTTAGGCCGCGCTTTAGGCCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC$	$\begin{array}{c} \text{GentreGecGec}\\ \textbf{A} & \textbf{E} & \textbf{A} & \textbf{R}\\ \textbf{TGCGCACGTGT}\\ \textbf{ACGCGTGCAA}\\ \textbf{R} & \textbf{V} & \textbf{N}\\ \textbf{TGCGCGCGCGAC}\\ \textbf{ACGCGCGCGAC}\\ \textbf{ACGCGCGCGAC}\\ \textbf{ACGCGCGCGAC}\\ \textbf{ACGCGCGCGAC}\\ \textbf{ACGCGCGCGAC}\\ \textbf{CCTAAGGACGACGA}\\ \textbf{CCTAAGGACGACGACGAC}\\ \textbf{CCTCCGGTACTGGCGCACGAC}\\ \textbf{CCTCCGGTACTGGCGCACGAC}\\ \textbf{CCCAAGGCATGACTTT}\\ \textbf{AGTCGCAAGAA}\\ \textbf{N} & \textbf{L} & \textbf{K} & \textbf{E}\\ \textbf{TGAGTGTTTT}\\ \textbf{AGTCGAAGACTAA}\\ \textbf{N} & \textbf{L} & \textbf{K} & \textbf{E}\\ \textbf{TGAGTGCACGATGACCTGAC}\\ \textbf{M} & \textbf{M} & \textbf{M} & \textbf{M}\\ \textbf{M} & \textbf{M} & \textbf{M} & \textbf{M}\\ TGGGTGGCGCGGCGCGACACCACGGCCCCCCCCCCCCCC$	$ \begin{array}{c} \textbf{GecGacGac}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GecGacGacAac}\\ \hline \textbf{CccGcGatTfG}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} \\ \hline \textbf{A} & \textbf{R} & \textbf{F} \\ \hline \textbf{A} & \textbf{R} & \textbf{F} \\ \hline \textbf{A} & \textbf{R} & \textbf{CccGGatTG}\\ \hline \textbf{TCcGGacGC}\\ \hline \textbf{TCcGGacGC}\\ \hline \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTACCCTAT}\\ \hline \textbf{GATCGGGATA}\\ \hline \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \textbf{GATCGGGATA}\\ \hline \textbf{R} & \textbf{T} & \textbf{K} \\ \hline \textbf{R} & \textbf{T} & \textbf{K} \\ \hline \textbf{R} & \textbf{T} & \textbf{K} \\ \hline \textbf{CacGGGGTC}\\ \hline \textbf{GATCGGGTC}\\ \hline \textbf{GATCGGGTC}\\ \hline \textbf{GATCGGGTC}\\ \hline \textbf{GATCGGCCAT}\\ \hline \textbf{F} & \textbf{R} & \textbf{T} \\ \hline \textbf{F} & \textbf{R} \\ \hline \textbf{T} & \textbf{K} \\ \hline \textbf{R} \\ \hline \textbf{CacCGGTCC}\\ \hline \textbf{GGGTCCCACAG}\\ \hline \textbf{H} & \textbf{T} \\ \hline \textbf{K} \\ \hline \textbf{E} \\ \hline \textbf{CCCCCACAG}\\ \hline \textbf{GGCTCTCC}\\ \hline \textbf{GGCCTCTAC}\\ \hline \textbf{GGGCGCTCTAC} \\ \hline \textbf{GGCCTCTAC} \\ \hline \textbf{GGCTCTAC} \\ \hline \textbf{GGCTTAC} \\ \hline \textbf{GGTTTAC} \\ \hline \textbf{GGTTTACCTAC} \\ \hline $
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8101 -2 8201 -2 8201 -2 8301 -2 8301	$\begin{array}{c} \textbf{CtCtraceGatter} \\ \textbf{GtCtrtaceG} \\ \textbf{GtCtrtaceG} \\ \textbf{GtGtrtccaGC} \\ \textbf{CacaaGGtG} \\ \textbf{GtGtrtccaFCa} \\ \textbf{GtGtrccatca} \\ \textbf{GtGtrcccacGC} \\ \textbf{GtGtcGcGtGCG} \\ \textbf{GtGtcGcGtaCG} \\ \textbf{GtGCGCGTGCG} \\ \textbf{GtGCGCGTGCG} \\ \textbf{GtGCGCGTACG} \\ \textbf{GacGtracta} \\ \textbf{GtGGtCGTACGC} \\ \textbf{GtGGtGCGTACG} \\ \textbf{GacGtracta} \\ \textbf{GtGGtGCGTACG} \\ \textbf{GtGGtGCGTACG} \\ \textbf{GtGGtGCGTACG} \\ \textbf{GtGGtGCAGTAT} \\ \textbf{TCGCGCATTA} \\ \textbf{L} \\ \textbf{D} \\ \textbf{T} \\ \textbf{GtGCGCATTA} \\ \textbf{CtGGCGTACTA} \\ \textbf{CtGCGCATAT} \\ \textbf{CtGCGCCATAT} \\ \textbf{CtGCGCGCATAT} \\ \textbf{CtGCGCGCCATAT} \\ \textbf{CtGCGCGCATAT} \\ \textbf{CtGCGCGCATAT} \\ \textbf{CtGCGCGCCATAT} \\ \textbf{CtGCGCGCCCATAT \\ \textbf{CtGCGCGCCATAT} \\ \textbf{CtGCGCGCCATAT} \\ CtGCGCGCCATAT \\ \textbf{Ct$	$\begin{array}{c} \text{CATTICGCCG}\\ \textbf{Y} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$\begin{array}{c c} \textbf{GacGetTAAA}\\ \hline & \textbf{R} & \textbf{I} & \textbf{E} \\ \hline \textbf{GAGEGTCGGC}\\ \textbf{CTTCGAGGCGCG}\\ \textbf{CTTCGAGGCGGC}\\ \hline \textbf{L} & \textbf{E} & \textbf{A} \\ \hline \textbf{GGTATCCAGT}\\ \textbf{GGTATCGGAAC}\\ \hline \textbf{CTATGGTAG}\\ \hline \textbf{GTATGGGAAC}\\ \hline \textbf{CTATGCTTG}\\ \textbf{S} & \textbf{V} & \textbf{S} & \textbf{R} \\ \hline \textbf{GTATAGGGAC}\\ \hline \textbf{CTATGCTTG}\\ \hline \textbf{GTATAGGCGT}\\ \hline \textbf{GTATAGGCGT}\\ \hline \textbf{GTATAGGCGT}\\ \hline \textbf{GTATAGGCGT}\\ \hline \textbf{GCGCGGAACGG}\\ \hline \textbf{GCGCGGTGCC}\\ \hline \textbf{A} & \textbf{R} & \textbf{V} & \textbf{T} \\ \hline \textbf{GCGCGCGTTGCC}\\ \hline \textbf{A} & \textbf{R} & \textbf{V} & \textbf{T} \\ \hline \textbf{GCGCGGGACGGC}\\ \hline \textbf{CGCGCTTGCC}\\ \hline \textbf{A} & \textbf{A} & \textbf{D} & \textbf{K} \\ \hline \textbf{GTCAAGGCCCG}\\ \hline \textbf{CGGCGTCGCC}\\ \hline \textbf{D} & \textbf{L} & \textbf{A} \\ \hline \textbf{CCGTAGGCCCG}\\ \hline \end{array}$	CATECEGATA CETTECEAA CETTECEAA CETTECEAA CETTECEAA CETTECEAA CETTECEAA CETTECEAA CETEACEC AGCEACTECE E S V G GGTAAACCAT CCATTEGETE CCATTEGETE CATECEACEC CCATTEGETE CATECEACEC CAECEACECEE ACCCATEGETE ACCCACECEE ACCCACECEE ACCCACECEE ACCCACECEE ACCCACECEE ACCCACECEE ACCCACECEE ACCCACECEE ACCCACECEE ACCCACECEE CAECECEE CAECECEE CAECECEE CAECECEE CETCAECCC CETCAECCC	CCGCCGTTCC CAGCCGCTGT CCGCCGTGCC CAGCCGCCTCC GTCGGCGAGG CAGCCGCCAAGG CCGCGCTCC CGGCGCGAGG CCGCGCTTCC CCGGAGCAAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGATAGC CCGCGCTTC CCGGGCAAG CCGCGCTTC CCGGGCAAGC CCCGCCCCCC CCCGCCCCCCC CCCGCCCCCCCC	THECAGE TEATT TIGCAGECTAA F T S K CCGCCCAAATC GCCGGATAC GCCGGTTAG CCGGAAACGGT GCCTTIGCCA F F F ATAAAACTTA TATTITCGAT CAAAAACTA TATTITCGAT CAAAAAGCA GTTITCCTA CAAAAGGAAT N E K TIGCGGCCCCC Q R R S TAACCGCCCCC AACCGCCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{c} \textbf{ACCGARGCAA}\\ \textbf{A} & \textbf{E} & \textbf{N}\\ \hline \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} & \textbf{R} & \textbf{H}\\ \hline \textbf{CTCACGATAT}\\ \textbf{CTCACGATAT}\\ \textbf{CTCACGTATAT}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D}\\ \hline \textbf{ATGAGCACAG}\\ \textbf{GAGTCAATAT}\\ \textbf{CTCGGTATAT}\\ \textbf{S} & \textbf{C} & \textbf{L}\\ \hline \textbf{GGCTTCACC}\\ \textbf{CCGCAAGTGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T}\\ \hline \textbf{CGCAAGGGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T}\\ \hline \textbf{ATAAGACATTT}\\ \textbf{ATAAAAATTTT}\\ \textbf{ATAAAAATTTT}\\ \textbf{ATAAAAATTTT}\\ \textbf{ATAAGATCGT}\\ \textbf{ATATTTAAAA}\\ \textbf{L} & \textbf{L} & \textbf{K}\\ \hline \textbf{GGCATGACCGC}\\ \textbf{ATTTCAGCCA}\\ \textbf{GTATGTCACCC}\\ \textbf{CCGTAACTGGC}\\ \textbf{ATAAGATCGT}\\ \textbf{TATTTTAACA}\\ \textbf{L} & \textbf{L} & \textbf{K}\\ \hline \textbf{GTACGCGCCT}\\ \textbf{TTTCACCGCC}\\ \textbf{CCGCACGCC}\\ \textbf{CCGCCCCCCCC}\\ CCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \textbf{GCCTTTTCC}\\ \textbf{CCGGAAAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CGGTCAATT}\\ \textbf{GCAAGTTAA}\\ \hline \textbf{GCCAAGTTAA}\\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{GCCCTCCAG}\\ \hline \textbf{GGGGGGGGGGGGGGCC}\\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GGGGTCCTAT}\\ \hline \textbf{AACCAGGGTG}\\ \hline \textbf{TTGACCAGGGT}\\ \hline \textbf{TTGACCAGTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{ACCACTGT}\\ \hline \textbf{CGCCACACTGT}\\ \hline \textbf{ACCACTGT}\\ \hline \textbf{CGCCACCACTGT}\\ \hline \textbf{GCGTTGACTA}\\ \hline \textbf{GCGGTGAGTA}\\ \hline \textbf{CGCGACCGCCC}\\ \hline \textbf{TTATGCCGCC}\\ \hline \textbf{TTATGGCCGCC}\\ \hline \textbf{TTATGGCCGCC}\\ \hline \textbf{TTATGGCCCGCC}\\ \hline \textbf{TTATGCCCCCCC}\\ \hline \textbf{TTATGCCCCCCCCCCCCC}\\ \hline TTATGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} GCTTCGCGCC\\ G & E & A & R \\ TGCGCCGCGT\\ ACGCGTGCAA\\ R & \forall & N \\ R & Q & R \\ R & R & Q \\ R & Q & R \\ R & Q & R \\ R & Q & R \\ R & R R \\ R & R \\ R & R \\ R \\ R \\ R & R \\ \mathsf$	$ \begin{array}{c} \textbf{GCGCFGCCGGC}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \\ \textbf{GGCGCGAAAC}\\ \textbf{CCGCGCTTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} \\ \hline \\ \textbf{A} & \textbf{GCCTGCG}\\ \textbf{TCCGGACGCTTG}\\ \textbf{TCCGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \\ \textbf{CTACCCTAT}\\ \textbf{GATCGGGATA}\\ \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \\ \textbf{GAACCGGGTA}\\ \textbf{GATCGGCATA}\\ \textbf{GATCGGCATA}\\ \hline \\ \textbf{F} & \textbf{R} & \textbf{T} & \textbf{F} \\ \hline \\ \textbf{GAACCGGGTA}\\ \textbf{R} & \textbf{T} & \textbf{K} & \textbf{E} \\ \hline \\ \textbf{CACGGGTCCG}\\ \textbf{GTATGACCACAG}\\ \textbf{G} & \textbf{G} & \textbf{S} & \textbf{T} \\ \hline \\ \textbf{CCCCGCACATG}\\ \textbf{G} & \textbf{G} & \textbf{S} & \textbf{T} \\ \hline \\ \textbf{GCCGCGTCCC}\\ \textbf{G} & \textbf{A} & \textbf{S} & \textbf{T} \\ \hline \end{array} $
	-2 7601 -2 7701 -2 7801 -2 7801 -2 7801 -2 7901 -2 8001 -2 8101 -2 8201 -2 8301 -2 8301 -2 8401	$\begin{array}{c} \textbf{CtCabarbec}\\ \textbf{GTCTTACGG}\\ \textbf{GTGTTCCAGC}\\ \textbf{GTGTTCCAGC}\\ \textbf{GTGTTCCAGC}\\ \textbf{GTGTTCCAGC}\\ \textbf{GTGTCCATCA}\\ \textbf{GTGTCCATCA}\\ \textbf{GTGTCCATCA}\\ \textbf{GTGTGCGATGT}\\ \textbf{A} & \textbf{E} & \textbf{M} & \textbf{L}\\ \textbf{TTGTTGGCC}\\ \textbf{GTGCCCACGC}\\ \textbf{AACCAGCGGTGCG}\\ \textbf{AACCAGCGGTGCG}\\ \textbf{AACCGCGGTGCCGAGC}\\ \textbf{GTGCGCGCGTGC}\\ \textbf{GGTCGCGCGTGC}\\ \textbf{GGTCGCGCGTGC}\\ \textbf{GACGTGGCGTAT}\\ \textbf{TCGCGGCGTAA}\\ \textbf{CCGGCGTATA}\\ \textbf{L} & \textbf{D} & \textbf{I} & \textbf{D}\\ \textbf{TCGCGCGTACA}\\ \textbf{GCGCCGTAAC}\\ GGCGCGTAACACAACAACAACAACAACAACAACAACAACAACAACA$	$\begin{array}{c} \text{CATTICGCCG}\\ \textbf{Y} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$\begin{array}{c c} \textbf{GacGetTAAA}\\ \hline & \textbf{R} & \textbf{I} & \textbf{E}\\ \hline \textbf{GAGEGTCGGC}\\ \textbf{CTTCGAGGCGC}\\ \textbf{CTTCGAGGCGC}\\ \hline \textbf{L} & \textbf{E} & \textbf{A}\\ \hline \textbf{GGTATCCAGT}\\ \textbf{GATACGGAAC}\\ \hline \textbf{CATAGGTCA}\\ \hline \textbf{T} & \textbf{D} & \textbf{L}\\ \hline \textbf{GATACGGAAC}\\ \hline \textbf{CTATGCTTG}\\ \textbf{S} & \textbf{V} & \textbf{S} & \textbf{R}\\ \hline \textbf{GTGTAAGACG}\\ \hline \textbf{GATACGGAAC}\\ \hline \textbf{CTATGCTTG}\\ \hline \textbf{GATATCGCAA}\\ \hline \textbf{M} & \textbf{L} & \textbf{T}\\ \hline \textbf{GTGTAAGACG}\\ \hline \textbf{GCGCGGAACGG}\\ \hline \textbf{GCGCGCTTGCC}\\ \hline \textbf{A} & \textbf{R} & \textbf{V} & \textbf{T}\\ \hline \textbf{GCGCGGCATTGCCC}\\ \hline \textbf{GTCCAGGGACCGCC}\\ \hline \textbf{CCGCGGTCGCCC}\\ \hline \textbf{D} & \textbf{L} & \textbf{A}\\ \hline \textbf{CCGTGGAGCCT}\\ \hline \textbf{GCCCGGGACCGG}\\ \hline \textbf{GCCCTCGGA}\\ \hline \textbf{CCGGGGCCTGGAC}\\ \hline \textbf{GCCCTCGGA}\\ \hline \textbf{GCCCGGGACCGG}\\ \hline \textbf{GCCCTCGGA}\\ \hline \textbf{CCGTGGAGCCT}\\ \hline \textbf{GCCCTCGGACCT}\\ \hline \textbf{GCCCTCGGACCC}\\ \hline \textbf{GCCCTCGGACCCC}\\ \hline \textbf{GCCCTCGGACCC}\\ \hline \textbf{GCCCTCGGACCCC}\\ \hline \textbf{GCCCTCGGACCCC}\\ \hline \textbf{GCCCTCGGACCCC}\\ \hline \textbf{GCCCTCGGACCCC}\\ \hline \textbf{GCCCTCGGACCCC}\\ \hline \textbf{GCCCTCGGACCCC}\\ \hline \textbf{GCCCCCGACCCGC}\\ \hline \textbf{GCCCCCGCCCCGACCCC}\\ \hline \textbf{GCCCCCGCCCCCGACCCC}\\ \hline GCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATECEGATA CETTECEAA CETTECEAA CETTECEAA CETTECEAA CETTECEAA CETTECEAA CETTECEAA CETEACEC AGCEACTECE E S V G GGTAAACCAT CCATTEGETE CCATTEGETE CATECEACEC CCATTEGETE CATECEACEC CATEC	CCGCCGTTCC CAGCCGCTCC CAGCCGCTCC CAGCCGCCTC CGCCGCCTCC CGCCGCCGCGC CGCCGCCGCC CGCCGCCGCC CGCGCGCGAGG ACCGCTTCC CCGGGCGAGG ACCGCTTCC CCGGGCCAACG CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGATAGC CGCGGCTTC TATCCGTCG ATAGCGCGC CGCGGCGAC TACCGCGCA ATAGCGCGC CGCGGCGATAGC CGCGGCCTTC CCGGGGGAAA CGCCCTTT CCGGGGGAA CGCCCTTC CCGGGGGAA CCCCTTC CCGGGGGAA CCCCTTCCCCCC CGCGGGGCAA CCCCTTCCCCCC CGCGGGGGAA	THECAGE TEATT TIGCAGECTAA F T S K CCGCCCAAATC GCCGGATTAG CCGCAAACGGT GCCTTIGCCA F F F ATAAAACTTA TATTITGAAT I F S L GTITICCAT CAAAAGCAA TATTITGAAT K E I GTITICCTA CAAAAGGAAT N E K TGGCGGCCGCCC Q R R S TAACCGCAG ACCGCCGCCA ACCGCCGCAAACCGA	$\begin{array}{c} \textbf{ACCGAGGGGAA}\\ \textbf{A} & \textbf{E} & \textbf{N} \\ \hline \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} & \textbf{R} & \textbf{H} \\ \hline \textbf{CTCACGATAT}\\ \textbf{CTCACGATAT}\\ \textbf{CTCAGTATAT}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \hline \textbf{ATGAGCACAG}\\ \textbf{GAGTCAATAT}\\ \textbf{CTCGGTATAT}\\ \textbf{CCGGTATAT}\\ \textbf{CCGGTACGCC}\\ \textbf{CCGCAAGTGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \hline \textbf{CGCATGACGGGGCT}\\ \textbf{CCGAAACGGGGGCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \hline \textbf{ATAAGAATTT}\\ \textbf{ATAAGAATTT}\\ \textbf{ATAAGAATGCC}\\ \textbf{T} & \textbf{TATTCTAGCA}\\ \textbf{A} & \textbf{V} & \textbf{V} \\ \hline \textbf{ATAAGATCGT}\\ \textbf{TATTCTAGCAC}\\ \textbf{I} & \textbf{L} & \textbf{D} & \textbf{Y} \\ \hline \textbf{TCTGGCGGCCT} \\ \textbf{AGACCGCCCGA} \\ \textbf{AGACCGCCCGA}\\ \textbf{AGACCGCCCGA}\\ \textbf{AGACCGCCCC}\\ \textbf{AGACCGCCCCC}\\ \textbf{AGACCGCCCCC}\\ \textbf{AGACCGCCCCCCCC}\\ AGACCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CGCGAAAAGG CCGGAAAAGG CCGGAAGTTA CCGGTCAATT GCCAAGTTA R N L K CGCCTCCAG GCGGCCCCCAG GCGGCCCCCAG GCGGCCCCAG GCGGCCCCAT CGCCAGGAA CGCCAGGAA TGACCAGGCTG TGACCAGCTGA TGGTCAGAC CCCCACGCGC TGGCTGAGTA AACCGCGCC TTATGGCCGCG TTATGGCCGCG TTATGGCCGCG TTATGGCCGCG TAACCGCGCC	$\begin{array}{c c} GCTTCGCGCC\\ G & E & A & R \\ TGCGCACGTT\\ ACGCGTGCAA\\ R & \forall & N \\ R & R & R \\ R & Q \\ R & R & R \\ R & Q \\ R & R \\ R \\ R & R \\ R$	$ \begin{array}{c} \textbf{GCGCCFGCCG}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \\ \textbf{GGCGCGAAAC}\\ \textbf{CCGCGATTG}\\ \textbf{CCGCGCTTTG}\\ \hline \\ \textbf{A} & \textbf{F} & \textbf{F} \\ \hline \\ \textbf{A} & \textbf{GCCTGCG}\\ \textbf{TCCGGACGCT}\\ \hline \\ \textbf{TCGGACGCTTGC}\\ \textbf{TCGGACGCCTAT}\\ \hline \\ \textbf{GATCGGGATA}\\ \hline \\ \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \\ \textbf{GAACCGGGTA}\\ \hline \\ \textbf{GATCGGCTAT}\\ \hline \\ \textbf{GAACCGGGTA}\\ \hline \\ \textbf{GATCGGCTAT}\\ \hline \\ \textbf{GAACCGGGTA}\\ \hline \\ \textbf{GATCGGCTAT}\\ \hline \\ \textbf{GAACCGGGTA}\\ \hline \\ \textbf{GTTATTTTCT}\\ \hline \\ \textbf{AAATAAAAGA}\\ \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \\ \textbf{CACGGGTGTC}\\ \hline \\ \textbf{GTGTATGGGTGT}\\ \hline \\ \hline \\ \textbf{V} & \textbf{Y} & \textbf{E} & \textbf{L} \\ \hline \\ \textbf{CCCCGCAATG}\\ \hline \\ \textbf{GGCGCGTCACG}\\ \hline \\ \hline \\ \textbf{GCACGGTGACG}\\ \hline \\ \hline \\ \textbf{GCACGGTGAC}\\ \hline \\ \hline \\ \hline \\ \textbf{GCACGGTGACG}\\ \hline \\ \hline \\ \hline \\ \textbf{GCGCGGTGAC}\\ \hline \\ \hline \\ \hline \\ \hline \\ \textbf{GCGCGGTGAC}\\ \hline \\ \hline$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8101 -2 8201 -2 8201 -2 8301 -2 8401 -2	$\begin{array}{c} \text{Crectadatice} \\ \text{GTGTTACGG} \\ \text{GTGTTCCAGC} \\ \text{GTGTTCCAGC} \\ \text{GTGTTCCAGC} \\ \text{GTGTCCAGC} \\ \text{GCACAGGTGG} \\ \text{H} \\ $	$\begin{array}{c} \text{CATTICGCCG}\\ \textbf{Y} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$\begin{array}{c c} \textbf{GACGCTTAAA}\\ \hline \textbf{CACGCTTAAA}\\ \hline \textbf{CACGCTTAA}\\ \hline \textbf{CACGCTAAA}\\ \hline \textbf{CACGCTAAGGTCCGC}\\ \hline \textbf{CTTCGAGGCGC}\\ \hline \textbf{CTTCGAGGCAGT}\\ \hline \textbf{CATAGGTCA}\\ \hline \textbf{T} D L \\ \hline \textbf{GATACGGAAC}\\ \hline \textbf{CATAGGTCA}\\ \hline \textbf{CATAGGCAC}\\ \hline \textbf{CATAGGCAC}\\ \hline \textbf{CATAGGCAC}\\ \hline \textbf{CACATCTGC}\\ \hline \textbf{GATACGGAACGGCT}\\ \hline \textbf{GATACGGAACGGCT}\\ \hline \textbf{GCGCGCTTGCC}\\ \hline \textbf{A} D K \\ \hline \textbf{GCGCGGACGCC}\\ \hline \textbf{GCCCGCGACCGG}\\ \hline \textbf{CCGGGGGCCT}\\ \hline \textbf{GCCCGGGACCGG}\\ \hline \textbf{GCCCTCGGA}\\ \hline \textbf{GCCCGGGACCGG}\\ \hline \textbf{GCCCCCGGAGCCT}\\ \hline \textbf{GCCCGGGACCGGC}\\ \hline \textbf{GCCCCCGGACCCGGCC}\\ \hline GCCCCCGGACCCGGCCCCGGACCGGCCCCGGACCCGGCCCCGGAGCCCCGGAGCCCCGGACCCGGGCCCCGGAGCCCCGGGGCCCCGGGGCCCCGGGGCCCCGGGGCCCCGGGG$	CATECEGETAA CETTEGEAA CETTEGEAA CETTEGEAA CETTEGEAA CETTEGEAA CETTEGEAA CETEACEC AGCEACTEGE E S V G GGTAAACCAT CCATTEGETE A L G D CECTACCCA GCGATEGET ACTCTACCAG CETACCCA CCATTEGETE A L G D CECTACCCA CCATTEGETE A CTCACAGAG C D D CECTACCCA CACATEGETAA ATCCTACAGAG C D D CECTCAGCCC CCACCCETAA CCCECCETAA CCCECCETAA CCCECCETAA CCCECCETAA CCCECCETAA CCCECCETAA CCCETCACCC	CCGCCGATCC CAGCCGCTCC CAGCCGCTCC CAGCCGCCTC CGCCGCCTCC CGCCGCCGCGCGCG CGCCGCCGCCTC CGCGCGCG	$\begin{array}{c c c c} \text{THCGAFCATA}\\ \hline T & TGCAGCTAA\\ \hline T & TGCAGCTAA\\ \hline T & TGCAGCTAA\\ \hline CGCGCAAACCGG\\ \hline GCCGTTAGCA\\ \hline & L & D\\ \hline & L & D\\ \hline & CGGAAACGGT\\ \hline GCCTTTGCCA\\ \hline & F & F\\ \hline & TAAAAACTTA\\ \hline T & F & S & L\\ \hline & ATAAAACTTA\\ \hline T & F & S & L\\ \hline & ATAAAACTTA\\ \hline T & F & S & L\\ \hline & ATAAAACTTA\\ \hline T & F & S & L\\ \hline & ATAAAACTTA\\ \hline & TATTTTGAAT\\ \hline & AAAAACTTA\\ \hline & TATTTTGAAT\\ \hline & CAAAAGCAA\\ \hline & F & F & L\\ \hline & CGAAAGCGCCCC\\ \hline & G & S & E\\ \hline & GCCTCCGCAA\\ \hline & G & S & E\\ \hline & GCCCCCGCAA\\ \hline & CGGAGCCCTT\\ \hline & CGGAGCCCTT\\ \hline & CGCACCCCCCA\\ \hline & CGGAGCCCTT\\ \hline & CGCGCCCCCCA\\ \hline & CGCGCCCCCCA\\ \hline & CGCGCCCCCCA\\ \hline & CGGAGCCCTT\\ \hline & L\\ \hline & CCGCACCCCCCA\\ \hline & CGCGCCCCCCCA\\ \hline & CGGAGCCCT\\ \hline & L\\ \hline & L & L\\ \hline &$	$\begin{array}{c} \textbf{ACCGACGTGTT}\\ \textbf{A} & \textbf{E} & \textbf{N} \\ \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} & \textbf{R} & \textbf{H} \\ \textbf{CTCACGATAT}\\ \textbf{CTCACGTATAT}\\ \textbf{CTCAGTATAT}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{ATGAGCACAG}\\ \textbf{GAGTCATAT}\\ \textbf{CCGGTATAT}\\ \textbf{CCGGTATCC}\\ \textbf{CCGCAAGTGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{CGCAAGGGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{ATAAGAATTT}\\ \textbf{ATAAGAATTT}\\ \textbf{ATATATAAAATTT}\\ \textbf{ATAAGATCGT}\\ \textbf{TATTTAAAA}\\ \textbf{L} & \textbf{I} & \textbf{K} \\ \textbf{GGCATGACCGC}\\ \textbf{GCTTGACCC}\\ \textbf{CCGAAACGGG}\\ \textbf{GCTTGACCC}\\ \textbf{CCGAAACGGG}\\ \textbf{GCTTGACCC}\\ \textbf{CCGAAACGGG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{ATAAGATTTT}\\ \textbf{ATAAGATCGT}\\ \textbf{TATTTAAAA}\\ \textbf{L} & \textbf{L} & \textbf{K} & \textbf{K} \\ \textbf{GGCCATGCCCC}\\ \textbf{A} & \textbf{K} & \textbf{K} \\ \textbf{CCGCGCCCCC}\\ \textbf{A} & \textbf{K} & \textbf{K} \\ \textbf{CCCCCCCCCC}\\ \textbf{CCCCCCCCC}\\ \textbf{CCCCCCCCCCCC}\\ CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \textbf{GCCTTTTCC}\\ \textbf{CCGGAAAAGG}\\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CGGTCAATT}\\ \textbf{GCAAGTTAA}\\ \hline \textbf{GCCAAGTTAA}\\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{GCCCTCCAG}\\ \hline \textbf{GGGGCCTCCAG}\\ \hline \textbf{GGGGTCCTAT}\\ \hline \textbf{A} & \textbf{L} & \textbf{L} \\ \hline \textbf{GCGCCTCAG}\\ \hline \textbf{GCGGTCCAG}\\ \hline \textbf{GCGGTCCAG}\\ \hline \textbf{GCGGTCCAG}\\ \hline \textbf{GCGGTCCAG}\\ \hline \textbf{TGACCAGGTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{ACCACTGT}\\ \hline \textbf{ACCACTGT}\\ \hline \textbf{ACCACTGTAGT}\\ \hline \textbf{GCGTTGACTA}\\ \hline \textbf{GCGTTGACTA}\\ \hline \textbf{GCGTTGACTA}\\ \hline \textbf{GCGTGAGTA}\\ \hline \textbf{AACTACCGCC}\\ \hline \textbf{TTATGGCCGCC}\\ \hline \textbf{TTATGGCCGC}\\ \hline \textbf{TAACCGCC}\\ \hline \textbf{TAACCGCC}\\ \hline \textbf{TAACCGCC}\\ \hline \textbf{AACTGCCGCC}\\ \hline \textbf{TAACCGCC}\\ \hline \textbf{AACTGCCGCC}\\ \hline \textbf{AACTGCCGCC}\\ \hline \textbf{AACTGCCGCC}\\ \hline \textbf{AACCGCCCACACCGC}\\ \hline \textbf{AACCGCCCCACACCGCC}\\ \hline \textbf{AACTGCCGCC}\\ \hline \textbf{AACTGCCGCC}\\ \hline \textbf{AACCGCCCCACACCCCCC}\\ \hline AACCGCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c} \textbf{GCGCCTGCCG}\\ \textbf{GCGCGACGGC}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \\ \textbf{GGCGCGAAAC}\\ \textbf{CCGCGCTTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} \\ \textbf{A} \\ \textbf{AGGCCTGCG}\\ \textbf{TCCGGACGC}\\ \textbf{TCCGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \\ \textbf{CTAGCCTAT}\\ \textbf{GATCGGGATA}\\ \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \\ \textbf{GATCGGGCTAT}\\ \textbf{GATCGGGCTA}\\ \textbf{GATCGGGCTA}\\ \textbf{T} & \textbf{K} & \textbf{E} \\ \hline \\ \textbf{GATCGGGTGTC}\\ \textbf{GATCGGGTGTC}\\ \textbf{GATCGGGTGTC}\\ \textbf{GTGTATGAGT}\\ \textbf{V} & \textbf{Y} & \textbf{E} \\ \hline \\ \textbf{CACCGGTTACGGGTGTC}\\ \textbf{GGACCGGGTGAGG}\\ \textbf{GTGTATGAGT}\\ \textbf{V} & \textbf{Y} & \textbf{E} \\ \hline \\ \textbf{CCCCGCACATG}\\ \textbf{GGCGGCTCACG}\\ \textbf{G} & \textbf{A} & \textbf{S} & \textbf{T} \\ \hline \\ \textbf{GCAGCGTGAGG}\\ \textbf{GCGCGCTCACG}\\ \textbf{G} & \textbf{GCGCGCTCACG}\\ \textbf{GCGCGCGACTC}\\ \textbf{H} & \textbf{H} & \textbf{H} \\ \hline \\ \textbf{GCAGCGTGAGGC}\\ \textbf{GTGCACGCTC}\\ \textbf{H} & \textbf{H} & \textbf{H} \\ \hline \end{array} $
	-2 7601 -2 7701 -2 7801 -2 7801 -2 7801 -2 8001 -2 8101 -2 8201 -2 8301 -2 8301 -2 8401 -2 8501	$\begin{array}{c} \text{Crecharter}\\ \text{GTGTTACGG}\\ \text{GTGTTCCAGC}\\ \text{GTGTTCCAGC}\\ \text{CACAGGTGC}\\ \text{GTGTCCAGC}\\ \text{GCACAGGTGAGT}\\ \text{H} \\ H$	CATTICECCE CATTICECCE T L P ACCASCEGAT TGGTCECGAT TGGTCECGAT CTGCCGACTA CTGCCGACTA CACCGGCGAC CACCGGGGCGG CACCGGGCGGC CCATTACCG CCATTACCG CCGCTATACCG CCGCCGCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCGCCA CCGCCCCCCCC CCGCCCCCCC CCGCCCCCCCC CCGCCCCCCCC	$\begin{array}{c c} \textbf{Caccertraa} \\ \textbf{Caccertraa} \\ \textbf{T} \textbf{R} \textbf{I} \textbf{E} \\ \hline \textbf{GAGEGTCGGC} \\ \textbf{CTTCGAGGCGC} \\ \textbf{CTTCGAGGCGC} \\ \textbf{CTTCGAGGCAC} \\ \textbf{CTTCGAGGCAC} \\ \textbf{CTTDCGCTG} \\ \textbf{S} \textbf{V} \textbf{S} \textbf{R} \\ \hline \textbf{GTATCGGAAC} \\ \textbf{CACACTCTGC} \\ \textbf{GTATCGCAA} \\ \textbf{GTATCGCAA} \\ \textbf{GTATCGCAA} \\ \textbf{GTATCGCAA} \\ \textbf{GTATCGCAA} \\ \textbf{GTATCGCAA} \\ \textbf{GTCAAGGCGTT} \\ \hline \textbf{GCGCGGCTGCC} \\ \textbf{A} \textbf{R} \textbf{V} \textbf{T} \\ \hline \textbf{GCGCGCGTGCC} \\ \textbf{A} \textbf{R} \textbf{V} \textbf{T} \\ \hline \textbf{GCGCGCGCTGCC} \\ \textbf{A} \textbf{CCGGGACGCG} \\ \hline \textbf{CCGGGAGCCG} \\ \textbf{GCCGTGAGCCCG} \\ \textbf{GCCGTGAGCCCG} \\ \textbf{GCCGTGAGCCCG} \\ \textbf{GCCGTGAGCCCG} \\ \textbf{GCCGTGAGCCCG} \\ \textbf{GCCGTGTGCCCGA} \\ \textbf{G} \textbf{H} \textbf{A} \textbf{C} \\ \hline \textbf{CCGTGAGCCCG} \\ \textbf{GCCTGTGCCCGA} \\ \textbf{G} \textbf{H} \textbf{A} \textbf{C} \\ \hline \textbf{CCGTGAGCCCCGA} \\ \textbf{G} \textbf{H} \textbf{A} \textbf{C} \\ \hline \textbf{CCGTGAGCCCCGA} \\ \textbf{G} \textbf{H} \textbf{A} \textbf{C} \\ \hline \textbf{CCGTGAGCCCCGA} \\ \textbf{G} \textbf{H} \textbf{A} \textbf{C} \\ \hline \textbf{CCTGTGAGCCCCGA} \\ \textbf{G} \textbf{H} \textbf{A} \textbf{C} \\ \hline \textbf{CCTGTTGCCCCA} \\ \textbf{G} \textbf{C} \\ \hline \textbf{C} \hline \hline \textbf{C} \\ \hline \textbf{C} \\ \hline \textbf{C} \hline \hline \textbf{C} \hline \hline \textbf{C} \\ \hline \textbf{C} \hline \hline \textbf{C} \\ \hline \textbf{C} \hline \hline $	CATECECTATE TH S M CETTTECEAA GCAAACECTT T Q S TCGCTGACGCC AGCGACTGCG E S V G GGTAAACCAC CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATCGCTG CCATCGCG CCATCGCG CCACCCG CCGCCGGCAACCCA CCGCCGGCAACCCA CCGCCGCGCAACCCA CCCCCCCGTA CCCCCCCGTA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCTACCA CCCCCCCCTACCA CCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCTACCA CCCCCCCCCC	CCGCCGCTCC CGCCCGTCC CGCCCGTCC CGCCCCTGT CGCCCCGT CGCCCCCT CGCCGCCGCGC CGCCGCCGCGCGCG	THECAGE TEATT TIGCAGECTAA F T S K CCGCCCAATC GCCGATTAG CCGCAACGGT GCCTTTGCCA F F F ATAAAACTTA TATTTTGAAT I F S L GTTTTCCAT CAAAAGCAAT TATTTTGAAT I F S L GTTTTCCTA CAAAAGCAA TGCGGCGCGCA ACCGCCGCA ACCGCCGCAA G S E GCCTCCGCAA GCGGAGCGCTT A E A I TTTCTGCCG	$\begin{array}{c} \textbf{ACCGATGATAT}\\ \textbf{ACCGATGATAT}\\ \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACCC}\\ \textbf{E} \textbf{R} \textbf{H} \textbf{H} \textbf{CTCACGATAT}\\ \textbf{GAGTGCTACC}\\ \textbf{GAGTGCATAT}\\ \textbf{GAGTGATAT}\\ \textbf{S} \textbf{D} \textbf{I} \textbf{D} \textbf{I} \textbf{D}\\ \textbf{ATGAGCACAG}\\ \textbf{GAGTGATAT}\\ \textbf{GCGTTCACCC}\\ \textbf{CCGAAGTGGTCCC}\\ \textbf{R} \textbf{F} \textbf{L} \textbf{T}\\ \textbf{GGCGTTAAGAGGG}\\ \textbf{GGTTGTCCC}\\ \textbf{R} \textbf{F} \textbf{L} \textbf{T}\\ \textbf{ATAAGATGTT}\\ \textbf{ATAAGATGCC}\\ \textbf{GCGTAGCGCGAACGGG}\\ \textbf{A} \textbf{N} \textbf{V}\\ \textbf{ATAAGATCGT}\\ \textbf{TATTTAAAA}\\ \textbf{L} \textbf{L} \textbf{D} \textbf{Y}\\ \textbf{TCTGGCGGCCT}\\ \textbf{AGACCGCCGAACGGG}\\ \textbf{A} \textbf{N} \textbf{V}\\ \textbf{GGCATGACCGT}\\ \textbf{TATTCTAGCC}\\ \textbf{GGCGTAGCCC}\\ \textbf{G} \textbf{CGCAACGGC}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}\\ \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G} \textbf{G}$	$\begin{array}{c} \textbf{GCCTTTTCC}\\ \textbf{GCCGTTAAAAGG}\\ \hline \textbf{C}\\ \textbf{CCGGAAAAGG}\\ \hline \textbf{C}\\ \textbf{CGCATCAATT}\\ \textbf{GCAAGTTAA}\\ \hline \textbf{GCCATCCAG}\\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGCCTCCAG}\\ \hline \textbf{GCGGCCCTCAG}\\ \hline \textbf{GCGGCCCTCAT}\\ \hline \textbf{CGCCAGGAAGTT}\\ \hline \textbf{A} \ \textbf{L} \ \textbf{L} \ \textbf{L} \\ \hline \textbf{A} \ \textbf{L} \ \textbf{L} \\ \hline \textbf{I} \ \textbf{G} \ \textbf{E} \ \textbf{L} \\ \hline \textbf{CGCCAGGAT}\\ \hline \textbf{TGACCAGTGT}\\ \hline \textbf{ATGGCCACA}\\ \hline \textbf{CGCCATCACAGTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{CGCCACACTGT}\\ \hline \textbf{CGCCACACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{CGCCACACTGT}\\ \hline \textbf{AACCACTGT}\\ \hline \textbf{CGCCACACTGT}\\ \hline \textbf{CGCCACACTGT}\\ \hline \textbf{AACCACCGCCTTTATGGCCGCCG}\\ \hline \hline \textbf{I} \ \textbf{G} \ \textbf{AACCCGCCTTTATGGCCGCG}\\ \hline \hline \textbf{I} \ \textbf{G} \ \textbf{A} \\ \hline \textbf{AACCGCCGCT}\\ \hline \hline \textbf{CGCCACACTGT}\\ \hline \textbf{GAGGTATAAA}\\ \hline \end{array}$	$\begin{array}{c c} GCTTCGCGCC\\ C & CGAAGCCGCG\\ A & E & A & R \\ TGCGCACGTT\\ ACGCGTGCAA\\ R & \forall & N \\ TGCGGCGCGC\\ ACGCTGCA\\ L & R & Q \\ CTAAGATCA\\ CCTAAGATCA\\ G & L & I & V \\ CTCCGGCACA\\ C & CCTAAGATCA\\ G & L & I & V \\ CTTCCGTACT\\ G \\ GGATTCTAGT\\ G \\ CCTAAGATCA\\ CACCGCGCACA\\ C \\ C \\ C \\ T \\ T \\ C \\ GGCTGTACTGATT\\ AGTCGAAAA\\ N & L & K \\ K \\ C \\ CCCCCGCCC\\ ACCCCCCCCCC\\ Q & T \\ A \\ TGGCTGGCGCG\\ ACGCACGCGCG\\ C \\ Q & L \\ CCCCCTGCTC\\ CCCCCCTGCTC\\ CCCCCCCCCCCCC\\ C \\ C \\ C \\ CCCCCCCCCCCCCCC\\ C \\ C \\ C \\ CCCCCCCCCCCCCCC\\ C \\ C \\ C \\ C \\ CCCCCCCCCCCCCCCC\\ C \\ C \\ C \\ C \\ C \\ CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$ \begin{array}{c} \textbf{GCGCGACGGC} \\ \textbf{GCGCGACGGCGAAA} \\ \textbf{CCGCGGCTTTG} \\ \textbf{GCGCGCGTTTG} \\ \textbf{GCGCGCTTTG} \\ \textbf{GCGCGCTTTG} \\ \textbf{GCGCGCTTGCG} \\ \textbf{TCCGGACGC} \\ \textbf{TCCGGACGC} \\ \textbf{GATCGGGATA} \\ \textbf{GATCGGGGTA} \\ \textbf{GACCGGGTA} \\ \textbf{GACCGGGTA} \\ \textbf{GACCGGGTA} \\ \textbf{GACCGGGTA} \\ \textbf{GGCGGGGTC} \\ \textbf{GGTGTAGGGTG} \\ \textbf{GCGCGCGTAG} \\ \textbf{GCGCGCGTAG} \\ \textbf{GCGCGCGTAG} \\ \textbf{GCGCGCGACTC} \\ \textbf{GCTCGACCT} \\ \textbf{GCTCGACCTC} \\ \textbf{GCTCAAGCCC} \\ \textbf{GCTCGACCTC} \\ \textbf{GCTCAAGCCC} \\ \textbf{GCTCGACCCC} \\ \textbf{GCTCGACCCC} \\ \textbf{GCTCGACCCC} \\ \textbf{GCTCGACCCC} \\ \textbf{GCTCGACCCC} \\ \textbf{GCTCGCCCCC} \\ \textbf{GCTCGCCCCC} \\ \textbf{GCTCGCCCCC} \\ \textbf{GCTCGCCCCC} \\ \textbf{GCTCGCCCCC} \\ \textbf{GCTCGCCCCCC} \\ \textbf{GCTCGCCCCCC} \\ \textbf{GCTCGCCCCCC} \\ \textbf{GCTCGCCCCCC} \\ \textbf{GCTCGCCCCCCC} \\ \textbf{GCTCGCCCCCCCCC} \\ \textbf{GCTCGCCCCCCCCCCC} \\ \textbf{GCTCGCCCCCCCCCCCC} \\ \textbf{GCTCGCCCCCCCCCCCC} \\ GCTCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8001 -2 8101 -2 8301 -2 8301 -2 8301 -2 8301 -2 8301	$\begin{array}{c} \text{Chechartec}\\ \text{GTCTTAC66}\\ \text{GTCTTAC66}\\ \text{GTGTTCCA6C}\\ \text{GTGTTCCA6C}\\ \text{GCAA66GTCG}\\ \text{H} \\ \text{E} \\ \text{L} \\ \text{GCTTCCATCA}\\ \text{GCAA66GTCG}\\ \text{GT} \\ \text{GCTTCCATCA}\\ \text{GCAA66GTCG}\\ \text{GT} \\ \text{GTTTGTTGCGC}\\ \text{AACG6GGTCG}\\ \text{GT} \\ \text{GTGCCAC6C}\\ \text{GTGCCCAC6C}\\ \text{GTGCCCAC6C}\\ \text{GTGCCCCAC6C}\\ \text{GTGCCCCAC6C}\\ \text{GTGCGCCCAC6C}\\ \text{CTCGGCCCAC6C}\\ \text{CTCGGCCCAC6C}\\ \text{CTCGGCCCAC6C}\\ \text{CTCGGCCCAC6C}\\ \text{CTCGGCCCAC6C}\\ \text{CTCGGCCATCA}\\ \text{GGGCCGTACA}\\ \text{GGGTCGATAT}\\ \text{TCCCACCCTAA}\\ \text{CTCGGCCCAACC}\\ \text{TCCGGCCCAACC}\\ \text{TCCGGCCCAC6C}\\ \text{CTCGGCCCAC6C}\\ \text{CTCGGCCCAC6C}\\ \text{CTCGGCCATCA}\\ \text{CTCGGCCCAC6C}\\ \text{TCCGCCCACCC}\\ \text{CTCGGCCCACCC}\\ \text{CTCGGCCCACCCC}\\ \text{CTCGGCCCACCCC}\\ \text{CTCGGCCCACCCC}\\ \text{CTCGGCCCACCCC}\\ \text{CTCGGCCCACCC}\\ \text{CTCGGCCCCACCCC}\\ CTCGGCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATTICECCG CATTICECCG T L P ACCAGCGCTA TGGTCGCGGTA CTGCCGGCTA CTGCCGGCTA CTGCCGGCTA CTGCCGCCCC L P R TGACCGGCGGC GTCGCCCCC CAGCGGGCGGC GTCGCCCCC CAGCGGCGCGC GTCGCCCCC CAGCGGCGCGC CCGATAACCG CCGATAACCG CCGATAACCG CCGATAACCG CCGATAACCG CCGATAACCG CCGCTAACCC CCAATCAATCG GTCGCGACA S L G CCGCCACCC A CCGCGCCCC A CCGCGCCCC A CCGCCCCC A CCGCCCCCC A CCGCCCCCCCCC A CCGCCCCCCC A CCGCCCCCCC A CCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{c c} \textbf{CaCcotartial} \\ \hline \textbf{CACcottara} \\ \hline \textbf{CCTTCEAGECG} \\ \hline \textbf{CCTTCEAGECG} \\ \hline \textbf{CCTTCEAGECG} \\ \hline \textbf{CCTTCEAGECG} \\ \hline \textbf{CCTTCEACTG} \\ \hline \textbf{CCTTCCTGC} \\ \hline \textbf{CCTTCCTGC} \\ \hline \textbf{CCTTCCCTG} \\ \hline \textbf{CCCTTCCCCTG} \\ \hline \textbf{CCCTTCCCCC} \\ \hline \textbf{CCCCTCCCCGC} \\ \hline \textbf{CCCCTCCCCGGACCGG} \\ \hline \textbf{CCCCTCTGCCCC} \\ \hline \textbf{CCCCTCGCCC} \\ \hline \textbf{CCCCTCCCGGACCGG} \\ \hline \textbf{CCCCTCCGGACCCGG} \\ \hline \textbf{CCCCTCCGGACCCGG} \\ \hline \textbf{CCCCTCCGGACCGGG} \\ \hline CCCCTCCGGACCGGGGACCGGGCCCCCGGACACCGGG \\ \hline \textbf{CCCCTCCGGACACCGGCCCCCGGGACACCGGGCCCCCGGGACACCGGGCCCCCGGACCGGGCCCCCC$	CATEGECTAT GTACCGAGATA GCATACGAACGCTT T \rightarrow \rightarrow s M CGTTGCGAA GCAAACGCTT T \rightarrow \rightarrow s TCGCTGACGCC AGCGACTGCG CCATTGGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTCA CCATTGGTG CCATTGGTCA CCATTGGTCA CCATTGGTCA CCATTGGTCA CCATTGGTCA CCATTGGTCA CCATCGACCC CCGTCAGACC CCGTCAGCC CCGTCAGCC CCGTCAGCC CCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA	CCGCCGTTCC CAGCCGCTCC GCGCCGTGACAA R V Q CAGCCGCTCC GTCGGCGAGA A A E CCGCCGCCTC GTCGGCGAGG ACCCGTTCC CCGGGCGAGG ACCCGTTCC CCGGGCAAGC GCCCGTTCC GCCGGATAGC GCCGCATAGC GCCGGATAGC GCCGGATAGC GCCGGATAGC GCCGGATAGC GCCGGATAGC GCCGGATAGC GCCGGCTACC GCCGGATAGC GCCCCTTC CCGGGGGAA ATATCGCGCG ATATCGCGCG ATATCGCGCG ATATCGCGCG CCGAAGCCTTC CCCGAGGCGAA I L L P GCCCCTTC CCCGGGGAA CCCCGTTCCCCCC GCCGGGCAAGC GCCCCTTC CCCGAGGCGAA CCCCTTCCCCCC CCGAAGCCGC S E A GCCCCTTCCCCCC CCGAAGCCGC S E A CCCGGGCGAAGCC GCCCATCCCTC CCGAGGCCGCC CCGAAGCCGC S E A CCCGACCTCCCCC CCGAAGCCGC S E A CCCCCCTCCCCCCCCCCC CCGAAGCCGC S E A CCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TIGCAGCTAA F T S K CCGCCAAATC GGCGGATTAG GGCGGTTAG GGCGGTTAG GGCGAACGGT GCCTTGCCA P F P ATAAAACTA TATTTTGAAT I F S L GTTTTCGAT CAAAAGCAA GTTTTCTGAT CAAAAGCAA GTTTTCTAA CAAAAGCAACGGC GTTTCCTAA CAAAAGCAACGGC G S E GCCTCCGCAA ACCGCCGCCA GGCGGAAGCGC G S E GCCTCCGCAA ACCGCCGCCA CGGGGGCGCGT A CGGGGGGCGTT A E A I TTCTTGCCG CGGAAGCGCT A E A I TTCTTGCCG CGGAAGACGCGC	$\begin{array}{c} \textbf{ACCGACTGUT}\\ \textbf{A} & \textbf{E} & \textbf{N} \\ \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} & \textbf{R} & \textbf{H} \\ \textbf{CTCACGATAT}\\ \textbf{CTCAGTATAT}\\ \textbf{CTCAGTATAT}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{ATGAGCACAG}\\ \textbf{GAGTCATAT}\\ \textbf{CTCAGTATAT}\\ \textbf{CTCAGTATAT}\\ \textbf{CCGATGTGTC}\\ \textbf{C} & \textbf{C} & \textbf{L} \\ \textbf{GGCGTTGTCC}\\ \textbf{CCGAACGGGGCTTGTCCC}\\ \textbf{CGCAACGGGGCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{CGCAACGGGGCTTGACCC}\\ \textbf{CCGAACAGGG}\\ \textbf{GCTTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{GGCATGACCGGCCCC}\\ \textbf{CCGAACGGGC}\\ \textbf{ATAGATGTT}\\ \textbf{ATAGAATGTT}\\ \textbf{TATTTTAAAA}\\ \textbf{L} & \textbf{L} & \textbf{L} & \textbf{M} & \textbf{Y} \\ \textbf{GGCATGGCCCC}\\ \textbf{CCGTAACTGGC}\\ \textbf{I} & \textbf{L} & \textbf{D} & \textbf{Y} \\ \textbf{TCTGGCGCCT}\\ \textbf{GGCGATGGCCCCCC}\\ GGCGATGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$ \begin{array}{c} \textbf{GCCTTTTCC} \\ \textbf{CCGGAAAAGG} \\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CGGGTAAATT} \\ \textbf{GCCAAGTTAA} \\ \hline \textbf{GCCAAGTTAA} \\ \hline \textbf{GCCACTCAG} \\ \hline \textbf{GGCGTCCAG} \\ \hline \textbf{GGCGTCCAT} \\ \hline \textbf{AACTGCGACA} \\ \hline \textbf{TGACCAGGTAT} \\ \hline \textbf{AACTGCGTCAC} \\ \hline \textbf{TGGTCACA} \\ \hline \textbf{GCGTTGAGT} \\ \hline \textbf{AACTGCGGCG} \\ \hline \textbf{TGGTTGACT} \\ \hline \textbf{R} & \textbf{L} & \textbf{E} \\ \hline \textbf{M} \\ \hline \textbf{GCGTTGAGT} \\ \hline \textbf{R} \\ \hline \textbf{L} & \textbf{G} \\ \hline \textbf{A} \\ \hline \textbf{CGCCACGCCAC} \\ \hline \textbf{TTGGTCGCC} \\ \hline \textbf{TGGTTGACT} \\ \hline \textbf{AACTGCGGCG} \\ \hline \textbf{TATGGCGCGC} \\ \hline \textbf{TTATGGCGCG} \\ \hline \textbf{TATGGCGCGC} \\ \hline \textbf{TATGGCGCG} \\ \hline \textbf{TATGGCGCG} \\ \hline \textbf{TATGGCGCG} \\ \hline \hline \textbf{TATGGCGCG} \\ \hline \hline \textbf{TGCTTGACTA} \\ \hline \textbf{GGGTTTAAG} \\ \hline \textbf{CCCCACATGT} \\ \hline \textbf{ATTGGCCGCC} \\ \hline \hline \textbf{TATGGCGCG} \\ \hline \hline \textbf{TATGGCGCG} \\ \hline \hline \textbf{TATGGCGCG} \\ \hline \hline \textbf{TGCTTGATTA} \\ \hline \textbf{CCCCACATAGT} \\ \hline \textbf{CCCCACATAGT} \\ \hline \textbf{CCCCACATAGT} \\ \hline \textbf{CCCCACATGCTG} \\ \hline \hline \textbf{CCCCACATGCTG} \\ \hline \hline \textbf{CCCCACATGCTG} \\ \hline \hline \textbf{CCCCACATGCTG} \\ \hline \hline \textbf{CCCCACATAGT} \\ \hline \textbf{CCCCACATAGT} \\ \hline \hline \textbf{CCCCACATGCTG} \\ \hline \hline \textbf{CCCCACATGCTG} \\ \hline \hline \textbf{CCCCACACGCGC} \\ \hline \hline \hline \textbf{CCCCACATGCTG} \\ \hline \hline \hline \textbf{CCCCACATGCTG} \\ \hline \hline \hline \hline \textbf{CCCCACACGCGC} \\ \hline \hline \hline \hline \hline \textbf{CCCCCACACCGCC} \\ \hline \hline \hline \hline \hline \textbf{CCCCCACACCGCC} \\ \hline \hline \hline \hline \hline \hline \hline \textbf{CCCCCACACCGCC} \\ \hline $	$\begin{array}{c c} GCTTCGGCGC\\ A & E & A & R\\ \mathbf{TGCGCGCGCGT}\\ A & CGAAGCCGTGT\\ A & CGCGTGCAA\\ R & V & N\\ R & R & R\\ R & CGCTGTCGGCG\\ R & CCCAAGACAA\\ N & L & K & E\\ R & CCCACGCGCCGC\\ Q & R & R\\ N & L & K & R\\ R & CCCCGCCCCC\\ Q & R & R\\ R & CCCCGCCCCC\\ Q & Q & A & R\\ R & R & CCCCGCCCCCCCCCC\\ Q & Q & A & R\\ R & R & CCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$ \begin{array}{c} \textbf{GCGCFGCCGGC}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GGCGCGAAAC}\\ \textbf{CCGCGACTTG}\\ \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{A} & \textbf{GCCTGCGCTTG}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{A} & \textbf{GCCTGCGC}\\ \hline \textbf{TCCGGACGC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTACCCTAT}\\ \textbf{GAAGCGGGTA}\\ \hline \textbf{L} & \textbf{G} & \textbf{T} \\ \hline \textbf{GATGGGCGATA}\\ \hline \textbf{CTCGCCCAT}\\ \hline \textbf{GAAGCGGGTA}\\ \hline \textbf{CTCGCCCAT}\\ \hline \textbf{GAAGCGGGTA}\\ \hline \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \textbf{CACGGGTGTC}\\ \hline \textbf{GAAGCGGGTA}\\ \hline \textbf{GATGGGCGCAT}\\ \hline \textbf{CTTGATTTCT}\\ \hline \textbf{AATAAAAGA}\\ \hline \textbf{N} & \textbf{I} & \textbf{K} & \textbf{E} \\ \hline \textbf{CACGGTGTC}\\ \hline \textbf{GTACCCACAG}\\ \hline \textbf{D} & \textbf{T} & \textbf{D} \\ \hline \textbf{CCCCGCAGATG}\\ \hline \textbf{GTACCCACAG}\\ \hline \textbf{G} & \textbf{T} & \textbf{S} & \textbf{T} \\ \hline \textbf{GCAGCGTGACG}\\ \hline \textbf{G} & \textbf{CCCCCCACAG}\\ \hline \textbf{G} & \textbf{T} & \textbf{S} & \textbf{T} \\ \hline \textbf{GCCGCGCAATG}\\ \hline \textbf{GGCGCGTAACG}\\ \hline \textbf{GCCGCGCACTC}\\ \hline \textbf{L} & \textbf{T} & \textbf{L} \\ \hline \textbf{GCTTAATCCC}\\ \hline \textbf{GCTTAATCCC}\\ \hline \textbf{GCTTAATCCCG}\\ \hline \textbf{GCTTAATCCCC}\\ \hline \hline \textbf{GCTTAATCCCC}\\ \hline \textbf{GCTTAATCCCC}\\ \hline \hline \hline \textbf{GCTTAATCCCC}\\ \hline \hline \hline \textbf{GCTTAATCCCC}\\ \hline \hline \hline \hline \textbf{GCTTATTCCCC}\\ \hline \hline$
	-2 7601 -2 7701 -2 7801 -2 7901 -2 8001 -2 8001 -2 8101 -2 8301 -2 8301 -2 8301 -2 8401 -2 8401 -2 8501	$\begin{array}{c} Choman focus of the construction of t$	CATTICECCG CATTICECCG CATTICECCG CATTICECCG CATTICECCG CATTICECCG CACCAGCGCTA CTGCCGCGAT CTGCCGGCGA CAGCGGGCGGG GTCGCCCCCC CAGCGGGCGGG GTCGCCCCCC CACCAGCGCC CCAATAACCG CCAATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATACCGCC CACCGCCACCC CACCGCCCCC CACCGCCCCC CACCGCCCCC CACCGCCCCC CCAATCACCGC CCCATACCG CCCCCCCCC CCCACCCCCC CCCATCCCCCC CCCCCCCCC CCCATCCCCCCC CCCCCCCCC CCCATCCCCCCC CCCATCCCCCCC CCCCCCCCCC	$\begin{array}{c} \textbf{GacGetTTAAA}\\ \textbf{T} \textbf{R} \textbf{I} \textbf{E}\\ \textbf{GAGEGTCGGC}\\ \textbf{CTTCGAGGCG}\\ \textbf{CTTCGAGGCG}\\ \textbf{GGTATCGAGTC}\\ \textbf{GGTATCGGGAAC\\ \textbf{CTATGCCTTG}\\ \textbf{S} \textbf{V} \textbf{S} \textbf{R}\\ \textbf{GTGTAAGGCAT}\\ \textbf{GTGTAAGGCAT}\\ \textbf{GTGTAAGGCAT}\\ \textbf{GTGTAAGGCGTGCC}\\ \textbf{GTGTAAGGCGTGCC}\\ \textbf{GTGTAAGGCGTGCC}\\ \textbf{GTCAAGGCGTTGCC}\\ \textbf{GTCAAGCGCTGGAC\\ \textbf{A} \textbf{R} \textbf{V} \textbf{T} \textbf{G}\\ \textbf{GCGCGTGTGCCC}\\ \textbf{GTCAAGCGCTGGAC}\\ \textbf{GTCAAGCGCTGGAC}\\ \textbf{GTCAAGCGCTGGC}\\ \textbf{GTCAAGCGCCGGACCGG}\\ \textbf{GCGCTGTGCCCGGC}\\ \textbf{G} \textbf{L} \textbf{A} \textbf{D} \textbf{K}\\ \textbf{GTCAAGCGCTGGAC}\\ \textbf{GCGCTGTGCCGGC}\\ \textbf{G} \textbf{L} \textbf{A} \textbf{D} \textbf{K}\\ \textbf{GTCAAGCGCGG}\\ \textbf{G} \textbf{H} \textbf{A} \textbf{D} \textbf{K}\\ \textbf{GTCAAGCGCGG}\\ \textbf{GCGCTGTGCCGGC}\\ \textbf{GCGCTGGCCCGGACCGGG}\\ \textbf{GCGCTGGCCCGGACCGGG}\\ \textbf{GCGCTGTGCCCG}\\ \textbf{GCGCACCGGGACGGGG}\\ \textbf{G} \textbf{H} \textbf{A} \textbf{D} \textbf{K}\\ \textbf{GCGGCACCGGG}\\ \textbf{GCCCTGTGCCCG}\\ \textbf{GGCACCCGGAACCGGG}\\ \textbf{GCCCTGTGCCCC}\\ \textbf{GGCACACCGGGC}\\ \textbf{G} \textbf{H} \textbf{A} \textbf{D} \textbf{K}\\ \textbf{GTCAAGCCTCGAACGGG}\\ \textbf{GTCAAGCGCCTGGAACCGGC}\\ GCCCTGTGCCCCGGAACCGGGCCCCGGAACCGGGCCCCGGAACCGGGGCCCCCGGAACCGGGCCCCCGGAACCGGGCCCCCGGAACCGGGCCCCCGGAACCGGGCCCCCC$	CATEGECTAT T G S M CGTTGCGAACGCTT T G S M CGTTGCGAACGCTT T G S M CGTTGCGACGCC AGCGACTGCG E S V G GGTAACCAC CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATCGACCA CCATCGACCA CCATCGACCA CCATCGACCA CCATCGACCA CCATCGACCA CCGTCAGCAC CCGTCAGCCA CCGTCAGCCC CCGCCGTA CCGTCGGCCA CCCCCCCGTA CCGTCGGCCA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCGTCGACCA CCCGCCGTA CCGTCGACCA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCGTCGACCA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGAACGC CCCGCCGAACGC CCCGCCGAACGC CCCGCCGTA CCCGCCGAACGC CCCGCCGAACGC CCCGCCGAACGC CCCGCCGAACGC CCCGCCGTA CCCGCCGAACGC CCCGCCGAACCA CCCGCCGTA CCCGCCGTA CCGTCGACCA CCCGCCGAACGC CCCGCCGAACCCA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGAACCCA CCCGCCGAACCCA CCCGCCGAACCCA CCCGCCGCAACCCA CCCGCCGTA CCCGCCGTA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCGCAACCA CCCGCCCGCAACCA CCCGCCGCAACCA CCCGCCCGCCAACCA CCCGCCCGCAACCA CCCGCCCGCAACCA CCGCCCGCAACCA CCGCCGCAACCA CCGCCGCAACCA CCGCCGCAACCA CCGCCCGCAACCA CCGCCGCAACCA CCGCCGCAACCA CCGCCGCAACCA CCGCCGCAACCA CCGCCGCAACCA CCGCCCGCAACCA CCGCCCGCCAACCA CCGCCCGCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCCAACCA CCGCCCGCAACCA CCGCCCGCCAACCA CCGCCCCCCCCCC	CCGCCGATCC CAGCCGCTCC GCGCCGTCC GCGCCGTGACAA R V V CAGCCGCCTC GTCGGCGAGA A A CGGCCGCTCC GCGCGCAAGA A CCCGCGTTCC CCGGACAAGA A CCCGGTTCC CCGGACAAGA A CCCGGTTCC CCGGACAAGA A CCCGGTTCC GCCGGCTAGC GCGCCATGC GCGCGATAGC GCGCGATAGC GCGCGATAGC GCGCGATAGC GCGCGATAGC GCGCGATAGC GCGCGATAGC GCGCGCTTC CCGGGGGAAA I L L D G GCTCGTCA A TATACGCTGC GCCCGCTTC CCGGGGGAA A TATGAGCAGC GCCCTTT CCCGGGGGAA A CCCGGGGCAAGC GCCCCTTT CCCGGGGGAA A CCCGGGGCAAGC GCCCTTC CCGAAGGCGG S S E A CCCAAGGCGA A CCCGGGCAAGC CCGAAGGCG S CCCAAGGCCA CCCGAAGGCGG S S E A CCCAAGCCTTT CCCGGGGGAA A CCCTCTTC CCCAAGGCCA A CCCGGCCAACC CCCGCA CCCGCC CCCGCACC CCCGCC CCCGCACC CCCGCAACC CCCGGCAAGC CCCGCCC CCCGCACC CCCGCACC CCCGCACC CCCGCACC CCCGCACC CCCGGCCACC CCCGGCCACC CCCGGCCACC CCCGGCCACC CCCGGCCACC CCCGGCCACC CCCGCCC CCCGGCCACC CCCGCCCC CCCGCCCCC CCCGCCCC CCCGCCC CCCGCCCCCC	TIGCAGCTAA F T S K CCGCCAAATC GGCGGATAGC GGCGATAGC GGCGAACGGT GCCTTGCCA F F P ATAAAACTTA TATTTTGAAT I F S L GTTTTCCAT CAAAAGCAAT R E I GTTTTCTGAT CAAAAGCAAT R E K GTTTCCTAA CAAAAGCAAT R E K GTTTCCTAC CAAAAGCAAT R E K GTTTGCCGCA ACCGCCGCCC G S E GCCGCCGCCA CGCGAGCCCA CGCGAGCCCA	$\begin{array}{c} \textbf{TCGCCTCGTT}\\ \textbf{TCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} & \textbf{R} & \textbf{H} \\ \hline \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{GAGTCATAT}\\ \textbf{TCCAGTTATA}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{GAGTCATAT}\\ \textbf{TCACGGTTATA}\\ \textbf{S} & \textbf{D} & \textbf{I} & \textbf{D} \\ \textbf{ATGAGCACAG}\\ \textbf{GGCGTTCACC}\\ \textbf{CGCAAGTGG}\\ \textbf{CCGAAGTGG}\\ \textbf{GCCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{GGCGATGACCG}\\ \textbf{GCTTGTCCC}\\ \textbf{R} & \textbf{F} & \textbf{L} & \textbf{T} \\ \textbf{ATAAGAATTT}\\ \textbf{TATTTTATAAA}\\ \textbf{L} & \textbf{I} & \textbf{K} \\ \textbf{GGCATTGACC}\\ \textbf{GCCATGACCG}\\ \textbf{GCTTGACCGC}\\ \textbf{GCCATGACCG}\\ \textbf{GCCATGACCGC}\\ \textbf{GCCCCCCAACTGGC}\\ \textbf{GCCATGACCGC}\\ \textbf{GCCATGACCGC}\\ \textbf{I} & \textbf{L} & \textbf{D} & \textbf{Y} \\ \textbf{GGCCATGGCCCC}\\ GCCGATGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$ \begin{array}{c} \textbf{GCCTTTTCC} \\ \textbf{CCCGARAAGG} \\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CGCTTCAATT} \\ \textbf{GCCAAGTTAA} \\ \hline \textbf{GCCAAGTTAA} \\ \hline \textbf{GCCATCCAG} \\ \hline \textbf{GGCCTCCAG} \\ \hline \textbf{GGCGTCCAG} \\ \hline \textbf{GGCGTCGAGT} \\ \hline \textbf{GGCGGCCGC} \\ \hline \textbf{GGGGTATAG} \\ \hline \textbf{GGCGGCCATT} \\ \hline \textbf{GGCGGCCATT} \\ \hline \textbf{GGCGGCCATT} \\ \hline \end{array} $	CCTAGCGCGCG A E A R TGCGCGCGCG A E A R TGCGCGCGCGT ACGCGTGCA R V N ACGCGTGCA ACGCGCGCGT ACGCGCGCGC CCTAGGCGACA L R Q CCTAGGCACA CCTAGGCACA G L I V CTTCCGTACT GAAGCATGA CCCAGGCACA CCCAGGCACA CCCAGGCACA CCCACGCCGC C T Y Q CTTCCGTACT GAAGCATGA CCCACAGGCACA N L K E TCACCTGGTC TTAGGTTAT TGGGTGGCG ACCCCCCCCC Q T A TGCGTGCGCG ACCCCCCCCC C C G AGCACCCC Q Q A R ACGCACCGCC C CCCAGGCCGCGCC C C C AGCACCCCCCC C C C AGCCCCCCCCC C C C C C C C C C C C C AGCACCCCCCCCCC	$ \begin{array}{c} \textbf{GCCGCTACCG} \\ \textbf{GCCGCGACGC} \\ \textbf{GCCGCGACTTG} \\ \textbf{GCCGCGCTTG} \\ \textbf{GCCGCGCTTG} \\ \textbf{GAGCCGGCTTG} \\ \textbf{GAGCCGGCTTG} \\ \textbf{GAGCCGGGTA} \\ \textbf{GATGCGGGTA} \\ \textbf{GAGCGGGTA} \\ \textbf{GGTGGGGTG} \\ \textbf{GGGGGGTA} \\ \textbf{GGGGGGTA} \\ \textbf{GGGGGGTA} \\ \textbf{GGGGGGTA} \\ \textbf{GGGGGGTA} \\ \textbf{GGGGGGGTA} \\ \textbf{GGGGGGTA} \\ \textbf{GGGGGGGGA} \\ GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG$
	-2 7601 -2 7701 -2 7801 -2 7801 -2 8001 -2 8001 -2 8101 -2 8301 -2 8301 -2 8301 -2 8301 -2 8401 -2 8501 -2 8601	$\begin{array}{c} \text{Cacabar bec}\\ \text{GTGTTAC66}\\ \text{GTGTTAC66}\\ \text{GTGTTAC66C}\\ \text{CACAA6GTCC}\\ \text{H} \\ \text{E} \\ \text{L} \\ \text{GTTCCATCA}\\ \text{GCACA6GTCC}\\ \text{A} \\ \text{CGAA6GTCAGTC}\\ \text{A} \\ \text{CGAA6GTCAGTC}\\ \text{A} \\ \text{CGAACACCCCG}\\ \text{T} \\ \text{Q} \\ \text{Q} \\ \text{A} \\ \text{CGACCACCCC}\\ \text{GTCCCACCC}\\ \text{GTCCCACCC}\\ \text{GTCCCACCC}\\ \text{GTCCCACCC}\\ \text{GTCCCCCCCC}\\ \text{GTCGCCCACG}\\ \text{C} \\ \text{L} \\ \text{A} \\ \text{CTGCCCCCCC}\\ \text{GTCGCCCACGC}\\ \text{C} \\ \text{C} \\ \text{C} \\ \text{CCCGCCCCACGC}\\ \text{C} \\ \text{C} \\ \text{CCCGCCCCACGC}\\ \text{C} \\ \text{C} \\ \text{CCCGCCCCACGC}\\ \text{C} \\ \text{C} \\ \text{C} \\ \text{CCCGCCCCCCCCCCCC}\\ \text{C} \\ \text{C} \\ \text{C} \\ CCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c} \text{CATTICGCCG}\\ \textbf{CATTICGCCG}\\ \textbf{CATTICGCCGAT}\\ \text{TGGTCGCGGAT}\\ \textbf{TGGTCGCGGAT}\\ \textbf{CTGCCGGCTA}\\ \textbf{CTGCCGGCTA}\\ \textbf{CTGCCGGCTA}\\ \textbf{CTGCCGGCCBC}\\ \textbf{L} & \textbf{F} & \textbf{S} & \textbf{I} \\ \textbf{GACGGGCGGG}\\ \textbf{GTCGCCGCCC}\\ \textbf{CACCGGCGCGC}\\ \textbf{CACCGGCGCGC}\\ \textbf{CACCGCGCCCC}\\ \textbf{CACCGCGCCCCC}\\ \textbf{CACCGCGCCCCC}\\ \textbf{CACCGCCCCCC}\\ \textbf{CACCGCCCCCCC}\\ \textbf{CACCGCCCCCC}\\ \textbf{CACCGCCCCCC}\\ \textbf{CACCGCCCCCC}\\ \textbf{CCCCCCCCCCC}\\ \textbf{CACCGCCCCCC}\\ \textbf{CCCCCCCCCCCC}\\ \textbf{CCCCCCCCCCCCCC}\\ \textbf{CCCCCCCCCCCCCCC}\\ \textbf{CCCCCCCCCCCCCCCC}\\ CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} \textbf{CaCcortial}\\ \textbf{CaCcortial}\\ \textbf{CaCcortial}\\ \textbf{CaCcortial}\\ \textbf{Caccortial}\\ \textbf{Caccortial}\\ \textbf{Catacosc}\\ Catacosc$	CATEGECTAAT GTACCGAGATA CETTTGCGAA GCAAACGCTT T	$\begin{array}{c} \textbf{GCCGCGGTTCC}\\ \hline \textbf{A} & \textbf{A} & \textbf{L} \\ \hline \textbf{CGCACCTETT}\\ \textbf{GCGTGGACAA}\\ \hline \textbf{GCGTGGCCTC}\\ \textbf{GTCGCCAAGA}\\ \hline \textbf{A} & \textbf{A} & \textbf{E} \\ \hline \textbf{CGCCGCTTCC}\\ \hline \textbf{GCCTCGTTTC}\\ \hline \textbf{CGCCAAGA}\\ \hline \textbf{A} & \textbf{A} & \textbf{E} \\ \hline \textbf{GCCTCGTTC}\\ \hline \textbf{CGCCAAGA}\\ \hline \textbf{A} & \textbf{E} & \textbf{N} & \textbf{F} \\ \hline \textbf{GCCTCGTTC}\\ \hline \textbf{GCCCCGTTC}\\ \hline \textbf{GCCCCGTTC}\\ \hline \textbf{GCCCCGTTCC}\\ \hline \textbf{GCCCCGTTCC}\\ \hline \textbf{GCCCCCTTC}\\ \hline \textbf{GCCCCCTTT}\\ \hline \textbf{GCCCCCTT}\\ \hline \textbf{GCCCCCTTT}\\ \hline \textbf{GCCCCCCTT}\\ \hline \textbf{GCCCCCCTT}\\ \hline \textbf{GCCCCCCTT}\\ \hline \textbf{GCTCCCCCTC}\\ \hline \textbf{GCATCCCCCC}\\ \hline \textbf{GCAACCCTCT}\\ \hline \hline \textbf{GCAACCCCCTCCCCTT}\\ \hline \hline \textbf{GCCAACCCCTCT}\\ \hline \hline \textbf{GCAACCCCCTCCCCTT}\\ \hline \hline \textbf{GCCAACCCCTCCCCTT}\\ \hline \hline \textbf{GCAACCCCCTT}\\ \hline \hline \textbf{GCAACCCCCCTT}\\ \hline \hline \textbf{GCAACCCCCCTT}\\ \hline \hline \textbf{GCAACCCCCCTT}\\ \hline \hline \hline \textbf{GCAACCCCCTCT}\\ \hline \hline \hline \hline \textbf{GCAACCCCCTT}\\ \hline \hline \hline \hline \textbf{GCAACCCCCCTT}\\ \hline \hline \hline \hline \textbf{GCAACCCCCTT}\\ \hline \hline \hline \hline \hline \textbf{GCCCCCCTT}\\ \hline \hline \hline \hline \textbf{GCAACCCCCCTT}\\ \hline \hline \hline \hline \hline \textbf{GCAACCCCCCTCT}\\ \hline \hline$	THECAGE TEATT TIGCAGECTAA F T S K CCGCCCAAATC GGCGGTTAG GGCGGTTAG GGCGGTTGCA P F P T S L CGGAAACGGT GCCTTTGCCA ATAAAACTA TATTITGAAT TATTITGAAT TATTITGAAT TATTITGAAT TATTITGAAT TATTITGAAT TATTITGAAT TATTITGAAT TATTITGAAT CAAAAGGAA CAAAAGGAAT TGGCGGGGCGGG ACCGCCTCCGCA ACCGCTTCGCA ACCGCTTCGCAAAGC TACCGCCTCCGCAA GGGGCGCTT A E A L TTCCTGCCC AAAGAACGGC C E Q R	CCCCCCTCAC CCCCCCCCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{c c} GCCCTTTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} GCTTCGCGCC\\ G & CGAAGCCGCCG\\ A & E & A & R \\ TGCGCGCGCT\\ ACGCGGCGCT\\ ACGCGGCGCGT\\ ACGCGGCGCGT\\ ACGCGGCGCGC\\ CCTAAGGATCA\\ G & L & I & V \\ CTCCGTACTG\\ G & L & I & V \\ CTCCGTACTG\\ G & L & I & V \\ CTCCGTACTG\\ G & CCACATGACT\\ G & CCACATGACT\\ G & CCACATGACT\\ TTAAGTTTTT\\ AATCCAAA\\ N & L & K & E \\ CGTCGGCGCCGC\\ ACGGCACCGCC\\ ACGGCACCGCC\\ ACGGCACCGCC\\ ACGGCACCGCC\\ ACGGCACCGCC\\ G & C & A \\ R \\ ACCCCCTGCTC\\ TGGGGGCGACG\\ G & G \\ G \\$	$\begin{array}{c} \textbf{GCCCTGCCG}\\ \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \\ \textbf{GCCCGCGAAAC}\\ \textbf{CCGCGGAAAC}\\ \textbf{CCGCGCTTGCG}\\ \textbf{TTCCGGACGCG}\\ \textbf{TTCCGGACGCC}\\ \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \\ \textbf{AGGCCTGCG}\\ \textbf{TTCCGGACGCC}\\ \textbf{CTAACCCTAT}\\ \textbf{GATGGGATA}\\ \textbf{GATGGGATA}\\ \textbf{GATGGGATA}\\ \textbf{GATGGGATA}\\ \textbf{GATGGGATA}\\ \textbf{GATGGGGGTA}\\ \textbf{GACGGGGTA}\\ \textbf{GACGGGGTA}\\ \textbf{GACAGGGGGTA}\\ \textbf{GTTCGCCCACAG}\\ \textbf{GTTCGCCCACAG}\\ \textbf{GTATGGGTGTC}\\ \textbf{GTATGAGGT}\\ \textbf{GCACTACTCA}\\ \textbf{GGGGGTCTCAC}\\ \textbf{GCGCGTCACC}\\ \textbf{GCGCGCTCAC}\\ \textbf{GCGCGCTCAC}\\ \textbf{GCGCGCGACTC}\\ \textbf{GCGCGCGCCC}\\ \textbf{S} & \textbf{L} & \textbf{A} \\ \textbf{Q} \\ \textbf{GCGGCCCCC}\\ \textbf{S} & \textbf{L} & \textbf{A} \\ \textbf{GCGGCCCCC}\\ \textbf{GCGACTC}\\ \textbf{GCGACTCC}\\ \textbf{GCGACTCC}\\ \textbf{GCGACTCC}\\ \textbf{GCGACTCC}\\ \textbf{GCGCGGGACC}\\ \textbf{S} & \textbf{L} & \textbf{A} \\ \textbf{Q} \\ \textbf{GCGGCCCCC}\\ \textbf{S} & \textbf{L} \\ \textbf{A} \\ \textbf{Q} \\ \textbf{GCGGCCCCC}\\ \textbf{S} & \textbf{C} \\ \textbf{CGCGGACCC}\\ \textbf{CGAATACGGC}\\ \textbf{CGCGACCC}\\ \textbf{CGAATACGGC}\\ \textbf{CGCGACCC}\\ \textbf{CGAATACGGC}\\ \textbf{CGCGGCCCC}\\ \textbf{S} & \textbf{L} \\ \textbf{A} \\ \textbf{Q} \\ \textbf{GCGGCCCCC}\\ \textbf{S} \\ \textbf{C} \\ \textbf{CCCGCAGACC}\\ \textbf{CGAATACGGC}\\ \textbf{CGCGACCC}\\ \textbf{CGAATACGGC}\\ \textbf{CGCGACCC}\\ \textbf{CGAATACGGC}\\ \textbf{CGCGACCCC}\\ \textbf{CGAATACGCG}\\ \textbf{S} \\ \textbf{C} \\ \textbf{CCCCCGCAGACC}\\ \textbf{CGCGCCCCCC}\\ \textbf{CGAATACCGGC}\\ \textbf{CGCGACCCC}\\ \textbf{CGAATACCGGC}\\ \textbf{CGCGACCCC}\\ \textbf{CGAATACCGGC}\\ \textbf{CGCGCCCCCC}\\ \textbf{CGAATACCGC}\\ \textbf{CGCGCCCCCC}\\ \textbf{CGAATACCGC}\\ \textbf{CGCGCCCCCC}\\ \textbf{CGAATACCGGC}\\ \textbf{CGCGCCCCCCC}\\ \textbf{CGAATACCGGC}\\ \textbf{CGCGCCCCCC}\\ \textbf{CGAATACCGGC}\\ \textbf{CGCGCCCCCCC}\\ \textbf{CGAATACCGCC}\\ \textbf{CGCGCCCCCCC}\\ \textbf{CGAATACCGCC}\\ \textbf{CGCGCCCCCCCC}\\ \textbf{CGAATACCGCC}\\ \textbf{CGCGCCCCCCC}\\ \textbf{CGAATACCGCC}\\ \textbf{CGCGCCCCCCC}\\ \textbf{CGCCCCCCCCCCCC}\\ \textbf{CGCCCCCCCCCCCCCCC}\\ CGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$
	-2 7601 -2 7601 -2 7801 -2 7801 -2 7801 -2 8001 -2 8101 -2 8301 -2 8301 -2 8401 -2 8501 -2 8601 -2 8601	$\begin{array}{c} Creation for the second for$	CATTICECCG CATTICECCG TGATTCECCG TGGTCECCGAT TGGTCECCGAT TGGCCGCGAT TGGCCGCGAT CTGCCGACTA CTGCCGACTA CAGCGGGCGGG GCGGCGGCGG GCGCGGCGGC CAGCGGCGCGC CAGCGGCGCGC CAGCGGCGCGC CAGCGGCGCGC CAGCGGCGCGC CCAATAACCG CCAATAACCG CCAATAACCG CCAATAACCG CCAATCAATCG S L G Q CAACCAGCGTAACCG CAACCGGAC S L G Q CAACCAGCGCA AACCGGCGTC TGGCCGGCAA CCCACGGACTT AACCGGCGTT GGTTAATTTAGC CGATTAAAA CGGCTTAATTTAGC CGATTAAAA CGGCAATTGCCGGAA CCCACGGACT CCAATCAATCG CCAATCAATCG CCCGGCAA CCCACGGCGTC CCCACGGCGTC CCCATTAAAC CCCACGGCGTC CCCATTAAAC CCCACGGCCTC CCCATTAAAC CCCACGGCCTC CCCATTAACCG CCCATTACCGCGAC CCCATTACCGCCCC CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATAACCG CCCGGCAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATAACCG CCCATCAACCG CCCATAACCG CCCATAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATAACCGGCCTT CCCACTAACCGGCCTC CCCATCAACCG CCCATCAACCG CCCATAACCG CCCATCAACCG CCCATCAACCG CCCATAACCG CCCATCAACCG CCCATAACCG CCCATCAACCG CCCATCAACCG CCCATCAACCG CCCATCACCGCCTT CCCATCACCGCCTT CCCATCACCGCCTC	CACCETTAAA T R I E GAAGCTCCGC CTTCCAGGCC CTTCCAGGCC CTTCCAGGCC CTTCCAGGCC CCTTCCAGGCC CCTTCCAGGCC CCTTCCAGCTC S V S R GTGTAAGCCTT GCGCGACCGC CACATCTGC T L R CATAAGCCTT GCGCGCTCCC A R V T GCGCGCTCCC A R V T CCCGTAGCCCC GCACTCCGCA CCGCTGCCCCGA GCCCCCGAGAAA CCCGTTGCCCCG GCACCCGGA CCCGTTGCCCCG GCACCCGGA CCCGTTGCCCC GGACAACGGG CCCGTTGCCCC GGACAACGGG CCCGTTGCCCC GGACAACGGG AGGCCTTGCC CCCGTTGCCCC GGACAACGGG AGGCCTTGCC CCCGTTGCCCC GACACCGGA CCCGTTGCCCCGA G H A C CCCGTTGCCCCGA G H A C CCCGTTGCCCCG GACACCGGA CCCGTTGCCCC GACACCGGA CCCGTTGCCCC GACACCGGA CCCGTTGCCCC GACACCGGA CCCGTTGCCCC GACACCGGA CCCGTTGCCCC CCCGTTGCCCC CCCGTTGCCCC CCCGTTGCCCCCGA CCCGTTGCCCCCGA CCCGTTGCCCCCGA CCCGTTGCCCCCGA CCCGTTGCCCCCGA CCCGTTGCCCCGA CCCGTTGCCCCGACGC CCCGTTGCCCCGACGC CCCGTTGCCCCGACGCCCCCGACCGC CCCGTTGCCCCGACGCCCCGACCGC CCCGTTGCCCCGACCGCCCCGACCGC CCCGTTGCCCCGACCGCCCCGACCGCCCCGACCGCCCCGACCGCCCCGACCGCCCCGCACCGCCCCGACCGCCCCGCCCCGCACCGCCCCGCACCGCCCCCGCCCCGCCCCCGCCCCGCCCCGCCCCCGCCCC	CATEGECTAA GTACCGAGATA CCATTGCGAA GCAAACGCTT T O S TCGCTGACGCT ACCCACTGCG E S V G GGTAACCAC CCATTGGGG C Y V V V CGCTAACCCA CCATTGGGG ACTCGACCAC ACTCGACCAC ACTCGACCAC ACTCGACCAC ACTCCGACGCA ACTCCGCGTA CCACTCACCA ATCCCACGGT TAAGGTCCAC CCGCGCGCAA CCCGCCGTA CCGCCGTA CCCGCCGTA CCCGCCGTA CCCGCCGTA CCGCCGCCGA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCCGTA CCGCCGCCGTA CCGCCCGCCGTA CCGCCGCCGA CCGCCCGCCG CCGCCCGTA CCGCCCGCCGA CCGCCCGCCG CCGCCCGTA CCGCCGCCGCA CCGCCCGCCG CCGCCCGCCG CCGCCCGCCG CCGCCCGC	CCGCGGTTCC CGGCGGTTCC CGGCGGTTCC GCGTGGCGACA CGGCGGCCTC GTCGCGCGCC GTCGCGCGCC GTCGCGCGCC GTCGCGGCGAAG A A E CCGCGGCGAAG A E N F CCCGGGCAAG A E N F CCGGACAAG CGGGCCAAG CGGGCCAAG CGGGCCAAG A E N F CCGGGCGAAG CGGGCCAAG A E N F CCGGGCGAAG GCCTCTTC CGGACGCGCC CGAAGCGCGC CCGGGCGAAG CCCGGCCACCG ATAACCGTCG ATGACCAGCCC CGAAGCGCG CCCGGAAG CCCGGAAGCGCG CCCGGAAGCGCG CCCGGAAGCGCG CCCGGAAGCGCG CCCGGAAGCGCG CCCCGAAGCCCC CGAAGCCCCT CCCGGGCAAG CCCCCCTT CCTTCCGCCG CCCGGACAGCCC CGAAGGCGGC CCCGGAAGCGCG CCCCCCTT CCTTCGCCCT CCCCGGAAGCCG CCCGGAAGCGCG CCCCCCTT CCCCCCT CCCCCCT CCCCCCCT CCCCCCCT CCCCCCCC	THE CAPTER THE THE CAPTER THE THE CAPTER THE	CCCCTCATT A E N CTCACGATGG GAGTGCTACC E R H GAGTCAATAT CTCAGTTATA S D I D ATGAGCACAG TACTCGTGTC ATGAGCACAG TACTCGTGTC CCCCCAACTGG GCTTTGTCCC GCGTTACTCG GCGTTGTCCC CCCCAACTGG CCCCCAACTGG A N V CCGAAACAGGG GCTTTGTCCC GGCATTGCCC CCCCTAACTGG A N V ATAAGATCGT ATATCTCAGCA I L D Y CTGGCGCGCCT AGACCCCCGA A GCCCCCCACCGG A I A ACGCCCTCAC CCCGCACCGG A I A CCCCCTACCGG A I A CCCCCTACCGG A I A CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCGCAAAAGG CCCGCAAAAGG CCGCAAAAGG CCGCAAGTTAA CCGCTCCAG CCGCCACGAG CCGCCACGAG CCCCACGAC CCCCACGCACTAA CCCCCACGACTAA CCCCCACGACTAA ACTGCCGCA TTGACGCCGTA TTGACGCCGTA CCCCACGCTGA TTGACGCCGTA CCCCACACTGA TTGACGCCGTA CCCCACACTGA TTGACGCCGTA CCCCACACTGA TTGACGCCGTA AAATACCGCGTA AACTGCCGTA AACTGCCGCAT ACTGCCGCAT ACTGCCCGTA CCCCACACTGA CCCCACCCGTA CCCCACCGTA CCCCACCCGTA CCCCACCGTGA CCCCACCCCAC CCCCACCCCAC CCCCACCCCCAC CCCCACCCCCCAC CCCCACCCCCCCAC CCCCACCCCCCCC	CTTCGCCGCCGCG A E A R TGCCGCCGCG A CGCGCGCGTT ACGCGGCGCGTT ACGCGGCGCGT ACGCGGCGCGT ATCGGCGCGCG CCTAAGGATCA GGATTCTAGT G L I V CTTCCGTACTG GAAGGCATGA CCACATGACT ACGCGCGCGCA GGTGAGGACAGC GGTGAGGACTAA M Q N TTGAGGTGGCGC ACGCACCGCC ACGCACCGCC ACGCCCGCCCA ACCCCCCCCC CGGCCAGATT CCAGCTCTAAA T L N ACCCCCCCCC	$\begin{array}{c} \textbf{GCCCFGCCGGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGACGGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGACGCC}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{AGGCCTGCG}\\ \hline \textbf{TTCCGGACGC}\\ \hline \textbf{TTCCGGACGC}\\ \hline \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{GATGGGATA} \\ \hline \textbf{GATGGGGGTA} \\ \hline \textbf{GATGGGGGTA} \\ \hline \textbf{GATGGGGGTA} \\ \hline \textbf{GAACGGGGTA} \\ \hline \textbf{GAACGGGGTA} \\ \hline \textbf{GAACGGGGTA} \\ \hline \textbf{GTTCGCCCACA} \\ \hline \textbf{GTTCGCCCACA} \\ \hline \textbf{GTATGGGTGTC} \\ \hline \textbf{GTGTATGACT} \\ \hline \textbf{GCCCGCAGATG} \\ \hline \textbf{GCCCGCGAGAT} \\ \hline \textbf{GCTTAATGCC} \\ \hline \textbf{GCTTAATGCCC} \\ \hline \textbf{GCTTAATGCCC} \\ \hline \textbf{GCTTAATGCCC} \\ \hline \textbf{GCTTAATGCCC} \\ \hline \textbf{GCTTCATCGCC} \\ \hline \textbf{GCTTCTCT} \\ \hline \textbf{GCCTTCATCGCC} \\ \hline \textbf{GCTTCTCTTCCCCCGAATTACGG} \\ \hline \textbf{GCTTCTCTTCCCCCGAATTACGC} \\ \hline \textbf{GCTTCTCTTCCCCCGAATTACGGC} \\ \hline \textbf{GCTTCTCTTCCCCCGAATTACGGC} \\ \hline \textbf{GCTTCTCTTCCCCCGACTC} \\ \hline \textbf{GCTTCTCTTCCCCCCGAATTACGGC} \\ \hline \hline \textbf{GCTTCTCTTCTCCCCCGAATTACGGC} \\ \hline \hline \textbf{GCCTTCTCCCCCCGAATTACCGG \\ \hline \hline \hline \textbf{GCCCCCTCTTC} \\ \hline \hline \textbf{GCCCCCCTCTC} \\ \hline \hline \hline \hline \textbf{GCCCCCCTCTC} \\ \hline \hline \hline \hline \textbf{GCCCCCCCCC} \\ \hline $
	-2 7601 -2 7701 -2 7801 -2 7801 -2 8001 -2 8101 -2 8301 -2 8301 -2 8301 -2 8301 -2 8401 -2 8501 -2 8601 -2 8701	$\begin{array}{c} \text{Creation at be constructed} \\ \text{GTCTTACGG} \\ \text{GTGTTCCABCG} \\ \text{GTGTTCCABCG} \\ \text{GTGTTCCATCA} \\ \text{GCGTTCCATCA} \\ \text{GCGTTCCATCA} \\ \text{GCGTTCCATCA} \\ \text{GCGTCCATCA} \\ \text{GCGTCCATCA} \\ \text{GCGTCCGACGT} \\ \text{ACCGGCTCCG} \\ \text{ACCGGCGTCG} \\ \text{GGCCGTACG} \\ \text{GGCCGTACG} \\ \text{GGCCGTACG} \\ \text{GGCCGTACG} \\ \text{TCCCACCCCCGT} \\ \text{AGCCGCCGTA} \\ \text{GCCGCGCATG} \\ \text{TCCCACCCCCGT} \\ \text{GGCCGTACG} \\ \text{TCCCACCCCGT} \\ \text{GGCCGTACG} \\ \text{TCCCACCCCGT} \\ \text{GGCCGTACG} \\ \text{TCCCACCCCCGT} \\ \text{GGCCGTACG} \\ \text{TCCCACCCCGT} \\ \text{GGCCGTACG} \\ \text{TCCCACCCCGT} \\ \text{GGCCGTACG} \\ \text{TCCCACCCGT} \\ \text{GGCCGCGTACG} \\ \text{TGCCACCTTC} \\ \text{TGCCAACGTTCC} \\ \text{TGCCAACGTTCC} \\ \text{TGCCAACGTTCC} \\ \text{TGCCAACGTTCC} \\ \text{TGCCACCTGTA} \\ \hline \\ \text{TCCCACCCCGTACG \\ \hline \\ \text{TCCCACCCCGTACG} \\ \hline \\ GCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATTICECCE CATTICECCE CATTICECCE CATTICECCE CACCEGCEA CGECEGCEGE CGECEGCEGE CGECEGCEGE CGECEGCEGE CGECEGCEGE CGECEGCEGE CGECEGCEGE CCEATABCCE CCACCES CCATABCCE CCACCES CCACES CCCACES CCACES C	CACCETTAAA T R I E GAAGCTCGCC CTTCGAGGCG CTTCGAGGCG CCTTCGAGGCG CCTTCGAGGCG CCTTCGAGCTG S V S R GTGTAAGCCTG CACATCTGC T L R CCTAGCCTTG CACATCTGC T L R GTGTAAGCCG CACATCTGC A R V T GCGCGACGGACGG CGCGCTCCCGA CAGGTCCGCGC CGGCACCGGG CCGCTGCCCG GGCACCCGGGACGC CCTGTGCCCC GGCACCCGGGACGC CCTGTGCCCC GGCACCCGGGACGC CCTGTGCCCC GGCACCCGGGACGC CCTGTGCCCC GGCACCCGGGACGC CCTGTGCCCC GGCACCCGGACCGC CCTGTGCCCC GGCACCCGGACCGC CCTGTGCCCC GAGGCCTTGCCC GAGGCCTTGCCC CCGCTGACCT CCCGTAACCCCG AAGCCTTGCCC CCGCTCGCGC CCCGTTGCCCC GGCACCCGGAC CCCGTTGCCCC GGCACCCGGACCGC CCCGTTGCCCC GGCACCCGGACCGC CCCGTTGCCCCC CCCGCTCGCCC CCCGCTCGCCC CCCGCTCGCCC CCCGCTCGCCC CCCGCTCGCCC CCCGCTCCCGCC CCCGCTCCCGCC CCCGCTCCCGCC CCCGCTCCCCCC CCCCTCGCCCCCCCC CCCCCTCCCCCCC CCCCCTCCCCCCC CCCCCTCCCCCCC CCCCTCCCCCCCC	CATECECARTA GTACCEGAGTA CCETTTGCEAA GCAAACGCTT T $ \odot$ S TCGCTGACGCC ACCCATTGCGTG CCATTTGGTG CCATTTGGTG CCATTTGGTG CCATTTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATTGGTG CCATCGACCA ACTCCGACCA ACTCCGACCA ACTCCGCGCGTA CCCCGCGTTA CCGCGCGCGTA CCGACCCGCGTA CCGACCCGCGTA CCGACCGCGTA CCGACCGCGTA CCGACCGCGTA CCGACCGCGTA CCGCGCGCGTA CCGACCGCGTA CCGCGCGCGTA CCGCGGCGGTA CCGCGGCGGTA CCGCGGCGGTA CCGCGGCGGTA	CCGCGGTTCC CGCCGGTTCC CGCCGGTTCC CGCCGGTTCC GCGTGGCGACA CGCCGCCTC GTCGCGGCGCTC GTCGCGGCGCC GTCGCGGCGCC CGGGCGAAG A A E CGCGGCGAAG A C N F CGCGGCGAAG A C N F CGCGGCGAAG CGGGCCAAG A C N F CGCGGCGAAG CGGGCCAAG A C CGGATACC GGCCCCTTTC CGGACGCTG A TATACGTGC A TGCCGAAG A C CGGACGCC CCGAGCGCC A TGCCCCTT TACTCGCGA CCCGGGAAG CCCGGGAAG CCCGGACGCC CGAAGCCGC CCGAAGCCGC CCGAAGCCGC CCGAAGCCGC CCGAAGCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CCGGAGGCCC CCGAGGGCC CCCCCTTCCCCCC CGAAGCCCCC CGAAGCCCCC CGAAGCCCCC CCGAGGCCCCC CGAAGCCCCC CGAAGCCCCC CCGAGGCCCCC CGAAGCCCCCCCCCC	THECAGECTEANT TIGCAGECTAA F T S K CCGCCCAAATC GGCGGTTAG GGCGGTTAG GGCGGTTAG GGCGGTTGCCA P F P ATAAAACTA TATATAGAACTA TATATAGAACTA TATATAGAACTA TATATAGAACTA CAAAAGCAA CAAAAGGAAC K E I GTTTTCCTA ACCACCCACCA ACCGCCGCCGCC C F R L CAAAAGGAACGCA ATTGGCGCGC AACGCCGAA GGCGCCGCCCC C F R L CCCCCCGCCA CCCCCGCCA CCCCCGCCA CCCCCGCGTTC C C C C C C C C C C C C C C C C C C	$\begin{array}{c} \text{ACCGCACGTGAT}\\ \textbf{A} \textbf{E} \textbf{N} \\ \hline \textbf{CTCACGATGG}\\ \textbf{GAGTGCTACC}\\ \textbf{E} \textbf{R} \textbf{H} \\ \hline \textbf{GAGTCATAT}\\ \textbf{CTCAGATAT}\\ \textbf{CTCAGATAT}\\ \textbf{CTCAGATAT}\\ \textbf{CTCAGTTATA}\\ \textbf{S} \textbf{D} \textbf{I} \textbf{D} \\ \textbf{A} \textbf{N} \textbf{V} \\ \hline \textbf{GGCATCGTGTC} \\ \hline \textbf{GGCATCGTGTC} \\ \hline \textbf{GGCATTGGTCC} \\ \hline \textbf{R} \textbf{T} \textbf{L} \\ \hline \textbf{CCCGCAAGTGG} \\ \hline \textbf{GCCTTGTCCC} \\ \textbf{R} \textbf{T} \textbf{L} \\ \hline \textbf{C} \\ \textbf{TATTATAAAATTT} \\ \textbf{TATTATAAAATTT} \\ \textbf{TATTAAAATTT} \\ \textbf{TATTAAAATTT} \\ \textbf{TATTAACCGCCC} \\ \hline \textbf{CGCATAGGCCC} \\ \hline \textbf{CGCATAGGCCC} \\ \hline \textbf{CGCATCGCC} \\ \hline \textbf{CGCATCGCC} \\ \hline \textbf{CGCCATCGCC} \\ \hline \textbf{CGCCATCGCC} \\ \hline \textbf{CGCCATCGCC} \\ \hline \textbf{A} \textbf{N} \textbf{V} \\ \hline \textbf{TCTGGCGGCT} \\ \hline \textbf{AGCCCCCCAC} \\ \hline \textbf{GGCCATCGCC} \\ \hline \textbf{A} \textbf{L} \textbf{A} \\ \hline \textbf{ACCCCCTCAC} \\ \hline \textbf{CGCGCATAGGC} \\ \hline \textbf{V} \textbf{G} \textbf{E} \textbf{S} \\ \hline \textbf{CTGCGGAATTG} \\ \hline \textbf{CCCCCATAA} \\ \hline \textbf{CCCCCTACCGCC} \\ \hline \textbf{A} \textbf{L} \textbf{A} \\ \hline \textbf{ACCCCCCCAC} \\ \hline \textbf{CCGCCCTACCGC} \\ \hline \textbf{CCGCCATAA} \\ \hline \textbf{CTCCGCGTATA} \\ \hline \textbf{CCGCCATAA} \\ \hline \textbf{C} \textbf{CCCCCCAC} \\ \hline \textbf{CCGCCCTACCCC} \\ \hline \textbf{CCGCCCTACCCC} \\ \hline \textbf{CCGCCCTACCCCC} \\ \hline CCGCCCTACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CCGCAACGAAAGG CCCGCAAAAGG CCCGCAAAAGG CCGCAAGTTAA CCCCTCCAG CCGCCACGCAG CCCCACGCAGCAGTCA CCCCCACGCAGGAGGTC L Q S V TAACCGCCTCAT AACTGCCACA ATTGGCCCCA ATTGGCCCCAG CCCCAGCGTGAA TGGCGCTCAT AACCACCGTGA ATTGGTCACA CCCCACGTGA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCCACCTCA CCCACCTCAC CCCACCTCACCT	CGAGCCGCCG CGAAGCCGCCGT ACGCGCGCGTT ACGCGCGCGTT ACGCGGCGCACA L R Q CCTACGGCGCACA L R Q CCTACGGCGCACA CCTAAGATCA GGATICTAGT GAAGCCATGAT TTGGGTGCGA GAAGCCATGA TTAAGTTTTT AATTCAAAAA N L K E CCACATGACTAA M Q N TTGGGTGCGCG ACGCACACGCC Q T A TGGGTGGCGCA ACCCCCGCC GGTCAGATCA GGTCAGACTAA CCACTCGTCT GGGCAGCAGC GGTCAGACTCCC GGTCAGACTCCC GGTCAGACTCCCC CCACTCCCTC CGGCGCCGCA CCACTCCCCCC CGGCCAGCAC CCACTCCCCCCC CGGCCAGCCGC CCACTCCCCCCCCC CGGCCAGCACC CCACTCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{c} GCCCFGCCGGCGCGGCGGCGGGCGGGGCGGGGGGGGGG$
	-2 7601 -2 7701 -2 7801 -2 7801 -2 8001 -2 8101 -2 8301 -2 8301 -2 8401 -2 8501 -2 8601 -2 8601 -2 8701	$\begin{array}{c} Creation for the second for$	CATTICECCE CATTICECCE CATTICECCE CATTICECCE CATTICECCE CACCEGCE CECCECE CECCECE CECCECE CECECE CECECE CACCEGECE CACCEGECE CECECECE	CACCETTAAA T R I E GAAGECTCGCC CTTCEAGGCG CTTCEAGGCG CCTTCEAGGCG CCTTCEAGGCG CCTTCEAGCTG GATACGCAAC CTTCEAC T D L T D L CCTATAGGCTG S V S R GTGTAAGACCG CCACTTCFGC T L R CCTATAGCCTTG GTATTCGCAA M L T GCGCGAACGG CCGCGTCTCTC CGCGCAACCGG CCGCTCCGGA CCGCTCCGGAACG CCGCTCCGCC D L A CCGTAAGCCCT GGCAACCGGC CCGCTCCGCGC CGCCTCCCGGAACG CCGCTTCCCC GGCAACCGGC CCGCTTCCCC GGCAACCGGC CCTCTTGCCCC GGCAACCGGC CCTCTTGCCCC GAAGCCTTGCCC GAAGCCTTGCCC GGCAACCGGC CCTCTTGCCCC GGCAACCGGC CCGCTTACCCC GGCAACCGGC CCTCTTGCCCC GGCAACCGGC CCTCTTGCCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCTCTTGCCC GGCAACCGGC CCCGTTACCCC GGCAACCGGC CCCGTTACCCC GGCAACCGGC CCCGTTACCCC GGCAACCGGC CCCGTTACCCC GGCAACCGGC CCCGTTACCCC GGCAACCGGC CCCGTTACCCC GGCAACCGGC CCCGTTGCCC GGCAACCGGC CCCGTTGCCC GGCAACCGGC CCCGTTGCCC GGCAACCGCC CCCGTTGCCC GGCAACCGCC CCCGTTGCCC GGCAACCGCC CCCGTTGCCC GGCAACCGCC CCCGTTGCCC GGCAACCGCC CCCGTTGCCC GGCAACCGCGC CCCGTTGCCC GGCAACCGCG CCCGTTGCCC GGCAACCGCGC CCCGTTGCCC GGCAACCGCGC CCCGTTGCCC GGCAACCGCG CCCGTTGCCC GGCAACCGCG CCCGTTGCCC GGCAACCGCGC CCCGTTGCCC GGCAACCGGC CCCGTTGCCC GGCAACCGGC CCCGTTGCCC GGCAACCGGC CCCGTTGCCC CCCGTTGCCC CCCGTTACCC CCCGTTACCCC CCCGTTGCCC GGCACCGCGC CCCGTTGCCC GGCACCGCGC CCCGTTGCCC GGCACCGCGC CCCGTTGCCC CCCGTTGCCC CCCGTTGCCC CCCGTTGCCC CCCGTTGCCC CCCGTTGCCC CCCGTTGCCC CCCGTTGCCC CCCGCTTGCCC CCCGCTTGCCC CCCCCTCGCCC CCCCCCCCCC	CATECECARTA GTACCEGAGTA CCETTGCEAA GCAAACGCTT T	CCGCGGTTCC CCGCCGTTCC CCGCCGTTCC CCGCCGCTCC GCGTGGCGACA CCGCCGCTC GTCGCGACAA A A E CCGCGCCTC GTCGCCGAAGA AGCGGTTCC CCGGACAAGA A E N E CCGCGACAAGA A E N E CCGGACAAGA CCGGACAAGA CCGGACAAGC CCGGACAAGC CCGGACAACC ATGACCAAGA A E N E CCCGGACAAC CCGGACAACC ATGACCAAGA A E N E CCCGGACAAC CCGGACAACC CCGACCCTTC CCGACCCCTT TACTCGCGA CCCCGTTCC CCGAAGCCGC ATGACCACCCC CCAAGCCGCA CCCCGAAGC CCCGGACACC CCCGACCCCCCC TACTCCCCCC CCAAGCCGCC CCAAGCCGCC CCCGACCCCC CCAAGCCCCC CCAAGCCCCCC CCAAGCCCCC CCAAGCCCCC CCAAGCCCCC CCAAGCCCCC CCCCCTTCCCCCC CCAAGCCCCC CCCCCTTCCCCCCC CCCCCCCCCCCCCCCC	AACCFICATIN TIFGCAGCTAA F T CCGCCAAATC GGCGGTTAG GGCGGTTAG GGCGGTTAG GGCGTTGCCA P F T T AL D CGGAAACGGT GCTTTGCCA P F ATAAAACTA TATATTGAAT TATATAAGACTA CAAAAGGAA H F CAAAAGGCAA ACCGCCGCCCC Q R CGAAAGGCAGC ATTTGGCGCGGA ATTGGCGCGGC ATTGGCGCGCGCCCC Q R CGCGCCGCCCCC Q R GCGCCCGCCCCC Q R CCCCCGCCGCCCC Q R CCCCCGCCGCCCC Q R CCCCCCGCCGCCC Q R GCGCCCCCCCCCCC A L A L CCCCCCGCGCCCCCC A L<	$\begin{array}{c} \text{ACCGCACGTGTT}\\ \text{ACCGACGATAG}\\ \text{GAGTGCTACCG}\\ \text{GAGTGCTACCG}\\ \text{GAGTGCTACCGTGTG}\\ \text{GAGTCATAT}\\ \text{CTCAGTTATA}\\ \text{S} D I D I \\ \text{D} I \\ \text{CCCGCAATGG}\\ \text{GCTTCACCGTGTC}\\ \text{CCCGCAAGTGG}\\ \text{CCCGCGCAGCGCG}\\ \text{A I A V V}\\ \text{CCGGCAGCGCGA}\\ CCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$ \begin{array}{c} \textbf{GCCTTTTCCC} \\ \textbf{CCCGCAAAAGG} \\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CCGCTCAAT} \\ \textbf{GCCAGTTAA} \\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{CGCCTCCAG} \\ \hline \textbf{GCGGGAGCT} \\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{G} & \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{G} & \textbf{G} & \textbf{G} \\ \hline \textbf{G} \hline \textbf{G} \hline \ \textbf{G} \hline \textbf{G} \hline \textbf{G} \hline \textbf{G} \hline \ \textbf{G} \hline \textbf{G} \hline \ \textbf{G} \hline \ \textbf{G} \hline \textbf{G} \hline \ \textbf{G} \hline \textbf{G}$	CGAAGCCGCG A E A R TGCGCGCGCGT ACGCGGCGCGT ACGCGGCGCGT ACGCGGCGCACA L R Q CTACGGCGCACA L R Q CCTAAGATCA GGATICTAGT G L I V CTTCCGTACTA GAAGCCACGCA GAAGCCACGATGA TTAAGTTACAAAA N L K E CCACATGACTA AGTAGATCA GGTACGAGATGA TTGGGGGCGCA ACCCCCCC CGACGCCCAC CACCGCCCAC CACCGCCCAC CCACACGCC CCACACGCCGC CCACCGCCCAC CCACCGCCCAC CCACCGCCCAC CCACCCCCC CGGCACGACCGCC CCACCCCCC CGGCACGACCGCC CCACCCCCC CGGCACGACCGCC CCACCCCCC CGGCACGCCGAC CCACCCCCC CGGCACGCCGC CCACCCCCCC CGGCACGCCGCA CCACCCCCCC CGGCACGCCGC CCACCCCCCC CGGCACGCCGC CCACCCCCC CGCCCCCCCC CGCCCCCCCC CGCCCCCCCC	$\begin{array}{c} \textbf{GCCCFGCCGGC}\\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGACGGC}\\ \hline \textbf{GCCGCGACGCC}\\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{GCCGCGCTGCG}\\ \hline \textbf{TCCGCGCGTGCG}\\ \hline \textbf{TCCGCGCACGC}\\ \hline \textbf{L} & \textbf{A} & \textbf{Q} & \textbf{T} \\ \hline \textbf{CTAGCCGGTA}\\ \hline \textbf{GATGGGATA}\\ \hline \textbf{GATGGGATA}\\ \hline \textbf{GATGGGATA}\\ \hline \textbf{GATGGGGTA}\\ \hline \textbf{GATGGGGTA}\\ \hline \textbf{GATGGGGTA}\\ \hline \textbf{GTTGGCCCACAG}\\ \hline \textbf{D} & \textbf{T} & \textbf{D} \\ \hline \textbf{CACGGGTC}\\ \hline \textbf{GTAGCCGCGTC}\\ \hline \textbf{GTGATGCCCACAG}\\ \hline \textbf{D} & \textbf{T} & \textbf{D} \\ \hline \textbf{CACCGGGTC}\\ \hline \textbf{GGCGTGTC}\\ \hline \textbf{GGCGTGTC}\\ \hline \textbf{GGCGTGTC}\\ \hline \textbf{GGCGCGTAC}\\ \hline \textbf{GGCGTCACCAGGGGA}\\ \hline \textbf{GGCGCTCAC}\\ \hline \textbf{GGCGCGCACC}\\ \hline \textbf{GCGCGCGGAA}\\ \hline \textbf{ACGGCCCCAG}\\ \hline \textbf{S} & \textbf{L} & \textbf{A} & \textbf{Q} \\ \hline \textbf{GCCGCGGAGA}\\ \hline \textbf{ACGGCCCCCAG}\\ \hline \textbf{A} & \textbf{S} & \textbf{T} \\ \hline \textbf{GCCGCGGAGA}\\ \hline \textbf{ACGCGCCCCC}\\ \hline \textbf{GCCGCGGAC}\\ \hline \textbf{ACGCGCCCCC}\\ \hline \textbf{GCCCCCCCG}\\ \hline \textbf{ACGCCCCCC}\\ \hline \textbf{GCCGCGGAC}\\ \hline \textbf{ACGCCCCCCC}\\ \hline \textbf{GCCGCGGAC}\\ \hline \textbf{ACGCCCCCC}\\ \hline \textbf{GCCCCCCCAG}\\ \hline \textbf{ACGCCCCCCC}\\ \hline \textbf{ACCCCCCCC}\\ \hline \textbf{ACCCCCCCCC}\\ \hline \textbf{ACCCCCCCCCC}\\ \hline \textbf{ACCCCCCCCCC}\\ \hline \textbf{ACCCCCCCCCC}\\ \hline \textbf{ACCCCCCCCCC}\\ \hline \textbf{ACCCCCCCCCC}\\ \hline \textbf{ACCCCCCCCCC}\\ \hline \hline \hline \hline \hline \textbf{ACCCCCCCCCC}\\ \hline \hline \hline \hline \hline \textbf{ACCCCCCCCCCC}\\ \hline \hline \hline \hline \hline \hline \hline \textbf{ACCCCCCCCCC}\\ \hline \hline$
	-2 7601 -2 7701 -2 7801 -2 7801 -2 8001 -2 8101 -2 8301 -2 8301 -2 8401 -2 8501 -2 8601 -2 8701 -2 8801	$\begin{array}{c} \text{Creation at be constructed} \\ \text{GTGTTACAGG} \\ \text{GTGTTACAGG} \\ \text{GTGTTCCAGCG} \\ \text{GTGTTCCAGCG} \\ \text{GTGTTCCATCA} \\ \text{GCTTCCATCA} \\ \text{GCTTCCACGCC} \\ \text{GTGCCACGCC} \\ \text{GTGCCCACGC} \\ \text{GTGCCGTCG} \\ \text{GTGCCGTATA} \\ \text{GTTCCGCGCTATA} \\ \text{GTCCGCCCATA} \\ \text{GTCGCCCATA} \\ \text{GTCGCCCCATA} \\ \text{GTCGCCCCCACGC} \\ \text{TGCGCCCATA} \\ \text{GTCGCCCCATA} \\ \text{GTCGCCCCATA} \\ \text{GTCCACCCCT} \\ \text{GTGCAAAGATA} \\ \text{GCGCCTGCA} \\ GTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CGATTICGCCG CATTICGCCG CATTICGCCG CATTICGCCGCT ACCAGCGCGAT CGCCGCGCG CGCCGCCGAT CGCCGGCGG CGCCGGCGG CGCCGGCGG CGCCGGCGG CGCCGCCGCC CGCCGCCGCC CGCCGCCGCC CGCCGCCGCC CGCCGCCGCC CGCCGCCGCC CGCCGCCGCC CCACTGGCGCG CCACTGGCGCGA CCGATAACCG CCACTGGCGGAA CCGACCGCGC CGTAATTTGC CCACTGCCGCGA CCGATCACCGA CGCGGCGGGA CCGACCGCGC CGTAATTTGCC CCATTAAAA CCGACCGCCGCC CGGTAATTTGC CCGATAACCG CCGATAACCG CCACTGCCGCA CCGCCGCCGA CCGCCGCCCGA CCGCCGCCGA CCGCCGCCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGA CCGCCGCCGC CCCCCCCCCC	CACCETTAAA T R I E GAAGCTCGCC CTTCGAGGCG CTTCGAGGCG CCTTCGAGGCG CCTTCGAGGCG CCTTCGAGCTG GATACGCGAC CTTCGCAG T D L T D L CCACATCTGC T L R CCACATCTGC T L R CCACATCGCCT GTATCGCAA M L T GCGCGAACGG CGCCGCTCCCGA A R V T GCGCGAACGG CGCCGCTCTTC CGCGCAACGGC CCAGTTCCGCC D L A CCTGTAGCCCT GGCAACCGG CCCGTTACCCC GGCACCCGGA CCCTTGCCCC GGCACCCGGA CCCTTGCCCC GGCACCCGGA CCCTTGCCCC GGCACCCGGA CCCTTGCCCC GGCACCCGGA CCCTTGCCCC GGCACCCGGA CCCTTGCCCC GGCACCCGGA CCCCTTGCCCC GGCACCCGGA CCCCTTGCCCC GGCACCCGGA CCCCTTGCCCC GGCACCCGGA CCCCTTGCCCC GGCACCCGGA CCCCTTGCCCC GGCACCCGGA CCCCTTGCCC GGCACCCGGAC CCCCTTGCCCC GGCACCCGGAC CCCCTTGCCCC GGCACCCGGAC CCCCTTGCCCC GGCACCCGGAC CCCCTTGCCCC GGCACCCGGACCG CCCCTTGCCCC GGCACCCGGACCGC CCCCTTGCCCC GGCACCCGGACCGC CCCCTTGCCCC GGCACCCGGACCCG CCCCTTGCCCC GGCACCCGGACCC CCCCTTGCCCC GGCCTTGCCCC GGCACCCGGACCC CCCCTTGCCCC GGCCTTGCCCC GGCCTTGCCCC GGCCTTGCCCC GGCCTTGCCCC GGCCTTGCCCC GGCCTCGCGC CCCCTTGCCCC GGCCTTGCCCC GGCCTTGCCCC GGCCTCGCCCC CCCCTTGCCCCC GGCCCCCCGCCC CCCCTTGCCCCC GGCCCCCCCGCCCC GGCCCCCCCGCCCCC GGCCCCCCCC	CATECECARTA GTACCEAGATA CCATTEGCAA GCAAACGCTT T $ \odot$ S TCGCTGACGC ACCCATTEGCA GGTAACCAC CCATTEGTG CCATTEGTG CCATTEGTG CCATTEGTG CCATTEGTG CCATTEGTG CCATTEGTG CCATTEGTG CCATCCACCA ACTCCACCA CCCATTEGTCAC CCCCATTAGCC CCCCATTAGCC CCCCCTTA ACTCCCCCCTTA ACTCCCCCCTTA CCCCCCCTTA ACTCCCCCCTTA ACTCCCCCCCTTA ACTCCCCCCCTTA ACTCCCCCCCTTA ACTCCCCCCCTTA ACTCCCCCCCTACAC CCCACCCCCTACAC CCCACCCCCTACCAC CCCACCCCCTTA ACTCCCCCCCCTACCAC CCCACCCCCTACCAC CCCACCCCCTACCAC CCCACCCCCTACCAC CCCACCCCCTACCAC CCCACCCCCTACCAC CCCACCCCCTACCAC CCCCCCCTTACCCC CCCACCCCCTACCAC CCCCCCCTTACCCC CCCACCCCCTACCAC CCCCCCCTTACCCC CCCACCCCCTACCAC CCCCCCCCTACCCCCCCCCC	CCGCCGCTCC CGCCCGTTCC CGCCGCTCC GCGCGCCTC GCGCGCCCC GTCGCCGC GTCGCCGCCC GTCGCCGCCC GTCGCCGCCCC CGGCGCGCCCC CGGACGCTCC CGGACGCAAG A E N F CGCCGGATAGC GCGCCTGTTCC CGGACGCAAG A E N F CGCGGCGAAG A E N F CGCGGCGAAG CGGCCTATCG GCCCCTTTC CGGACGCGCC ATGACGCGCC ATGACGCGCC CGAAGCCGC GCCCCCTTC GCCCCCCTTC CCCGGGCGAA C CCCGGCGCCCC ATGACGCGCC CGAAGCCGCC CGAAGCCGCC CGAAGCCCCC CGAAGGCGCC CCGGAGGCGCC CCGGCGCCCCCCC CGAAGGCGCC CCGGAGGCGCC CCGGCCCCCCTTC CCCGGGCGAA C CCCCCCCCCCCCCCCCCCCCCCCCC	AACGFICATI TIGCAGCTAA F T TIGCAGCTAA GGCGGCAAACC GGCGGTTAG GGCGGTTAG GGCGGTTAGC GGCGGTTAGC GGCGTTGCCA P F TATATACGAT TATATAGAT I F STATAAACTTA CAAAAGCAA CAAAAGCAA H F STATTTCCATA CAAAAGGCAA H TIGCGCGCGCGC ACCCCTTTGCCA ATTAGGCCGCG ATTAGGCCGCGCGCC C TITCTTGCCG AAAGGGCCGCGCCGCC C CGGAGGCGCGCCGCCCCCCCCGCA CGGCCGCGCCCCCCCCCCCCCCCA CGGGCGCGCCCCCCCCCCCCCCA CGGGCCGCGCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ACCECTCGTT ACCECTCGTG GAGTGCTACC E R H GAGTGCTACC E R H GAGTGCTACC E R H GAGTCAATAT CTCAGTTATA S D I D ATGABCACAG TACTCGTGTC CCCCCAATGG CCCCCAATGG CCCCCAATGG CCCCCAATGG CCCCCAATGG CCCCCAATGG CCCCCAATGG CCCCCAATGGC CCCCAATGGCC CCCCAATGGCC CCCCAACTGG A N V CCCCCAACTGG CTATCTGCCCC CCCCAACGCCGCA A N V TCCGGCGCGCA A N V CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	$ \begin{array}{c} \textbf{GCCTTTTCCC} \\ \textbf{CCCGCAAAAGG} \\ \hline \textbf{A} & \textbf{K} & \textbf{E} \\ \hline \textbf{CCGTTCAATT} \\ \hline \textbf{GCCAAGTTAA} \\ \hline \textbf{R} & \textbf{N} & \textbf{L} & \textbf{K} \\ \hline \textbf{CGCCTCCAG} \\ \hline \textbf{GCGGGAGGTC} \\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{G} & \textbf{CGCCTCCAG} \\ \hline \textbf{GCGGGTCCTAT} \\ \hline \textbf{A} & \textbf{L} & \textbf{L} \\ \hline \textbf{G} & \textbf{E} & \textbf{L} \\ \hline \textbf{A} & \textbf{L} & \textbf{L} \\ \hline \textbf{G} & \textbf{CGCGTCCTAT} \\ \hline \textbf{A} & \textbf{L} & \textbf{L} \\ \hline \textbf{G} & \textbf{CGCGTCCTAT} \\ \hline \textbf{A} & \textbf{L} & \textbf{L} \\ \hline \textbf{A} & \textbf{L} & \textbf{L} \\ \hline \textbf{A} & \textbf{CGCGTCCTAT} \\ \hline \textbf{A} & \textbf{L} & \textbf{L} \\ \hline \textbf{A} & \textbf{ACCGATGT} \\ \hline \textbf{AATGGCGCTGT} \\ \hline \textbf{ATTGGCCACA \\ \hline \textbf{ATTGGTCGCACA \\ \hline \textbf{TGACCGATGT} \\ \hline \textbf{ATGGTGCACA \\ \hline \textbf{TGACCGTGTGACCT} \\ \hline \textbf{GCGTTGGTCACCT} \\ \hline \textbf{GCGTTGACCCTGT} \\ \hline \textbf{AATGGCTGCTGT} \\ \hline \textbf{AATACCGCGCT} \\ \hline \textbf{AAATGGCCGCT} \\ \hline \textbf{AAATGGCCGCTAT} \\ \hline \textbf{GGGTTGGCTCCT} \\ \hline \textbf{GGGTTGGCCCCT} \\ \hline \textbf{ACGCCCAAATTGACCGGCGT} \\ \hline \hline \textbf{ACGCCCGAACTGA} \\ \hline \textbf{CCCGGACTTATAA} \\ \hline \textbf{CCCCAACGGGGATTAAATT} \\ \hline \textbf{CCCCAACGGGGGCT} \\ \hline \hline \textbf{CCCGGACGTGG} \\ \hline \hline \textbf{CCCCAACGGAGG} \\ \hline \hline \textbf{CCCCGACGGGCGTG} \\ \hline \hline \hline \textbf{CCGGTCGGTG} \\ \hline \hline \textbf{CCGGCCCGGGCGTG} \\ \hline \hline \hline \textbf{CCGCTCGCGCGTA} \\ \hline \hline \hline \textbf{CCGCTCGCTGT} \\ \hline \hline \textbf{CCGCTCCGGCGCTC} \\ \hline \hline \hline \textbf{CCCCCTACCGGCGCTCCCT} \\ \hline \hline \hline \textbf{CCCCCTACCGGCGCTC} \\ \hline \hline \hline \hline \textbf{CCCCCTACCGGCGCTC} \\ \hline \hline \hline \hline \hline \hline \textbf{CCCCCTCCTCCCCTC} \\ \hline \hline \hline \hline \textbf{CCCCCTCCTC} \\ \hline \hline \hline \hline \textbf{CCCCCTCCTC} \\ \hline $	CGAAGCCGCG A E A R TGCGCGCCGTT ACGCGGCGCACT ACGCGGCGCACA L R Q CCTAAGGATCA G L I V CTTCCGTACTA GAAGCCATGA TAAGTAGATCA GAAGCCATGA TTAGGTGCGCA CCACATGACTA CCACATGACTA AGTAGACTAA M Q N TTGGGTGCGCG ACGCACACGCC Q T A TGGGTGCGCA ACCCCCGCC G Q E GGTCAGATCA GGTCAGACTAA M Q N TGGGTGCGCG ACGCACCGCC CCACTGACTAA CCACTGACTAA ACCCCGCCTCA GGTCAGACTG GGTCAGCTGA CCACTGACTAA ACCCCGCCTCA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA ACCCCGCCGCA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA CCACTGACTAA	$\begin{array}{c} \textbf{GCCTGCCGGCGAAAC} \\ \hline \textbf{GCCGCGACGGC} \\ \hline \textbf{A} & \textbf{Q} & \textbf{R} \\ \hline \textbf{GCCGCGAAAC} \\ \hline \textbf{CCGCGCGTTG} \\ \hline \textbf{A} & \textbf{R} & \textbf{F} & \textbf{A} \\ \hline \textbf{AAGGCTGCG} \\ \hline \textbf{TCCGGGACGGC} \\ \hline \textbf{TCCGGGACGGC} \\ \hline \textbf{CTAACCCTAT} \\ \hline \textbf{GATGGGATA} \\ \hline \textbf{GATGGGATA} \\ \hline \textbf{GATGGGGTA} \\ \hline \textbf{CTTGGCCCAT} \\ \hline \textbf{GAACCGGGTA} \\ \hline \textbf{CTTGGCCCACAG} \\ \hline \textbf{D} & \textbf{T} & \textbf{D} \\ \hline \textbf{CACATGTCA} \\ \hline \textbf{GACATGTCA} \\ \hline \textbf{GTGATGAGG} \\ \hline \textbf{V} & \textbf{Y} & \textbf{E} \\ \hline \textbf{CCCCGCGACTG} \\ \hline \textbf{GGCGTGTC} \\ \hline \textbf{GGCGTGTACGG} \\ \hline \textbf{GGCGTGTC} \\ \hline \textbf{GGCGCGTGAC} \\ \hline \textbf{GGCGTCTAC} \\ \hline \textbf{GGCGTCTACG} \\ \hline \textbf{GCGCGCGAAT} \\ \hline \textbf{GGCGCGGAA} \\ \hline \textbf{ACGGCCCTCA} \\ \hline \textbf{GCCGCGGAAA} \\ \hline \textbf{ACGCGCCCCG} \\ \hline \textbf{A} & \textbf{S} & \textbf{T} \\ \hline \textbf{GCCGCGGAAA} \\ \hline \textbf{ACGCGCCCC} \\ \hline \textbf{C} \\ \hline \textbf{TGCGCGGGAAA} \\ \hline \textbf{ACGCGCCCCC} \\ \hline \textbf{CGCTTAATGCC} \\ \hline \textbf{GCCCCGGAAA} \\ \hline \textbf{ACGCGCCCTC} \\ \hline \hline \textbf{CCTTATGCCC} \\ \hline \textbf{CATTTATGCC} \\ \hline \hline \textbf{CATTTATGCCG} \\ \hline \hline \textbf{CATTTATGCCG} \\ \hline \hline \end{array} $

Amino acid sequences:

1- *dgoT*: >tr|B5QUN8|B5QUN8_SALEP D-galactonate transporter OS=*Salmonella* enteritidis PT4 (strain P125109) GN=dgoT PE=4 SV=1

MDISVTAAQPGRRRYLTLVMIFITVVICYVDRANLAVASMHIQKEFGITKAEMGYVFSAF AWLYTLCQIPGGWFLDRIGSRLTYFIAIFGWSVATLLQGFATGLLSLIGLRAITGIFEAP AFPANNRMVTSWFPEHERASAVGFYTSGQFVGLAFLTPLLIWIQEMLSWHWVFIVTGGIG IIWSLVWFKVYQPPRLTKSLSQAELEYIRDGGGLVDGDAPAKKEARQPLTKADWKLVFHR KLVGVYLGQFAVNSTLWFFLTWFPNYLTQEKGITALKAGFMTTVPFLAAFFGVLLSGWLA DKLVKKGFSLGVARKTPIICGLLISTCIMGANYTNDPLWIMALMAIAFFGNGFASITWSL ISSLAPMRLIGLTGGMFNFIGGLGGISVPLVIGYLAQSYGFAPALVYISVVALLGALSYI LLVGDVKRVG

2- *dgoD*: >sp|B5QUN9|DGOD_SALEP D-galactonate dehydratase OS=*Salmonella* enteritidis PT4 (strain P125109) GN=dgoD PE=3 SV=1

MKITHITTYRLPPRWMFLKIETDEGVVGWGEPVIEGRARTVEAAVHEFADYLIGKDPARI NDLWQVMYRAGFYRGGPIMMSAIAGIDQALWDIKGKVLNAPVWQLMGGLVRDKIKAYSWV GGDRPADVIDGIEKLRGIGFDTFKLNGCEEMGVIDNSRAVDAAVNTVAQIREAFGSEIEF GLDFHGRVSAPMAKVLIKELEPYRPLFIEEPVLAEQAEYYPRLAAQTHIPIAAGERMFSR FEFKRVLDAGGLAILQPDLSHAGGITECYKIAGMAEAYDVALAPHCPLGPIALAACLHID FVSRNAVFQEQSMGIHYNKGAELLDFVKNKEDFSMDGGFFKPLTKPGLGVDIDEARVIEL SKSAPDWRNPLWRHADGSVAEW

3- SEN3645: >tr|B5QUP0|B5QUP0_SALEP 2-dehydro-3-deoxy-6-phosphogalactonate aldolase (Ec 4.1.2.21) (6-phospho-2-dehydro-3-deoxygalactonate aldolase) (2-oxo-3-deoxygalactonate 6-phosphate aldolase) OS=*Salmonella* enteritidis PT4 (strain P125109) GN=SEN3645 PE=4 SV=1

MQWQTNLPLIAILRGITPDDALAHVGAVVDAGFDAIEIPLNSPQWEKSISFVVKAYGGRA LIGAGTVLKPEQVDQLAGMGCKLIVTPNIQPEVIRRAVSYGMTVCPGCATATEAFSALDA GAQALKIFPSSAFGPGYISALKAVLPPDVPLFAVGGVTPENLAQWIKAGCVGAGLGSDLY RAGQSVERTAQQAAAFVNAYREAVK

4- *dgoK*: >tr|B5QUP1|B5QUP1_SALEP 2-dehydro-3-deoxygalactonokinase (Ec 2.7.1.58) (2-keto-3-deoxygalactonokinase) (2-oxo-3-deoxygalactonate kinase) OS=*Salmonella* enteritidis PT4 (strain P125109) GN=dgoK PE=4 SV=1

MTARYIAIDWGSTNLRAWLYQGDKCLESRQSEAGVTRLNGKSPDAVLAEVTTHWRDSATP VVMAGMIGSNVGWQNAPYLPVPALFSAIGEQLTAVGDNIWIIPGLCVSREDNHNVMRGEE TQLLGARELSPSSVYVMPGTHCKWVQTDTQQIHDFRTVMTGELHHLLLRHSLVGAGLPEQ EVSGDAYAAGLERGLNSPAVLPSLFEVRASHVLGHLAREQVSDFLSGLLIGAEVASMSES FAAQQAITLVAGPALISRYQQAFSAIGRDVSTVDGDMAFQAGIRSIAHAVAN

5- *dgoR*: >tr|B5QUP2|B5QUP2_SALEP Galactonate operon transcriptional repressor OS=Salmonella enteritidis PT4 (strain P125109) GN=dgoR PE=4 SV=1

MTLNKTDRIVITLGKQIVSGKYVPGSALPAEADLCEEFETSRNIIREVFRSLMAKRLIEM KRYRGAFIAPRNQWNYLDTDVLQWVLENDYDPRLISAMSEIRNLVEPAIARWAAERATSS DLAEIESALNDMIANNQDREAFNEADIRYHEAVLQSVHNPVLQQLNVAISSLQRAVFERT WMGDAANMPKTLQEHKALFDAIRHQDGDAAEQAALTMIASSTRRLKEIT

6- *yidA*: >tr|B5QUP3|B5QUP3_SALEP Uncharacterized protein OS=*Salmonella* enteritidis PT4 (strain P125109) GN=yidA PE=4 SV=1

MAIKLIAIDMDGTLLLPDHTISPAVKNAIAAAREKGVNVVLTTGRPYAGVHSYLKELHME QPGDYCITYNGALVQKAGDGSTVAQTALSYDDYRYLEKLSREVGSHFHALDRNTLYTANR DISYYTVHESYVATIPLVFCEAEKMDPNTQFLKVMMIDEPAVLDRAIARIPAEVKEKYTV LKSAPYFLEILDKRVNKGTGVKSLAEALGIKPEEVMAIGDQENDIAMIEYAGMGVAMDNA IPSVKEVANFVTKSNLEDGVAWAIEKFVLNPDHSSGHFPAR **7-** *torS*: >tr|B5QUN7|B5QUN7_SALEP Two-component sensor protein histidine protein kinase OS=*Salmonella* enteritidis PT4 (strain P125109) GN=torS PE=4 SV=1

MSTPSLTRRLWLAFALMAALTLLSTVIGWISLRVISQVEQTNTQALLPTMNMARQLSEAS AYELFSAONLTNADSEGVWLAOGKMLKAQSLKINHLLOALSEOGFNTSAIARQEKEIAOT LGQQGTLVGEILTLRAQQQQLSRQIAEAAESIAAQAHGQANNAATSAGATQAGIYDLIES GKGDQAERALDRLIDIDLEYVNQMNELRVNALRFKLLIVTLKDAQGLSDAEDTDEKLNQL VKILSRRQQRIEDPTVRAQIADALEKINQYTTLVTLFRKENAIRDQLQTLMANNLFQFTR FSTEVSQLVNAIEKRNEAGLARLTHASQRGQIGLVILGILALCSLSFILWRVVYRSVSRP LAQQTQALQRLLEGDIDSPFPEAAGVSELDTISRLMEAFRANVRKLNRHREDLAEQVRSQ TAELHALVLEHRQARAEAEKANEAKSTFLAAMSHEIRTPLYGILGTVQLLADKPLMANYR DDLQAINDSGESLLAILNDILDYSAIEVGGTNVSISEEPFEPRQLLNSALHLMHSRVQVA LIADFSEQLPSTLQGDPRRIRQIVINLLSNAAKFTDRGSIVLRTFCDDQSWFIEVEDTGC GIPEAKLTAIFKPFVQATGRRGGTGLGLAISASLAEAMGGTLTVTSTLHVGSCFRLQLPV RHPKPASKSAFRKPINLNGLRLLLIEDNMLTQRITAEMLTGKGVKVSVAESANDALRCLA EGESFDVALVDFDLPDYDGLTLAQQLMSQYPAMKRIGFSAHVIDDNLRQRTAGLFCGIIQ KPVPREELYRMIAHYLQGKSHNARAMLNEHQLAGDMASVGPEKLRQWIALFKDSALPLVE EIEAARAMNDDVNIKRLAHKLKSGCASLGMTQATEACRELELQPLSDIDIKTIVTQGVTA LDAWIADHPSP

Appendex 2: Translated sequence of induced genes (using Vector NTI program)

_	301	AATGAAGACG TTACTTCTGC	GCCACCCGAA CGGTGGGCTT	GGTGGCCGTT CCACCGGCAA	AGAACCCGGT TCTTGGGCCA	AGCGCTTCGC TCGCGAAGCG	TTATCAGGCC AATAGTCCGG	AGTGAAACCG TCACTTTGGC	AGACAGCTTT TCTGTCGAAA	AGCGGTGAGA TCGCCACTCT	CGCCAGCGGC GCGGTCGCCG
_	-1	TETTOCOTON	TTEECCE	GTOCTTTTCT	CCATCTTTAA	теретрессер	CACCATCATC	CCCCCC0T00	TATCORDOR	R H S	A L P Q
	101	ACAAGGCAGT	AACCGCCCAC	CACGAAAAGA	CCTAGAAATT	AGTCATGGGT	GTCCTACTAC	CGCGGCTATT	ATAGCTTGTC	TGACTCACGT	TACTTCTTGC
-	-1	Q E T M GGCCGTAACC	GATAATAGCG	H K E ACCAATGCCC	P D K I CCATAAACAG	GTTGAAGCTT	L I I AACTGTCCCA	A G I I TCCATGCCGC	D F L AGATCCTGCC	S L A AGACCGGCAA	I F F P CCGTCGCGAC
	-1	CCGGCATTGG	CTATTATCGC	TGGTTACGGG	GGTATTTGTC M F L	CAACTTCGAA	TTGACAGGGT	AGGTACGGCG	S G A	TCTGGCCGTT	GGCAGCGCTG
_	601	CTCATTTTTC GAGTAAAAAG	TTGAACAGGT AACTTGTCCA	CAGCGCTCAT GTCGCGAGTA	GGTAATCACG CCATTAGTGC	ACGGTAGACA TGCCATCTGT	GCGTCTGGTG CGCAGACCAC	CGCAAAACCG GCGTTTTGGC	CCAATACTCA GGTTATGAGT	TCAATGCAAT AGTTACGTTA	GGCGACATAC CCGCTGTATG
_	-1	E N K	K F L D) A S M	TIV	V T S L	T Q H	A F G	G I S M	í L A I	
	-1	CCCAATCAGC	ATTACGACTG	TTTTGGCTAG	CTTTAGTAGT	TTTGTCGCGG	CTACCATTTC	GATGTTGCTG	CGCGTAACTA	GCACCAGTAC	ACGTAGAAAA
_	801	CCATGAAGAA	CTTCGCCAGG	AAACCGCCTG	CAACGCAGCC	AAAGTCCGCC	GCCAGGAACG	GCAGCCAGGC	AAACATGGCG	ATCTCTTTTA	ACGGCAGATG
_	-1	E M F F	K A L	F G G A	V C G	F D A	A L F F	L W A	F M A	I E K L	P L H
	901	CATCACGTTA GTAGTGCAAT	ATCAGATACA TAGTCTATGT	GCGGCATCCA CGCCGTAGGT	GAAACTCAGC CTTTGAGTCG	GTTCCCCATG CAAGGGGTAC	CCGGGTCAGC GGCCCAGTCG	CAGAAAACGG GTCTTTTGCC	GTAATCGCCA CATTAGCGGT	ACGCCCAGAA TGCGGGTCTT	GTTACGCTTT CAATGCGAAA
_	-1	M V N	I L Y L	P M W	F S L	T G W A	P D A	LFR	T I A L	A W F	N R K K
	1001	AATTGCTAGA	GAAATTGCCG	GCCAAAAAAAT	AATAGCAGGA	CCTCTATCGA	AAGGACCGGC	AGAAGCTATA	TTGCGTCGAG	AAACACCCAC	TAGGTGCCCA
SEN2977	-1	GTTTATTTGG	E K V A CGAGTTATAA	P K K ATCAGGAACC	N D D C AGGTGATGGC	GAACAGCACG	E Q G CCGATCCCAC	D E I Y CGGTAATCAC	R L E AAACGCCATT	K H T TCAGTACCGA	I W P H TGCCGCTATC
434 aa-1305 b	p1	CAAATAAACC	GCTCAATATT	I L F W	TCCACTACCG	ETTGTCGTGC	GCTAGGGTG	GCCATTAGTG	TTTGCGGTAA	AGTCATGGCT	G S D
382-1683 Fr; 3,182,378	1201	AGCAAAGGTC TCGTTTCCAG	AGCATCGCCC TCGTAGCGGG	AGACCACCAG TCTGGTGGTC	CGGCGGCGCCC	AGCATCGCGC TCGTAGCGCG	CAATCGAAGT GTTAGCTTCA	ACCGATATTA TGGCTATAAT	AACAGTCCCC TTGTCAGGGG	CCGCAATACC GGCGTTATGG	GCGTTCTTTC CGCAAGAAAG
To; 3,183,682	-1	A F T	L M A V	GGCTTTAATC	P P A	L M A G	I S T	G I N	F L G G	A I G	R E K T
B5QYA9 Hexuropate	-1	CAGCCCTTGG	TAAGCCGCGA		GGCCGGCCTT	AGCGACTGCG	AAGCCAATCG	GEGTAGTCTG	GTGCGTCTTT	ACGGTCTGAA	TAGGTCGGCG
transporter	1401	CGGCCAGCGC	GTGCGCCATG	TTGATCAGCG	ACCATAACAA	GGCAAAGATA	AAGAAGCCGA	TTTTCAGTCC	GATGACATCC	ATCAAATAGC	CGGTGATCGG GCCACTAGCC
_	-1	G A L A	H A M	N I L S	W L L	A F I	F F G I	K L G	I V D	M L Y G	TIP
	1501	CTGCGCAATG GACGCGTTAC	GTGTAACAAA CACATTGTTT	GCTGGAAGGC CGACCTTCCG	GCTCACCACC CGAGTGGTGG	CAGGAATATT GTCCTTATAA	GTTGCTCATC CAACGAGTAG	AAAATGGAGT TTTTACCTCA	TCTTTCATCA AGAAAGTAGT	TCGCTGGCGC AGCGACCGCG	CGCAACGCTA GCGTTGCGAT
_	-1 1601	Q A I AGCGAGCTAC	T Y C L GGGACAGGTA	GTTTACGATG	G CCCTACGC	AGACCAGGCC	GATAAT CAC	F H L CATCTTAATT	E K M M TTGTCATCTT	1 A P A CATGATAAAG	A V S L CCTCGTTAAA
	-1	L S S R	CCCTGTCCAT		CAGGGATGCG	TCTGGTCCGG		GTAGAATTAA	AACAGTAGAA	GTACTATTTC	GGAGCAATTT
_	1701	AAACATTAGC	TCCGCGCCGT	CTGACCTGTA	GCGGGAAAAC	TGAAATCACA	AAAATTGGTT	GACCAATCAA	TATTTTCTCA	TCCGGACCGG	TCAAAATCTG
				I S D I S I S S S D I S D I S	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ACTTTACTCT	1-1-1-1-1-0 0(1(1-0.0	Configuration of Configuration	ATAAAAGAGT	leccercece	0.02010101010.0020.000
-	1801	GCGGCAACTC	CCGCCCTTAA	TATTCTTGAA	ATAAATACAT	ACTTTAGTGT CAACTCATTT	CAAGTTGAAT	GGTGAATGCT	ATAAAAGAGT TTTCACCCAG	AGGCCTGGCC CAAAAATTGA	CACAGATCAA
_	1801	GCGGCAACTC CGCCGTTGAG ATAAATAAAA	CCGCCCTTAA GGCGGGGAATT TTGGCAGACC	TATTCTTGAA ATAAGAACTT GTGAATGATG	CGCCCTTTTG ATAAATACAT TATTTATGTA ATATTTGTGA	ACTTTAGTGT CAACTCATTT GTTGAGTAAA TTTGGTTAGC	CAAGTTGAAT GTTCAACTTA CAATTTAGAC	GGTGAATGCT CCACTTACGA AATAAGGTTG	ATAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGAGAAA	AGGCCTGGCC CAAAAATTGA GTTTTTAACT AATTTTGGTT	AGTTTTAGAC CACAGATCAA GTGTCTAGTT GAATTTAGGG
-	1801 1901 +3	GCGGCAACTC CGCCGTTGAG ATAAATAAAA TATTTATTT	CCGCCCTTAA GGCGGGGAATT TTGGCAGACC AACCGTCTGG	GACIGGACAT TATTCTTGAA ATAAGAACTT GTGAATGATG CACTTACTAC	CGCCCTTTTG ATAAATACAT TATTTATGTA ATATTTGTGA TATAAACACT	ACTITAGIGI CAACTCATTT GTTGAGTAAA TTTGGTTAGC AAACCAATCG MGQ	CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG F R R	CTGGTTAGTT GGTGAATGCT CCACTTACGA AATAAGGTTG TTATTCCAAC R Y R	ATAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGAGAAA TGTACTCTTT K K R K	AGGCCTGGCC CAAAAATTGA GTTTTTAACT AATTTTGGTT TTAAAACCAA C R D M	AGTITITAGAC CACAGATCAA GTGTCTAGTT GAATTTAGGG CTTAAATCCC
-	1801 1901 +3 2001	GCGGCAACTC CGCCGTTGAG ATAAATAAAA TATTTATTTT GTCAGAACGG CAGTCTTGCC	AGCGCGCGCA CCGCCCTTAA GGCGGGAATT TTGGCAGACC AACCGTCTGG GTTTACAAAG CAAATGTTTC	CACTGGACAT TATTCTTGAA ATAAGAACTT GTGAATGATG CACTTACTAC CACGTTTCGA GTGCAAAGCT	CGCCCTTTTG ATAAATACAT TATTTATGTA ATATTTGTGA TATAAACACT TGTTGGCTGG ACAACCGACC	ACTITAGIGI CAACTCATTI GTIGAGTAAA TTIGGTIAGC AAACCAATCG M G Q CAATGGGICA GTIACCCAGI	TTTTAACCAA CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG P R R ACCACGAAGA TGGTGCTTCT	CTGGTTAGTT GGTGAATGCT CCACTTACGA AATAAGGTTG TTATTCCAAC R Y R CGTTATCGCA GCAATAGCGT	ATAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGAGAAA TGTACTCTTT K K R K AGAAGAGGAA TCTTCTCCTT	AGGCCTGGCC CAAAAATTGA GTTTTTAACT AATTTTGGTT TTAAAACCAA C R D M ACGAGATATG TGCTCTATAC	CACAGATCAA GAGTCTAGTT GAATTTAGGG CTTAAATCCC K Q T W AAACAAACCT TTTGTTGGA
-	1801 1901 +3 2001 +3	GCGCCAACTC CGCCGTTGAG ATAAATAAAA TATTTATTTT GTCAGAACGG CAGTCTTGCC W R W Y	CCGCCCGCA GCCGCCCTAA GGCGGGAATT TTGGCAGACC AACCGTCTGG GTTTACAAAG CAAATGTTCC C G F N	CACGTTCGA TATTACTAC ATAAGAACTT GTGAATGATG CACGTTACTAC CACGTTTCGA GTGCAAAGCT D P V	CGCCCTTTIG ATAAATACAT TATTATGTA ATATTATGTA ATATTAGTGA TATAAACACT TGTTGGCTGG ACAACCGACC T L S L	ACTITAGTGT CAACTCATTT GTTGAGTAAA TTTGGTTAGC AAACCAATCG CAATGGGTCA GTTACCCAGT V R Q	CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG P R R ACCACGAAGA TGGTGCTTCT À G À	CTGGTTAGTT GGTGAATGCT CCACTACGA AATAAGGTTG TTATTCCAAC R Y R CGTTATCGAA GCAATAGCGT T G V V	ATAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGAGAAA TGTACTCTTT K K R K AGAAGAGGAA TCTTCTCCTT T A L	AGGCCTGGCC CAAAAATGA GTTTTAACT AATTTTGGTT TTAAAACCAA C R D M ACGAGATATG TGCTCTATAC H H I	AGTITTAGAC CACAGATCAA GTGTCTAGGG CTTAAATCCC K Q T W AAACAAACCT TTTGTTTGGA P N G E CCC
-	1801 1901 +3 2001 +3 2101	GCGCCAACTC CGCCGTTGAG ATAAATAAAA TATTTATTTT GTCAGAACGG CAGTCTTGCC W R W Y GGCCCTGGTA CCGCGCACCAT	CCGCCCGCAT GCCGCGGAATT TTGGCAGACC AACCGTCTGG GTTTACAAAG CAAATGTTTC CG P N CGGACCTAAC GCCTGGATTG	CACCITCAAAAGAACTT GTGAATGATG CACCTTACTAC CACGTTTCGA GTGCAAAGCT D P V GACCCGGTAA CTGGGCCATT	CGCCCTTTIG ATAAATACAT TATTTATGTA ATATTTGTA TATAAACACT TGTTGGCTGG ACAACCGACC T L S E CGCTGTCAGA GCGACAGTCT	ACTITAGTGT CAACTCATTT GTTGAGTAAA TTTGGTTAGC AAACCAATCG CAATGGGTCA GTTACCCAGT V R Q TGTACGCCAG ACATGCGGTC	TTITAACCAA CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG F R R ACCACGAAGA TGGTGCTTCT A G A GCTGGCGCAA CGACCGCGTT	CTGGTTAGTT GGTGAATGCT CCACTTACGA AATAAGGTTG TTATTCCAAC R Y R CGTTATCGCA GCAATAGCGT T G V V CCCCCCTCCT GGCCCACCA	ATAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGAGAAA TGTACTCTTT K K R K AGAAGAGGAA TCTTCTCCTT T A L AAGCCCCTT1 T C CCCCAAT	AGGCCTGGCC CAAAAATTGA GTTTTTAACT AATTTTGGTT TTAAAACCAA R D M ACGAGATATG TGCTCTATAC H H I CACCATATAC GTGGTATAG	AGTITTAGAC CACAGATCAA GTGTCTAGTT GAATTTAGGG CTTAAATCCC K Q T W AAACAAACT TTTOTTTGGA P N G E CGAATGGAGA GCTAACTCT
-	1801 1901 +3 2001 +3 2101 +3 2201	GGGGCAACTC CGCCGTTGAG ATAAATAAAA TATTATTTT GTCAGAACGG CAGTCTTGCC W R W Y GGCCCTGTA CCGCGCACCAT E I W S AATTTGGTCG	CGGCCGCCTTAA CGCCGGCATTA TTGGCAGAACC AACCGTCTGG GTTTACAAAG CAAATGTTTC 7 G P N CGGACCTAAC GCCTGGATTG CCGCACCGATG I D E ATAGACCAGA	CACITGARCAT TATTACTGAA ATTAGAAACTT GTGAATGATG CACITTACTAC CACGTTTCGA GTGCAAAGCT D P V GACCCGGTAA CCGGGCCATT I Q K R TCCAGAAACG	CGCCCTTTG ATAAATAGT TATTATGTA ATATTTGGA TATAAACACT TOTTGGCTGG ACAACCGACC T L S E CGCGTCGCAGA GCGACAGTCT K A I TAAAGCTATC	ACTITATOTGT CAACCAATCG GTTCAGTAAA TTTGGTTAGC AAACCAATCG CAATGGGTCA GTTACCCAGT O V R O TGTACGCCAG ACATGCGGTC V E E GTTGAAGAGG	TTTTAACCAA CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG P R R ACCACGAAGA TGGTGCTTCT A G A GCTGGCGCAA CCGACCGCGTT A G L E CGGGCTCTGGA	GGTGATGTT GGTGAATGCT CCACTTACGA AATAAGGTG TTATTCCAAC GGTTATCCACA GGTATCGCA GGTCGCACCA GGCGCACCA GGCGCCACCA GGCGCCACCA	ATAAAAAAGAGT TTTCACCCAG AAAAGTGGGTC ACATGAGAAA TGTACTCTTT K K R R K AGAAGAGGAA TCTTCTCCCTT T A L HAGGGCTTA TTGCCGCAAT TTGCCGCCAT CTAGAGAGGCG	AGGCCTGGCC CAAAAATTGA GTTTTAACT AATTTTGGTT TTAAAACCAA C R D M ACGAGATATG TGCTCTATAC H H I ACGAGATATG GTGGTATAG V P I H TACCTATCA	AGTITTAGAC CACAGATCAA GGTGTCTAGGT GAATTTAGGG CTTAAATCCC K Q T W AAACAAACCT TTTGTTTGGA P N G E CGAATGGAGA GCTATGCCCT E D I CGAAGATATC
-	1801 1901 +3 2001 +3 2101 +3 2201 +3	GGGGCAACTC GGGGCAACTC GGCGAACTC GCCGTTGAG ATAAATAAAA TATTTATTTT GTCAGAACGG CAGTCTTGCC GGGCGTGGTA CCGGGACCAT E I W S AATTTGGTCG TTAAACCAG K T H	CGGCCGCCTTAA GCGGGGAATT TTGGCAGACC AACCGTCTGG GTTTACAAAG CAANGGTTC CGGACCTAAC GCGTGGATTG I D E ATAGACGAGA TATCTGCTCT T G Q Y	CACTIGARAT TATTCTTGAA ATAAGAACTT GTGAATGATG CACTTACTAC CACGTTCGA GTGCAAAGCT D P V GACCCGGTAA CTGGGCCATT I Q K R TCCAGAAACG TCCAGAAACG TCCAGAAACG	CGCCCTTTIG ATAAATACAT TATTATTAGTA ATATTTGGA TATAAACACT TGTTGGCTGG ACAACCGACC T L S L CGCTGCAGA GGGACAGTCT TAAAGCTATC ATTTCGATAG I K N	ACTITATORGY CAACCAATCG GTTCAGTAAA TTTGGTTAGC AAACCAATCG CAATGGGTCA GTTACCCAGT V R Q TGTACCGGTC V E E GTTGAACAGG GTTCAACAGG CAACTGCCCC Y Q Q T	TTTTAACCAA CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG P R R ACCACGAAGA TGGTGCTTCT A G A A G A CGACGGCGCAT A G L E CGGCCGGGTT CGGCCCAGACCT L R N	CTGGTTAGTT GGTGANTCT CCACTTACGA ANTAAGGTTG TTATTCCAAC GGTTATCGCA GGTTATCGCA GGTCACGGT CCCCCCCACTCC CACCAGACAC L A Q	ATAAAAAAGAGT TTTCACCCAG AAAACTGGGTC ACATGAGAAA TGTACTCTT K & K & K AGAAGAGGAA TCTTCTCCCTT T & L TTGCCGCAAT V E S GTAGAGACG CATCTCTCGC C G I I	AGGCCTGGCC CAAAAATTGA GTTTTAACT AATTTGGT TTAAACCAA ACGAOATATG TGCTCTATAC H H I ACGAOATATG GTGCTATAG V P I H TACCTATCCA ATGGATAGGT	AGTITTAGAC CACAGATCAA GGTCTAGAG GAATTAGGG CTTAAATCCC K Q T W AAACAAACCT TTTGTTTGGA P N G E CGATGGAGA GCTTACCCTCT CGAAGATATC GCTTATAG Y N F M
	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301	GCGCGAACTC CGCCGTTGAG ATAAATAAAA TATTTATTTT GTCAGAACGG CAGTCTTGCC W R W Y GCCCTGGTA CCGCGACCAT E I W S AATTTGCTCG TTAAACCAGC K T H AAAACCCACA	CGCCCGCCATAA GCCGGCATTA TTGCCAGACC AACCGCCTGG GTTTACAAAG CAAATGTTC CGGACCTAAC GCCTGCATAG TD E ATACACGAGA TATCGCCTT T G Q Y CCGGCCGTCATA	CACTOGARAT TATTCTTGAA ATAAGAACTT GTGAATGATG CACTTACTAC CACGTTTCGA GTGCAAAGCT I Q K R CGACCGGCAAA CTGGGCCATT I Q K R CCGGAAAGC CD L W CCGATTATGG GGTLAATACC	CGCCCTTTG ATAAATACAT TATTATGTA ATATTATGTA ATATTAGTA ATATTGGCTGG ACAACCGACC T L S L CGCTGTCAGA GCGACAGTCT K A I TTAAAGCATC ATTTGGATAG I K N ATCAAAAACT	ACTTAGTGT CAACTCATTT GTGAGTAAA TTTGGTTAGC AAACCAATCG CAATGGTCA GTAACCAATCG O V R O O V R O V R C TGTAGCCAGT V E E GTGAAGAGG CAACTGCGGTC Y Q Q T ACCAGCAAAC TGGTCGTTTGC	TTTTAACCAA CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG P R R ACCACGAGAGA TGGTGCTCTG A G A CGCTGGCGCAA CGACCCGCGTT L R N GCTGGCGGTATG GCCAGCCATTG	CTGGTARGT GGTGARTGCT CCACTTACGA AATAAGGTG TTATTCCAAC CGTTATCGAA GCAATAGCGT GGTCGCACCA W S V GTGGCCACCA W S V GTGGCCGCAACA L A Q CTGGCGGCAAT	$\begin{array}{c c} \mathbf{ATAAAAAGAGT}\\ \mathbf{TTTCACCCAG}\\ \mathbf{AAACTGGGTC}\\ \mathbf{ACATGAGAAA}\\ \mathbf{TGTACGAGAA}\\ \mathbf{TGTACGAGAA}\\ \mathbf{TGTACGAAA}\\ \mathbf{TGTACGAAA}\\ \mathbf{TGTCTCCTT}\\ \mathbf{T} & \mathbf{A} \\ \mathbf{T}\\ \mathbf{T} & \mathbf{A} \\ \mathbf{TGCCGCAAT}\\ \mathbf{V} & \mathbf{E} \\ \mathbf{S}\\ \mathbf{GTAGGAGACC}\\ \mathbf{C} & \mathbf{G} & \mathbf{I} & \mathbf{Y}\\ \mathbf{GCGCTATCAA}\\ \mathbf{GCGCTATCAA}\\ \mathbf{T} \\ \mathbf{T} & \mathbf{A} \\ \mathbf{GCGCTATCAA}\\ \mathbf{T} \\ \mathbf{GCGCTATCAA}\\ \mathbf{T} \\ \mathbf{T} $	AGGCCTGGCC CAAAAATTGA GTTITAACT AATTTGGT TTAAAACCAA CGAGATATG TGCTCTATAC H H I CGGGTATAG GTGGTATAG Y P I H TACCTATCCA ATGGATAGGT TACGGTTGC TACGGTTGC	AGTITTAGAC CACAGATCAA GAATTTAGGG CTTAAATCCC K Q T W AAACAAACT TTTOTTTGGA P N G E CGAATGGAGA GCTTACCTCT GCGAATGGAGA GCTTACCTCT CGAAGGATATC GCTTATAG Y N F M TATATGAAAT
SEN2978 394 aa- 1185 b 2043-3269	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 p. +3 +3	GGGGCAACTC GGGGCAACTC GGGGCAACTC GCGCGTTGAG ATATATATATTT GTCAGAACGG CAGTCTTGCC W R W Y GGCGCTGGTA CCGCGACCAT E I W S AATTTGGTCG TTAAACCACGC K T H AAAACCACAC	CCGCCCTTAA GCCGGCATTA AACCGTCTGG GTTTACAANG CAAAAGTATC CAAAAGTATC CCGACCTGATTG I D E ATAGACGAGA TATCTGCTCT T G Q Y CCGGTCAGTAG GCCCAGTCAT CCCGGTCAGTA	TATTCTTGAR TATAGAACTT GTGAATGATG CACTTTCGA GTGCAATGATG GTGCAAAGCT D F V GACCCGGTAA CTGGGCCATT I O K R I CCAGAAACG AGGTCTTTGC CGATTATGC GCTAAATACC R T D	CGCCCTTTIG ATAAATACAT TATTATAGAT ATAATACAT TATTATAGAT ATAATACAT TOTTGGCTGG ACAACCGACC T L CGCTGTCAGA GCGACAGTCT TAAACCTATC ATTTGGATAG I K ATCAAAAACT TAGTTTTGA L E E Y	ACTITAGIGI CAACTCATTT GTTGAGTAAA TTTGGTTGGC AAACCAATCG M G Q CAATGGGTCA GTTACCCAGT V R Q TGTACGCCAG V F E GTTGAACGCCAG CAACTGCCCC Y Q Q T ACCAGCAAAC TGGTCGTTTG L P D	TTTTACCAA GAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG P R R ACCACGAAGA TGGTGCTTCT A G A GCTGGCGCAA CGACGCGCT A G L R A G A GCCGGCTCTGGA GCCCAGACCT L R N GCTGCCGTAAC GCCCCGACCATTG G S K	CTGGTCATT CCACTTACGA AATAAGGTTG TTATTCCAAC R Y R GGTTATCCCAC GGTATACGCA GGTATACGCA GGCGCACCAC L A Q CTGGTCTCTG CACCAGACAC L A Q CTGGCCCATA A L R F	ATAAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGACAAA TGTACTCATTT K K R R AGAAGAGGAA TCTTCTCCTTT T A L HCCCCCCAT Y E S GCGCATCTA GCGCGTACTA B Q I D Q I	AGGCCTGGCC CAAAAATTGA GTTTTTAAACT AATTTGGTT TTAAAACCAA C R D M ACGAGATATG TGCTCTATAC H H I H GTGTTATAG GTGGTATAG TGCTTTTAATTG TGCTATAG TGCTATAG TGCTTTAATTG TGCTATAG TGCTTTAATTG TGCTTTTAATTG TGCTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTAATTG TGCTTTTG TGCTTTTTG TGCTTTTTG TGCTTTTTG TGCTTTTG TGCTTTTG TGCTTTTG TGCTTTTG TGCTTTTG TGCTTTTG TGCTTTTG TGCTTTTG TGCTTG TGCTTG TGCTTTG TGCTG TGCTG TG	$\begin{array}{c} \textbf{AGTITTAGAC} \\ \textbf{CACAGATCAA} \\ \textbf{GTGTCTAGTT} \\ \textbf{GAATTTAGGG} \\ \textbf{CTTAAATCCC} \\ \textbf{K} & \textbf{Q} & \textbf{T} & \textbf{W} \\ \textbf{AAACAAACCT} \\ \textbf{TITOTTGGA} \\ \textbf{F} & \textbf{N} & \textbf{G} \\ \textbf{F} & \textbf{CGAATGGAGA} \\ \textbf{GCTTACTCT} \\ \textbf{GCTTACCCCT} \\ \textbf{GCTTACCTT} \\ \textbf{GCTATAGGAGA} \\ \textbf{GCTTCATAGGAGA} \\ \textbf{GCTTCATAGGAGA } \\ GCTTCATAGGA $
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To: 3,185 271	1801 1901 +3 2001 +3 2101 +3 2201 2301 	GGGGCAACTC GGGGCAACTC CCCCCTTAG ATATATATAT TATTTATTTT GTCAGAACGG CAGTCTTGCC W R W Y GGCGCCGGTA CCGCGCCGCTA CCGCGCCCGCA ATTTGGTCG TTAAACCACC K T H AAACCCACC K T H AAACCACCACC M P V L M P V L M P V C ACGGCCACGA	CCGCCCTTAA CCGCCCCTTAA TTGCCAGACC AACCGTCTGG GTTTACAAAG CAAATGTTC CGGACCTAAC GCCTGGATTG GCCTGGATTG GCCTGGACTAC CCGGCCAGTAAT CCCGGCCAGTAAT GCCCGGCCAGTAAT GCCCGGCCAGTAAT GCCCGGCCAGTAAT	CACTOGARAT TATTCTTGAA ATAAGAACTT GTGAATGATG CACTTACTAC CACTTACTAC D P V CACCGGTAA CTGGGCCATT I Q K R TCCAGAAACG AGGTCTTTGC CGATATTACG GCTAATACC R T D CCTACGGATC GCTACGGATC	CGCCCTTTIG ATAAATACAT TATTATGTA ATATTTGGA TATAAACACCT TGTTGGCTGG ACAACCGACC T L S L CGCGCGCCAGA GCGACAGTCT K A I TAAAGCTATC ATTTCGATAG L E Y V TGGAATACGT ACCTTATGCA	ACTITATION THAN THE CAACCATCA GTTCAGTAAA TTTGGTTAGC AAACCAATCG CAATGGGTCA GTTACCCAGT OTV R Q CACATCGGCCA CACATCGGCCA V E E GTTCAACAGG CCAACTTCTCCC Y Q Q T ACCAGCAAAC TOGTCGGTTC L P D ATTGCCGGAT TAACGGCCTA	TTTTACCAA CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG P R R ACCACGAAGA TGGTGCTTCT A G A GCTGGCGCAA GCCGGCGCAA GCCCGGCGTT GCCCGGCAT CGACCGCATG GCCCAGACCT G S K GGTCCAAAAA CCAAGTTTC	CTGGTAGTT GGTGANTGCT CCACTTACGA ANTAAGGTTG TTATTCCAAC GCATTACGA GCATTACGCA GGTCGCTC CCCCCCAGACCCA CCCCCAGACACCA L A Q CCCCCCGCGTTA GACCGCATT GCCCCCCGCTTT GCACCGCAAA	ATAAAAAAGAGT TTTCACCCAG AAACTGGGTC ACATGAGAAA TGTACACTATT AGAAGAGGAA TCTCTCCCTT T T ACACGAGAA TTGCCGCAAT Y S GTACAGAGCC CATCTCCCCTT T A TGCCGCAAT Y S GTACAGAGCC C G T CACCCATAGAACC CGCCATACT CGCCATACAT CACCGCACATT ACTGGTCTAA	AGGCCTGGCC CAAAATTGA GTTTTAACT AATTTTGGT TTAAACCAA ACGAGATATG TGCTCTATAC H H I ACGAGATATG GTGGTATAG V P I H TACCTATCCA ATGGATAGGTTTGC ATGCCAAACG E F A GAATTCGCCG CTTAAGCGGC	XGTITTAGAC CACAGATCAA GGTGTCTAGTT GAATTTAGGG CTTAAATCCC K Q T W AAACAAACCT TTTGTTGGA P N G E CGAATGGAGA GGTTACCTCT CGAATGGAGA GGTTACCTCT GGTTACCTCT GGTTACCTCT GGTTACCTCT GGTTACTTATAG GCTTATAG Y N F M TATAATTTA ATATTAAAAT A F E L GGTAGCTGA
SEN2978 394 aa- 1185 b 2043-3269 Fr; 3,185,271 B5QYB0	1801 1901 +3 2001 +3 2101 +3 2201 +3 2201 +3 2301 +3 2401 +3 2501	GGGGCAACTC GGGGCAACTC GGGGCAACTC GGCGCTGAG ATAAATAAAA TATTTATTTT GTCAGAACGG CAGTCTGCC W R W Y GGCCTGGTA CGGCCACGA K T H AAATTGGTCG TTAAACCAAC K T H AAATCCACA M P V L TGCCGGTGCT ACGGCCACGA L H L GGATATCCTG	CGCCCTTAA GCCGGCATTA TTGCCAGACC AACCGTCTGG GTTTACAAAG CAAATGTTC CGGACCTAAC GCCTGGATTG T D E ATAGACGAGA TATCGCCTT T G Q Y CCGGCCATCAT GGCCAGTCAT GGCCAGTCAT GGCCAGTCAT CCGGACTGGACA	CACTIGARAT TATTCITGAA ATAAGAACTT GTGAATGATG CACTTACTAC CACGTTTCGA GTGCAAAGCT I Q K R TCCAGAAAGC CACCGTAA CTGGGCCATT I Q K R TCCAGAAAGC CCTACTACTAC CGATTATGAC CGATTATGAC CGACCGGAAC	CGCCCTTTIG TATAATACAT TATTATGTA ATATTATGTA ATATTAGTA TGTTGGCTGG ACAACCGACC T L S L CGCTGTCAGA GCGACAGTCT K A I TAAAGCATC ATTTGGATAG L E Y V TGGAATACGT ACCTTATGCA D Y T CGACTATACG	ACTTACTGT CAACTCATTT GTGAGTAAA TTTGGTTAGC AAACCAATCG CAATGGGTCA GTTAACCAGTC V E C GTGAGCAGC ACATGCGGTC V E E GTGAGAGAGAT ACCACCAATC CAACTGCGCTTC L P D ATTOCCGGTT TAACGGCCTA CCACACCAACAA	TTTTAACCAA CAAGTTGAAT GTTCAAACTTA CAATTTAGAC GTTAAATCTG P R R ACCACGAGAGA TGGTGCTCTGT A G A CGTCGCCCCA CGACCGCGTT L R N GCTCGCGCAAC GCCCAGCATTG GCCCAGCATTG CCAACGCATTG CCAAGTTTTC	CTGGTARGTT GGTGARTGCT CCACTTACGA AATAAGGTG CGTATCCAAC CGTTATCCAAC CGTTATCGAA GCAATAGCGT GGCGCACCA W S V GTGGTCGTG GGCGCACCA L A Q CTGGCGCAACA CACCAGCAA A L R F CGTTCGCGTTA GCACCGCGTTA GCACCGCATA E R R AGACGCACGCGT	$\begin{array}{c c} \mathbf{ATAAAAAGAGT}\\ \mathbf{TTTCACCCAG}\\ \mathbf{AAACTGGGTC}\\ \mathbf{ACATGAGAAA}\\ \mathbf{TGTACAGAAA}\\ \mathbf{TGTAGAAAA}\\ \mathbf{TGTAGAAA}\\ \mathbf{TGTAGAAA}\\ \mathbf{TTTCTCTCT}\\ \mathbf{T} & \mathbf{A} & \mathbf{L}\\ \mathbf{T} & \mathbf{A} & \mathbf{L}\\ \mathbf{T} & \mathbf{T} & \mathbf{A} & \mathbf{C}\\ \mathbf{C} & \mathbf{G} & \mathbf{I} & \mathbf{Y}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{G} & \mathbf{G} & \mathbf{G} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T}$	AGGCCTGGCC CAAAAATTGA GTTTTAAAT TTAAAACCAA C R D M ACGAGATATG TGCTCTATCA H H I TACCTCTATCA GTGGTATAG V P I H TACCTATCCA ATGCATAGGTTTGC TTC C TTC C TC C TTC C TACGGTTGC C TTC C TACGGTTGC C TACGGTTGC C TACGGTTGCC C TAAGCAGC TC C TC	XGTITTAGAC CACAGATCAA GAATTTAGGG CTTAAATCCC K Q T W AAACAAACT TTTOTTTGGA P N G E CGAATGGAGA GCTTACCTCT GCGAATGGAGA GCTTACCTCT CGAATGAGAA GCTTACCTCT CGAATGAAAATTAA A F E L CGTTCGAACT A F E L CGTTCGAACT A F E L CGTTCGAACT A F K A AGACAAAGCA
SEN2978 394 aa- 1185 b 2043-3269 Fr; 3,184,087 To; 3,185,271 B5QYB0 Mannonate	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 p. +3 2401 +3 2501 +3 2501	GGGGCAACTC GGGGCAACTC GGGGCAACTC GCGCGTTGAG ATAATAAAAA GTTTATTTT GTCAGAACGG CAGTCTTGCC W R W Y GGCGCTGGTA CCGGCACCAT E I W S AATTTGGTCG TTAAACCACC K T H AAAACCACAC M P V L TTGGGGGTGT M P V L TGCGGTGCT M P V L GGCATACGAC CGTAACGAC CGCATACGAC CGTAAGGAC CGTAAGGAC CGTAAGGAC	CCGCCCTTAA GCCGGCATTA AACCGCAGACC AACCGCCTCG CAAATGTTCC CAAATGTTCC CCGACCTAAC GCCTGGATGCC T D E ATAGCGAGA TATCCGCGAG CCCGACCTCAT CCGGCCGCTCATA GCCCGGCCAGTCA CCCGGCCGAGCCA K R P AAGCGCCGCG K R P AAGCGCCGCG R N I I	TATTCTTGAA ATAAGAACTT TATTCTTGAA ATAAGAACTT GTGGAATGATG CACTTACTAC CACGTTTCGA GTGCAAAGG TGCGAAAGG AGGTCTTTGC CGAAACGAAA	CCCCCTTTIG ATAAATACAT TATTATAGAT ATATTATGTA ATATTTGGATAGA TGTTGGCTGG ACAACCGACC T L CCCTGTCAGA GCGACAGTCT TAAACCAACC K A TAAACCTATC ATTTGGATAGA CCCTGTCAGA GCGACAGTCT TAAACCTATC ATTTGGATAGC L E Y TGGAATACGT CCATATAGC D Y CGCACTATAGG CCATATACG F G	$\begin{array}{c} \text{ACTTATGTGT}\\ \textbf{CAACTGATTT}\\ \textbf{CAACTGATTAGTAGA}\\ \textbf{TTTGGTTAGC}\\ \textbf{AAACCAATCG}\\ \textbf{AAACCAATCG}\\ \textbf{CAATGGGTCA}\\ \textbf{GTTAACCCAGTC}\\ \textbf{GTTAACCCAGCAATCGC}\\ \textbf{CATTGCGCCAG}\\ \textbf{ACATGCGCCAG}\\ \textbf{ACATGCGCCAG}\\ \textbf{CAACTGCGCCAG}\\ \textbf{CAACTGCGCCAG}\\ \textbf{CAACTGCGCCAG}\\ \textbf{CACTTCCCC}\\ \textbf{Y} & \textbf{E} & \textbf{E} \\ \textbf{GCAGAGGCAAA}\\ \textbf{CAGCCTTCTTCT}\\ \textbf{A} & \textbf{E} & \textbf{E} \\ \textbf{GCAGAAGAAA}\\ \textbf{GCAGAAGAAA}\\ \textbf{CACCTCTCTCTT}\\ \textbf{E} & \textbf{E} & \textbf{G} & \textbf{Y} \end{array}$	TTTTACCAA GAGTTGAAT GTTCAACTTA CAATTTAGAC TTAAATCTG P R R ACCACGAAGA TGGTGCTTCT A G L E GCTGGCCCAA CGACCGCGTA CGGCTCTGGA GCCCAGGACCT G S K GGTTCAAAAG GCTCCAGGACATTTCC I A Q A TTCCACGGAC	CTGGTCATTOCCA GCAATAGGTTG TTATTCCAAC R Y R GCATATCCAAC CCACTATCCAAC CCACTATCCCAC GCAATACGCA GCAATACGCA T G V V CCCCCCCCCCCAC CACCAGACAC L A Q CTGGCCTCTG CACCGCCATA A L R F CGTACGCCAAA CCCCCCCTAA A L R F CGCACGCACAT GCAACGCACGAC CCCCCCCCACA Q F R	ATAAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGACAAA TGTACACCAGT TGTACACAAA TGTACACAAA TTTCACCTTT TTTCACCAGAAA TCTTCTCCTT TTTCACCAGAAA TTTCACCGCAAT V GCGGCTACTAA GCGCGTACTAA CGCGCTATCAA CGCGCTATCAA ACTGGCCACATAAAT D <q< td=""> TGACCAGACTAAA TGACCCAGACTAAAT D<q< td=""> F ACTGGCCACACA AAGCGGCGTGCTA AAGCGGCGACCA AAGCGGCGTGGT</q<></q<>	AGGCCTGGCC CAAAAATTGA GTTTTAAACT AATTTGGTT TTAAAACCAA C E D M ACGAGATATG TGCTCTATAC H H I H GAGATAGGT TGCTATACA ATGGATAGGT TACCGTTTGC ATGGATAGGT GAATTGCCCG M S E F A GAATAGCCAACG M S E F A GAATAGCGAGGA ACTCGCCCCT T Y K	XGTITTAGAC CACAGATCAA GTATTAGGG CTTAAATCCC K Q K Q W AACAAACCT TTTOTTGGA T P N GGAATGAGAG GCTTACCTCT CGAATGAGAGA GCTTACCTCT GGAATGACAA GCTCTATAG Y N TATAACAAAT A A F E CGTACGAAT GCAACTGAA GCAACAAACAAT A A F E CGTTCGAACT GCAACAAGCA TCTGTTGA D D K A ACAAAACCT
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 +3 2401 +3 2401 +3 2501 +3 2601	GGGGCAACTC GGGGCAACTC GGGGCAACTC GGGGCAACTC GCAGTCTTGCC W R W Y GGCGCTGGTA CCGGGCCGGTA CCGGGCCGGTA CCGGGCCGGACCAT TTTGGGTGT TTAAACCACC K T H AAAACCACC K T H AAAACCACCACA TTTTGGGTGT TTTGGGTGT TTTGGGTGT CGGCACACCACA CGGCACCACA CGGCACCACA CGTATAGGAC R L T CCTCTTGGCCC R L T	CCCCCCTTAA CCCCCCCTTAA CCCCCCCTTAA ACCCCCCTTAA CAAACGTACC CAAACGTATTCC C G P N CCCGACCTGATTG C G P N CCCGACCAGTAT CCCGACCAGTA CCCGCCCCGATA GCCCACCTGATA GCCACCTGATA CCCGACCTGACA CCCGACCCCGCC R N I I CCTTCCCAGCCC	TATTCTTGAA ATAAGAACTT TATTCTTGAA ATAAGAACTT GTGAAATGATC GTGCAAATGATC GTGCAAAGGT D P V GACCCGGCAAT TCCAGGAAACG AGGTCTTGC CGATTATGG GCTAAATACC R T D CCAGAAACG AGGTCTTAGG GCAACCTAG G A E A GAGCGAGAAGC CTCGGCTTCG CTCCGGTTTC GCACCCTACCTAG G A C A GAGCGAGAAGC	CCCCCTTITG ATAAATACAT TATTATAGAT ATAATACAT TATTATAGAT ATAATACAT TATTATAGAT ATAATACAT TOTTGGTGGA ACAACCGACC T ACAACCGACC CGCTGTCAGA GGGACAGTCT K TAAAACTATC ATCAAAAACT TAGATATGCA L E L T CGACTATAGCA ACCTTATAGCA D Y CGACTATATAGC F G CCTGGCGCGC C	ACTITACIGIT CAACTCATTT GTTCAGTAAA TTTGGTTAGC AAACGATCG CAATGGGTCA GTTAACCAGT V E C GTTCAACAGG CACATGCGGTC V E E GTTCAACAGG CAACTGCGTC CACATCGCCAC TGTCGCGTTC C L P D ATTGCCGGAT TAACGGCCTA ACCAGCAAAA TGGCGCTTC C L P D ATTGCCGGAT TAACGGCCTA CACCAGCAAAA	TTTTACCAA CAAGTTGAAT GTTCAACTTA CAATTTAGAC GTTAAATCTG P R R ACCAGGAGA TGGTGCTTCT A G A GCTGGCGCAA CGACCGCGAT CGACCGCATTG G S K CGACCGCATTC CAACTTCAAAAG GCTCAGAAG CCAAGTTCCAAAG CCAAGTTCCAAAG CCAAGTTCCAAGAC TTGCCACGCGCT T L D TACCGCGCAT	CTGGCGCATAGGA AATAAGGTTG CCACTTACGA AATAAGGTTG TATTCCAAC R Y R GGTATACCCA GCAATAGCGT GGLGCACCA W S V CCCCCCCAGACAC W S V GGGCGCACCA CACGCGCAT A C CTGGCCCACA CACGCGCAA E R R AGACCGACGA CACGCCGCAA CACGCCGCAA CACGCCGCCAC	ATAAAAAGAGT TTTCACCCAG AAACTGGGTC ACATGAGAA TGTACACAAA TGTACACAAA TGTACACAAA TGTACACAAA TTTCACCTT TA T AAAGAGGAAT T T GCGGTATCTAC CCCCATAGAT ACCGGTCTAA F A T TGACCACATT ACCGGTCTAA F A T TGCCCACACTACA AAACCGCTGGT Q H AAACCCGTGGT Q H AAACCCGTGACGAC <th>AGGCCTGGCC CAAAATTGA GTTTTAACT AATTTGGT TTAAACCAA C R D M ACGACATATG TOCTCTATAC H H I ACGACATATG GTGGTATAG T V C TACGATAGA ATGGATAGG T V C TACGGTTGC ATGCCAACG CTTAAGCGAGGA ACTCGCTGCT T Y K GACGTATAAA</th> <th>$\begin{array}{c} \textbf{AGTITTAGAC} \\ \textbf{AGTITTAGAC} \\ \textbf{GTGTCTAGTT} \\ \textbf{GAATTTAGGG} \\ \textbf{CTTAAATCCC} \\ \textbf{K} & \textbf{Q} & \textbf{T} & \textbf{W} \\ \textbf{AAACAAACCT} \\ \textbf{TTTGTTTGGA} \\ \textbf{P} & \textbf{N} & \textbf{G} \\ \textbf{E} \\ \textbf{CGAATGGAGA} \\ \textbf{GCTTACCTCT} \\ \hline \textbf{E} & \textbf{D} & \textbf{I} \\ \textbf{CGAAGAGAGA} \\ \textbf{GCTTACTTAA} \\ \textbf{GCTTCTATAG} \\ \textbf{Y} & \textbf{N} & \textbf{F} \\ \textbf{M} \\ \textbf{TATACTTAA} \\ \textbf{ATATTGAAAT} \\ \textbf{A} & \textbf{F} & \textbf{E} & \textbf{L} \\ \textbf{CGTTGGAACTTGA} \\ \hline \textbf{CGTAGAAACGA} \\ \textbf{TCTGTTTGAT} \\ \textbf{D} & \textbf{K} & \textbf{A} \\ \textbf{AGACAAAGCA} \\ \textbf{TCTGTTTGT} \\ \textbf{D} & \textbf{I} & \textbf{D} & \textbf{K} \\ \textbf{GATATCGATACCTTGA} \\ \hline \textbf{CTTTAGAATACCAAAGCA} \\ \textbf{CTTTTGTTGAACCAAAGCA} \\ \textbf{CTTTTGTTGAACGAAAGCA} \\ \textbf{CTTTTGTTTGAAAACGAAAGCA} \\ CTTTTGTTTGAAACGAAAGCAAAGCAAAGCAAAGCAAAG$</th>	AGGCCTGGCC CAAAATTGA GTTTTAACT AATTTGGT TTAAACCAA C R D M ACGACATATG TOCTCTATAC H H I ACGACATATG GTGGTATAG T V C TACGATAGA ATGGATAGG T V C TACGGTTGC ATGCCAACG CTTAAGCGAGGA ACTCGCTGCT T Y K GACGTATAAA	$\begin{array}{c} \textbf{AGTITTAGAC} \\ \textbf{AGTITTAGAC} \\ \textbf{GTGTCTAGTT} \\ \textbf{GAATTTAGGG} \\ \textbf{CTTAAATCCC} \\ \textbf{K} & \textbf{Q} & \textbf{T} & \textbf{W} \\ \textbf{AAACAAACCT} \\ \textbf{TTTGTTTGGA} \\ \textbf{P} & \textbf{N} & \textbf{G} \\ \textbf{E} \\ \textbf{CGAATGGAGA} \\ \textbf{GCTTACCTCT} \\ \hline \textbf{E} & \textbf{D} & \textbf{I} \\ \textbf{CGAAGAGAGA} \\ \textbf{GCTTACTTAA} \\ \textbf{GCTTCTATAG} \\ \textbf{Y} & \textbf{N} & \textbf{F} \\ \textbf{M} \\ \textbf{TATACTTAA} \\ \textbf{ATATTGAAAT} \\ \textbf{A} & \textbf{F} & \textbf{E} & \textbf{L} \\ \textbf{CGTTGGAACTTGA} \\ \hline \textbf{CGTAGAAACGA} \\ \textbf{TCTGTTTGAT} \\ \textbf{D} & \textbf{K} & \textbf{A} \\ \textbf{AGACAAAGCA} \\ \textbf{TCTGTTTGT} \\ \textbf{D} & \textbf{I} & \textbf{D} & \textbf{K} \\ \textbf{GATATCGATACCTTGA} \\ \hline \textbf{CTTTAGAATACCAAAGCA} \\ \textbf{CTTTTGTTGAACCAAAGCA} \\ \textbf{CTTTTGTTGAACGAAAGCA} \\ \textbf{CTTTTGTTTGAAAACGAAAGCA} \\ CTTTTGTTTGAAACGAAAGCAAAGCAAAGCAAAGCAAAG$
SEN2978 394 aa- 1185 b 2043-3269 Fr; 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 +3 2501 +3 2501 +3 2601 +3 2501	GGGGCAACTC GGGGCAACTC GGGGCAACTC GGCGCTGAG ATAAATAAAA TATTTATTTT GTCAGAACGG CAGTCTGCC W R W Y GGCCTGGTA CGGCCACGG K T H AAATTGGTCG TTAAACCAAC K T H AAATCCACA K T H AAATCCACC K T H AAATCCACC CGTATAGGAC CGCACGGGCCCA R L T CGCCTCGACCC CGCACACTGGG K A K L	CGGCCGCTTAA GCCGGCATTA TTGGCAGACC AACCGTCTGG GTTTACAAAG CAAATGTTC CGGACCTAAC GCCTGGATTG CGGACCTAAC GCCTGGATTG T G Q Y CCGGCCATAC GGCCAGTCAT GGCAACGGCA K R P AAGCGTCCGG TTCGCAGGCC R N I I GCAACATTAT CCTTCTAATA 2 R E H	CACTOGARAT TATTCITGAA ATAAGAACTT GTGAATGATG CACTTACTAC CACGTTCGAA GTGCAAAGCT I Q K R CACCGGCAAA CTGGGCCATT I Q K R CCGGCAAAACC R T D CGATTAGGAACC R T D CGATAGAAACC R T D CGATAGAACC R T D CGATAGAACC R T D CGACCGGATA GCAAAGCC R T D CGACCGGACC R T D CGACCGGACC R T D CGACCGGACC R T D CGACCGGACC CCGTCCTCG C A G L ACGGCCAAAT F A Y	CGCCCTTTIG ATAAATACAT TATTATAGTA ATATTTATGTA ATATTTGGA TGTTGGCTGG ACAACCGACC T L CCCTCTCAGA GCGACAGTCT K A TAAATACGATAG ATTGATAGCAACCGACC K A TAAAGCAATC ATAGTTTTGA ATTTGGATAG L E ACCTAAGAACGT CGACTATACGT CGGCGCCCC F G CCGGCGCCC F C CCGCGCCCC F L C K CCGCGCCCC	ACTTACTGT CAACTCATTT GTTGAGTAAA TTTGGTTAGC AAACGATCG M G Q CAATGGGTCA GTTAACCAGT V E C GTTGAGCCAG ACATGCGGTCA V E E GTTGAAGAGG CAACTGCGGTC V E C CACATGCGGTCA CAACTGCGGTCA CAACTGCGGTCA CAACTGCGGTCA CAACTGCGGTCA CAACTGCGGTCA CAACTGCGGTCA CAACTGCGGTCA CAACTGCGGCA ACTGCCGGTCA CAACTGCGGCA CAACAGGCCA CAACGGCCA CAACGGCCA CAACGGCCA CAACGGCCA CAACGGCCA CAACGGCCA CACGCCACGACAACA CGCCTCCTTTT E E G V AAGAAGGCCA	TTTTAACCAA CAAGTTGAAT GATCAAACTTA CAATTTAGAC GTTAAAATCTG P R R ACCACGAAGA TGGTGCTCTC A G A GCTGGCGCAA CGACCGCGTT L R N GCTGCGGTAAC GCCCGAGACT CGACCGCATC CGACCGCATC CCAAGTTTC CCAAGTTTC CCAAGTCCAAG TTGCCCAAGT TACCGCGAT TACCGCGACT	CTGGTAGTT GGTGATTGCT CCACTTACGA AATAAGGTTG CGTATCCAAC CGTTATCGAA GCAATAGCGT T G V V GTGGTCGTGGC GGCGCACCA W S V GTGGTCGTGG GGCGCACCA L A Q CTGGCGCAAC A L R F CGTTGCGTTT GCAACGGCAA A L R R AGACGGCAAC E R R AGACGCACGCAT TCTCGCTGCA Q F R CACTGCGCCAC	$\begin{array}{c c} \textbf{ATAAAAAGAGT}\\ \hline \textbf{TTTCACCCAG}\\ \textbf{AAACTGGGTC}\\ \textbf{ACATGAGAAA}\\ \hline \textbf{TGTACACTGT}\\ \hline \textbf{ACATGAGAAA}\\ \hline \textbf{TGTACCTTT}\\ \hline \textbf{T} & \textbf{A} \\ \hline \textbf{AGAAGAGGAA}\\ \hline \textbf{TTTCTCCTT}\\ \hline \textbf{T} & \textbf{A} \\ \hline \textbf{CGCCGCAAT}\\ \hline \textbf{V} & \textbf{E} \\ \hline \textbf{S} \\ \hline \textbf{GTAGGAGACC}\\ \hline \textbf{C} & \textbf{G} & \textbf{I} \\ \hline \textbf{Y} \\ \hline \textbf{GCGGTATCTA}\\ \hline \textbf{CGCGCATAGAT}\\ \hline \textbf{ACCCCACATT}\\ \hline \textbf{ACCGGCTAA}\\ \hline \textbf{CGCGCATAGAT}\\ \hline \textbf{CGCCACATGAT}\\ \hline \textbf{CGCCCACATGAT}\\ \hline \textbf{CGCCCACATGAT}\\ \hline \textbf{CGCCCACATGAT}\\ \hline \textbf{ACCCCCCCCACA}\\ \hline \textbf{R} & \textbf{M} \\ \hline \textbf{A} \\ \hline \end{array}$	AGGCCTGGCC CAAAAATTGA GTTTTAACT AATTTTGGTT TTAAAACCAA C R D M ACGAGATATG TGCTCTATCA H H I H H I H CGAGTATGG GTGGTATAGG TACGGTTAGC TACGGTTAGC TACGGTTAGC TACGGTTAGCGC CTTAACGAGA AGCCCAACCG TGCACTAACAA ACTCGCCCCCT T Y K GACGTATAAT V H F	XGTITTAGAC CACAGATCAA GATTTAGGG CTTAAATCCC K Q T W AAACAAACCT TTTOTTTGGA P N G E CGAATGGAGA GCTTACCTCT CGAATGGAGA GCTTACCTCT CGAATGGAGA GCTTACCTCT CGAATGAAATTA A F E L CGTTCGAACT A F E L CGTTCGAACT A GACAAAGCA TCTGTTCGGT D K A AGACATAGCA CTTATCGATA CGTTCGAACT
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 p. +3 2401 +3 2501 +3 2601 +3 2601 +3 2701	GGGGCAACTC GGGGCAACTC GGGGCAACTC GCGCGTTGAG ATAAATAAAA GTATATATTT GTCAGAACGG CAGTCTTGCC W R W Y GGCGCTGGTA CCGGCACCAT E I W S AATTTGGTCG TTAAACCACG K T H AAAACCACAC K T H AAAACCACAC CGTTAAACCACG CGATATCCTC CGTATAGACC GGAACTGG K A K I AAGCAAAACT TTCGTTTTGA	CCGCCGCTTAA GCCGCCGCTTAA AGCCGCCGCTTGA GTTTACAAAG GTTTACAAAG GCTTACAAAG CCGCCTGATTG I D E ATAGCGAGA TATCGCGAGCC CCGGACCTGCATCA CCGGACGTCAGTA AGCCTCCGCG TTCGCAGCCC R N I I GCAACATTAT CCTGCAACCT CCGCAACTTAT CCGCAACAT CCGCAACTTAT	CACTGGACAT TATTCTTGAA ATAAGAACTT GTGAATGATG CACTTACTAC CACGTTCGA GTGCAATGATG D P V GACCGGCAAACG AGGTCTTTGC CGACGGCAAACG AGGTCTTTGC CGAAACGAAACG AGGTCTTTGC G A E A GGACGCCTAG G A E A GGACGCCTAG G A E A GACCGGCTAAT ACGGCCAAT F A Y TTGCCGATTA	CCCCCTTTIG ATAAATACAT TATTATAGAT ATATTATGTA ATATTTATGTA ATATTTGGATGA TGTTGGCTGG ACAACCGACC T L CGCTGTCAGA GCGCACAGTCT K A TAAAGCAACC K A TCAGCTATC ATTTGGATAG ATCCAAAAACT TAGTTTTGA ACCTAAAAAACT TGGAATACGT CGGACTATAGC GCTGGCCGGG GCTGGCCGCGC F G CCTGGCCGCGG GGACCGCCCC F K TCCTGAAAACC AGGACCGCCCC F K CCTGCCGAAACCAACC GGACCGCCCC F K TCCTGAAAAC AGGACTTCC	ACTTATGGT CAACTCATTT GTTGAGTANA TTTGGTTAGC MACCAATCG AAACCAATCG GTTAGGGTCA GTTACCCAGT TOTACGCCAGC V R O V TOTACGCCAGC V R O V GTTAACCAGC ACACGCCAC Q Q TOTACGCCAGC Y Q ACCAGCAAACG CAACTGCGCTTCC Y Q ACCAGCAAAACGCCTTCTCC TATCGCCGGAT A E GCACAAGAAA CGTCTCTTCTTCTCCAT T I AGAAGGCTA TTTCCCCAT I I AGAAGGCTA TTTCCCGAT I I T I CATTATCCCG GTATAAGGC	TTTTAACCAA GAGTTGAAT GAATTTAGG GTTAAATCTG F R R ACCACGAAGA TGGTGCTTCT A G L E GCTGGCGCAA CGACCGCCAT GCCCAGACCT L R N GCTCGCGTAAC GCCCAGACCT G S K GGTTCAAAAG TTGCTCAAAG TTGCTCAGGC AACGAGTCCA I A Q A TTGCTCAGGC AACGAGTCCA	CTGGTCATT CCACTTACGA AATAAGGTTG TTATTCCAAC R Y R CGTTATCGAA GCAATACGGA GCAATACGGA GCGCACCACA W S V CCCCCCCCCC GGCGCACCA W S V CCCCCCCCCCC CACCAGACACA A L R F CGTTCCGTCTC GCCACGCATA GCCACCGCATA CCCCCCCCCC	$\begin{array}{c c} x \text{TAAAAAAGAGT}\\ \hline \text{TTTCACCCAG}\\ AAAGTGGGTC\\ ACATGAGAAA\\ TGTACGAGAA\\ TGTACGAGAAG\\ TCTTCTCTTT\\ T A L\\ T $	AGGCCTGGCC CAAAAATTGA GTTTTAAACT AATTTAGGT TTAAAACCAA C R D M ACGAGATATG TGCTTCTATAC GTGCTATAG C P I B C P I C TACCGATATCA ATGGATAGGT T V C TACGGTTGCC ATGCCAAACG E F A GAATTGCCGG CATTAGCGACGA ACGCGACGA CT V C TACGGTTGCC M S E E TGAGCGACGA ACTGCCATATT V H P GTCACCCTG CAAGGAC	$\begin{array}{c} \textbf{AGTITTAGAC} \\ \textbf{AGTITTAGAC} \\ \textbf{GAATTTAGGG} \\ \textbf{CTTAAATCCC} \\ \textbf{CTTAAATCCC} \\ \textbf{K} Q T W \\ \textbf{AACAAACT} \\ \textbf{TTTOTTGGA} \\ \textbf{F} N \textbf{G} \textbf{E} \\ \textbf{CGAATGGAGA} \\ \textbf{GCTTACCTCT} \\ \textbf{GCTAATGGAGA} \\ \textbf{GCTTACCTCT} \\ \textbf{GCTAATGAGAGA} \\ \textbf{GCTTACTTA} \\ \textbf{GCTTATAGGAGA} \\ \textbf{GCTTATAGGAGA} \\ \textbf{GCTTCATAGGAGA} \\ \textbf{GCTCTATAGGAGA} \\ \textbf{GCTCTATAGGAGA} \\ \textbf{GCTTCATAGGAGA} \\ \textbf{GCTACTCTATAGGAGA} \\ \textbf{GCTACCTGAACTGAGAGA} \\ \textbf{GCTACCTGAACTGAGAGA} \\ GCAAGCTGGAGAGAGCAAGGCAAGGCAAGGCAAGGCAAG$
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 +3 2301 +3 2401 +3 2401 +3 2401 +3 2501 +3 2501 +3 2001 +3 -3 -3 -3 -3 -3 -3 -3 -3 -3 -	GGGGCAACTC GGGGCAACTC GGGGCAACTC GCGCGTTGAG ATAATAATAAT GTCAGAACGG CAGTCTTGCC W R W Y GGCGCTGGTA CCGCGCCGGTA TTTGGGTGT TTAAACCAG K T H AAAACCACA TTTTGGGTGT TTTGGGTGT TTTGGGTGGT CGTATAGGAC CGCACAGGG CAGACTGGG CAGACAGGG K & K L AAGCAAAACT TCGTTTTGAACCAG R L T CGTGTTGACCC	CCGCCCTTAA GCCGGCATTAA GCCGGCATTAC AACCGTCTGG GTTTACAAAG GCTAATGTTC CAAATGTTC CAAATGTTC CAAATGTTC CCGGACCTAAC GCCTGGATG GCCAGTCATA GGCCAGTCATA GGCCAGTCATA GGCCAGTCATA GGCCAGCCCGG R N I I GCAACTGTAATA CTTGGAAGGCC R N I I GCAACTTGTAATA CTTGGAAGGCC R N I I GCAACATTAT CCCACCTTGTAATA CTTGGAAGCCC R N I I GCAACATTAT CCCACCTGTAATA CTTGGAAGCCC	CACITGARAT TATTCITGAA ATAAGAACTT GTGAAATGATC CACTTTCGAA GTGCAAATGATC D P V GACCCGGTAA CTGGGCCAAT TCCAGAAACG AGGTCTTGC CGATTGATAATACC R T D L W TCCAGAAACG AGGTCTTAGC GCATCCTAG GCATCCTAG GACCCGGTTA ACGGCCAAAT TTGCCGGTTA ACGGCCAATGT	CGCCCTTITG ATAAATACAT TATTATAGAT ATAATACAT TATTATAGAT ATAATACAT TATTTGTAGA ATAATACAT TOTTGGCTGG ACAACCGACC T L CGCTGTCAGA GGGACAGTCT TAAACCTATC ATTTGGATAG I K ATACAAAAACT TACAAAAACT TACAAAAACT TGGAATAGCT P GCAATATAGC P GGAATAGCGGCG P GGAACGCGCCC P C TCCTGAAAGCATTTGCA TCCTGAAAGCAACTTTG S T CTCTACCAT	ACTITATGTGT CAACTCATTT GTTGAGTANA TTTGGTTGAC M G CAATGGTCA GTTACCAGTC GTTACCCAGTC GTTACCCAGTC GTTACCCAGTC GTTACCCAGTC GTTACCCAGTC V E GTTGAACAGG CAACTCCGCAC ACCACCAACCA TGGTCGTTTC L P ATGCCCGAACA TACCGCCAAC TACCGCCAAC ATGCCCGCAAC ATGCCCGCTTC L P AAGAAGGCTA TTCTCCGAT A E E CAACAACAACAAC TTCTCCGGCT I I CAATTATCCC GAAACATATACGC	TTTTACCAA GAAGTTGAAT GAATTTAGAC GTTAAATTAGAC GTTAAATTAGAC P R R ACCACGAAGA TGGTGCTTCT A G A GCTGGCGCAA CGACCGCGTTAC CGGCTCGGA GCCCGCGCAAC CGACGCATTC CAACGAGCCAA TTGCCGACGCAT TTGCCGACGCAA ATGCGACGCGCAA CGACGCTGAA ATGCGACGCGCAA CGACGCTGAA CGACGCTGAA CGACGCTGAA CGACGCTGAA CGACGCTGCAA CGACGCTGCA ATGCGACGCTGA	CTGGTCATT CCACTTACGA AATAAGGTTG TTATTCCAAC R Y R GGTTATCCCAC GCAATAGCGT T G V V CCCCCCCCCCCC GGCGCACCAC C	ATAAAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGAGAAA TGTACTGATAA TGTACTGATAA TGTACTGAGAAA TCTTCTCCTT T & A L HOCCCGCAAT T & A L HOCCCGCAAT TGACGAGACG CATCTCTCGC C G I Y GCGGTATCTAA CGCCATAGAT TGACCAGATT TGACCAGATT TTGCCCACAA F A T TTGCCCACAA AAGCGGTGT Q H L A AAAGCCGGGT Q H L A AACCCTGGC C GCGTATCGCC C GCATACCGG R M A GCGTATCGCCC C GCATACCGG R M A	AGGCCTGCCC CAAAAATTGA GTTTTAAATTGA AATTTGATT TTAAAACCAA C R D M ACGAGATATG TGCTCTATAG TGCTCTATAG H H I H TACTATCCAA ATGGATAGGT TACGTATAG TACGATAGGT TACGTATCCA ATGGATAGGT CTTAAGGGAGA AATCGCCCT CATGCCAACG CTTAACGGAGA ACCGCGTATAA CTGCAATGGAC V H F GAACTGGCAC CAACGGCTT	$\begin{array}{c} \textbf{XGTTTTAGAC} \\ \textbf{CACAGATCAA} \\ \textbf{GTGTCTAGTT} \\ \textbf{GAATTTAGGG} \\ \textbf{CTTAAATCCC} \\ \textbf{K} & \textbf{Q} & \textbf{T} & \textbf{W} \\ \textbf{AACAAACCT} \\ \textbf{TTTOTTGGA} \\ \textbf{F} & \textbf{N} & \textbf{G} \\ \textbf{E} \\ \textbf{CGAATGGAGA} \\ \textbf{GCTTACTCT} \\ \textbf{GCTTACTCT} \\ \textbf{GCTACTCTACCCC} \\ \textbf{GCTTACTCT} \\ \textbf{GCTACTCTAAC} \\ \textbf{GCTTCATAC} \\ \textbf{GCTTCATAC} \\ \textbf{GCTTCATAC} \\ \textbf{GCTTCATAC} \\ \textbf{GCTCCTTAAC} \\ \textbf{GCTCCTTAAC} \\ \textbf{GCTCCTAAC} \\ \textbf{GCTCCTAAC} \\ GCTCCAACAACCAACCAACCAACCAACCAACCAACCAACC$
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 +3 2401 +3 2401 +3 2401 +3 2401 +3 2501 +3 2401 +3 2301 +3 2301 +3 2301 +3 2301 +3 2301 +3 2301 +3 2301 +3 2001 +3 -3 -3 -3 -3 -3 -3 -3 -3 -3 -	GGGGCAACTC GGGGCAACTC GGGGCAACTC GGGGCAACTC GGGGCAACTT GGCGCCTGTA CGGCCTGTA CGGCCGGCTGTA CGGCGCGGCTA K T H AAATCGCACA TTTGGCTGT TAAACCACA K T H AAACCACCACA TTTGGCGGTGT R L T CGCTATAGGAC R L T CGCCCCTATT CGCGGGATAACT TCGCGGGATAACT TCGCGGCTATT	CCGCCCTTAA GCCGGGAATT TTGGCAGACC AACCGTCTGG GTTTACAAAG GCTAATGTTCC CAAATGTTCC G P N CGGACCTGATTG I D E ATAGACGAGA TATCGCCTCT T G P N CCGGCCAGTCAT GCCCGCCAGTA GGCCAGTCATA GGCCAGTCATA CCCGGCCAGTCA K R P AAGCGTCGGA CCTGACCTGGT R I I I GCAACATTAT CCGGACCTGGACA CCTGACCTGTAA CCTGACCTGA	CACTOGARAT TATTCTTGAA ATAAGAACTT GTGAATGATG GTGCAAAGAT GTGCAAAGAT D P GTGCAAAGCT GTGCAAAGCT GTGCAAAGCT GTGCAAAGCT GTGCAGAAGCT C D P Q K TCCAGAAACC AGGTCTTTGC CGATTATAGG GCATCTATATGC G A CGACCGGATC CGCATTATGG GACCGGGTAA CGCACCTAG GACCGGGTAA ACGGCCAAAT ACGGCCAAT ACGGCCAAT ACGGCCAAT ACGGCCAAT AAGGGCTACA P P Q N	CCCCCTTTIG ATAAATACAT TATTATAGTA ATATTTATGTA ATATTTATGTA ATATTTAGGA TGTTGGCTGG ACAACCGACC T L CCCTCTCAGA GCGACAGTCT K A TAAAAACGATC TAAAGCTATC ATTTGAAAAACT TAGATATTGAAAAACT AATCGAAAAACT ACCTAACAACGT QCATATACG ACCTATACG GGCGAATATCG F G CCTGGAATACG GGCGAATATCG GGACTATACG GGACTATACG GGACTATACG GGACCGCCCC F L TCCTGAAAGC CTCTGAAAGCCCCCC S T CTCTAACAATTCG AGGACTTTCG AGGACTTCCAATTCG AGAATGCTAA	ACTTAGTGT CAACTCATTT GTTCAGTAAA TTTGGTTAGC AAACGATCG M G Q CAATGGGTCA GTTAACCAGT V E C TGTAGCCAGT V E E GTTGAAGAGG CAACTGCGGTC V E E GTGAACAGCAAAC TGGTCGTTTCTCC Y Q Q T ACCAGCAAAC ACATGCGGTC L P D ATTGCCGGAT TAACGGCCTA CCAACAACAAAA CGTCTTCTTT E E G Y AAGAAGGCTA I I P CATTATTCCGG GTAATAAGGC CTTCTTATTCCG GTAATAAGGC CTTCTTGTACG D M I K	TTTTAACCAA CAAGTTGAAT GTTCAAACTTA CAATTTAGAC GTTAAAATCTG P R R ACCACGAAGA TGGTGCTCTCT A G A GCTGGCGCAA CGACCGCGTCTGGA CGACCGCATG GCCCGGCAAC CGACCGCATG G S K GGTTCAAAAG CCAAGGTCCG TACCTGGAT TTGCTCAAGC AACGAGTCCG AACGAGTCCG AACGAGTCCG TACCTGGAT ATGCGCCGCATG CAACGGCTGC Q W M V AATGGATGGT TTACCCACCG	CTGGTAGTT GGTGATGCT CCACTTACGA AATAAGGTTG CGTATCGAA GGTATACGA GGTATCGA GCAATAGGGT T G V V GTGGTCGTG GGCGCACCA W S V GTGGTCGTCGTG GGCGCACCA L A Q CTGGCGCAAT GACCGCGAA A L R F CGTTGCGTT GCAACGGAAA E R R AGACGGCAGCA TCTCGCTGCGC E V G V AGGTTGCCGT TCCCACCGCA E T V GGAACCGTT CCCACCGCGT	ATAAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGAGAAA TGTACCCTT T A L AGAAGAGGAA TCTTCTCCTT T A L TGCCGCAAT V E S GTAGAGAGCG C G I Y GCGGTATCTAC CGCCATAGAT ACGGCTATGACC C G I Y GCGGTATCTAC CGCCATAGAC F A T TTCGCCACAT AAGCGGTGGT CGCATACCTGG C G I Y GCGGTATCACC C G I Y GCGGTATCACC C G I Y CGCGTATCGACC C G I Y C GCGTATCGACC C G I Y C GCGCATCGAC C G I Y C GCGCACACACC C G I Y C GCGCACACAC AAGCGGTGGT C G I Y C GCGCACACACAC C G I Y C GCGCACACACACACACACACACACACACACACACACAC	AGGCCTGGCC CAAAAATTGA GTTTTAACT AATTTTGGTT TTAAACCAA C R D M ACGAGATATG TGCTCTATAC H H I CCCCTTATAC GTGGTATAG V P I H TACCTATCCA ATGGATAGGT C T V C TACGGTTGCAAACG M S E F GAACTACCAA GAACTACCAA CCCCTATCAA ACGCAAGAA ACTCGCCTCT V H F GTCACCGAGGA ACTCGCTATAAA CAACGGCTC CAAGGGGACCA A N G F CGAACGGCCT CCTACCAAGGCCT CCTACCAAGGCCT	AGTITTAGAC CACAGATCAA GATTTAGGG CTTAAATCCC K Q T W AAACAAACCT TTTOTTTGGA P N G E CGAATGGAGA GCTTACCTCT CGAATGGAGA GCTTACCTCT CGAATGGAGA GCTTACCTCT CGAATGGAGA GCTTACCTCT CGAAGATAC GCTTCTACGT GCTCGAATGGAGA GCTTCGAACT GCAACTTGAAT A F E L CGTCGTACGACT GCAACTTGATA A F E L CGTTCGAACT GATATGGATA A GACAAAGCA TCTGTTCGT D D F P ACGACATGGCC TGCTAGGCGG T M C CGACATGGTG GTGTCACACA
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 +3 2401 +3 2501 +3 2601 +3 2601 +3 2601 +3 2601	GGGGCAACTC GGGGCAACTC GGGGCAACTC GGGGCAACTC GCGCGTGAG ATAAATAAAA W R W Y GGCGCTGGTA CGGCGCGGTA CCGGGACCAT E I W S AATTTGGTCG TTAAACCACG K T H AAATACCACG K T H AAAACCACG K T H AAAACCACG C K T H AAAACCACG C K T H AAAACCACG C K T H C C C C C C C C C C C C C C C C C C C	CCGCCCTTAA GCCGGCACCTTAA ACCGTCTGG GTTTACAAAG CAAATGTTC CCGACCTAAC CCAAATGTTC CCGCACTAAC CCCCTGATTG T D E ATAGACGAGA TATCACGAGA TATCGCGCTCT CCGGACCTCAT CCGGCTCGCA CCTGCATGT CCGCACCTGT R N I I GCAACATTAT CCGTGTAATAA R E H CCGGACTGTAATA CCGTGAACATTAT CCCTGCAGCCCG R N I I CCCCCTTGTAATAA CCGTGCAACATTAT CCCCCTTGTAATAA CCCCCTTGTAATAA CCCCCTTGTAATAA CCCCCCTGCA	CACCIGGACAT TATTCITGAA ATAAGAACTT GTGAATGATG CACTTACTAC CACGTTCGA GTGCAAAGCT D P V GACCGGCAAACCT CACGAAAGCA AGGTCTTTGC CCACGAAAGCA AGGTCTTTGC GCATAATACC R T D CCTAGGAATCA CCCACGACAACCA GACCGCCTAG GACCGCTACA CCCCCATGT P A I V CCCCCCATGT CCCCCCATGCTACA CCCCCCATGCAAACCA AAACGGATAAA	CGCCCTTTIG TATATATGA TATTATGA TATTATGA TATTATGA TATATAGA TATATAGA TGTGGCTGG ACAACCGACC T CGCTGTCAGA GCGACAGCGTC CGCGCTCAGA GCGACAGCACC TAAAGCAACC ATTICGATAG ATTICGAAAGCAC L E Y V TGGACTATACG ACCTTATGCA TAGTTATGC CGCGCGCGCG GGACCGCCCC F L K A TCCTGGAAGC GGACTATACG GGACTATACG GGACTATACG GGACTATACG GGACTATACG GGACTATACG GGACTATACG GGACTATACG GGACTATACG GGACTATACG GGACTATACG CCTGGCGCG GGACCGCCCC S T I CCTGCAAGCAA	ACTTAGTGAT CAACTCATTT GTTGAGTAAA TTTGGTTAGC AAACCAATCG CAATGGGTCA GTTACCCAGT O V R Q O V R Q CAACTGCGCAG CAACTCTCCC V Q Q T ACCAGCAACA CGACTTCTCC CAACTCCCGAT TAACGGCCTA TAACGGCCTA TAACGGCCTA TAACGGCCA CAACTCCCGAT TAACGGCCA CAACTCCCGAT CAACACAGGCCA CAACTCCCGAT CAACAACATGC CAACTACCCGAT CAACAACATGC CTCCTGTACG D M I K	TTTTAACCAA GAAGTTGAAT GAAGTTGAACTTA CAATTTAGAC GTTAAAATCTG P R R ACCACGAAGA CGTGGCCCA CGCTGGCCCAA CGCCGCCCCA CGCCCGCCTTGA GCCCACGCCAA CGCCCAGACCT L R N GCTCGCGCAA CGCCCAGACCT CGCCCAGACCT CGACCGCATG GCCACGCTCGA GTTCCAAAG CCAACGTCCGCA T L D TACCCCGAC CCAACGTCCGCAA ATGCGACCTA V A D GTTGCCCGCCAC CAACGCCCCACCTA CAACGTCCCAACA V A D GTTGCCCGCCAC CAACGTCCCACACA V A D GTTGCCCACCACACACACACACACACACACACACACACAC	CTGGTTAGTT CCACTTACGA AATAAGGTTG CCACTTACGA CCACTTACGA CGTATCCAAC CGTTATCGAA GCAATACCGT GGTCGCACCA CACCACCACA CACCACCACACA CACCACCACACA CCACCA	ATAAAAAGAGT TTTCACCCAG AAAGTGGGTC AAAGTGGGTC AAAGTGGACAA TGTACACAAA TGTACACAAA TGTACACAAA TGTACACAAA TGTACACAAA TTTCCCTT TA TTGCCGCAAT TGCGCTACTA ACGCCTACTA ACACCACATTA ACACCACACTACAA AAACACCTGCC TGACACAGGTGGCT AACACCTGCC AACACCTGCC AACACCTGCC AACACCTGCGC N <s< td=""> AATAGCATCGG TTATCGTACCAC AATAGCATCGC TACTTACCAA</s<>	AGGCCTGGCC CAAAAATTGA GTTITTAAACT AATTTAGGT TTAAAACCAA C R D M ACGAGATATG TGCTCTATACA ACGAGATATG TGCTCTATCA H H I CACGATTGC GGGTATAG T V C TACGGTTGC T V C TACGGTTGC M S E E TGAGCGAGGA ATGCCAAACG M S E E TGAGCGAGGCA CAAGTGGCACCTC A N G F CGAGCGATATAA CTGCATAACCTC A N G F CGAGCGCCCC	XGTITTAGAC CACAGATCAA GAATTTAGGG CTTAAATCCC K Q T W AAACAAACT TTTOTTIGGA F N G E CGAATGAGAGA GCTTACTCT CGAATGAGAGA GCTTACCTCT CGAATGACAA GCTTACCTCT CGAATGACAA GCTTACCTCT CGAATGACAA GCTTACTCT GCTTACAAT A F E L CGTTCGAACT GCATACGATA GCAAGCTTGA CGTTCGAACT GCAAGCAAAGCA TCTOTTTCGT D I D K GAATATCCGCC TCATAAGCTAT D I D K GCATATCGATA CTATAGCTAT D I D K GCACCATCCTCT CACCATCCCCC T M C CACCATCCCCC T M C CACCATCCCCCC T M C CACCATCCCCCC T M C CACCATCCCCCCC
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 +3 2301 +3 2501 +3 2501 +3 2701 +3 2601 +3 2701 +3 2801 +3 2801 +3 2901 +3 2901 +3 2001 +3	GGGGCAACTC GGGGCAACTC GGGGCAACTC GGGGCAACTC GCAGTCTTGCC W R W Y GGCGCTGGTA CCGCGCCGTTA CCGCGCCGTTA AATTGGTCG TTAAACCAGC K T H AAATCGGTCG TTAAACCAGC K T H AAAACCAGC K T H AAAACCAGC CTAAACCAGC CTAAACCAGC CTAAACCAGC CTAAACCAGC CTAAACCAGC CTAAACCAGC CTAAACCAGC CTAAACCAGC CTAAACCAGC CTAACCAGC CTAACCAGC CTAACCAGC CTAACCAGC CTAACCAGC CTAACCAGC CGCCCCACGA P R P I CGCGCGCTATT CGCCGGGATAA T G S ACCGGCACTAGA CGCGCCTAGAA	CCCCCCACC CCCCCCTTAA GCCGGCATTACAAAG CCAAATGTTTC CAAATGTTTC CAAATGTTCC CAAATGTTCC CAAATGTTCC CAAATGTTCC CCCGACCTAAC GCCAGCTAAC GCCAGCTCAGTA GCCAGCTCAGTA GCCAGTCAGTA CCCGACCTGCA TCCCCACCGC CCCGACCTGCA CCCGACCTGCACA CCCGACCTGCACA CCCGACCTGTA CCCGACCTGCACA CCCGACCTGTA CCCGACCTGTA CCCGACCTGTA CCCGACCTGTA CCCGACCTGTA CCCGACCTGTA CCCGACCTGTA CCCGACCTGTA CCCGACCTGCCACA CCCGACCTGTA CCCGACCTGTA CCCGCCCCCCC C C CCCGCCCACCC C C CCCGCCCACCC	CACTOGARAT TATTCTTGAA ATAAGAACTT GTGAATGATG CACTTTGCA GTGCAATGATG GTGCAATGATG GTGCAAAGCT CACGTTTGCA GACCCGGCAAT TCCAGAAACG AGGTCTTTGC CGATCGCTAT GCCAGCACAACG GACCCGCTATG CCAGCAAACG GACCCGCTATG CCAGCAGAACG CCTCTTCTCTC GACCGCCAAT F A Y TTGCCGGCTAT AACGGCTATA AACGGCTATA AACGGCTATA CCGCGCAAACG CCGCGCAAACG A D N CCCCGCAAACA CCCGCCTATG CCCGCCATGT CCCGCCAATG CCCCGCCAATG	CCCCCTTTIG ATAAATACAT TATTATAGAT ATAATACAT TATTATAGAT ATAATACAT TATTATAGAT ATAATACAT TGTGGCTGG ACAACCGACC K CCCTGTCAGA CGCACTATCA ATTTGGATAGC L K ATTTCGATAG ACTAAAAACT TAGTTATGCA ACCTAATACG CCGACTATACG GGAATACGT F CCTGGCGCGCG GGACATATGCA CCTGGCGCGCGCG S T ACCTATATGCA CCTGGCGCGCGCG S T CCTGACGAAGC S S T CCTGACAAGCGAACC S T CCTGACAGCGCGCC S T GACGACGGCCC S CTAGACCATC A <	ACTTAGTGT CAACTCATTT GTTGAGTAAA TTTGGTTGGC AAACCAATCG MAACCAATCG GTTACCCAGTCA GTTACCCAGTCA GTTACCCAGTCA CAATGGGCAA ACATCGCGCAA CATCCCGCAA CAACTCCCCC Y Q Q T ACCAGCAAACA GTTGAAGAGGCAA CAACTCCCCC Y Q Q T ACCAGCAAACA CAACTCCCCC TGGTCGTTTC L F D ATGCCGGAATA CAACGCAAACA CAACTCCCCCTT TAACCGCCTTCTTT E E G Y AAGAAGGCTA TTCTCCGGAATA CCATTATCCC GTAATAAGGC CAACTCCCCCTCTTTC I I P GAAGACATCC CTCCTGTACC D M I K ATATGATCAA	TTTTAACCAA GAGTTGAAT GTTCAACTTA GAATTTAGG F R R ACCACGAAGA TGGTGCTTCT A G L E GCTGGCCCAA CGACCGCGCT A G L R N GCTGCCCGCAA CGACCGCGTT G S K GGTTCCAAAG GCTCCGTAAC CGACGCTTCGA G S K GGTTCCAAAG GCTCCGACGCATT I A Q A TTGCTCAAGGC AACGAGTCCG AACGAGTCCG AACGAGTCCG CAACTTCGAT TTACCTCCACG Q M Y AATGCATCGAT AATGCATCGAT TTACCTACCA Q F G ACACTTTGCT	CTGGTTAGTT CCACTTACGA AATAAGGTTG CCACTTACGA TTATTCCAAC R Y CGTTATCGAC GCAATACGT GGCAATACGGT GGCGCACCA W Y CCCCCCCCCC GGCGCACCA L A CTGGCTCTGT CACCAGACAC L A CTGGCCCCTTA A L CACCGCCACAT CACGCCCCAA CACGCGCACCT CAGTTCCGTC GCAACGCACCAT CGAGTCCGTC GCAACGCCAAC C CAGTTCCGTC GCAAGCGACGT TCCAACCGCA E T GGAAACCGTT CCCCGCCATLGA F R CCCTTGCGCTAGA F R CCCCGCCACCTT CCCTTGCCAACCTT CCCCGCCATLA F R CCCCGCCATLA <	ATAAAAAAGAGT TTTCACCCAG AAAAGAGGAC ACATGAGAAA TGTACACCATA TTTCACCCAG AGAAGAGGAA TCTTCTCCTT T A L T T T A L T T A L T T A L T T A L T T T T T T A L T T A L A A C A C T T T A L T T T T T A L T T T T T A L T T T A L T T T T T T A L T T T T T T A L T T T T T A L T T T T T T A L T T T T T T A L T T T T T A L T T T A L T T T A L T T T T A L T T T A L T T T A L T T T T A L T T T A L T T T A L T T T T A L T T T T T T T T T T T T T T T T T T T	AGGCCTGCCC CAAAAATTGA GTTTTTAAAT AATTTGGTT TTAAAACCAA C E D M ACGAGATATG TGCTTCTATAC H H I CTGCTTATAC H H I CTGCTATAG TGCGTATAG TACGGTTGCG ATGCGAAACG E F A GAATTGCCCG CTTACGGTTGC CAACGGAGA ACTCGCTCT V H F CCAACGGAGA ACTCGCCCT V Y K GACGTATAAA CTGCATATAA CTGCATATAAA CTGCATATAAA CTGCATATAAA CTGCATATAAA CTGCATATAAA CTGCATATAAA CTGCATATAAA CTGCATATAAA CTGCATATAAA CTGCATATAAA CTGCCATATAAA	$\begin{array}{c} \textbf{XGTITTAGAC} \\ \textbf{CACAGATCAA} \\ \textbf{GTGTCTAGTT} \\ \textbf{GAATTTAGGG} \\ \textbf{CTTAAATCCC} \\ \textbf{K} Q T W \\ \textbf{AAACAAACCT} \\ \textbf{TTITTTGGA} \\ \textbf{P} \textbf{N} \textbf{G} \textbf{E} \\ \textbf{CGAATGACA} \\ \textbf{GCTTACTCT} \\ \textbf{GCTTACTCT} \\ \textbf{GCTTACTCT} \\ \textbf{GCTTACTCT} \\ \textbf{GCTTACAATTAACTTAA} \\ \textbf{N} \textbf{F} \textbf{E} \textbf{L} \\ \textbf{GGTTCGAACT} \\ \textbf{GCTTGACAAT} \\ \textbf{GCTACGAAT } \\ \textbf{GCAACTAATAACTTAA} \\ \textbf{A} \textbf{F} \textbf{E} \textbf{L} \\ \textbf{CGTTCGAACTGAAT} \\ \textbf{GCAACTAACTGAAT} \\ \textbf{AATATGAAATAAT} \\ \textbf{A} \textbf{F} \textbf{E} \textbf{L} \\ \textbf{GCTACGAACTGAAT } \\ \textbf{GCAACTGAAT} \\ \textbf{GCAACTAACTTAACTTAA} \\ \textbf{A} \textbf{F} \textbf{E} \textbf{L} \\ \textbf{GCTACGAACAATCAAT} \\ \textbf{CGTTCGAACTGAAT} \\ \textbf{CGAACTAACTGAAT} \\ \textbf{D} \textbf{D} \textbf{P} \textbf{P} \\ \textbf{ACGATCGACA} \\ \textbf{CTATAGCTAT} \\ \textbf{D} \textbf{D} \textbf{P} \textbf{P} \\ \textbf{ACGATCGCCGC} \\ \textbf{T} \textbf{CCACTTGCTTCG} \\ \textbf{GTGGTACACA} \\ \textbf{T} \textbf{L} \textbf{R} \textbf{E} \\ \textbf{ACGCCCCCCCCGCG} \\ \textbf{T} CCACCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2301 +3 2301 +3 2401 +3 2501 +3 2601 +3 2601 +3 2801 +3 2801 +3 201 +3 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4	GGGGCAACTC GGGGCAACTC GGGGCAACTC GCGCGTTGAG ATAATAATAAT ATATTATTTT GTCAGAACGG CAGTCTTGCC W R W Y GGCGCTGGTA CCGGGCCGGTA CCGGGCCGGACCAT TTTGGGTGT TAAAACCACC F L T GGCACTGGG CAGACCACG CTATAGGAC CCTATAGAACCAC CCTATAGGAC CCTATAGAACCAC CCTATAGGAC CCTATAGAACCAC CCTATAGGAC CCTATAGAACCAC CCTATAGGAC CCTATAGAACCAC CCTATAGGAC CCTATAGAACCAC CCTATAGGAC CCTATAGAACCACACACACACACACACACACACACACACA	CCCCCCTTAA GCCGGCATTAA GCCGGCATTAA GCCGGCATCG GTTTACAAAG GCTGGATAC GCTAATGTTC CAAATGTTC GCACTGATTG T GCGGACCATAC GCCTGGATGACA TATGACGAGA TATGACGAGA TATGACGAGA TATGACGAGA TATGACGAGA CCCGGCCAGTGACA CCCGACCTGATA GCCACCTGATA GCCACCTGCA TTGCGAGGCC CCCGCCCGC R N I I GCAACATTAT CCCACCTGCA CCCGCCCCGC GACGCCCGCC G G V F ACCGCCTGCC TCCCGCCACCG CTCCGGCACCGC CTCCGGCACCGC	CACITGARAT TATTCITGAA ATAAGAACTT TATACITGAA ATAAGAACTT GGAATGATG GGCGAAAGCT GGCGCGTAA CTGGGCCATT I O K P TCCAGAAACG AGGTCTTTGC CGATTGAAACG AGGTCTTTGC GCATCACAGAACG GACCCGGTTA CGCGCCAACA CCGGCCAAA CCGGCCAAT TTGCCCGGTTA ACGGCCAAT TTGCCCGGTTA ACGGCCAATGT GCCCGCCAACA CCCCGCTACACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACACA CCCCGCCAACA CCCCGCCAACA CCCCGCCAACACA CCCCGCCAACACA CCCCGCCAACACA CCCCGCCAACACACACACACACACACACACACACACACA	CCCCCTTITG ATAAATACAT TATTATACAT TATTATACAT TATTATACAT TATTATACAT TATTATACAT TATTATACAT TATTTGTGA ATAATACAT TOTTGGTGG ACAACCGACC T L CCTTCAGA GCGACAGTCT TAAAACTAT TAAAACTATCAAAACT TAAAACTATATGCA D Y TGGAATAGGT CCTGACGCGCGC GGACAGCTATACC F G CCTGGCGCGCG S T CCTGGAAAGC AGACTATACCA T CCAAAACT C S T CCTGACACGGCC S T CCTCGACACGCGCC S T CCTCACACT GGACTCGCGTA D L CCTAACCAAT A H CCCATTTCCA CCCATTTCCA	ACTITATION CAACTCATTT GTTGAGTAAA TTTGGTTGAC AAACCAATCG M G Q CAATGGTCA GTTACCCAGTCA GTTACCCAGTCA GTTACCCAGTCA CAACTGCGCCA V E E GTTGAACAGG CAACTTCTCC V C E GTTGAACAGG CAACTTCTCC C T P D ATTGCCGATT TACGGCCAAAC ACCACCAAACAAAC CGCTCTCTTT E E G Y AAGAGGCTA TTCTTCCGAT TTCTTCCGAT C AACACATGC CTCTTGTACG D M I K ATATGATCAA	TTTTAACCAA GAGTTGAAT GAATTTAGCAC GTTAAATCTG P R R ACCACGAGAAGA TGGTGCTTCT A G A GCTGGCGCAA CGACCGCGTT CGGGCTCGGA GCCCGCGCAA CGACCGGCTT CACGACCCAGACCT CGACGCATAC CGACGCATTGC CCAAGTTCCAAAAC ACGACGCCGA T L D TACGCAGCAC AACGACTCCAC AACGACTCCAC AACGACTCCAC AACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC TTACCTACCA ATGGACTAC AATGGACTAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCAC CAACGACTCCACCAC CAACGACTCCACCAC CAACGACTCCACCAC CAACGACTCCACCAC CAACGACTCCACCACCACCACCACCACCACCACCACCACCACCAC	CTGGTAGTT CCACTTACGA AATAAGGTTG TTATTCCAAC R Y R GGTGATTACCCA GCAATAGCGT T G V V CCCCCCCCCCCC GGCGCACCA C A Q V CCCCCCCCCCCC C	ATAAAAGAGT TTTCACCCAG AAAGTGGGTC ACATGAGAAA TGTACTCCTTT K K R R AGAAGAGGAA TCTTCTCCTT T À L HOCCCGCATT T A L HOCCCGCATCTTC CGCGTATGAGAC CGCGTATGAGACC Q H L AAGCGGTGGT Q H L AAGCGGTGGT Q H L AAGCGGTGGT CGCGTATGGCC CGCGTATGGCC CGCGTATGGCC CGCGTATGGCC CGCGTATGGCC CGCATACCGC TGTGGAACCG T TATCGTACC A L TCGCTAACAC	AGGCCTGCCC CAAAAATTGA GTTTTAAAACCAA AATTTGATT TAAAAACCAA C R D M ACGAGATATG TGCTCTATAC T ACGATATAC TGCTCTATAC T V C GTGGTATAG T V C TACGTTTCC ATGGATAGGT T V C TACGTTTCC ATGGATAGGT CTTAAGGTTTCC ATGCAATTCCCA ATGCAAACGAC CTTAAGCGACAAC CTCACGAGAAGACCCT CTTACGCTCCAA CTGCAATGGAAC	$\begin{array}{c} \text{AGTITTAGAC} \\ \text{AGTGTCTAGTT} \\ \hline & \text{GAATTTAGGG} \\ \hline & \text{CTTAAATCCC} \\ \hline & \text{K} & \mbox{\square} & \mbox{\square} & \mbox{\square} \\ \hline & \text{AACCAAACCT} \\ \hline & \text{TTTTTGGA} \\ \hline & \text{F} & \mbox{\square} & \mbox{\square} \\ \hline & \text{CGAATGGAGA} \\ \hline & \text{GCTTATCTT} \\ \hline & \mbox{\square} & \mbox{\square} & \mbox{\square} & \mbox{\square} \\ \hline & \mbox{\square} & \mbox{\square} & \mbox{\square} \\ \hline & \mbox{\square} & \mbox{\square} & \mbox{\square} & \mbox{\square} \\ \hline & \mbox{\square} & \mbox{\square} & \mbox{\square} & \mbox{\square} \\ \hline & \mbox{\square} & \mbox{\square} & \mbox{\square} & \mbox{\square} & \mbox{\square} \\ \hline & \mbox{\square} & \square
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 -43 2401 +3 2401 +3 2401 +3 2501 +3 2501 +3 2501 +3 2501 +3 2401 +3 2501 +3 3001 +3 3001 +3 3001 +3 3001 +3 3001 +3 3001 +3 3001 +3 3001	GGGGCAACTC GGGGCAACTC CCCCCTTCAG ATAATAAAA ATATTATTTT TATTTATTTT GGCGCCGGTA CCGCGCTGGTA CCGCGCGGCGA E I W S AATTTGGTCG TTAAACCACC K T H AAACCCACC K T H AAACCACCACC CGTATAGGAC CGTATAGGAC R L T CGCGCGCTGCT ACGGCCACGA R L T CGCTGACCCC R L T CGCTGACCCC CGTATAGGAC R L T CGCTGACCCCAC R L T CGCCGCTGACCC CGCACACTGG CGCACCTGG CGCACCTGG CGCACCTGG CGCACCTGG CGCCCCTATT CGCCGCCTATT CGCCGCCTATT CGCCGCCTATT CGCCGCCTATT CGCCGCCTATT CGCCGCCTATT CGCCGCCTATT CGCCCCCCCTCT CGCCCCCCCCC CTTCTTTGG C S ACCCGACCTCT	CCGCCCTTAA GCCGCCCTTAA ACCGCAGACC AACCGCTCGG GTTTACAAAG CAAATGTTC CCGACCTAAC GCCTGGATTG I D E ATAGACGAGA TATCGCCGTCT T G Q Y CCGGCCTAAC GCCCAGCTCAT CCGGCCTGCACT CCGCCCGCCCGC R N I I GCAACATTAT CCGTGTAACAT CCGTGTAACAT CCGTGTAACAT CCGTGTAACAT CCGCCGCCGCC R N I I CCCACTTGT CCGCACTCT CCCCCTTGTAATA CCCCCCCCCC	CACIGGALAI TATTCITCAA ATAAGAACTT GTGAATGATG CACTTTCGA GTGCAAAGCT D P V GACCCGGTAA CTGGGCCATT I Q K R TCCAGAAACG AGGTCITTGC GCATACTTATGG GCATATTATGG GCATAATACC R T D CGTAAAATACC R T D CGTAAAATACC R T D CGTAGGATA GGCCGACAG GCAGCCTAG GCAGCCTAG GCAGCCAAAT F A Y TTGCCGATAT AACGGCCAAAT F A Y CGCCGATAA P R I V CGCCGATAA CGCCGACAAC CGCCGACAAC CGCCGACAAC CGCCGACAAC CGCCGACAAC CGCCGACAAC CGCCGACAAC CGCCGACAAC CGCCGACAAC CGCCGACAAC CGCCGCAC CGCCGCAC CGCCGCAC CGCCGCAC CGCCCGAC CGCCCAAC CGCCGCAC CGCCGCAC CGCCCCAC CGCCCCAC CGCCCCAC CGCCCCAC CGCCCCAC CGCCCCAC CGCCCCAC	CGCCCTTITG TATATATGTA TATATATGTA TATATATGTA TATATAGAT TGTTGGCTGG ACAACCGACC T L S L CGCTGTCAGA GCGACAGCCTC C K A I TAAAGCATC ATTTGGATAG I K N ATCAAAAACT L E Y V TGGAATACGT ACCTATATGG GCTGAAAGCC F L K A CCTGGCGCG GGACCGCGCC F L K A CCTGGCGCG GGACCGCGCC S T I CTCTACCATT CTCTACCATTA AGAATACGT AGAATACGT AGAATACGT CCTGGCGCG GGACCGCCCC S T I CTCTACCATT CCTACCATTCG AGAATCGTA AGAATCGTA AGAATCGTA CCTGGCGCGC CTAACCAACAAC A H L H CCCATTGCA	ACTTAGTGT CAACTGATTT GTTGAGTAAA TTTGGTTAGC AAACCAATCG M G Q CAATGGGTCA GTTAACCAGT V E C TGTAGGCCAG ACATGCGGTC V E C GTTGAAGAGG CAACTGCGGTC V C CACTCTCC V Q Q T ACCAGCAAACA CGACATCTCC L P D ATTGCCGGCAT TAACGGCCAA CGACATCTCTTT E E G Y CAATGATCAA GAACAACATGC TL I P CAATGATCAA GAACACATGC CTTCTTCCGAT I I P GAAGAAGCTA TTCTTCCGAT CATATAACGAC CTTCTTCTCCGAT I I P GAAGACATGC CTTCTTCTCTCCGAT CATATAACGAC CTTCTTCTCCGAT CATATAACGAC CTTCTTCTCCGAT CATATAACGAC CTTCTTCCTACCA CTTCTTCTCCAC D M I K ATATGATCAA TTATCCACATTCC CCGCCCTACCA	TTTTAACCAA CAAGTTGAAT GATTTAACCA GTTCAAATCTG P R R ACCACGAAGA TGGTGCCTCT A G L CGCTGGCCCAA CGACCGCGCT L R N GCTCGCGCAA CGACCGCGTTCGA GCCCGGACCT CGACCCGCATG GCCCAGACCT T L D TTGCTCAAAAG CCAAGTTCC A G R GTTCCAAACA ATGCGACCTA TTGCTCAGGC AACGGTCTGA CAACGGCTGC Q W M V AATGGATGGT TTACCTACCAA CAACTTTGCT TACCTACCAACCA D M Y GATATGTAG CTAACTACAACCA	CTGGTTAGTT CCACTTACGA AATAAGGTTG TTATTCCAAC CGATTACGA GCAATACGAT GCGTATCGA GCAATACGAT GGCCACCA CACCAGACAC L & Q V V GTGGTCTGTG GACCACGACAC L & Q CTGCGCACCA L & Q CTGCGCACCA CACCAGACAC L & Q CTGCGCGCAAT CACCAGACAC CACCAGACCGTAA CAGTTGCGCT TCCAACGCAAA F R I CCCTTGGCAA CCTTGGCAAC CCTTGGCAAC CCTTGGCAAC CCTTGGCAAC F R I CCCCCGCGACAC F R I CCCCCGCGACAC	$\begin{array}{c c} \mathbf{ATAAAAAGAGT}\\ \mathbf{TTTCACCCAG}\\ \mathbf{AAACTGGGTC}\\ \mathbf{ACATGAGAAA}\\ \mathbf{TTTCACCCAG}\\ \mathbf{ACATGAGAAG}\\ \mathbf{AGAAGAGGAA}\\ \mathbf{TCTTCTCTT}\\ \mathbf{T} & \lambda & \mathbf{L}\\ \mathbf{T} & \lambda & \mathbf{L}\\ \mathbf{T} & \mathbf{T} & \lambda & \mathbf{L}\\ \mathbf{T} & \mathbf{T} & \lambda & \mathbf{L}\\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} \\ \mathbf{T} & $	AGGCCTGGCC CAAAAATTGA GTTITTAAAT AATTTAGGTT TTAAAACCAA C R D M ACGAGATATG TGCTCTATCA ACGAGATATG TGCTCTATCA H H I TACCACGTTGC T V C TACGGTTGC T V C TACGGTTGC ATGCCAAACG CTTAAGCGAC CTTAAGCGAC ACTGCCAAACG ACTGCCAAACG ACTGCCAAACG ACTGCCAACG ACTGCCCAACG ACTGCCCAACG ACTGCCCAACG ACTGCCCAACG A N G F CCAACGGCCAA CAACTGCCCAA CAACTGCCCAA CAACTGCCCAA CAACTGCCCAA CAACTGCCCAA ATGCCCCCAACG A N G F CCAACGGCCAA A N G F CCAACGCCCAA ACTGCCCCAACG ACTGCCCCAACG A N G F CCAACGCCCAA ACTGCCCCAACGCC ACTCCCCCAACGCCCAA CAACTGCCCAACGCCCAA C L R S TCTCCCCCCAACGCCCAA C A G C C C C C C C C C C C C C C C C C	XGTITTAGAC CACAGATCAA GAATTTAGGG CTTAAATCCC K Q T W AAACAAACCT TTTOTTTGGA P N G E CGAATGAGAGA GGTTACCTCT E D I CGAATGAGAGA GCTTACCTCT CGAATGACAA GCTTACCTCT CGAATGACAA GCTTCTATAG Y N F M TATACATTA ATATGAAAACA AGACAAAGCA CCTTCGAACT GCATATCGATA CCTCTTTCGT D I D K AGACAAGCA CTCTTCGACAT CATATCGATA CATATCGATA CATATCGATA CACCATCCGCC T L R E ACCATCCGCG TGCCACCACTT GGGATACCACA T L R E ACCCTCCGACCACC H R R K ACCCTCCGATA TCCGCACCATT
SEN2978 394 aa- 1185 b 2043-3269 Fr; 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2201 +3 2201 +3 2301 p. +3 2301 +3 2401 +3 2501 +3 2601 +3 2701 +3 2701 +3 2801 +3 2801 +3 2301 +3 2301 +3 2301 +3 201 +3 3001 +3 3001 +3 3001 +3 3001	GGGGCAACTC GGGGCAACTC GGGGCAACTC GGGGCAACTC GCGCGTTGAG ATATATATTAT GTCAGAACGG CAGTCTTGCC W R W Y GGCGCTGGTA CCGGCACCAT E I W S AATTTGGTCG TTAAACCACC K T H AAATACCACC K T H AAATACCACC K T H AAATACCACC K T H AAATACCACC K T H AAATACCACC K T H AAAACCACCAC CGTCTGACCCG GCACACTGGG K A K I AAGCAAAACT TTGGCTTACAAA F G S AACCGACAAACT TGGCCTACAA E E N F AAGCAAAACC TTGGCTACAAACC TCGCCGGATAA T G S AACCGACAGA CCCGGATCAT CCCCGGATCAT CCCCGGATCAT	CCGCCCTTAA AGCGGCGCCTTAA AACCGTCTGG CAAATGTTTC CAAATGTTTC CAAATGTTC CAAATGTTC CAAATGTTC CCGACCTAAC CCGACCTAAC GCCTGGATTG I D E ATAGCGCGAC CCGACCTGGATTG CCGGCTCGCT CCTGCCGCGCCGC	CACCIGGACAT TATTCITCAA ATAAGAACTT GTGAATGATG CACTTACTAC CACTTACTAC GTGCAATGATG GTGCAAAGCA ACTGGCCATAT I Q K R CGACTCGCCAAACG AGGTCTTTGC CGACAGCACCA AGGTCTTTGC G A E A GACCAGCATGC CTCCAGAAACG AGCACACAAACG GCATAGTTATGG G A E A GACCAGCATAG F A Y TTGCCGATAA P R I V CGCGCTAACA CGGCGCTAACA CGCCGCAAACG CGCCCTACT AAACGGATAA P R I V CGCCCAATGC CGCCCTACCAAT ACGGCCTACCAAT ACGGCCTACCAAT ACGGCCTACCAAT CGCCCTACCAAT CGCCCCACACCACCGCCACACCACCACCACCACCACCACC	CGCCCTTTIG TATATATCAT TATTATGTA ATATTATGTA ATATTATGTA TATTATGTA TATTATGTA TATAGAC TATTGGCTGG ACAACCGACC CGCTGTCAGA GGGACAGCTC ATTTGGATAG I K N I ATCTAAAAACT I K N ATCTAAAAACT I K N ATCTAAAAACT CGCTGTCGG GGACACTATCG F G Å ACCTTATGCA F G Å CCTGGCCGGG GGACCGCGCC F L K A ACCTATATGC AGGATCATATG CGCCAAAGC AGGACTTTCG S T I CTCTGAAAGC AGGACTTCG GGACAGCTAC CAGACGCAAC CCAGCCACCAC A CTACCCAATCG CTGCAAAGC AGGACTTCG CTGCAAAGC AGGACTTCG CTGCCAAAGC AGGACTATCG CTGCAAAGCAAC A CTCTGCAATCG CTGCCAAAGCAAC AGGACTTCG CTGCAAAGCAAC A CTCTGCAAAGCAAC A CTCCCAAAGCAAC A CTCCCAAACCAAC A CTCCCAAAGCAAC A CTCCCAAAGCAAC A CTCCCAAAGCAAC A CTCCCAAACCAAC CCCCGCCCGCAC CCCCGCCCGAC CCCCGGACCCGCCC CCCCGCGAC CCCCGCGCCCGAC CCCCGCGCCCGAC CCCCGCGCCCGAC CCCCGCGCCCCCCCGAC CCCCGCGCCCCCCCC	ACTTAGTGT CAACTCATTT GTGAGTAAA TTTGGTTAGC MAACCAATCG CAATGGGTCA GTTACCCAGT V R Q V R Q TOTACGCCAG ACATCCGCCA CAATGCCAG CAACTCCCCC V C E GTGAAGAGG CAACTCCCCC Y Q Q T ACCAGCAAACA CGTCTACGCAA CGTCGTCGTTCC L P D ATTGCCGGAT ACCAGCAACAGG CAACTCCCCGT L P D ATTGCCGGAT CGCGCATTGC CTTCTCCGAT CACGCCACTAG D M I K ATATGATCAA CGCCGCACTAC CGCCGCCACCAG CACGCCACTAC CGCGCCACTAC	TTTTAACCAA GAGTTGAAT GAATTTAGG GTTAAATCTG F R R ACCACGAAGA CGTGACCGCAA CGACGCCCAA CGCCGCCAA CGCCGCCAC CGGCTCTGAA GCCCAGGCCAA GCCAGGCCAA CGACCGCATTG G S K GGTTCCAAAG CCAAGTTTCC I A Q A TTCCTCAACAGTCTC I A Q A TTCCTCAACGCCAA CCAAGTCCGC AACCAGTCCGC AACCAGTCCGA CCAAGTCCGAC CAACGGCTGCA Q W M V AATGGATGGT TTACCTCACAA Q F G ACACTTTGCTAACAA Q F G ACACTTTGCTAACAA Q M V CTATCCTAACAA	CTGGTCATTACGA CCACTTACGA AATAAGGTTG CCACTTACGA TTATTCCAAC CCACTTACGA GCAATACCGA GCAATACCGA GCACACCACA W S V CCCCCCCCC GGCCACCACA W S V CCCCCCCCCC CACCAGACACC L A Q CTGGCCCTTGC CACCAGCACAC CCCCCCCCCCC CCCCCCCCCC	$\begin{array}{c c} \mathbf{ATAAAAAGAGT}\\ \mathbf{TTTCACCCAG}\\ \mathbf{AAAGTGGGTC}\\ \mathbf{ACATGAGAAA}\\ \mathbf{TTTCACCCATA}\\ \mathbf{TCACCAGAAA}\\ \mathbf{TCTTCTCCTT}\\ \mathbf{T} & \mathbf{A} & \mathbf{L}\\ \mathbf{T} & \mathbf{CGCGAAAT}\\ \mathbf{T} & \mathbf{CGCGATACTA}\\ \mathbf{T} & \mathbf{CGCCAATA}\\ \mathbf{T} & \mathbf{CGCCAATAGAT}\\ \mathbf{T} & \mathbf{CGCCAACAT}\\ \mathbf{T} & \mathbf{A} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{CGCCACAT}\\ \mathbf{A} & \mathbf{A} & \mathbf{CGCGTACTA}\\ \mathbf{A} & \mathbf{A} & \mathbf{CGCGTACCA}\\ \mathbf{A} & \mathbf{A} & \mathbf{CGCGTACCA}\\ \mathbf{T} & \mathbf{A} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{CGCACCGG}\\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{C}\\ \mathbf{C} & \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{A} & \mathbf{A} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{A} & \mathbf{C} & \mathbf{C} & \mathbf{T} & \mathbf{T}\\ \mathbf{C} & \mathbf{C} & \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{C} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{C} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T} & \mathbf{T}\\ \mathbf{T} & \mathbf{T} \\ \mathbf{T} & \mathbf{T} &$	AGGCCTGGCC CAAAAATTGA GTTTTAAAT AATTTGGTT TTAAAACCAA C E D M ACGAGATATG TGCTTCTATAG C B D M ACGAGATATG TGCTATAG C B D M ACGAGATATG TGCTATAG C B D M ACGAGATATG C C T V C TACCTATCA ATGGATAGGT C T V C TACGTTGCC ATGCCAAACG E F A GAATTGCCG CATTAGCGAGCA ACTGCCAAACG C T V C T V C TACGTATAAA CTGCATATT V H P C CGAAGCGATAT C C C C C C C T V C C C C C C C C C C C C C C C C C C C	AGTITTAGAC CACAGATCAA GTGTCTAGTT GAATTTAGGG CTTAAATCCC K Q T W AACAAACCT TTITTTTGGA P N G E CGAATGACA GCTTACCTCT CGAAGATATC GCTTACTCT GCTTCTATAG Y N F M TATATGAAT A F E L CGTTCGAACT GCAACGTGACA CGAAGCTGAC CACAGTCGACA CGATACCACT D K A AGACAAACCA TCTCTTCGT D K A AGACAAACCA TCTCTTCGT D F P ACGATCCGCT CACCATGTC CACCATGTC CACCATGCCC TCTCACCCC TCCTACGCCC TCCCACCGCC TCCCACCGCC A CCCTCCACT CACCATCGCCC TCCCACCCCT CACCATCGCCC TCCCACCCCT CCCCACCCCCT CCCCACCCCT CCCACCCCCCT CCCACCCCCCT CCCACCCCCCT CCCACCCCCT CCCACCCCCCT CCCACCCCCCT CCCACCCCCCT CCCACCCCCC CCCCCCCCCC
SEN2978 394 aa- 1185 b 2043-3269 Fr, 3,184,087 To; 3,185,271 B5QYB0 Mannonate dehydratase	1801 1901 +3 2001 +3 2101 +3 2201 +3 2301 +3 2401 +3 2501 +3 2601 +3 2601 +3 2801 +3 2801 +3 3001 +3 2801 +3 2801 +3 201 +3 -3 201 +3 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4	GGGGCAACTC GGGGCAACTC GGGGCAACTC GGGGCAACTC GCAGTCTTGAG ATATATATTT GTCAGAACGG CAGTCTTGCC W R W Y GGCGCTGGTA CCGCGCACCAT E I W S AATTTGGTCG TTAAACCACA TTTGGGTGT TAAACCACA TTTGGGTGGT M P V L TGCCGGGGCACGA L H I L GCATATCCTG GCAGACGGGG CAGACTGGG CAGACACGGG CAGACACGGG CAGACACGGG P R P I ACGGCCACGA CATTTGGCTTGA P R P I CGCGCCTACTA CGCGGACTAGA CGCGCACTAGA CGCGCCACTAGA CGCGCCACTAGA CGCGCCACTAGA CGCGCCACTAGA CGCCCCTCGA	CCGCCCTTAA GCCGGCATTAA AGCCAGACC AAACGTATAC CAAATGTTACAAAG GCTTACAAAG GCTAGATATC CAAATGTTACAAAG I D E ATAGACGAGA TATCACGAGA TATCACGAGA TATCACGAGA TATCACGAGA TATCACGAGA TATCACGAGA CCTGACCTCCT K R P AACGCACGCACT CCTGACGAGCCATC CCTGACGAGCCTCCA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGACCTCCAACA CCTGCCCCAACACA CCTGCCCCAACACA CCTGCCCCAACACA CCTGCCCCAACACACACC CCTGCCCCAACACAC CCTGCCCCCACC S K T F GAAGACCTTCC AAGGCCTGCCAACAC S D D ACCGACGACCTCCAACAC	CACITICAA ATAAGAACTT TATTCTTGAA ATAAGAACTT CACTTTGAA GTGCAATGATGA GTGCAAAGCT D F V GACCCGGTAA CTGGGCCATT CCAGAAACG AGGTCTTTATGC GCTAATATACC CCATTTATGC GCATCCTTA GCACCCTAG CCATCTTATGC GCATCCTTG CCACCGGTAA CTGCGGCAAACG AG L TGCCGGCAGAAGC CTCGTCTTCG CCACCTTG CCCGCCAATGA CCCGCCAATGA CCCGCCAATGA CCCGCCAATGA CCCGCCAATGA CCCGCCAATGA CACCGGCAAACG A D N CCCCGCAATGA CCCGCCAATGA CCCGCCAATGA CACCGGCAAACG C A D N CCCCGCAATGA CCCGCCAATGA CACCGGCAAACG C CCCGCCAATGA CCCGCCAATGA CCCCGCCAATGA CCCGCCAATGA CCCCGCAATGA CCCCCCAATGGGTAA CCCCGCCAATGA CCCCCCAATGGGTAA CCCCCCAATGGGTAA CCCCCCAATGGGTAA CCCCCCAATGGGTAA	CCCCCTTTIG TATATATACAT TATTATAGAT ATATTATGA ATATTATGA ATATTAGA ATATTAGA ATATTAGA ATATTAGA ATATTAGA ACCTACAGA CCCCCCAGA CCCCCCCCCC	ACTITACIGATI CAACTCATTT GTTGAGTAAA TTTGGTTGGCA M G Q CAATGGGTCA GTTACCCAGT V R Q TOTACGCCAG ACATGCGCCA ACATGCGCCA CAACTGCGCCA V E E TTTAACGGCAA CAACTGCGCAT CACCAGCAAACA ACCAGCAAACA CAACTGCGCAT C L P D ATTGCCGAATA ACCAGCAACAA ACCAGCAACAA CACGCAACAA ACCAGCAACAA CACGCAACAA CACCAGCAACAA C CAGCAACAA C CAGCAACAA C C C C C C C C TACCGCCATCC C L P D ATTGCCGCAT C C C C C C C TACCGCCATCC C L P D ATTGCCGCAT C C C C C C C TACCAGCCAACAA C C C C C C C C C C C C C C C C C C C	TTTTACCAA GAAGTTGAAT GAATTTAGCA GTTAAATCG GTTAAATTGG P R R ACCACGAAGA TGGTGCTTCT A G A GCTGGCCCAA CGGCTCTGGA GCCGCCCAGACCT L R N GCTGCCGAA GCCCGGACCA GCCCGGACCA GCCCGGACCATG G S K GGTTCCAAAAG GCTACCCGACA ACGACCTCGAA ACGACCTCGAA ACGACCTCGAA ACGACCCAC AACGAGTCCG AACGAGTCCG AACGAGTCGA T L D TACCCTCGAA ATGGACTAA ATGGACTAA ATGGACTGA AATGGACTGA CTAACCATCGA CTATACTACCAA D M Y GATAGTCTGGA CTAACCACATCC Q M L D AGATCCTGGA CTAACCACACACACA CTATACCACACA CTATACCACACA CTATACCACACACACACACACACACACACACACACACAC	CTGGTCATT CCACTTACGA AATAAGGTTG TTATTCCAAC R Y R GCTTATCCAAC CCACTACGA GCAATAGCGT GGCGCACCAC A Y C CTGGTCTGTG CACCAGACAC L A Q TGGTCTGTG CACCAGACAC L A Q CTGGTCTGTG CACCAGCACA A L R F CGTGCGCTAA A L R F CGTGCGCTAA A L R F CGTGCGCTAA A L R F CGTGCGCTAA CGCACGCACA CACGCGCACAC CTGCGCGCACA C F R CACGCGCACG CTCAAGGCAC CTTGGCCGTAA E V V K AGGTTGCGTTAA E V V K AAGTCGTTAA CCCCCGCAAT CCCCCGCAAT CCCCCGCAAC CTTGGCCAA F R I CCCCCGCAAC C F R CCCCCGCAAC C F R CACGCGCAAC C F R CACGCGCACA C F R CACGCGCACAC C C C C C C C CCCCGCGCAAC C C C C C C C C CCCCGCGCAA C C F R CCCCCGCGCAAC C C C C C C C C C C C C C C C C C C	ATAAAAAGA TTTCACCCAG AAAGTGGGTC ACATGAGAAA TGTACTACTATT K K R K AGAAGAGGAA TCTTCTCCTT T A L T C C C C A L T Y C C G I Y G C G T A C A GACC C C C A L T Y C C C C A L T Y T C C C C C A L T Y T C C C C C A L T Y T C C C C C C A L A A C C C C C C C A L A A C C C C C C C C C C C C C C C C C C	AGGCCTGGCC CAAAAATTGA GTTTTAAATTGA AATTTGGTT TTAAAACCAA C R D M ACGAGATATG TGCTCTATAC H H I H GTGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCTATAG TGCGATAG TT V C TACGTTCA ATGCAAAGG CTTAAGCGACAG TT V C TTACGTATAAA CTGCATATT V H F TTCACCGAC TT V C CCAACGGCAG A N G F CAACAACAGC CCAACGGCAG TCTCCCGAC TCTCCCGAC TCTCCCGAC TCTCCCGAC TCTCCCGAC TCTCCCGAC TCTCCCGAC TCTCCCGAC TCTCCCGAC TCTCCCGAC A N G F CAACAACAGC CTTCCCCGAC N F G Y ATCCGGGTA	$\begin{array}{c} \textbf{XGTITTAGAC} \\ \textbf{XGTTTAGAC} \\ \textbf{GAATTTAGGG} \\ \textbf{CTTAAATCCC} \\ \textbf{K} & \textbf{Q} & \textbf{T} & \textbf{W} \\ \textbf{AAACAAACCT} \\ \textbf{TITOTTGGA} \\ \textbf{F} & \textbf{N} & \textbf{G} \\ \textbf{F} \\ \textbf{CGAATGGACA} \\ \textbf{GCTTACTCT} \\ \textbf{GCAATGGACA} \\ \textbf{GCTTACCTCT} \\ \textbf{GCAATGACA} \\ \textbf{GCTTACTCT} \\ \textbf{GCAATGACAATC} \\ \textbf{GCTTACTTAAG} \\ \textbf{GCTTCATAG} \\ \textbf{GCTTCATAG} \\ \textbf{GCAACTGAAAT} \\ \textbf{A} & \textbf{F} & \textbf{E} \\ \textbf{L} \\ \textbf{GCTCGAACTGAAT} \\ \textbf{GCAACTGAAT} \\ \textbf{GCAACTGCGCAT} \\ \textbf{GCAACTGCGCAT} \\ \textbf{GCGCGCAT} \\ \textbf{GCACGCGCAT} \\ \textbf{GCACGGCCAT} \\ \textbf{GCACGCGCAT} \\ \textbf{GCACGCGCAT} \\ \textbf{GCACGCCAT} \\ \textbf{GCACCCAT} \\ \textbf{GCACCAT} \\ \textbf{GCACCCAT} \\ \textbf{GCCACCAT} \\ \textbf{GCACCCAT} \\ \textbf{GCACCCCAT} \\ \textbf{GCACCCAT} \\ \textbf{GCACCCCCAT} \\ \textbf{GCACCCCAT} \\ \textbf{GCACCCCCAT} \\ \textbf{GCACCCCAT} \\ \textbf{GCACCCCCCCCAT} \\ \textbf{GCCACCAT} \\ \textbf{GCCACCAT} \\ \textbf{GCACCCCCCCCCCCCAT} \\ \textbf{GCCACCCCCCCCCCCCC} \\ GCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$

Appendix

	+2									M E Q	N I A
	3301	GCGATTTCCC CGCTAAAGGG	CTACTCAATG GATGAGTTAC	CAATAGCAAC GTTATCGTTG	ATGCCTCGCC TACGGAGCGG	CCGGAGATCG GGCCTCTAGC	GCCCGCTTCT	CGTCGAATGA GCAGCTTACT	CAGGAGTTTG GTCCTCAAAC	CAATGGAACA GTTACCTTGT	GAATATCGCC CTTATAGCGG
-	+2	ΤΑΟ	V S V A	RPN	WDK	S R L V	SRI	VHL	G C G A	FHR	A H Q A
	3401	ACCGCCCAGG	TTTCCGTCGC	CCGCCCAAAC	TGGGACAAAT	CACGTCTGGT	ATCCCCTATT	GTOCATCTOC	ectecienic	GTTTCACCGC	GCGCACCAGG
_		A L F 7	T H H L	L E K	ACCCTGTTTA S D S D	GTGCAGACCA W G I	C E V	CACGTAGACC	CGACGCCCCG	A R L I	CGCGTGGTCC
	3501	CGCTCTTTAC	CCATCATTTA	CTGGAAAAGA	GCGACAGCGA	CTGGGGCATT	TGTGAAGTGA	ACCTGATGCC	GGGTAACGAC	GCGCGGCTGA	TCGCGAACCT
_		GCGAGAAATG	GGTAGTAAAT	GACCTTTTCT	CGCTGTCGCT	GACCCCGTAA	ACACTTCACT	TGGACTACGG	CCCATTGCTG	CGCGCCGACT	AGCGCTTGGA
	+2		N L L	Y T V A	E R G		T E L K	I I G	S M K	E A L H	P E F
	3001	CTTTCGCGTT	TTAGACGACA	TGTGGCAACG	TCTTGCGCCG	CGTCTTTCGT	GGCTTGACTT	TTAATAGCCA	AGTTACTTTC	TTCGCGACGT	GGGACTTAAG
_	+2	DGH	A G I L	A A M	A R P	ETAI	V S L	TVT	E K G Y	CTD	PASG
	3701	GACGGCCATG	CAGGGATTCT	GGCGGCAATG	GCGCGCCCGG	AAACCGCCAT	CGTCTCCTTA	ACCGTGACCG	AAAAAGGCTA	CTGCACCGAC	CCCGCCAGCG
_	+2	G E L I	D V N N	PLI	Q N D L	A H P	Q Q P 1	K S A I	GYI	V E A I	, N M R
	3801	GCGAGCTTGA	TGTCAATAAC	CCGCTGATCC	AAAACGATCT	TGCCCATCCA	CAGCAGCCTA	AATCCGCCAT	TGGCTATATT	GTCGAAGCGC	TAAACATGCG
_		CGCTCGAACT	ACAGTTATTG	GGCGACTAGG	TTTTGCTAGA	ACGGGTAGGT	GTCGTCGGAT	TTAGGCGGTA	ACCGATATAA	CAGCTTCGCG	ATTTGTACGC
SEN2979	+2	CCGGGIGCII	GULTGING	CTTTTACCGT	GTTGTCATGC	GITILCETIC	GCGIGIITGG	TCACGTCGCG	NGGCTGCCG	тестевесст	GCLINIGCT
490 aa-1473 bp.		GGCCCTCGTT	CCTGACTTTC	GAAAATGGCA	CAACAGTACG	CTATTGCATG	CGCTCTTACC	AGTGCAGCGC	TTCCGACGGC	AGGAGCCGGA	CCGTTTTCGA
3383-4852	+2	R D A	A L A A	W I A	DNV	T F P C	TMV	D R I	V P A A	TEE	T L Q L
Fr; 3,185,382	4001	GCGGATGCCG	GGGAGCGCGC	ATGGATTGCC	GACAATGTGA	GGNGGGGRC	GIGGIACCAL	GACCGCATTG	TTCCGGCGGC	GACCGAAGAG	ACGCTACAAC
DSOVD1	+2	LVAI	D Q L G	V Y D	P C A I	A C E	PFR	Q W V I	EDN	F V N G	RPD
D-mannonata	4101	TGGTTGCCGA	CCAGTTAGGG	GTTTATGATC	CTTGCGCTAT	CGCCTGCGAA	CCCTTCCGCC	AGTGGGTCAT	CGAAGATAAC	TTTGTGAATG	GTCGCCCGGA
ovido reductas		ACCAACGGCT	GGTCAATCCC	CAAATACTAG	GAACGCGATA	GCGGACGCTT	GGGAAGGCGG	R M L	GCTTCTATTG	AAACACTTAC	CAGCGGGCCT
oardo reductus	4201	CTGGGACACC	GTCGGCGCGC	AATTCGTTGC	AGATGTCGTA	CCGTTTGAAA	TGATGAAGCT	GCGTATGCTT	AACGGCAGTC	ACTCTTTCCT	CGCCTATCTG
_		GACCCTGTGG	CAGCCGCGCG	TTAAGCAACG	TCTACAGCAT	GGCAAACTTT	ACTACTTCGA	CGCATACGAA	TTGCCGTCAG	TGAGAAAGGA	GCGGATAGAC
	+2	GGTTNTCTCC	GCGGGTATCA	CACTATOGOC	D T M	T N P A	TRR	A A L	COCTONTOCT	TGATGAACAA	A P T L
	4301	CCAATAGAGC	CGCCCATACT	GTGATAGCGG	CTATEGTACT	GGTTGGGCCG	AATGGCAGCG	CGCCGCGACC	GCGACTACGA	ACTACTTGTT	CGCGGCTGCA
_	+2	LSMI	PEGT	D L E	G Y A N		A R F	T N P S	L K H	R T W C	I A M
	4401	TATCAATGCC ATAGTTACGG	GGAAGGCACC	GATCTGGAAG	GGTATGCGAA CCATACGCTT	TTTGTTGATC AAACAACTAG	GCGCGTTTCA	CTAACCCCTC GATTGGGGAG	GTTGAAACAC CAACTTTGTG	CGCACCTGGC GCGTGGACCG	AGATCGCGAT TCTAGCGCTA
-	+2	MDGS	QKL	P Q R L	LDP	VRL	H L Q Q	GDD	YRR	L T L G	VAG
	4501	GGATGGCAGT	CAGAAACTGC	CGCAGCGTTT	GTTGGACCCG	GTACGTCTGC	ACTTGCAACA	GGGCGATGAC	TATCGCCGCC	TGACGCTGGG	CGTCGCCGGA
_	12	W M R	Y V G G	GCGTCGCAAA	Q G K	T I D V	V D P	L L A	ATAGCGGCGG	I H Q	Q Y Q T
	4601	TGGATGCGTT	ACGTCGGCGG	TATCGATGAA	CAAGGTAAAA	CCATTGATGT	CGTCGATCCG	CTGCTCGCGC	AGTATCAGGC	GATTCATCAG	CAATATCAGA
_		ACCTACGCAA	TGCAGCCGCC	ATAGCTACTT	GTTCCATTTT	GGTAACTACA	GCAGCTAGGC	GACGAGCGCG	TCATAGTCCG	CTAAGTAGTC	GTTATAGTCT
	+2 4701	CGCCGGAAGA	ACGCGTTCGC	GGGCTACTGG	CCATCGAGTC	TATCTTTGGC	AGCGACCTGC	CGALGALCCA	CGAATTTGTG	CAAGCCGTTA	CCGACGCTTA
_		GCGGCCTTCT	TGCGCAAGCG	CCCGATGACC	GGTAGCTCAG	ATAGAAACCG	TCGCTGGACG	GCTTCTTGGT	GCTTAAACAC	GTTCGGCAAT	GGCTGCGAAT
	+2	YQQL		G A K A	T V E		K	м	3 T F M		FI. I. K
	+1 4801	CCAACAGCTA	TTGCAGAACG	GCGCGAAAGC	CACGGTAGAA	GCGCTGGCTA	AGTAAGGAGA	TAAATCATGG	CTACGTTTAT	GACTGAAGAT	TTTCTACTTA
_		GGTTGTCGAT	AACGTCTTGC	CGCGCTTTCG	GTGCCATCTT	CGCGACCGAT	TCATTCCTCT	ATTTAGTACC	GATGCAAATA	CTGACTTCTA	AAAGATGAAT
	+1	K N D I	IART	LYH	K Y A A	P M P	I Y D	F H C H	L S P	QEIB	D D R
	4901	TTTTGCTGTA	ACGGGCGTGC	GACATGGTAT	TTATGCGGCG	CGGGTACGGC	TAAATACTGA	AGGTGACGGT	AAATTCAGGC	GTCCTTTAGC	GGCTGCTAGC
	+1	R R F D	N L G	QIWI	EGD	H Y K	W R A L	R S A	GVD	E S L I	T G K
	5001	GGCAAAGCTA	AACCTCGGTC TTGGAGCCAG	AGATCTGGCT TCTAGACCGA	GGAAGGCGAC CCTTCCGCTG	CACTATAAAT GTGATATTTA	GGCGAGCGCT	ACGAAGCGCA	GGCGTGGATG	AGTCGCTGAT TCAGCGACTA	CACCGGCAAA GTGGCCGTTT
_	+1	ETS	D Y E K	YMA	WAN	т у р к	TLG	NPL	Y H W T	HLE	LRRP
	5101	GAGACCAGCG	ATTATGAAAA	ATATATGGCC	TGGGCCAATA	CCGTACCAAA	AACGCTGGGC	AATCCGCTGT	ATCACTGGAC	GCACCTTGAA	CTACGCCGTC
-	±1	PFG	I T G T	L F G	P D T A	E S I	W T Q	C N E K	L A T	P A F S	A R G
	5201	CATTTGGCAT	TACAGGTACG	CTGTTCGGAC	CGGATACCGC	GGAAAGTATC	TGGACGCAGT	GTAATGAGAA	ACTGGCGACG	CCGGCCTTTT	ccececece
_		GTAAACCGTA	ATGTCCATGC	GACAAGCCTG	GCCTATGGCG	CCTTTCATAG	ACCTGCGTCA	CATTACTCTT	TGACCGCTGC	GGCCGGAAAA	GGCGCGCGCC
	+1	TATTATGCAG	CAGATGAATG	TGCGGATGGT	CGGAACCACC	GACGACCCGA	TAGATTOTOT	GGAATATCAC	CGCCAGATAG	CCGCCGACGA	CAGCATTGAT
SEN2980		ATAATACGTC	GTCTACTTAC	ACGCCTACCA	GCCTTGGTGG	CTGCTGGGCT	ATCTAAGAGA	CCTTATAGTG	GCGGTCTATC	GGCGGCTGCT	GTCGTAACTA
470 aa-1149 b	p. +1	I E V	A P S W	R P D	K V F	K I E L	DGF	V D Y	L R K L	E A A	A D V S
4807-0270 Fr: 3 186 866	5401	TAACTTCAGC	GCGGTTCGAC	CGCGGGGGGCTG	TTTCAAAAGT	TTTAGCTTGA	CCTGCCGAAA	CAGCTAATGG	ACTCCTTTGA	CCTTCGCCGC	CGTCTACAGT
To; 3,188,278	+1	SITE	RFDD	LRQ	A L T R	RLD	HFA.	A C G C	R A S	D H G J	ETL
B50YB2	5501	GCATTACCCG	TTTCGACGAT	TTACGTCAGG	CGCTCACTCG	CCGCCTCGAC	CATTTCGCCG	CCTGCGGCTG	CCGCGCGTCG	GATCATGGCA	TTGAAACGCT
Glucuronate	+1	LR F A	P V P	D D A G	LDA	I L G	K R L A	GET	LSE	LEIA	QFT
isomerase	5601	GCGATTTGCG	CCGGTGCCCG	ACGACGCGCA	GCTTGACGCC	ATTCTGGGCA	AACGTCTGGC	TGGCGAAACG	CTGAGCGAAC	TTGAGATCGC	CCAGTTTACC
isomerase		CGCTAAACGC T A V	GGCCACGGGC	G R O	Y A A	TAAGACCCGT R G W V	M O L	ACCGCTTTGC H I G	GACTCGCTTG A I R N	AACTCTAGCG N N T	GGTCAAATGG R M F R
	+1 5701	ACGGCGGTGC	тестстест	GGGCCGTCAG	TACGCCGCGC	GCGGCTGGGT	GATGCAGCTA	CATATTGGCG	CGATCCGTAA	TAATAATACC	CGGATGTTCC
_		TGCCGCCACG	ACCAGACCGA	CCCGGCAGTC	ATGCGGCGCG	CGCCGACCCA	CTACGTCGAT	GTATAACCGC	GCTAGGCATT	ATTATTATGG	GCCTACAAGG
	+1 5801	GCCTGCTGCG	GCCGGATACC	GGCTTTGACT	CCATTGGCGA	TAATAACATT	AGCTGGGCGC	тетессотт	GCTCGACAGT	ATGGATGTGA	CCAATGAACT
_		CGGACGACCC	CGGCCTATGG	CCGAAACTGA	GGTAACCGCT	ATTATTGTAA	TCGACCCGCG	AGAGGGCAAA	CGAGCTGTCA	TACCTACACT	GGTTACTTGA
	+1	LPKT	I L Y	C L N F	R D N	EVL	A T M I	G N F	Q G P	G I A G	K V Q
	5901	GCCCAAGACT CGGGTTCTGA	ATCCTCTATT TAGGAGATAA	GCCTGAACCC CGGACTTGGG	ACGTGATÀÀC TGCACTATTG	GAAGTCCTGG CTTCAGGACC	CGACCATGAT GCTGGTACTA	CGGTAACTTC GCCATTGAAG	CAGGGGCCGG GTCCCCGGCC	GAATTGCCGG CTTAACGGCC	RRAAGTGCAG TTTTCACGTC
_	+1	FGS	G W W F	NDQ	K D G	MLRQ	LEQ	LSQ	MGLL	SQF	VGML
	6001	TTTGGTTCCG	GCTGGTGGTT	TAACGATCAG	AAAGACGGTA	TGCTGCGCCA	ACTEGACCAA	CTGTCGCAAA	TGGGACTGTT	AAGTCAGTTT	GTCGGGATGC
_	+1	L T D S	S R S F	L S Y	T R H E	Y F R	R I L	C N L L	G Q W	A Q D G	E I P
	6101	TGACCGACTC	CCGCAGTTTC	СТТТСТТАТА	CGCGACATGA	ATATTTCCGT	CGTATTCTCT	GTAACCTGCT	GGGACAGTGG	GCGCAGGACG	GAGAGATTCC
_		ACTGGCTGAG	A M L	GAAAGAATAT S R M V	GCGCTGTACT	C F N	GCATAAGAGA	CATTGGACGA Y F T	CCCTGTCACC I K	CGCGTCCTGC	CTCTCTAAGG
	+1 6201	TGATGATGAA	GCGATGCTAA	GCCGTATGGT	TCAGGATATC	TGCTTCAATA	ATGCCCAGCG	TTATTTCACG	ATTAAATAAT	CGCTATTAAT	CTGTGAGCAC
		ACTACTACTT	CGCTACGATT	CGGCATACCA	AGTCCTATAG	ACGAAGTTAT	TACGGGTCGC	AATAAAGTGC	TAATTTATTA	GCGATAATTA	GACACTCGTG

	+1	PDDE	AML	SRMV	QDI	CFN	N A Q R	YFT	IK			
	6201	TGATGATGAA	GCGATGCTAA	GCCGTATGGT	TCAGGATATC	TGCTTCAATA	ATGCCCAGCG	TTATTTCACG	ATTAAATAAT	CGCTATTAAT	CTGTGAGCAC	
		ACTACTACTT	CGCTACGATT	CGGCATACCA	AGTCCTATAG	ACGAAGTTAT	TACGGGTCGC	AATAAAGTGC	TAATTTATTA	GCGATAATTA	GACACTCGTG	
	6301	AGGTCTCAGG	TATTTCGAGG	GGCAGGAAGG	CGGTAAGAGA	GTGAGAAATT	TAGCAGTAAC	GCACTTTTCC	AGCCAAAGTC	TGGCCGCCGA	TAAAAAGCAA	
		TCCAGAGTCC	ATAAAGCTCC	CCGTCCTTCC	GCCATTCTCT	CACTCTTTAA	ATCGTCATTG	CGTGAAAAGG	TCGGTTTCAG	ACCGGCGGCT	ATTTTTCGTT	
	6401	AGATGTCGCT	GATTAATCCC	GAATCACTTA	CTAAGGTAAG	TGATTCGGGT	GAACAAACAC	AGCCATCTTC	TGCGTCATGC	TTCTTTTCTC	TGTCCGGAAT	
		TCTACAGCGA	CTAATTAGGG	CTTAGTGAAT	GATTCCATTC	ACTAAGCCCA	CTTGTTTGTG	TCGGTAGAAG	ACGCAGTACG	AAGAAAAGAG	ACAGGCCTTA	
	6501	AACTCCAGTG	CAGAAAGTGC	GATACATCAG	TCCTGTTGAT	GATGGCTCCT	GTTGGTTATA	CGGTAATAAC	TCATGACTTC	ACAGTGTGGT	TTCTTTTAAC	
		TTGAGGTCAC	GTCTTTCACG	CTATGTAGTC	AGGACAACTA	CTACCGAGGA	CAACCAATAT	GCCATTATTG	AGTACTGAAG	TETCACACCA	AAGAAAATTG	
	-2										F K L L	
	6601	ABATCGACAG	CCDCCDTTTC	CACCTECTTA	ACCACCTCAT	TACTOCATCO	CTCCDTTCTC	TCCDCTTCTT	CCCA CATCOC	ATTCATATCE	CCTECEECOD	
	0001	TTTAGCTGTC	GGTCGTAAAG	GTCGACCAAT	TECTECACTA	ATGACCTAGC	GAGGTAAGAC	AGCTCAAGAA	CECTETAECE	TAAGTATAGC	GGACGCCGTT	
	~					HIGHCCIMBC						
	-2	LUVA	LME	LUN	V V E N	S S R	E M R		5 M A	N M D	GAAF	
	6701	MATAGTTCAG	CGCAGAGTGG	CCAGCGGTAT	GCAUCTUTTT	ATGAGGCGCT	ACCREACTAC	GGAAACTGTA	GTAATTGUTA	AACTGCTGTC	CITCAAAACC	
		TIATCAAGTC	GCGTCTCACC	GUICUCAIA	CGIGGAGAAA	TACICCOCBA	AGGICIGATO	CUTITGACAT	CATTAACGAT	TIGACGACAG	GAAGITIIGG	
	-2	FYNL	A S H	G A T H	VEK	HPA	ELSR	FSY	Y N S	FQQG	EFG	
	6801	GTAGTACCAT	TTCCCCAGCC	GACACTGATC	GTGCATCGTT	ATTTTACTGT	TTATGTCTTT	ATTTAAGAGT	AATTTGTAGA	CTTCCATTTT	CCAGATAACA	
		CATCATEGTA	AAGGGGTCGG	CTGTGACTAG	CACGTAGCAA	TAAAATGACA	AATACAGAAA	TAAATTCTCA	TTAAACATCT	GAAGGTAAAA	GGTCTATTGT	
	-2	Y Y W	K G L F		HMT	I KSN	IDK	NLL	L KY V	EMK	WIVH	
	6901	TGGTCAACTT	TGACAATATT	CAAAAACTGC	ACCGTAGAGA	TGTACTGCAT	GACCGATTTC	ATGCTCAGCG	ATTTATCGAT	GACGATGCCA	ATAGACTCCA	
		ACCAGTTGAA	ACTGTTATAA	GTTTTTGACG	TGGCATCTCT	ACATGACGTA	CTGGCTAAAG	TACGAGTCGC	TAAATAGCTA	CTGCTACGGT	TATCTGAGGT	
SEN981	-2	н р у к	VIN	LFQ	V T S I	YQM	VSK	M S L S	KDI	VIG	ISEV	
351 aa-1056 bp.	7001	CAATTTGATT	TATGTTAGTC	GTAATATTAT	CAATAACCGG	CTGTTGGTTA	TCCAGTACCT	CACTCACTCT	GGCCGTATTA	TCTTTGATTA	CAGAGGTTAA	
6582-7634		GTTAAACTAA	ATACAATCAG	CATTATAATA	GTTATTGGCC	GACAACCAAT	AGGTCATGGA	GTGAGTGAGA	CCGGCATAAT	AGAAACTAAT	GTCTCCAATT	
FR; 3,188,578	-2	VIQN	INT	TIND	IVP	Q Q N	DLVE	SVR	ATN	D K I V	STL	
To; 3,189,633	7101	CGTACTGACG	CTTTTCGAAC	TGTGTTTAAC	ATCTTCGGAC	AGGTTTTTCA	CCTCTTTCGA	AATAACGCTA	AAACCACGCC	CGGCATCCCC	TACGCGCGCG	
DSOVD2		GCATGACTGC	GAAAAGCTTG	ACACAAATTG	TAGAAGCCTG	TCCAAAAAGT	GGAGAAAGCT	TTATTGCGAT	TTTGGTGCGG	GCCGTAGGGG	ATGCGCGCGC	
	-2	TSV	S K S S	HKV	DES	LNKV	EKS	IVS	FGRG	A D G	VRAA	
Uncharacterized	7201	GCCTCGATCG	CGGAGTTAAT	CGCTATCAAA	TTGGTCTGGT	TAGCGATCTT	CTGTATCTCT	TTAATACACG	CATTAATTTG	CGTTAAGGAG	GTATTTAAGT	
protein		CGGAGCTAGC	GCCTCAATTA	GCGATAGTTT	AACCAGACCA	ATCGCTAGAA	GACATAGAGA	AATTATGTGC	GTAATTAAAC	GCAATTCCTC	CATAAATTCA	
	-2	AEIA	S N I	A I L	N T O N	A I K	O I E	K I C A	NIO	T L S	TNLD	
	7301	CRCCCATACR	ACAGCCAATA	TCCGATGAAG	TCGTGCCAAT	TTCGGTGATA	AATTCCACGA	GCGATTTTAA	TGAACTTCTG	GCTTTATCAT	TCTGTTCGTT	
		GCGGGTATGC	TGTCGGTTAT	AGGCTACTTC	AGCACGGTTA	AAGCCACTAT	TTAAGGTGCT	CGCTAAAATT	ACTTGAAGAC	CGAAATAGTA	AGACAAGCAA	
	-2		C G T	D S S T	T G T	E T T	FEVI	SKL	S S R	A K D N	OEN	
	7401	AAGTTCATTA	ATGCGTTCTT	TTTCAATGTT	GAGCTGCTCA	CARGACTTAN	TGATGGCATT	АССЛАТСАТА	TCGATCETCG	AGATCCCCAT	CAGAATTTTT	
		TTCAAGTAAT	TACGCAAGAA	AAAGTTACAA	CTCGACGAGT	GTCCTGAATT	ACTACCETAA	TECTTACTAT	AGCTAGCAGC	TCTAGGGGTA	GTCTTAAAAA	
	-2						T 3 M			T 0 M		
	7501	TEACACACTA	ТССССТСАТА	ACCTCACCA	GTATCCACCA	CCACETCACA	TECATTACCO	CCACATTORS	CTRECCECCO	TATTTTACA	ACATTATTAT	
	7301	ACTETETCAT	ACCCCACTATA	TCCCACCA	CATACCACCA	CEACOLEAGA	ACCTABLICCO	CCAGATICIG	CIGGCGGCGG	ATAAAAATCT	TCTABTABTA	
		ACIDICICAI	AGOCOAGIAI		CAIAGOIGGI	GOIDCADICI	ACOTANICOC	BOICIANDAC	GACCOCCOCC		IGIAAIAAIA	
	-2	Q C L I	R E Y	P E P	TDVV	V D S	A N A	G S E A	P P P	IKV	V N N N	
	7601	TCCAGTGGCG	CAAGCCTAAT	ACTTTTTTAA	ACATATACCA	TICTCCTTAT	TATATGTGGA	CATCCTGATC	TATGCCTGGA	TAAAGTATTA	TCAACTCATG	
		AGGTCACCGC	GTTCGGATTA	TGAAAAAATT	TETATATGGT	AAGAGGAATA	ATATACACCT	GTAGGACTAG	ATACGGACCT	ATTTCATAAT	AGTTGAGTAC	
		that was to see them. I										

1- SEN2978 >sp|B5QYB0|UXUA_SALEP Mannonatedehydratase OS=Salmonella enteritidis PT4 (strain P125109) GN=uxuA PE=3 SV=1

MKQTWRWYGPNDPVTLSDVRQAGATGVVTALHHIPNGEIWSIDEIQKRKAIVEEAGLEWS VVESVPIHEDIKTHTGQYDLWIKNYQQTLRNLAQCGIYTVCYNFMPVLDWTRTDLEYVLP DGSKALRFDQIEFAAFELHILKRPGAEADYTAEEIAQAERRFATMSEEDKARLTRNIIAG LPGAEEGYTLDQFRQHLATYKDIDKAKLREHFAYFLKAIIPVADEVGVRMAVHPDDPPRP ILGLPRIVSTIEDMQWMVETVNSMANGFTMCTGSYGVRADNDLVDMIKQFGPRIYFTHLR STLREENPKTFHEAAHLHGDVDMYEVVKAIVEEEHRRKAEGSDDLIPMRPDHGHQMLDDL KKKTNPGYSAIGRLKGLAEVRGVELAIQRAFFSK

2- SEN2979 >tr|B5QYB1|B5QYB1_SALEP D-mannonateoxidoreductase OS=Salmonella enteritidis PT4 (strain P125109) GN=SEN2979 PE=4 SV=1

MEQNIATAQVSVARPNWDKSRLVSRIVHLGCGAFHRAHQALFTHHLLEKSDSDWGICEVN LMPGNDARLIANLKAQNLLYTVAERGAESTELKIIGSMKEALHPEFDGHAGILAAMARPE TAIVSLTVTEKGYCTDPASGELDVNNPLIQNDLAHPQQPKSAIGYIVEALNMRREQGLKA FTVLSCDNVRENGHVAKAAVLGLAKARDAALAAWIADNVTFPCTMVDRIVPAATEETLQL VADQLGVYDPCAIACEPFRQWVIEDNFVNGRPDWDTVGAQFVADVVPFEMMKLRMLNGSH SFLAYLGYLGGYDTIADTMTNPAYRRAALALMLDEQAPTLSMPEGTDLEGYANLLIARFT NPSLKHRTWQIAMDGSQKLPQRLLDPVRLHLQQGDDYRRLTLGVAGWMRYVGGIDEQGKT IDVVDPLLAQYQAIHQQYQTPEERVRGLLAIESIFGSDLPKNHEFVQAVTDAYQQLLQNG AKATVEALAK 3- SEN2980 >sp|B5QYB2|UXAC_SALEP Uronateisomerase OS=Salmonella enteritidis PT4 (strain P125109) GN=uxaC PE=3 SV=1

MATFMTEDFLLKNDIARTLYHKYAAPMPIYDFHCHLSPQEIADDRRFDNLGQIWLEGDHY KWRALRSAGVDESLITGKETSDYEKYMAWANTVPKTLGNPLYHWTHLELRRPFGITGTLF GPDTAESIWTQCNEKLATPAFSARGIMQQMNVRMVGTTDDPIDSLEYHRQIAADDSIDIE VAPSWRPDKVFKIELDGFVDYLRKLEAAADVSITRFDDLRQALTRRLDHFAACGCRASDH GIETLRFAPVPDDAQLDAILGKRLAGETLSELEIAQFTTAVLVWLGRQYAARGWVMQLHI GAIRNNNTRMFRLLGPDTGFDSIGDNNISWALSRLLDSMDVTNELPKTILYCLNPRDNEV LATMIGNFQGPGIAGKVQFGSGWWFNDQKDGMLRQLEQLSQMGLLSQFVGMLTDSRSFLS YTRHEYFRRILCNLLGQWAQDGEIPDDEAMLSRMVQDICFNNAQRYFTIK

4- SEN2977 >tr|B5QYA9|B5QYA9_SALEP Hexuronate transporter OS=Salmonella enteritidis PT4 (strain P125109) GN=SEN2977 PE=4 SV=1

MKMTKLRWWIIGLVCVGTIVNYLSRSSLSVAAPAMMKELHFDEQQYSWVVSAFQLCYTIA QPITGYLMDVIGLKIGFFIFALLWSLINMAHALAGGWISLAFLRGLMGLTEASAIPAGIK ASAEWFPTKERGIAGGLFNIGTSIGAMLAPPLVVWAMLTFADSGIGTEMAFVITGGIGVL FAITWFLIYNSPNKHPWITHKELRYIEDGQESYLQDDNKKPAVKEIVKKRNFWALAITRF LADPAWGTLSFWMPLYLINVMHLPLKEIAMFAWLPFLAADFGCVAGGFLAKFFMEKMHMT TINARRCSFTIGAVLMISIGFVSITTNPYVAIALMSIGGFAHQTLSTVVITMSADLFKKN EVATVAGLAGSAAWMGQLSFNLFMGALVAIIGYGPFFIALSLFDIIGAIILWVLIKDPEK HHPPMTEQPLASHR

5-SEN2981 >tr|B5QYB3|B5QYB3_SALEP Uncharacterized protein OS=Salmonella enteritidis PT4 (strain P125109) GN=SEN2981 PE=4 SV=1

MFKKVLGLRHWNNNVVKIPPPAESGANASDVVVDTPEPYERILCQKILMGISTIDIIRNA IIKSCEQLNIEKERINELNEQNDKARSSLKSLVEFITEIGTTSSDIGCRMGDLNTSLTQI NACIKEIQKIANQTNLIAINSAIEAARVGDAGRGFSVISKEVKNLSEDVKHSSKSVSTLT SVIKDNTARVSEVLDNQQPVIDNITTNINQIVESIGIVIDKSLSMKSVMQYISTVQFLNI VKVDHVIWKMEVYKLLLNKDINSKITMHDQCRLGKWYYGFEGQQFSNYYSFRSLEAPHKE VHTAGHSALNYFAAGDMNAMSQELDRMERSSNEVVNQLEMLAVDLLKETTL

Appendix 3:	Translated sequence	of induced genes	(using V	ector NTI program)
--------------------	---------------------	------------------	----------	--------------------

	13										
	901	CTGT1G11CC	CCATGCGGTG	GCGACGCTGT	CCTGTATTAC	TACCATTOTT	AGCGACAGCG	CGCTCCCGTC	TGCTATCGCC	CGACGTTATC	AGC AGGC GGG
		GACATCTTGG	GGTACGCCAC	CGCTGCGACA	GGACATAATG	ATGGTAAGAA	TCGCTGTCGC	GCGAGGGCAG	ACGATAGCGG	GCTGCAATAG	TCGTCCGCCC
_	+3										
	+2					_					
	1001	GTGCCGTCTC	ATCATGCCAG	ACCCCTCATC	AGGAGTGCGT	TAATGGACAT	CAGGCAAATG	AACAAAACCC	ATCTGGAGCA	CTGGCGCGGA	TTGCGAAAAC
		CACGGCAGAG	TAGTACGGTC	TGGGGAGTAG	TCCTCACGCA	ATTACCTGTA	GTCCGTTTAC	TTGTTTTGGG	TAGACCTCGT	GACCGCGCCT	AACGCTTTTG
SEN1437	+2										
145 aa-438 bp	1101	AGCTCTGGCC	CGGTCACCCG	GATGACGCCC	ATCTGGCAGA	CGGCGAAGAA	ATCCTGCAAG	CCGATCATTT	GGCATCATTT	GTTGCGATGG	CAGACGGGGT
1043 - 1480		TCGAGACCGG	GCCAGTGGGC	CTACTGCGGG	TAGACCGTCT	GCCGCTTCTT	TAGGACGTTC	GGCTAGTAAA	CCGTAGTAAA	CAACGCTACC	GTCTGCCCCA
Fr; 1,526,021 To: 1,526,458	+2	00003888000	MMM04003 M 0	0000330000	003.003.003.003.00	00033000000	0003/03/0000/	0000000000	MMM/MM/233 /	олаллала	mamaaaamaa
B5R542	1201	CCCCTAACCC	ANACECCTAC	CUTCAATCUG	CCACGATTAT	CACTTACCEA	CACTETCAAC	CEEECACCEA	ТТТСТТЬААС АЛАСААСТТС	CATAAAAAACA	ACACCCCTCG
Amino	+2	COULINACCO	AAACOCCIAC	UNAUT INDUC	UUIUCIAAIA	CHUIINCCUN	CACIDICANO	COOCACCOA	AAAOAACIIC	(A IAAAAA (A	AUNOUGUNUC
glycoside	1301	TTCCGTCAAC	GCGGCGTAGC	GAAACAATTG	ATTGCAGCGG	TGCAACGATG	GGGAACGAAT	AAAGGGTGTC	GGGAAATGGC	CTCTGATACC	ACGCCGGAAA
N (6')-acety		AAGGCAGTTG	CGCCGCATCG	CTTTGTTAAC	TAACGTCGCC	ACGTTGCTAC	CCCTTGCTTA	TTTCCCACAG	CCCTTTACCG	GAGACTATGG	TGCGGCCTTT
Transferase	+2								,		
	1401	ATACAATTTC	CCAGAAAGTT	CACCTEECAT	TAGGATTTGA	GGAAACAGAG	CGCGTCATTT	TCTACCGGAA	GCGTTGTTGA	GGGAAATATT	CAGGCTCCTT
_		TATGTTAAAG	GGTCTTTCAA	GTGGACCGTA	ATCCTAAACT	CCTTTGTCTC	GCGCAGTAAA	AGATGGCCTT	CGCAACAACT	CCCTTTATAA	GTCCGAGGAA
	1501	ATAGAGAAGT	GGTCACTGAT	TAAAAAGCAG	CCTCAATATA	CAATAATTCC	TGTTTCTAAA	ACGGAGTCCA	TCAGGACGTC	GGGAATAAGA	CCCGGCAGAT
_		TATCTCTTCA	CCAGTGACTA	ATTTTTCGTC	GGAGTTATAT	GTTATTAAGG	ACAAAGATTT	TGCCTCAGGT	AGTCCTGCAG	CCCTTATTCT	GGGCCGTCTA
	1601	GCAAGCCGGG	CCTTCATTGC	TTACGGTCTT	GAGGCCGTAC	CGTCTGGTGT	GCGCGCCATT	GTCCAGGATG	GGATTCCCCC	CTCGCACGGA	TATTTCGCCG
		CGTTCGGCCC	GGAAGTAACG	AATGCCAGAA	CTCCGGCATG	GCAGACCACA	CGCGCGGTAA	CAGGTCCTAC	CCTAAGGGGGG	GAGCGTGCCT	ATAAAGCGGC
_	-2			* P R	S A T G	DPT	R A M	TWSP	IGG	ECP	ΥΚΆΆ
	1701	CTTTGGCTTC	ATCAATGTCA	ATACCCAGAC	CCGGCTTATC	GTTCAGATAG	GCATAACCGT	GGTCAATCTC	AGGGCAGCCC	GGGAAGACAT	CACGCAGCGC
			TAGTTACAGT	TAIGGGICIG	GGUUGAATAG		CGTATIGGUA	CURGITAGAG			GIGUGIUGUG
-	-2	AKAE	DID	IGLG	F K D	N L Y	A Y G H	DIE	P C G	PFVC	RLA
	1801	GTCGTTCATT	GGCGTGTATT	CCTGAATGCC	AAAGTTCGGT	GAACTCAAAT	CCARATGCAT	ATTGGCGCAA	ACCCCAACTG	GAGAAATGTC	GCCCGGCCCG
		CAGCAAGTAA	CUGURURIAR	GGACITACGG	TITCARGOUR	CIIGAGIIIA	GGITTACGIA	TRACCOCGIT	TOOGGITGAC	CICILIACAG	CGGGCCGGGC
_	-2	D N M	P T Y E	Q I G	F N P	S S L D	L H M	N A C	V G V E	S I D	G P G H
	1901	ACCATCCCG	ICCCCACACC ACCCCACACC	CINGTONIC	GACATCTGTC	GTTCNNGGN	COCCOGCOTA	TACGGCGGCCGA	TGGTACTCAC ACCATGAGTG	CACCATCACT	ATGTAGTCGA
-	2001	TGAGCCTCTT	CTCCATCACC	COCTTOCACT	CATTCACATT	CICINCICC	TCACCATCO	T G G I	CONCICTOR	Teccertect	TTTTCACCA
	2001	ACTOGGACIA	CIGCTICTCG	CCGNGGTGN	GTANGTGTAN	GTGTTTGTCG	AGTGGGTACC	TATAGCCTCA	CCTTCTGACG	ACCGCGTCGT	ABBIGTOGGT
	_2			D V M D	N V N		E A M C		1 2 2 0		
SEN1426	2101	GTCGATGTTT	тессетессь	COGGATCOTO	AAGATAAAAT	AGTTGATACT	CTTCNGCGT	TTTGGCCAGA	TTNTGGCNG	TCACCCCCCT	CACCCCCTCA
3EN1430	2101	CAGCTACAAA	AGGCCACGGT	GGCCTAGGAG	TTCTATTTTA	TCAACTATGA	CAAGTTCGCA	AAACCGGTCT	AATTACCGTC	AGTGGCCCCA	GTGCGCGAGT
1621-2880	-2		E P & V	PDF	I. Y F	LOYO	ELT	K & L	NTAT	V P T	VREH
Fr; 1,524,621	2201	TGTACATCAT	GANTANTTC	GITGCCIIII	CCGAGCTTGT	TICGCIGGTG	GTCGLICIGI	CGCGGLLCGC	TTTTGGCGTA	GGCGTCCGGG	TCANAGTAGA
To; 1,525,880		ACATGTAGTA	CTTATTTAAG	CTACGGTTTT	GGCTCGAACA	ATGCGTCCAC	CAGCTTGTCT	GCGCCTTGCG	AAAACCGCAT	CCGCAGGCCC	AGTTTCATCT
B5R541	-2	H V D H		IGF	G L K N	R L H	DFL	R P V S	K A Y	A D P	DFYI
putative	2301	TGCCGGGCGT	TTTGCTGCGC	GGTGAACGTT	TCGGCTGAAT	GTTCTTCGCA	CGTGCCAGCT	GCGTGGCGAT	CAGTTTAAGG	TCATCGGTCC	CCGCACCGCC
dehvdratase		ACGGCCCGCA	AAACGACGCG	CCACTTGCAA	AGCCGACTTA	CAAGAAGCGT	GCACGGTCGA	CGCACCGCTA	GTCAAATTCC	AGTAGCCAGG	GGCGTGGCGG
	-2	IG P T	KSR	PSRF	C P Q I	NKA	R A L Q	TAI	LKL	D D T G	AGG
-	2401	ATACATTCCC	ATCTGGCAGC	GAACGTACTG	GTAACCCTCT	TCCATTCTGG	CTCGGATGTT	ATCTTCCACT	TCAACTTCAT	CACCACCATC	GGTGTGGCAG
		TATGTAAGGG	TAGACCGTCG	CTTGCATGAC	CATTGGGAGA	AGGTAAGACC	GAGCCTACAA	TAGAAGGTGA	AGTTGAAGTA	GTGGTGGTAG	CCACACCGTC
	-2	YMG	M Q C R	V Y Q	YGE	EMRA	RIN	DEV	EVEE	G G D	тнсү
-	2501	TACAGTGGGA	TCCCATCACG	ACATTTGCCG	CCCAGCAGAT	CATAAACTGG	CATCCCCGCT	AGCTTACCTT	TGATATCCCA	CAGTGCCATG	TCCACGCCGG
		ATGTCACCCT	AGGGTAGTGC	TGTAAACGGC	GGGTCGTCTA	GTATTTGACC	GTAGGGGCGA	TCGAATGGAA	ACTATAGGGT	GTCACGGTAC	AGGTGCGGCC
_	-2	YLP I	GDR	CKG	G L L D	Y V P	MGA	L K G K	'I D'W'	'L'A'M'	D V GS
	2601	ACAATGCATT	GTTCATGATC	GGGCCATTGC	GCCAGTAGCC	GCTCACCACG	CCTGACTGCC	AGATGTCCTC	AATACGGGTC	GGATCTTTGC	CAACCAGAAA
		TGTTACGTAA	CAAGTACTAG	CCCGGTAACG	CGGTCATCGG	CGAGTGGTGC	GGACTGACGG	TCTACAGGAG	TTATGCCCAG	CCTAGAAACG	GTTGGTCTTT
_	-2	SLAN	NMI	P G N F	ε W Y G	's'v'v'	<u>GʻSʻQʻ</u> W	I I D'E'	IRT	PDKG	V L F
	2701	AGGCGCCATG	TACTCATCGA	TCGCGCTTTT	TACCGCGAAG	ATACGCTGGG	TAAATGTCGC	ACATCCCAGC	CCATACAGCC	CTGGCTCGTT	GGTTTCTATC
		TCCGCGGTAC	ATGAGTAGCT	AGCGCGAAAA	ATGGCGCTTC	TATGCGACCC	ATTTACAGCG	TGTAGGGTCG	GGTATGTCGG	GACCGAGCAA	CCARAGATAG
_	-2	PAM	Y E D I	'A'S'K	VAF	I RQ T	FTA	C G L	G Y L G	PEN	TEIK
	2801	TTACGACTG	CCARATCAAT	GCCGCCCGGC	GCCGTCAGAA	TCGTTTTCAC	GTTGGTAATT	TTCAGGTTAC	TCACTTTCAC	TCCTTAACTC	AGCGAATTTA
		AATTGCTGAC	GGTTTAGTTA	CGGCGGGCCG	CGGCAGTCTT	AGCAAAAGTG	CAACCATTAA	AAGTCCAATG	AGTGAAAGTG	AGGAATTGAG	TCGCTTAAAT
_	-2	K V V A	L'D'I'	GGP	A TL I	TKV	NTI	K L N S	V K V		
	2901	CCGCATGTCA	TACCGTTGTG	GATGATACAA	CAGATATACG	TAAGTTGTCC	TACAACTTTA	TGCTGTTATA	TCACGCTTCT	TTTACAGCGG	ATCGCTGCGA
-		GGCGTACAGT	ATGGCAACAC	CTACTATGTT	GTCTATATGC	ATTCAACAGG	ATGTTGAAAT	ACGACAATAT	AGTGCGAAGA	AAATGTCGCC	TAGCGACGCT
	3001	GTGCCATTCT	GTGACAGCGA	TCAAACTGCG	TTAAATTTAA	ACCANATATA	TACATATAAT	TCAATTAATT	AATATTAATT	ACGCGGGTCGG	CATGTGGAGC
		CACGGTAAGA	CACIGICGCT	AGITIGACGC	AATTTAAATT	IGGITTATAT	AIGIATATTA	AGITAATTAA	TTATAATTAA	TACACCHACC	GIACACCICG

PhD Thesis

Appendix

-		
	+2	MTA LFD LTG KTAL VTG SAR G LGF AYA E GLA
	3201	GAGGAAACAA CATGACCGCT TTATTTGATT TAACTGGGAA AACGGCGCTG GTAACGGGTT CTGCACGAGG ACTGGGCTTT GCCTACGCAG AGGGTCTTGC
		CTCCTTTGTT GTACTGGCGA AATAAACTAA ATTGACCCTT TTGCCGCGAC CATTGCCAA GACGTGCTCC TGACCCGAAA COGATGCGTC TCCCAGAACG
-		
	+2	
	3301	CGCTGCGGGT GCACGGGTTA TCCTGAATGA TATTCGCGCC ACGCTGTTGG CCGAATCAGT GGATACGCTG ACCAGAAAAG GCTACGACGC GCATGGCGTG
		GCGACGCCCA CGTGCCCAAT AGGACTTACT ATAAGCGCGG TGCGACAACC GGCTTAGTCA CCTATGCGAC TGGTCTTTTC CGATGCTGCG CGTACCGCAC
-	1.2	AFDVTDELAIEAAFSKLDAEGIHVDILINNAGIQ
	+4	
	3401	GETTIGAEG TEREEGATGA REIGGEGATI GREGERGETI TIRGERREET IGREGERGAR GEGERTEERTG TIGATATEET GREGERATARE GEEGGTATEE
-		CGAAAACTGC AGTGGCTACT TGACCGCTAA CTCCGTCGAA AATCGTTTGA ACTACGTCTT CCCTAGGTAC AACTATAGGA CTAGTTATTG CGGCCATAAG
	+2	QYR K PMV ELE LENWQKV IDT NLT SAFL VSR SAAK
	3501	AUTOCICACAN ACCANTGETC GAGCTGEAGC TEGANALITE CONGALGETE ATGALACTA ACCEGACANA
SEN1435	- 3301	ANALONA ACCARDING CARCHORDO INCARACIÓ CONTRACIÓ ALCONACIA ACCIDACIÓN CONTRALA CONCECCARA
355 or 7(9 ho		
255 aa-768 bp.	+2	real arm see riin ies bis y aar rivarii taar
3212 - 3979	3601	ACGGATGATC GCCCGCAACA GCGGCGGCAA AATTATTAAT ATCGGCTCGC TTACCAGCCA GGCGGCCCGC CCGACTGTTG CCCCGTACAC GGCAGCAAAA
Fr: 1.523,522		TECCTACTAE CEGECETTET CECCECCETT TTAATAATTA TAECCEAECE AATGETCEET CCECCEGECE GECTEACAAC GEGECATETE CCETCETTTT
To: 1 524 289		GGIKMLTCSM AAEWAOFNIO TNAIGPGYIL TDMN
	+2	
B5R540	3701	GECEGCATCA ARATECTCAC CTECTCEATE ECTECTEAAT GEECECAETT TAATATCCAE ACTAACECCA TTEEACCTEE CTACATTCTE ACCEACATEA
outative		CCGCCGTAGT TTTACGAGTG GACGAGCTAC CGACGACTTA CCCGCGTCAA ATTATAGGTC TGATTGCGGT AACCTGGACC GATGTAAGAC TGGCTGTACT
patative	+2	NTALIED KQFD SWV KSS TPSQ RWG RPE ELIG TAI
hexonate	3801	NACCOCCT TATTOMAN MACAGETICS ATAGETISSET CAMAGEAGE ACCOUNTED ACCOUNTS ACCOUNTS SACTATOS
dehvdrogenas	e	
	_	
	+2	T T T T T T T T T T T T T T T T T T T
	3901	CTTCTTATCA TCCAAAGCAT CAGATTATAT CAACGGGCAG ATTATTACG TCGATGGTGG CTGGCTCGCA GTTTTATAAG TCGCCCGCAC TCTACCTACT
		GAAGAATAGT AGGTTTCGTA GTCTAATATA GTTGCCCGTC TAATAAATGC AGCTACCACC GACCGAGCGT CAAAATATTC AGCGGGCGTG AGATGGATGA
-	+1	MÀ CVY PDÀ CRYN TGIIMKÀ SRQ RLFILTL
	4001	COCCTITARA ATTOATOGCA TOTOTTACC CCGATGCATG CCGATATARC ACAGGTATTA TTATCAAGGC TTCAAGACAA CGACTCTTTA TTCCACATT
		GOGALLITT TARTACCT ACACALITIC GOTACTAC GOTATING GOTATING GOTATING AND ACACALITY ACAC
-		LI, F, T, Y, T, J, T, N, Y, M, D, R, J, N, J, Y, J, C, S, N, T, O, N, D, F, C, T, D, T, O, T,
	+1	
	4101	GCTGTTTATT GTGACTGCAA TTAATTATAT GGATCGCGCC AATCTTGCCG TCGCTGGTTC GAATATCCAG AATGATTTCA GTCTGACGCC AACACAACTG
_		CGACAAATAA CACTGACGTT AATTAATATA CCTAGCGCGG TTAGAACGGC AGCGACCAAG CTTATAGGTC TTACTAAAGT CAGACTGCGG TTGTGTTGAC
	+1	G L L F S M F T W A Y A A S Q I P V G Y V L D R I G S R I L Y G G A
	4201	GOTTIGETTI TETECATOTI TACCTOGOCE TACCETOCEA GTEAAATTEE TOTEGGETAT GTEETGATE GEATTGEETE ACGTATTETT TATGGTGGTG
		CCAAACGAAA AGAGGTACAA ATGGACCCGG ATGCGACGGT CAGTTTAAGG ACAGCCGATA CAAGAACTAG CGTAACCGAG TGCATAAGAA ATACCACCAC
-		ATTLWSTFTFMMGFASHHLFATATASFAMLLACR
	+1	
	4301	CGATTATICT GIGGAGIAIT THACCTHA IGAIGGGIT CGCCEACAC CATTAITCG CGACGGCAAC CGCTEAITT GGAAGCATE IGGCCIGCCG
-		GCTARTARGA CACCTCATAR ARATGGARAT ACTACCCCAR GCGGAGTGTG GTARATARGC GCTGCCGTTG GCGARGTARA CGTTACGARG ACCGGACGGC
	+1	RALI GVA E A P S F P S N T K I I A T W F P D H E R A R A T A I
	4401	CGCATTAATT GGTGTTGCCG AAGCGCCATC CTTCCCGTCT AACACAAAAA TAATCGCCAC CTGGTTCCCG GACCATGAAC GTGCCCGTGC GACCGCGATT
		GCGTAATTAA CCACAACGGC TTCGCGGTAG GAAGGGCAGA TTGTGTTTTT ATTAGCGGTG GACCAAGGGC CTGGTACTTG CACGGGCACG CTGGCGCTAA
	-2	
-	+1	YSS A QYIGLA LLT PALAFIVANYG WEMSFYLSGG
	4501	TATTCCARTE CACAATATAT CONTENESS CTECTERCE CTECCTERC CTTATTERE CONACTACE STIGGAAAT STORTTTAT CTETCEGERE
		ATAAGGTCAC GTGTTATATA GCCAGAGCGC GACGACTGCG GACGGGAGCG GAAATAACAC CGATTGATGC CAACCCTTTA CAGCAAAATA GACAGGCCAC
	-2	
SEN1434	1.1	GÀGILFG IYWL MYYRDP QHSTÀVN QÀE LDYI KÀG
469 ap 1410 b	4601	COCCURATE TETETITICS ANTITATION TRADUCTOR TRADUCTOR CACCALAGA STOCTETAN COCCURATE TANANGAS
403 aa-1410 0	. 1001	COCCOLUME ACAMPTOC ALCINCIC INTERIAL COCCUMPT COCCUMA CANCER CANCERANT COMMITTE ACTIVITY
4015 - 5424	2	COCOCCATA AGREARACCO TROPTORCO ARTACATARI GOCCTRODI GICUTOTO DECORCALITI GOTCOGOCTI DECCARATAT ROTTOGICO
Fr; 1,522,077		
To; 1,523,486	+1	
B5R539	4701	CGGCGGCTAT GGCTCGGAGA ACCAATCCTC CGTGAGTGCA AAAATCAGCT GGCAAAATAT TAAATTCTTC CTCAGCAAAA AAACGATTTG GGGTTTGTTC
Putative		GCCGCCGATA CCGAGCCTCT TGGTTAGGAG GCACTCACGT TTTTAGTCGA CCGTTTTATA ATTTAAGAAG GAGTCGTTTT TTTGCTAAAC CCCAAACAAG
L attained	-2	
nexonate	+1	I T Q F A C S S T L Y F F L T W F I V Y L E K G L H L S I S K A G I
sugar transpo	4801	ATCACCCAGT TTGCCTGCTC GTCTACGCTC TATTTTTTCC TGACCTGGTT TATTGTTTAC CTGGAAAAAG GACTGCATCT CTCTATTTCC AAAGCCGGGA
		TAGTGGGTCA AACGGACGAG CAGATGCGAG ATAAAAAAGG ACTGGACCAA ATAACAAATG GACCTTTTTC CTGACGTAGA GAGATAAAGG TTTCGGCCCT
	-2	
-	1.1	IGÀMLPY IMÀMLGVLCG GTLSDMLLKK GKSRTLÀ
	4001	TRECCEPT COTECTIVE ATTACCOL TOTTOCCCE COTACCETA COLOTATIC COTALANA COLUMNES CARACTER
	4501	INCONCERT CALCAGENT ATTAINCEST ICENTICUES CELEBROOCE CONTROLLAR CONTRELS CONCENTER CONTENTE CONTENTES CONCENTER
-		TREESE CONSISTING TRADECISE CONVERCES CONVERCES CONVERCES CONTINUES CONTINUES IN TRADECISE CONVERCES
	+1	
	5001	ACGAAAATTA CCCGTIAIGG CTGGCCTGTG CGTCACCATG ATTATTGGCC TGGTCAATTT CTTTGAAAAC CAGCCAGTTA TTGCGATTGT CATTCTGTCT
-		IGCITITIAAT GGGCAATACC GACCGGACAC GCAGTGGTAC TAATAACCGG ACCAGTTAAA GAAACTTTTG GTCGGTCAAT AACGCTAACA GTAAGACAGA
	+1	V A F F A N A F S N L G W V V W S D V I P R N F L G T M G G F L N I
	5101	бтібсейтест тібселлебе сіттітеллле стобостобо тобтетовая сблібтлатт сессотлатт тестебобле татоботобе ттітталата
		CARCGCARGA ARCGGTTGCG GARARGTTTG GACCCGRCCC ACCAGRCCTC GCTRCATTAR GGGGGCATTAR ARGAGCCCTG ATACCCRCCG ARARATTTAT
-	1.1	ICGNLSGIVSPIVI GVI LQRTQNFQYA MWYI AGV
	E201	TTEGEGEN CETTEGEGE MEGTINGE CNITETINT TEGEGEGEN CENNETT CONTRICE ATTEGES
	5201	Information and another control control of the second and the second contract to the second to the s
-		AAROOLOTT USAAADUUUU TAGUAATUGO GITAACAATA AUUUUAATAA GAUGIUGUGI GUGITITUGAA GETUATAUGO TAUKUUATAT AUUGIUUGUA
	+Z	
	+1	1 Å N A U Y I U Å N A A U 1 L Y 4 I A U Y U 1 I A U U 2 U Q Y Y I I A U U 1 I A U U V U V V V V V V V V V V V V V V V
	5301	CGCCGGACTG GGCTTATTAG CCTACATATT CCTGGTGGGC AAAATTGAAG TGATCCTGCC TGGAAAAAAG AATGCCGACA CTGTGGATAA GAATGCCATT
_		GCGGCCTGAC CCGAATAATC GGATGTATAA GGACCACCCG TTTTAACTTC ACTAGGACGG ACCTTTTTTC TTACGGCTGT GACACCTATT CTTACGGTAA
	+2	
	+2	NPATANK*
	+2 +1 5401	N P A T A N K *

PhD Thesis

Appendix

SIN 133 141 - 644 S		5501	TCACGTGAAA TCGAGCACAG TGAACATGAC GTCGTCGTGA AGGTCGCCTG CGGCGGCATC TGTGGCTCTG ATATTCACTA CTATCAGCAT GGCCGCGCGG
Standard		-1	AFIGCACITI AGETCOTOTIC ACTIVITACITO CAGCAGCACI TECAGOGGAC GEOCOCIGIAG ACACEGAGAC TATAAGIGAT GATAGITGAT CEOGOCOCOC
SIN143 317 BL01410 517 BL01410 517 BL01410 517 BL01410 517 BL014100 517 B		+2	GMSVLKH PMVIGHEFVGVISKVPAGSDLKVGQTV
Strukturg Contractive deservative deservative curvatures Curvatures <thcurvatures< th=""> <thcurvatures< th=""> Cu</thcurvatures<></thcurvatures<>		5601	ббатотстот сстелляелт себлюботоја теобесатол оттествос бтојателоса лабтоселос лобеслос стојаллотов беселолебот 📃
347 a. 1041 hp 547 a. 104 hp 548 a. 104 hp 547 a. 104 hp 547 a. 104 hp 548 a	SEN1433		CCTACAGACA GGAGTTCGTA GGCTACCACT AACCGGTACT CAAACAACCG CACTAGTCGT TTCACGGTCG TCGGTCGCTG GACTTTCACC CGGTCTGCCA
5411-644 (1-2007) 551 Construction Construction <th>347 aa-1044 bp.</th> <th>-1</th> <th>VAVN PSSPCNOCEMCLSGHONLCGSMRFMGSAOF</th>	347 aa-1044 bp.	-1	VAVN PSSPCNOCEMCLSGHONLCGSMRFMGSAOF
Pf: 1-23-001 Concentration 1: 1-22-001	5441-6484	+2	TREGETERIC CONTOLING CONTOLING TREGERENT TOCOTTOCE CONTOLINE TOTOTGOGOC TOCATOCOT TINTGGOLG COCOLUTIO
BYRS38	Fr; 1,521,017	0.01	ACGCCACTTG GGCAGGTCGG GTACGTTAGT CACGCTCTAC ACGGAAAGGC CGGTAGTTTT AGACACGCCG AGGTACGCGA AATACCCCGTC GCGCGTCAAG
12.1.2.00.016 13.1.2.00 13.1.2.00 14.1.2.00	D5D528	-1	
Sett Transcort Dist Transcort Dist Dis Dis Dist	L-idonate 5-	+2	N P H V N G G F S E Y V V V K P E Q C I P Y D R R V P A N V M A FS
1 1	dehvdrogenase	5801	ARTCOGCARE ICAARGOCGE TITTETGAA TACONGRIGG TAAAACCOGA GCAARGOTT CETTACGACA GGGGCONGCE CCCAAACGTG ANGGCETTT TIAGGGGTA CAATAGCGE AAAAAGACT ATGCACCACE ATTTTGGGCT CGTAGGTAA GGAATGGET CCGGGACGG GGGTTGGAC TACCGGAAAA
4:3 1		-1	
 Sent consector decontor function tubulation tubulation decontrol for the set of the se		+2	SEPLAVAIHAVKKAGQLTGKRVLVIGAGPIGCLI
 1. SURVEY CONTRACT CONTRAC		5901	CG6AGCCGCT GGCGGTCGCC ATTCATGCAG TAAAAAAGC AGGCCAGTTG ACCGGCAAAC GGCTACTGGT GATTGGCGCA GGCCCCATTG GTTGTTGAT
		-1	GCCTCGGCGA CCGCCAGCGG TAAGTACGTC ATTITITICCG TCCGGTCAAC TGGCCGTTTG CGCATGACCA CTAACCGCGT CCGGGGTAAC CAACAGACTA
6010 COTOCOGO CONCLUTE CONSTRUCT ALTONOMY COLLAGORIA CONSTRUCT COLLAGORI		+2	ILÀÀ À RS À G À S E L V À S D L S P R C L E L À R Q M G À T À V
• -1 • -1 • -1 • -1 • -1 • -1 • -1 • +2 • -1 •		6001	сстебесбесб бебесбелеть себебленте толлетобать белтесблате телебеское стотетоболь стобослебсе ллатбоббос следосботб
-1 M D P R D E E O V A P V O O R K O V P D V T E A S O A P I A V A P 42 M D P R D E E O V A P V O O R K O V P D V T E A S O A P I A V A P 43 ATÓSITCOCO SCÓLTOLAS ALLEGATICO CATINGO TACTALOS CACOLANCE ALCONOLOGIA CONCOLLA CONCOLLO CONCOLLA CONCOLLO CONCOLLA CONCO			GEAGCGCCGC CGCGCGTCLC GGCCTCGTAG ACTTGACCAT CGTAGGCTAG AGTCGGGTGC GACAGACCTT GACCGTGCGG TTTACCCCCG GTGTCGCCAC
101 ATGRATCOC CALIFICATION ALAGOTICOC CALIFICATION CONCERNING ACCOUNTING 1 1 1 1 1 1 1 0		-1	MDPRDEEQVA HYQQHKGYFD VVFEASGAPI AVAS
SEN143 239 BA-720 brief COCCUMPTOR CONTACTOR CONTACTOR CANCELATE CANCELATE CANCELATE CANCELATE CONCENCE CONCENCE <t< th=""><th></th><th>+2</th><th>ATOGATOCOC GEGATOANGA ACAGOTTOCC CATTATCAGE AGEATAAAGG CTACTITEAT GTOGTOTTTE AGGCCTCCGG COCCCCATT GCCGTOGCAT</th></t<>		+2	ATOGATOCOC GEGATOANGA ACAGOTTOCC CATTATCAGE AGEATAAAGG CTACTITEAT GTOGTOTTTE AGGCCTCCGG COCCCCATT GCCGTOGCAT
-1-1			TACCTAGGCG CGCTACTTCT TGTCCAACGG GTAATAGTCG TCGTATTTCC GATGAAACTA CAGCACAAAC TCCGGAGGCC GCGCGGCTAA CGGCACCGTA
422 2 2 2 2 2 1 2 4 2 5 1		-1	
		+2	
+2 L M W V S F F T A V N N N N N N P D R P L I S A R 6301 CAACTOOR OCCUTTAGE CONTRACE ACTUANCIA TO COOCCUTTAL ALL CONCOUNT ACTUATED CONTRACE ACTUANCIA TO COOCCUT ALL CONCOUNT ACTUATED CONCECT ALL CONCERNES ACTUANCIA TO COOCCUT ALL CONCECT ALL CONCERNES ACTUANCIA TO CONCECT ALL CONCECT ALL CONCERNES ACTUANCIA TO CONCECT ALL CONCERNES ALL CONCECT ALL CONCERNES ALL CONCERN		0201	GETGGCAGET GAAGTGGGGG GGACGGCCGET GGTAGCAAGT GCACCCCTAC CCCCCGGTCG GCCATCAC GGCACCAC TGCTACCAC ACATTGCTACA
6501 CLACTOGOT South Tick T N T N		+2	LNWV GSFRFIGEFI TAVRWLEDGR VDPRPLISAE
4 7 P P Q 1 K K T T T F N Y S R Y L T T T T N N Y S R Y L R F D Y S R Y L R F D Y S R Y L R F D Y S R Y L R R D Y T T D R N Y S R Y L R R D Y L R N Y S R Y L D S R N C R Y N D R R D P S R R D R R L D D R R D P R R R R R R R R R R R R R R		6301	caactegere geoteatree grittiatege teagtteate acegedate getegetea agategede geotegete geotegete
430 440 440 440 440 440 440 440			GTTGACCCAG CCCAGTAAGG CAARATAGCC ACTCAAGTAG TGGGGCCATG CGACCGACCT TCTACCCGCG CAGCTAGGAG CGGGCGAATA GTCGCGGGCC F P P O O T F D B L T T B T D K N Y S B K Y L T B F D *
ALGOGOGOGO TOSTITALCT TOTOGOGOLO TALTOGOGOSO OTCTOTITIT ACAGAGAC TOTOCATGAST ALGOALAGET ALTOGOLOT TITOGOGOGO 42 6501 GCCCGGCGGA LACAGARTT TOTOTITAL ALAGAGATTO TTATATATA ALAGCACTTCA GCGACATCT ALGOGATACC CATOTIGAGA ALTOGOLOT TITOGOGOGO 42 0 M V W M E I Y D U I S K L L 0 G S W A P G S W A P G S W A P G S E V E L T A 6601 GAATETTEGA TATATATATA TATATATAT TITOGTGAAGT COCGACCCE GEGGCGATE GAATATCE TATATAGATCA ALTOTAGAA ALGOCTTA ALGOGATACC CATOTIGGG ALGOCGE ALGOLACCE GEGGCGATE ALGOLACCE GEGGGGA GEGGGGGA GEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE TAGTAGACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG GICCACCE CEGGGGGGG CACCEGGGGGG CACCEGGGGGG CACCEGGGGGG CACCEGGGGGG CACCEGGGGGG CACCEGGGGGG CACCEGGGGGG CACCEGGGGGGGGGG		+2 6401	TTCCCCCCCC ACCANTTON ACACCCCCCG ATTACCCCCA CAGACAAAA TGTCTCTCCC AAGGACACA TTCCTTTCCA TTACCCCCCCG
+2 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC			AAGGGCGGGG TCGTTTAACT TCTGCGCGAC TAATGGCGGT GTCTGTTTTT ACAGAGACGA TTCCATGAGT AAGCAAAGCT AATTGCCACT TTTCGCGGGC
Sell GCCGGCCCCT MACAGAGATT THATATAT AGCATTCT ALGCATTCT		+2	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		6501	GCCGGCGCCT TIGTICITAA AAGAGAATIG TIATATAATA AAGGACTICA GCGACACTIT AACGGATACC CATCHTGAGC ATAAAATCCA TICAAAAACA CGGCCCGCG AACAAAGAATIT TICICITAAC AATATATATI TICGIGAAGT CGCTGTAGAA ITGCCTAAGTG GATATAAGTG TAATTATAGT AAGTTITTET
6601 GARTGTOTAL ATGALACTT ATGATGATTT ATGATGATTT ATGATGATTT ATGATGATTT ATGATGATCA CTGCTGGACG GGCGTCGC CGCCGGGTAGC CTTTCACCT GGAGTGGGC GGCGCTCGC CGGAGTGGGC GGCGGTCGC CGGAGTGGGC GGCGGTCGC GGCGGTCGGC GGCGGTCGC GGCGGTCGC GGCGGTCGGC GGCGGTCGG GGCGGGGGGGGG GGCGGGGGGGGGGGGGGGGGGGGGGGGGGG		+2	QNVVNEIYDQISSKLLDGSWAPGS RLPSEVELTA
$\frac{\operatorname{CTRACACAC}}{\operatorname{CTRACACAC}} \operatorname{TRACTTRAR} \operatorname{TRACEGTA} \operatorname{TRACTCGTT} \operatorname{TRACCGTEC} \operatorname{CFCRACCC} \operatorname{CGCCCCCEC} \operatorname{CGCCCCCEC} \operatorname{CGCCCACCE} \operatorname{CGCCCACCE} \operatorname{CGCCCACCE} \operatorname{CGCCCACCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCCE} \operatorname{CGCCCE} \operatorname{CGCCE}$		6601	GAATGTTGTA AATGAAATTT ATGATCAGAT AAGTAGCAAA CTGCTGGACG GCAGTTGGGC GCCGGGTAGC CETTECCCT CAGAAGTGGA ACTGACCGCC
$\frac{12}{575} + \frac{1}{8} + \frac$			CTTACAACAT ITACITIAAA TACTAGTCTA ITCATCGTITI GACGACCTGC CGTCAACCCG CGGCCCAT G GCAAACGGGA GTCTTCACCT TGG TGGGGGG
5/11 AGLABATIGE AGECUTICE AGEGUTICE AGECUTICE AGECUTICE		+2	
<pre>-1 +2 -2</pre>		6701	AGTAAATTGC ICAGCGGCCCA GTCGCCAAGGC TCGCGTCATG TCGCAAAAGC ACTGGACCCC TAACACCACT GCGCAGTCCC GTCGCCGTCG ATGCACTGCC
+2 E N F T P Q L S N D P R P I M H L S R E E F H D M M I F R Q T V E AGACCTCAC ATTAGAAGTE ATTAGAAGTE AGACCTCAC AATCCATEGE AATCCATEGE CATAGTEGEGE AATCATEGEGE AATCATEGEGE AATCATEGEGE CATAGEGEGE TTTCCAAAGE GTTAGAAGCA TCTGGGCACT TAGGGGGCGC TTGGGGCGC TTGGGGCGCC TTGGGGCGC TTGGGGCGCC TGGGGCGGC TGGGCGGCC TGGGCGGCC TGGGCGGGC TGGGCGGGC GCGGGGGCC CCTCGGAAG GGCGGGGCC CCTCGGAAG AGGGTGGGCC CCTCGGAAG AGGGTGGGCC CCTCGGAAGA TGGGCGGCC TGGCGGGCC CCTCGGAAGAAG TGGCGGGCC CCTCGGAGGCGC TGGCGGGGCC CCTCGGAAGAAG TGGCGGGCC CCTCGGAAGAAGAAGAAGAC CCTCGGAGGCC TGGCGGGCC TGGCGGGCC TTGTGGGGCC TGGCGGGCC TGGCGGGCC TGGCGGGCC TGGCGGGCC TGGCGGGCC TGGCGGGCC CCTCGGAGGCC TGGCGGGGCC TGGCGGGGCC TGGCGGGGCC TGGCGGGGCC		-1	+
SEN1432 AAAACTTACC CCCCGAGATE THAGTAAGE ATCCCCCCC AAAACTTACC CTTAGTAGE ATCCCCCCC AAAAATTTAC ATCCCCCCC AAAAATTTACC CTTAGGGGCGC CTTAGTAGAC TTGAGTAGE TTGAGGGCCG TTGAGTAGE TTGAGGCCG TTGAGTAGE TTGAGGGCCG TTGAGTAGE TTGAGGCCGC TTGAGTAGE TTGAGGCCGC TTGAGGCCGC TTGAGTAGE TTGAGGCCGC CTTAGTACAC TTGAGTAGE TTGAGGCCGC TTGAGTAGE TTGAGTAGE TTGAGTAGE TTGAGTAGE TTGAGGCCGC TTGAGTAGE TTGAGGCCGC TTGAGGCCGCG TTGAGTAGCAC CGGGTGAGTG TTGAGTAGCAC CGGGTGAGTG TTGAGGCCGGG TTGAGTAGA GCGCACCGCC TTGAGTAGE TGAGTAGCAC GGCGTGCGCG TGAGTAGCAC GGCGTGAGTGC GGCGTGAGTAG GGCGTGGCGC GGCGTGGCGGC GGCGTGGCGGC GGCGTGGGCGGC GGCGTGGCGGC GGCGTGGGCGC		+2	ENFTPQMLLSNDPRPIMHLSREEFHDMMIFRQTVE
239 aa-720 bp. -1 239 aa-720 bp. -1 6575 - 7291 FR: 1,520,207 6901 GTCRAATGC GGCGGCG CG CCGTCGACCA CGCCACCA CGCCACCA CGCCACCA CGCCACCA CGCCACCA CGCCCACCA CGCCCACCA CGCCCACCA CGCCCACCA CGCCCACCA CGCGCCGGGC CGTCGTGGCGGGGGGGGGG	SEN1432	6001	AMARCITAL CLECKBARD I DARMARCE ALCCENCE ANTARCALCE CHARCENCE AREADILLA CHARLEND ALLINE ANTARCE ARECEDEN
6575 - 7291 +2 E F K C V E L A V T H A T D D D D E E A L I H K G D Y H A T D <t< th=""><th>239 aa-720 bp.</th><th>-1</th><th></th></t<>	239 aa-720 bp.	-1	
FR: 1,520,307 6301 GTTCAAATGC GTGGAAGTCG CCGCCCCCCACA CGCCACCCGAT GGCCACCCGAT GGCCACCCGAT GGCCACCCGAT GGCCACCCGAT GGCCAGCCCCACA AGGTCATATAT To: 1,520,306 BSR 537 -1 </th <th>6575 - 7291</th> <th>+2</th> <th>EFKC VELAVTHATD DDIRQLEEAL NNM LIHKGDY</th>	6575 - 7291	+2	EFKC VELAVTHATD DDIRQLEEAL NNM LIHKGDY
BSR537 Putative GntR-family regulatory protein -1	FR; 1,520,207 To: 1.520.926	6901	GTICARATEC ETGERECICE CCETCACACA CECCACCERT EATERCATIC ECCAEGCICER GERRECATE AACAACATEC TEATCACAA AGETERITAT Cametriae caucificate eccaege conteners eccaeges contenense contenense internation contentates activity of the
Putative GntR-family regulatory protein +2 +2 + + + + + + + + + + + + + +	B5R537	-1	
GntR-family regulatory protein 7001 AAAAAATACT TITTTTAFGA CGGAAGGGGA GCCTCGCCT CATCGGATT CATCGGGCGA CATCGGGGGA GTGGTCGACG TTTTGCAGGGC AACGTCGAT AACGCTGAT ACACGTGAT GAGCTCGAT AAAGACATCT TTTTGTAGAGA -1	Putative	+2	KKYSEADYEF HLAIVRASHN SVFYNVM SSI KDIY
THIS IN THE ALL OF THE	GntR-family	7001	AAAAAATACT CEGRARGEGGA CTACEAGTTC CATCTEECCA TFETCAGEGE AFCECACEAC ACCESTETTCT ACAACETEAT GAGCTCCATT AAAGACATCT
+2 Y Y Y Y Y L E E L N R À L G I T L E S V E À H I K V Y M S I K N R D À 7101 ATTACTACTA TCTTGAAGAG CTAACCGTG GCGACCGATA -1 -1 +2 A S T À V E V L N E À M S G N I I À I E K I K S T E T S G T K * 7201 CAGCACGGCC GTCGAGGGC CAGGAGTAGCG GAGTGCAGGC GATGTCAGGC GAGTGCAGGC GAGTACTG GAGTAGCTG GAGTAGCAG ATATACTG CGATGAGAG ATACCGCTTG GTCGCGGCAGGA ATACCGCTTG GTCGCGGCAGGC CAGGTAGTGCG TTATAATAG 7301 GTTACAAGCT CAAGGTAGTAG AGCAATTTAA CATATCTGAA CATATCTGAA CTATATCAAC CTATATAGGT GAAATAGC CATAGTGCG ATATGTG CATAGTTGGAA TCGGGAATAG CTATATCGCAACGTTG GATAGTAG CACCGCAGGCG TTAAATAG 7301 GTTACAAGCT CAAGTAGTAG AGCAATTTAA CATATCTGAA TCCGAAATTG CTGCGAAATGT GATAGTAG GCCGCAGCGCG TTAAATACGAATGCCC ACTTATAATTAA CATATCTGAA CTATATAGGG ACTACT CATATTTAA CATATCGAA	regulatory	-1	TITITITATEA GUUTTUGUUT GATGUTUAAG GTAGAUUGUT AAUAGTUUUG TAGUGTGTTG TUGUAUAGA TETTGUAUTA UTUGAGUTAA TITUTGTAGA
7101 ATTACTACTA TCTTGAAGAG CTTGAACGAG CGCTGGGTA TACCCTGGA AGFGTGGGAAG CCCATATCAA GGTCTACATG TCGATAAAGA ATGCGGATGC -1 1 1	protein	+2	YYYYLEE LNR ALG I TLE SVE AHIKVYM SIKN RDA
TARTGATGAT AGAACTICTC GAATTGGCAC GCGACCCATA ATGGGAACTI TCACACCTTC GGGTATAGTT CCAGATGTAC AGCTATTTCT TAGCGCTACG -1 +2 A S T A V E V L N E A M S G N I I A I E K I K S T E T S G T K * 7201 CAGCACGCCG GTCGAGGGC TCAGGGGC GAGTGCG GAGTGCCG TTATAATAGC GCTAGGTTT TTAGTTAGA ATCAGATCT ACAGGGACAT CAGGGGACAAA ATAACCGTTG GTCGTGCCGG CAGCTTCACG AGTTATTCG CTACATCGGAA TCGGAATTG TTAGAATAGC GCTAGCTTT TTAGTTAGTG ACATGCTCG AGTCGCTGTG TTATAATT GTATGGCAACG 7301 GTTACAAGCT CAAGTAGTAG AGCAATTTAA CATATCGAA TCCGAAATAG TTGGCACAA CTATTTAGTG ACATAGCCG ACTTTAAAAT CGTGGCGATG CAAGTATCTCGA GTTCATCATC TCGTTAAATT GTATAGAATT ACAGTACTT AAGGTTATT ACGGTAGTT GATAATACCC TGTTTACGGG TGAAATTTTA GCACCGTCAC		7101	АТТАСТАСТА ТСТТБАЛБАБ СТТАЛССБТБ СБСТБББТАТ ТАСССТТБАЛ АБТБТББАЛБ СССАТАТСАЛ ББТСТАСАТБ ТСБАТАЛАБА АТСБСБАТБС
+2 +2 +2 +2 +2 +2 +2 +2 +2 +2		-1	TAATGATGAT AGAACTICIC GAATIGGCAC GCGACCCATA ATGGGAACIT ICACACCIIC GGGTATAGII CCAGAIGTAC AGCIAITICI TAGCGCIACG
7201 CAGCACGGCC GTCGAAGTGC TCAATGAAGC GATGTCAGGC AATATTATTG CGATCGAAAA AATCAAATCT ACAGAGACAT CAGGGACAAA ATAACCGTTG GTCGTGCCGG CAGCTTCACG AGTTACTTCG CTACAGTCCG TTATAATAAC GCTAGCTTIT TTAGTTTAGA TGTCTCTGTA GTCCCTGTT TATTGGCAAC 7301 GTTACAAGCT CAAGTAGTAG AGCAATTTAA CATATCTGAA TCCGAAATAG TTGCCATCAA CTATTTAGTG ACATAGTCCC ACTTTAAAAT CGTGGCAGTG CAATGTTCGA GTTCATCATC TCGTTAAATT GTATAGACTT AGGCTTTATC AACGGTAGTT GATAAATAC TGTATCAGGG TGAAATTTTA GCACCGTCAC		+2	ASTA VEVLNEAMSG NIIAIEKIKS TET SGT K*
GICGIGCCGG CAGCIICACG AGIIACIICG CIACAGICCG IIAIAAIAAC GCIAGCIIII IIAGIIIAGA IGICICIGIA GICCCIGIII IAITGGCAAC 7301 GIIACAAGCI CAAGIAGIAG AGCAAIIIAA CAIAICIGAA TCCGAAAIAG IIGCCAICAA CIAIIIIAGIG ACAIAGICCC ACIIIIAAAI CGIGGCAGIG CAAIGIICGA GIICAICAIC ICGIIAAAII GIAIAGACII AGGCIIIAC AACGGIAGII GAIAAAICAC IIGIACAGGG IGAAAIIIIA GCACCGICAC		7201	CAGCACGGCC GTCGAAGTGC TCAATGAAGC GATGTCAGGC AATATTATTG CGATCGAAAA AATCAAATCT ACAGAGACAT CAGGGACAAA ATAACCGTTG
7301 GTTACAAGCT CAAGTAGTAG AGCAATITAA CATATCTGAA TCCGAAATAG TTGCCATCAA CTATTTAGTG ACATAGTCCC ACTTTAAAAT CGTGGCAGTG CAATGTTCGA GTTCATCATC TCGTTAAATT GTATAGACTT AGGCTTTATC AACGGTAGTT GATAAATCAC TGTATCAGGG TGAAATTTTA GCACCGTCAC			GICGIGCGGG CAGCITCACG AGTIACTICG CTACAGICCG ITATAATAAC GCTAGCITIT ITAGITTAGA IGICICIGIA GICCCIGITI TAITIGCCAAC
		7301	GIRCERGE CREDIER SCHERTER CHERTERA CHERTERA ICCORARIAG TIGCCETCAR CHETTRAGE ACCHEGECC ACTITARART COTGCACH

1- SEN1433 >tr|B5R538|B5R538_SALEP Putative hexonate dehydrogenase OS=Salmonella enteritidis PT4 (strain P125109) GN=SEN1433 PE=4 SV=1

MEKITCNACLAHAEKDVRFESREIEHSEHDVVVKVACGGICGSDIHYYQHGRAGMSVLKH PMVIGHEFVGVISKVPAGSDLKVGQTVAVNPSSPCNQCEMCLSGHQNLCGSMRFMGSAQF NPHVNGGFSEYVVVKPEQCIPYDRRVPANVMAFSEPLAVAIHAVKKAGQLTGKRVLVIGA GPIGCLILAAARSAGASELVASDLSPRCLELARQMGATAVMDPRDEEQVAHYQQHKGYFD VVFEASGAPIAVASTVDFTRPAGTIVQVGMGASPVSWPVSTMLVKELNWVGSFRFIGEFI TAVRWLEDGRVDPRPLISAEFPPQQIEDALITATDKNVSAKVLIRFD

2- SEN1434 >tr|B5R539|B5R539_SALEP Putative hexonate sugar transport protein OS=Salmonella enteritidis PT4 (strain P125109) GN=SEN1434 PE=4 SV=1

MACVYPDACRYNTGIIMKASRQRLFILTLLFIVTAINYMDRANLAVAGSNIQNDFSLTPT QLGLLFSMFTWAYAASQIPVGYVLDRIGSRILYGGAIILWSIFTFMMGFASHHLFATATA SFAMLLACRALIGVAEAPSFPSNTKIIATWFPDHERARATAIYSSAQYIGLALLTPALAF IVANYGWEMSFYLSGGAGILFGIYWLMYYRDPQHSTAVNQAELDYIKAGGGYGSENQSSV SAKISWQNIKFFLSKKTIWGLFITQFACSSTLYFFLTWFIVYLEKGLHLSISKAGIGAML PYIMAMLGVLCGGTLSDMLLKKGKSRTLARKLPVMAGLCVTMIIGLVNFFENQPVIAIVI LSVAFFANAFSNLGWVVWSDVIPRNFLGTMGGFLNICGNLSGIVSPIVIGVILQRTQNFQ YAMWYIAGVAGLGLLAYIFLVGKIEVILPGKKNADTVDKNAINPATANK

3- SEN1435 >tr|B5R540|B5R540_SALEP Putative hexonate dehydrogenase OS=Salmonella enteritidis PT4 (strain P125109) GN=SEN1435 PE=1 SV=1

MTALFDLTGKTALVTGSARGLGFAYAEGLAAAGARVILNDIRATLLAESVDTLTRKGYDA HGVAFDVTDELAIEAAFSKLDAEGIHVDILINNAGIQYRKPMVELELENWQKVIDTNLTS AFLVSRSAAKRMIARNSGGKIINIGSLTSQAARPTVAPYTAAKGGIKMLTCSMAAEWAQF NIQTNAIGPGYILTDMNTALIEDKQFDSWVKSSTPSQRWGRPEELIGTAIFLSSKASDYI NGQIIYVDGGWLAVL

4- SEN1436 >tr|B5R541|B5R541_SALEP Putative dehydratase OS=*Salmonella* enteritidis PT4 (strain P125109) GN=SEN1436 PE=1 SV=1

MKVSNLKITNVKTILTAPGGIDLAVVKIETNEPGLYGLGCATFTQRIFAVKSAIDEYMAP FLVGKDPTRIEDIWQSGVVSGYWRNGPIMNNALSGVDMALWDIKGKLAGMPVYDLLGGKC RDGIPLYCHTDGGDEVEVEDNIRARMEEGYQYVRCQMGMYGGAGTDDLKLIATQLARAKN IQPKRSPRSKTPGIYFDPDAYAKSVPRLFDHLRNKLGFGIEFIHDVHERVTPVTAINLAK TLEQYQLFYLEDPVAPENIDWLKMLRQQSSTPISMGELFVNVNEWKPLIDNRLIDYIRCH VSTIGGITPARKLAVYSELNGVRTAWHGPGDISPVGVCANMHLDLSSPNFGIQEYTPMND ALRDVFPGCPEIDHGYAYLNDKPGLGIDIDEAKAAKYPCEGGIPSWTMARTPDGTASRP

5- SEN1432 >tr|B5R537|B5R537_SALEP Putative GntR-family regulatory protein OS=Salmonella enteritidis PT4 (strain P125109) GN=SEN1432 PE=4 SV=1

MSIKSIQKQNVVNEIYDQISSKLLDGSWAPGSRLPSEVELTASFNVSRVSVRSAVQRFRD LGIVVTRQGSGSYVSENFTPQMLSNDPRPIMHLSREEFHDMMIFRQTVEFKCVELAVTHA TDDDIRQLEEALNNMLIHKGDYKKYSEADYEFHLAIVRASHNSVFYNVMSSIKDIYYYYL EELNRALGITLESVEAHIKVYMSIKNRDASTAVEVLNEAMSGNIIAIEKIKSTETSGTK

```
6- SEN1437 >tr|B5R542|B5R542_SALEP Aminoglycoside N(6')-acetyltransferase type 1 OS=Salmonella enteritidis PT4 (strain P125109) GN=SEN1437 PE=3 SV=1
```

MDIRQMNKTHLEHWRGLRKQLWPGHPDDAHLADGEEILQADHLASFVAMADGVAIGFADA SIRHDYVNGCDSSPVAFLEGIFVLPSFRQRGVAKQLIAAVQRWGTNKGCREMASDTTPEN TISQKVHLALGFEETERVIFYRKRC

Appendix 4: Translated sequence of induced genes (using Vector NTI program)

	1	GCGCAACCCA	ATTGCCTAAA	TCCACTAGCC	GTTCATGTCC	CTGGTGGCTG	GCCATATTCG	CCAGCGCTTT	ACGCAACGCA	TCCGGGGGCGC	CTGCCGCGCC
	101	GGGACGCCCT	TAACGGATTT GCATTGCGTT	AGGTGATCGG TTACCCCCTC	CAAGTACAGG GTCGCAGGCA	GACCACCGAC AATCCCATTA	CGGTATAAGC ACGCGATCTT	GGTCGCGAAA TTGCTGCCAG	TGCGTTGCGT	AGGCCCCGCG	GACGGCGCGG GCTGCGCGTG
		CCCTGCGGGA	CGTAACGCAA	AATGGGGGAG	CAGCGTCCGT	TTAGGGTAAT	TGCGCTAGAA	AACGACGGTC	AAAAGCCCCC	TTTTCCAGAG	CGACGCGCAC
	201	ACGGTCTGAA TGCCAGACTT	ACAGACGCCG TGTCTGCGGC	CGCATCCGGC GCGTAGGCCG	GCTTCTATAC CGAAGATATG	TGTCATCGCG ACAGTAGCGC	CCCCTGCCAG GGGGACGGTC	AGCGCCGGAG TCGCGGCCTC	AAGCCGGATA TTCGGCCTAT	CCATTGCGTC GGTAACGCAG	ATGAGATTTG TACTCTAAAC
	-1		1000000000		1000000000					*	S I Q
ybhC	301	AGGAGACATA	TGGGCGACTA	ATAAAGGATT TATTTCCTAA	TGCGGGGGCCG	AGTATGTGTT	GGAGATGGCC	GACAAGTCGT	AGGGTGTGGT	GTTTTAGGCG	GGCCATCAGC
	J _1	GRY	V R Q Y	LPN	R G P	E Y V V	EVP	QEA	D W V V	FDA	RYDA
	401	GCCCTGAGCT	GCCCATGCGT	CGCCTGTCTT GCGGACAGAA	CCCAGCGCAC GGGTCGCGTG	GAGCGGCATG	GCGCGTAACG	CCCGCCCATG GGGCGGGTAC	CCTCTTCCGA GGAGAAGGCT	CGTCAGACCA GCAGTCTGGT	AACTGTACGC TTGACATGCG
	-1	ARL C	GHT	A Q R	G L A R	λλΗ	RTV	G A W A	EES	TLG	FQVC
	501	AGGCCATATT	CATCGCCAGA	TGCAAACTGC	AAAACGGGCT	GGTGCCGGGA	TTGAAATCGC	TGGCGACGGC	GACAGGCACC	TGATAGCGGC	GCAGCAGTTC
	-1	C A M N	M A L	H L S C	F P S	T G P	N F D S	ACCGCIGCCG	V P V	O Y R R	
	601	CACCGGCGGG	CGCTGCGTCT	CGCGCAGAAA	ATAAAACGCG	CCAGGCAACA	ACACGCCGAC	AGTACCGCCG	TCACGCATCG	CCGCGACGCC	CGCTTCATCA
		GTGGCCGCCC	GCGACGCAGA	GCGCGTCTTT	TATTTTGCGC	GGTCCGTTGT	TGTGCGGCTG	TCATGGCGGC	AGTGCGTAGC	GGCGCTGCGG	GCGAAGTAGT
	701	AGATATTCGA	TATGATCCGC	CGATAAACCC	TGATAGCGAC	G P L L TCACCAGCTG	CGCGCCGCCC	AACAGCGAAA	GCTGCTCAAC	A V G ATGGCCTTTA	ACGGGAATAC
		TCTATAAGCT	ATACTAGGCG	GCTATTTGGG	ACTATCGCTG	AGTGGTCGAC	GCGCGGCGGG	TTGTCGCTTT	CGACGAGTTG	TACCGGAAAT	TGCCCTTATG
	-1	LYEI	H D A	S L G	Q Y R S	CICITTING	A G G	L L S L	GTCTACCCCA	H G K	V P I G
	001	GATTGCGGAA	GCGGCAGACG	TTATGCGCGA	GTGAGACGCG	GTGTAATTTC	GGCTGCGAGA	GCGTTTTCTC	CAGATGGCGT	AGTTTATTGG	GAAAAACGGT
	-1	GLAK	A T Q	LVRE	SQA	VNF	3 V S E	CFL	D V A	DFLG	KQW
	901	GAGCTGCGGA CTCGACGCCT	ATCATCGTCT TAGTAGCAGA	CGCAGACCAG GCGTCTGGTC	AGTGATGTAG TCACTACATC	GGCAGGCCCA	CGTCACGATA GCAGTGCTAT	CTCCGCTGGC GAGGCGACCG	GTAGCATGAG CATCGTACTC	CGGCCAATAG GCCGGTTATC	CGTGGGGGCTA GCACCCCGAT
	-1	LQP	I M T E	C V L	TIY	G D P D	DRY	EAP	T A H A	à L L	TPSI
	1001	ATGTCGATAG	CGTTTTCGGC	GGCAAGTTTT	GCAGCAACGC	GCAACAGCTT	TTCTTCTGTC	GCCAGCTCCA	GGCCATAGCC	GGATTTAATC	TCCAGCAGCG
	-1	I D I A	N E A	A L K	A A V R	LLK	E E T	A L E L	G Y G	S K I	E L L T
	1101	TAACGCCTTC	GCTGGCAAGG	CGCATCATGC	GTTCGCGCGC	CAGCAGGTAG	AGCGTCTCCT	CCGCACAGGC	GCGGGTTGCT	GATACCGTCG	CGTTAATGCC
	_1	ATTGCGGAAG	CGACCGTTCC	GCGTAGTACG	CAAGCGCGCG	GTCGTCCATC	TCGCAGAGGA	GGCGTGTCCG	CGCCCAACGA	CTATGGCAGC	GCAATTACGG
	1201	GCCGCCCTGA	GCGCTAATAT	GCTGGTATGA	CGCGCCGTTA	AGCCGCTGCT	CCCACTCTGC	GGCGCGGTTA	CCGGCAAACA	CCAGATGCGT	GTGGCAATCG
		CGGCGGGACT	CGCGATTATA	CGACCATACT	GCGCGGCAAT	TCGGCGACGA	GGGTGAGACG	CCGCGCCAAT	GGCCGTTTGT	GGTCTACGCA	CACCGTTAGC
	1301	ATAAGTCCCG	A S I H GGGTTACCAG	TCGTCCCTGC	A G N ATATCATGGA	L R Q E	ACTGACAGGA	A R N AGCTGCGTCT	G A F V CTGGCACGAT	ATCGCAAATA	TGCCCTTCGC
		TATTCAGGGC	CCCAATGGTC	AGCAGGGACG	TATAGTACCT	ATAACAGGGG	TGACTGTCCT	TCGACGCAGA	GACCGTGCTA	TAGCGTTTAT	ACGGGAAGCG
	-1	ILG E	COCCTOCTTA	R G Q	M D H I	N D G	S V P	L Q T E	PVI	CIGICIGTIT	H G E R
	1401	CATGTTAGTC	GCGGACCAAT	AGGTGGTCGG	GCATGCCCCG	AACGGCGACG	CCCAGGTAAC	AGCGGTCGGA	CTACAAAGCG	GTCTGTCATA	GCGGGCCATT
	-1	RVIL	A Q N	D V L G	YPA	QRQ	P D M T	à L R	INR	W V T D	GPL
	1501	AAGTTGCCGC TTCAACGGCG	ATTCCTGTCG TAAGGACAGC	CTCTCTTGCC GAGAGAACGG	TGTCATGAGT ACAGTACTCA	TGTATAGACA ACATATCTGT	TTTATTTTCT AAATAAAAGA	TTCTGCTCCG AAGACGAGGC	GATTGTCAAC CTAACAGTTG	TCAAAGCGCG AGTTTCGCGC	AAAGTTGTTG TTTCAACAAC
	-1	LQR	MGTA	RKG	TML	Q I S M					
	-1 1601	CTTAATTGTG GAATTAACAC	M G T A ATAAAACTAT TATTTTGATA	R K G CTGATGCTAC GACTACGATG	T M L AGGTGTTTCC TCCACAAAGG	Q I S M GGCCTGAAAA CCGGACTTTT	GGAACTTTTT CCTTGAAAAA	ACCTTTTCGC TGGAAAAGCG	CTTCCCGTTT GAAGGGCAAA	CGTTCAACTT GCAAGTTGAA	АСТАТААААА ТСАТАТТТТТ
	-1 1601 +1	L Q R CTTAATTGTG GAATTAACAC	M G T A ATAAAACTAT TATTTTGATA	R K G CTGATGCTAC GACTACGATG	T M L AGGTGTTTCC TCCACAAAGG	Q I S M GGCCTGAAAA CCGGACTTTT	GGAACTTTTT CCTTGAAAAA V N T L	ACCTTTTCGC TGGAAAAGCG S V S	CTTCCCGTTT GAAGGGCAAA R L A	CGTTCAACTT GCAAGTTGAA L A L A	AGTATAAAAA TCATATTTTT F G V
	-1 1601 +1 1701	L Q R CTTAATTGTG GAATTAACAC AGCAGGCTTC	M G T A ATAAAACTAT TATTTTGATA AATGGATGTC	R K G CTGATGCTAC GACTACGATG	T M L AGGTGTTTCC TCCACAAAGG	GGCCTGAAAA CCGGACTTTT GAGCAACCTG	GGAACTTTTT CCTTGAAAAA V N T L TGAATACATT	ACCTTTTCGC TGGAAAAGCG S V S ATCGGTTTCC	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCGC GCAGCCGCC	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC	AGTATAAAAA TCATATTTTT F G V TTTTGCGGTG
	-1 1601 +1 1701 +1	L Q R CTTAATTGTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG	M G T A ATAAAACTAT TATTTTGATA AATGGATGTC TTACCTACAG A C S S	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P	T M L AGGTGTTTCC TCCACAAAGG TTCAAGCCCG AAGTTCGGGC	GGCCTGAAAA CCGGACTTTT GAGCAACCTG CTCGTTGGAC P S D Q	GGAACTTTTT CCTTGAAAAAA V N T L TGAATACATT ACTTATGTAA T A P	ACCTTTTCGC TGGAAAAGCG S V S ATCGGTTTCC TAGCCAAAGG G T A	CTTCCCGTTT GAAGGGCAAA A CGTCTGGCGC GCAGACCGCG S _ R _ P _ I	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC ACCGTGACCG L S A	AGTATAAAAA TCATATTTTT F G V TTTTGGCGTG AAAACCGCAC N E A K
	-1 1601 +1 1701 +1 1801	L Q R CTTAATTGTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACGCTGAGCG	M G T A ATAAAACTAT TATTTTGATA AATGGATGTC TTACCTACAG A C S S CCTGTAGCTC	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACGCCACCC	T M L AGGTGTTTCC TCCACAAAGG TTCAAGCCCG AAGTTCGGGC D Q I GATCAGATCC	GGCCTGAAAA CCGGACTTTT GAGCAACCTG CTCGTTGGAC P S D Q CTTCCGATCA	GGAACTTTTT CCTTGAAAAA V N T L TGAATACATT ACTTATGTAA T A P AACCGCGCCCT	ACCTTTTCGC TGGAAAAGCG S V S ATCGGTTTCC TAGCCAAAGG G T A GCCACCGCCT	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCGC GCAGACCGCG S R P I CGCGCCCAAT	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA	AGTATÀÀAAA TCATATITT F G V TTTTGGCGTG AAAACCGCAC N E A K AATGAAGCGA
	-1 1601 +1 1701 +1 1801 +1	L Q R CTTAATTGTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACGCTGAGCG TGCGACTCGC K N F V	M G T A ATAAAACTAT TATTTTGATA AATGGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG A A R	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACGCCACCC ATGCGCFGG Y F A	T T M L AGGTGTTTCC TCCACAAAGG TTCCAAGACCCG AAGTTCGGGC D Q I GATCAGATCC CTAGTCTAGG	Q I S M GGCCTGAAAA CCGGACTTTT GAGCAACCTG CTCGTTGGAC P S D Q CTTCCGATCA GAAGGCTACT N T A	GGAACTTTTT CCTTGAAAAAA V N T L TGAATACATT ACTTATGTAA AACCGCGCCT T GGCGGGGA F W S	ACCTTTTCGC TGGAAAAGCG S V S ATCGGTTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGA P S P I	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCGC GCAGACCGCC S R P I CGCGCCCAAT GCGCGGGTTA T L P	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGT A Q P	AGTATAAAAA TCATATTTT F G V TTTTGGGGGG AAAACCGGAC N E A K AATGAAGCGA TTACTTCGCT D F V V
	-1 1601 +1 1701 +1 1801 +1 1901	L Q R CTTAATTGTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACGCTGAGCG TGCGACTCGC K N F V AAAACTTCGT	M G T A ATAAAACTAT TATTTTGATA AATGGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG GGACATCGAG TGCGGCGCCC	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACGCCACCC ATGCGCTGGG Y F A TATTTTGCCT	T M L AGGTGTTTCC TCCACAAAGG TTCCACAAAGG TTCCACACAAAGG AAGTTCGGGC D Q I GATCAGATCC CTAGTCTAGG CTAGTCTAGG CCTCGACGCC	Q I S M GGCCTGAAAA CCGGACTTTT GAGCAACCTG CTCGTTGGAC P S D Q CTTCCGATCA GAAGGCTAGT A T A GAATACCGCC	GGAACTTTTT CCTTGAAAAAA V N T L TGAATACATT ACTTATGTAA T & P AACCGCGCCT TIGGCGGCGA E W S CCGTGGTCGCC	ACCTTTTCGC TGGAAAAGG S V S ATCGGTTTCC TAGCCAAAGG G T A GGCACCGCCC CCGTGGCGGA P S P I CCGTCGCCGAT	CTTCCCGTTT GAAGGCCAAA R L A CGTCTGGCGC GCAGACCGCG S R P I CGCGCGCAAT GCGCGGGTTA T L P TACCCTGCCC	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGCGA AGACAGCCGT A Q P GCACAGCCTG	AGTATAAAAA TCATATTTT F G V TTTTGGGGGTG AAAACCGCAC N E A K AATGAAGCGA TTACTTCGGT D F V V ACTTTGGT
	-1 1601 +1 1701 +1 1801 +1 1901	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACGCTGACCG TGCGACTCGC K N F V AAAACTTCGT TTTTGAGCA Y G P A	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG À C S S CCTGTAGCTC GGACATCGAG A A R TGCGGCGCGCC G T P	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCCCCC ATCCCCT GG Y F A TATTTGCCT ATAAAACGGA G V T H	TCAAGCCCG AAGTTCGGGC DQI GATCAGATCGGGC DQI GATCAGATCC CTAGTCTAGG CCCTGACGCC GGGACTCCGG TSI	GAGCATCGAAAA CCGGACTTTT GAGCAACCTG CTCGTTGGAC P S D Q CTTCGATCA GAAGGCTAGT GAAGGCTAGT GAATACCGCC CTTATGGCGG Q A A	GGAACTTTTT CCTTGAAAA T GAATACATT ACTTATGTAA T à P ACCGCGCCT F W S CCGTGGTCGC GGCACCACCC Y D à à	ACCTTTTCGC TGGAAAACCG S V S ATCGGTTTCC TAGCCAAAGG G T A GGCACCGCCT CCGTGGCGGAA GCACCGCCTA GCACCGCCTA M V K	CTTCCCGTTT GAAGGCAAA R L A CGTCTGCCGC GCAGACCGCG S R F I CGCCGCCCAAT T L P TACCCTGCCC ATGGACCGG R T N	CGTTCAACTT GCAAGTTGAA L À L À TGGCACTGGC ACCGTGACCG L S À TCTGTCGCAA AGACAGCCGT À Q P GCACAGCCGT CGTCTCGAAC K R Q Y	AGTATAAAAA TCATATTTT F G TTTTGGCGTG AAAACCGCAC N E AATGAAGCGA TTATCTGGCT D F V ACTITGTGGT TGAAACACCA I A
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2001	L Q R CTAATTGG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACCCTCACGC TGCGACTGCC K N F V AAAACTTGT TTTTGAAGCA V G P A	M G T A ATAAAACTAT TATTTGATAT AATGGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACACCAG A A R TOCGGCCGCGC G T F GGTACGCCCAG	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACOCCACCC ATGCCCACCC ATGCCCACCC ATGCCCACCC ATGCCCACCC ATGCCCACCC ATGAAACGGA G V T H GCGTTACGCA	T M L ASGTOTTICC TCCACAAAGC AAGTICGGGC D Q I GATCAGATCC CTATGTAGG GGACTACGGC GGGACTCCGG T S I CACCTCAAT	GGCCTGAAAA CCGGACCTTG GAGCAACCTG CTCGTGGAC P S D Q CTTCCGATCA GAAGCTAGT R T A GAATACCGCC CTTATGCGGG Q A A CAGGCCCGCGG	GGAACTTTTT CCTTGAAAAA V N T T T T ACTTATGTAA T A P A ACTAGCGCGCA F Y F Y CCGTGGCCCGA GGAACCACCCG GGCACCACCCG GGCACCACCCG GGCACCACCCG TCGATGCGCGCC TCGATGCGCGCC	ACCTTTTCGC TGGAAAAGCG S V S ATCGGTTTCC TAGCCAAAGG G T A GGCACCGCCT CCGTGGCGGA GCACCGCCT GCACCGGCT GCACCGGCT M V K AATGGTTAAA	$\begin{array}{c} \text{CTTCCCGTTT}\\ \hline \text{GAAGGGCAAA}\\ \hline \text{R} & \text{L} & \text{A}\\ \hline \text{CGTCTGGCGC}\\ \hline \text{GCAGACCGCG}\\ \hline \text{GCAGACCGCG}\\ \hline \text{GCCGCCCAAT}\\ \hline \text{GCGCGGGTTA}\\ \hline \text{T} & \text{L} & \text{P}\\ \hline \text{TACCCTGCCC}\\ \hline \text{ATGGGACGGG}\\ \hline \hline \text{R} & \text{T} & \text{N}\\ \hline \text{CGCACGGACAAC} \end{array}$	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC ACCGTGACCGC L S A TCTGTCGGCA AGACAGCCGT CGTGTCGGAC CGTGTCGGAC K R Q Y AACGCCAGTA	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACCGCAC N E & K AATGAAGCGA TTACTTGGCT D F V V ACTITGTGGT TGAAACACCA I & I CATTGCTATT
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2001	L Q R CTAATTGG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACCCTCACGC TGCGACTGCC K N F V AAAACTTGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCGGCA	M G T A ATAAAACTAT TATTTGATAT AATGGATGTC GGACGCCCC GGACATCGAC G T F GGTACGCCCAC CCTTGCGCCCAC CCTTGCGCCCAC CCTTGCGCCCAC CCTTGCGCCCAC	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCCACCC ATGCCCACCC ATGCCCACCC ATGCCCACCC G V T H G V T H GCGTTACGCA CGCAATGCGT	T M L ASGTOTTICC TCCACCAAAGC AAGTICGGGC D Q I GATCAGATCC CTATGTAGATCC CCCTGACGCC GGGACTCCGG T S I CACCTCGAT GTGCAGGCTAA	GGCCTGAAAA CCGGACCTTG GAGCAACCTG CTCGTGGAC P S D Q CTTCCGATCA GAAGCTAGT R T A GAATACCGCC CTTATGCGGG Q A A CAGCCCCGG GTCCGGCGCC	GGAACTTTTT CCCTTGAAAAA V N T L TGAATACATT ACTTATGTAA T À P ACCGCGCGCA T A P ACCGCCGCA CCCTGGCCCCA GGCACCACCG GGCACCACCG GGCACCACCG GGCACCACCG AGCTACGCCG AGCTACGCCG	ACCTTTTCGC TGGAAAAGCG S V S ATCGGTTTCC TAGCCAAAGG G T A GGCACCGCCT CCGT0GCGGA GCACCGCCT CCGT0GCGGA GCACCGCCT M V K AATGGTTAAA TTACCAATT V A T	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCGC GCAGACCGCG GCAGACCGCG GCAGACCGCG T L P T L P TACCTGCCC ATGGGACGGG R T N CGCACGAACA GCGTACTGTTN	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGT CGTGTCGGAC CGTGTGGGAC CGTGTCGGAC TTGCGGGTGAT	AGTATAAAAA TCATATTTT F G TTTTGGCCTC AAAACCGCAC N E AATGAAGCA TTTATGGCT D F ACTATGGCA TTATTGGCT CATTGTGGT GATGAACACCA I A CATTGCTATT GTAACGATAA K T
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2001 +1 2101	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACCCTGAGCG TGCGACTGGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCCGGCA CCCCGGCCGCA M F G ATGCCGGGCG	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TOCGGCCGCGC G T F GGTACGCCGCGC G T F GGTACGCCAG CCATOCGGTC D Y G G	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGTGG Y F A TATTTTOCCT ATAAAACGGA G V T H GCGTTAACGCA CCCAATGCGT T V Y	T M L ASGTOTTICC TCCACAAAGC AAGTICGGGC D Q I GATCAGATCC GGATCAGATCC CTATOTCTAGG T S I CACCTCAAT GTGCAGCTAA V P A GTTCCCGCCG	I S M GGCCTGAAAA CCGGACCTTG GAGCAACCTG GAGCAACCTG CTCCGATGAC GAAGCAACCTG P S D CTTCCGATCA GAAGCTAGT GAATACCGCC CTTATGCGGG CTATAGCGG GCCGGGCGCGG GTCCGGCGCC GCCCGGGCACG	GGAACTTTTT CCCTTGAAAAA V N T L TGAATACATT ACTTATGTAA T À P ACCGCCGCA T A P ACCGCCGCA SCCCCTGCC GGCACCACCG GGCACCACCG GGCACCACCG ACTACCCCG ACTACCCCCG	ACCTTTTCGC TGGANAAGCG S V S ATCGGTTTCC TAGCCAAAGG G T A GGCACCGCCT CCGT0GCGGA GCACCGCCT CCGT0GCGGA GCACCGGCTA M V K AATGGTTAAA TTACCAATTT Y G T TACGGGACCG	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGGCG GCAGACCGCG GCAGACCGCG GCAGCCCCAAT GCGCGGCTA T L P T A CCCCCCGAT GCGCGGCTA R T N CGCACGGACCA GCGTCCTGT G E K T G E K F GTGAAAAACC	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGT CGTGTCGGAC CGTGTCGGAC CGTGTCGGAC TTGCGGGTCAT I D V AACGCCAGTA	AGTATAAAAA TCATATTTT F G V TTTTGGCCTC AAAACGGCAC N E & K AATGAAGCGA TTACTTGGCT D F V V ACTITOTGGT TGAAACACCA I A I CATTGCTATT GTAACGATAA K I G M AAAATTGGGA
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2001 +1 2101	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAC TCGTCCGACCGC TGCGACTGGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA M F G ATGCCGGGCCGCA	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG C A A R TOCGGCCGCCG G T F GGTACGCCCG CCATGCGGCC D Y G G ACTATCAGG TGATAGTCCC	R K G CTGATGGTAC GACTACGATG GACTACGATGCTAC GACTATGAAA T P P TACCCCACCC ATGCGCACCCC ATGCGCACCC ATGCACCACCC T P P TACCCCACCC ATGCACCACCC ATGCACACCCACCC ATGCACACCC GCGTAAAACGGA GCGTTAACGCA GACCATGCAT T T V GACCATGCAT CGGCACATA	T M L ASGTOTTICC TCCACAAAGC AAGTICGGGC D Q I GATCAGATCC GGATCAGATCC CTACTCTAGG T S I CACCTCATT GTGCAGCTAA V P A GTTCCCGCCG CAAGGCCGCG	Q I S M GGCCTGAAAA CCGGACCTTG GAGCAACCTG CTCGTGGAC P S D Q CTTCCGATCA GAAGCTAGT R A A GAATACCGCC CTTATGCGG Q A A CAGCCCGGCGCC A P G S CCCCGGCAAC	GGAACTTTT CCCTTGAAAA V N T L TGAATACATT ACTTATGTAA T À P ACCGCCGCA T A P ACCGCCGCA CCCTGGCCCCA GGCACCACCG GGCACCACCG GGCACCACCG AGCTACGCCCA AGCTGGGAC	ACCTTTTCGC TGGANAAGCG S V S ATCGGTTTCC TAGCCAAAGG G T A GGACCGCCT CCGTGGCGGAT GCACGGCTA M V K AATGGTTAAA TTACCGATT Y G T TACGGGACCG	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCGC GCAGACCGCG GCAGACCGCG GCAGCCCCAT GCGCGCCCAT GCGCGGGCTA T L P T L P TACCTGCCC ATGGGACGGG R T N CGCACGAACA GCGTGCTGTT G E K P GCGAAAAACC CACTTTTTGG	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGT CGTGTCGGAC CGTGTCGGAC CGTGTCGGAC K R Q Y AACCCCAGTA TTGCGGGTCAT I D Y ACCGCTACACTTG CTAGCTACACTTG	AGTATAAAAA TCATATTTT F G V TTTTGGCCTC AAAACGGCAC N E & K AATGAAGCGA TTACTTCGCT D F V V ACTITOTGGT CATTGATACCA I A I CATGCATAT GTAACGATAA K I G M AAAATTGGGA TTTTTACCCT
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2001 +1 2101 +1	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACCCTGAGCG TGCGCACTGGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA ATGCCGGGCG TACGGCCCGCC M A TGCCGGCCG TACGGCCCCC	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TGCGGCCGCGCCG G T F GGTACGCCGC CCATCGGGC CCATCGGGCCAG CCATCGCGCC D Y G G ACTATCAGGG TGGTAAGTCCC G E M TGCGCAATCGAG	R K G CTGATGCTAC GACTACGATG GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCG ATGCCGTGG Y F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y AACGTCGCTG	T M L ASGTOTTICC TCCACAAAGC AAGTICGGGC AAGTICGGGC D Q I GATCAGATCC CTACTTCTAGG CCCTCACCC GGGACTCGGC T S I CACCTCCATT GTGCAGCTAA V P A GTTCCCGCGC CAAGGCCGCC D W R R ATTGCCCCC	Q I S M GGCCTGAAAA CCGGACCTTG GAGCAACCTG CTCGTGGACC P S D Q CTTCCGATCA GAAGCTAGT N A GAATACGCC CTTATGCGG Q A A CAGGCCGCG GTCCGGCGCC A F G S CCGCCGGGAAG GCGGCCCTC A V N	GGAACTITTT CCCTTGAAAAA V N T L TGAATACATT ACTTATGTAA T À P ACCGCCGCA T A P ACCGCCGCA CCCTGGCCCCA GGCACCACCG GGCACCACCG GGCACCACCG AGCTACGCCCA AGACTGGGAC P G G CCCGCGGGA	ACCTTTTCGC TGGANAAGCG S V S ATCGGTTCC TAGCCAAAGG G T A GGACCGCCT CCGTGGCGGAT GCACGGCTA M V K AATGGTTAAA TTACCGATT Y G T TACGGGACCG ATGCCCTGGC ATGCCCTGGC	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGGCG GCAGACCGCG GCAGACCGCG GCAGACCGCG T L P T L P TACCTGCCC ATGGGACGGG R T N CGCACGAACA GCGTGCTTGT G E K P GTGAAAAACC CACTTTTGG G K P TGTGAAAACC	CGTTCAACTT GCAAGTGAA GCAAGTGAA L A TGGCACTGGC ACCGCACTGGC ACCGCACTGGCA AGACAGCCGT GCAACAGCCTG CGTACGACCGG CGTACGCACAGCCTG CGTACGACCGGTATA A Q P AACGCCAGCCTG GCACAGCCTG CGTACGACAGTA TTGCGGGTCAT TGCGGGTCAT GATCGATOTG CATCGATATGC CTAGCTACACC A W Y	AGTATAAAAA AGTATAAAAA TCATATTTT F G TTTTGGCCTG AAAACGGCAC N E K AATGAAGCGA TTACTTGGCT D F V ACTATGTGGCT D F V A CATTGTTGGCT CO F V A CATTGTTGGT CATTGCTATT GTAACGATAA K I G M K I G M AAAATTGGGA TTTTAACCCT M F D N T
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2101 +1 2101 +1 2201	L Q R CTAATTGG GAATTAACAC GGATTAACAC AGCAGGCTTC TCGTCCGAACGC TCCGCCGACCGC TCCGCCGACCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGCCCGGCA CCCCGGCCGCA ATGCCGGCGCC TACGGCCCGCC M A T C TGCCCTTCA	M G T AAAAACTAT TATAAAACTAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TOCGGCCGCGC G A A R GT T F GGTACGCCGCCGC CCATOCGGTC CCATOCGGTC D Y G G ACTATCAGG TGATAAGTCCC O G E M TGGTGATAATC	R K G CTGATGCTAC GACTACGATG GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCG ATGCCGTGG Y F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y GACCGTGTAT CTGGCACATA S V A ACCGCCGTG	T M L ASGTOTTICC TCCACCAAAGC AAGTICGGGC AAGTICGGGC D Q I GATCAGATCC CTACTGCTAGG CCCTGACGCC GGGACTCGGC T S I CACCTCGACT GTGCCGCGCC CAAGGCCGCC D W R A ACTGCCCGCC	Q I S M GGCCTGAAAA CCGGACCTTG GAGCAACCTG CTCGTGGACC P S D Q CTTCCGATCA GAAGCTAGT R A A CAGGCCGCC CTTATGCGG Q A A CAGGCCGCGC GTCCGGGCCC A F G S CCCCGGGAAC GCGCCGCGAAT GCGCCGCGCATA	GGAACTTTT GCATGATAAAA V N T L TGAATACATT ACTTATGTAA ACTTATGTAA T À P ACCGCCGCA T A P ACCGCCGCA GGCACCACCG GGCACCACCG AGCTACGCCG AGCTGGGAC P G G AGCTGGGAC P G G CCCGCGGGA GGGCCCGCCT	ACCTTTTCGC TGGANAAGCG S V S ATCGGTTTCC TAGCCAAAGG G T A GGACCGCCT CCGTGGCGGAT GCACGGCTA M V K AATGGTTAAA TTACCGATT Y G T TACGGGACCG ATGCCCTGGC ATGCCCTGGC ATGCCCTGGC ATGCCCTGCC TATATATACG	$\begin{array}{c} \text{CTTCCCGTTT}\\ \hline \text{GAAGGGCAAA}\\ \hline \text{R} & \text{L} & \text{A}\\ \hline \text{CGTCTGGCGC}\\ \hline \text{GCAGACCGCG}\\ \hline \text{GCAGACCGCG}\\ \hline \text{GCAGACCGCCAAT}\\ \hline \text{GCGCGCCCAAT}\\ \hline \text{GCGCGGGCTA}\\ \hline \text{T} & \text{L} & \text{P}\\ \hline \text{TACCCTGCCC}\\ \hline \text{ATGGGACGGG}\\ \hline \text{R} & \text{T} & \text{N}\\ \hline \text{CGCACGAACA}\\ \hline \text{GCGCACGAACA}\\ \hline \text{GCGCACGAACA}\\ \hline \text{GCGCACGAACA}\\ \hline \text{GCGCACGAACA}\\ \hline \text{GCGACGAACA}\\ \hline \text{GCGACAACA}\\ \hline \text{CACATTTGGG}\\ \hline \end{array}$	CGTTCAACTT GCAAGTGAA L A TGGCACTGGC ACCGCACTGGC ACCGCACTGGC AGACAGCCGT GCAACGCCG CGTACGCACGCCG GCACAGCCGG CGTACGACCGCG GACGACAGCCGT GACGACAGTA TGCGCACAGTA GATCGACTGGAC CTAGCTACACCAC A Q P ACCCCCAGTA TGCCGCTATAC CGGACCATAT	AGTATAAAAA TCATATTTT F G TTTTGGCCTG AAAACGGCAC N E AATGAAGCGA TTACTTGGCT D F Q V ACTATGGCA TGAAACACCA I A CATTGCTATT GTAACGATAA K I K G AAAATGGGGA TTTTAACGATAA K F AAAATGGGA TTTTAACCCT M F TGTTCGGATATA ACAAGCTATAA ACAAGCATAA
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2101 +1 2201 +1 2201	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACCCTCAACGC TGCGCACTGGC TTTTTGAAGCA Y G P A AGGGCCCGGCA CCCCGGCCGCA CCCCGGCCGCA M F G ATGCCGGCCGCC TACGGCCCGCC ACCGCCAACT N C Q S	M G T A ATAAAACTAT TATTAGATOTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG C A A R TOCGGCCGCCCC A C T F GGTACGCCGCCCC G C T F GGTACGCCAG CCATCGCGCCC CCATCGCGCCCC CCATCGCGCCCC CCATCGCCCC CCATCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCCCCCC ATGCCCCCCCC ATGCCCCCCCC T V X GCGTTACGCA CGCCAATGCGT T V Y GACCGTCGTCA S V A ACCGTCGCTC	T M L ASGTOTTICC TCCACCAAAGC AAGTICGGGC D Q I GATCAGATCC GGACTAGATCC CTACTTCTAGG T S I CCCTCACCC GGGACTCAG T S I CACCTCCATT GTGCAGCCGAT GTTCCCGCGC CAAGGCCGCC D W R ACTGCCCGCGC CAAGGCCGCC V M C TGACCCGCGCC	Q I S M GGCCTGAAAA CCGGACCTTG GAGCAACCTG CTCGTGGACC P S D Q CTTCCGATCA GAAGCTAGT R A GAATACGCC CTTATGCGG Q A A CAGGCCGCGC GTCCGGCGCC A P G S CCCCGGGAAC GCGCCGCGAAC GCGCCGGCAAT GCGCCGCGCAT S A A	GGAACTITTT GCATGATAAAA V N T T ACTTATGTAA T A P A ACTTATGTAA T A P A CCGTGGCCGGA B W B W CCGTGGTCGC GGAACCACCG ACTATACCACC AGACTGGGAC P B CCGACGGGA CCGATGGGAC CCGGCGGGA P G CCCGGCGGGA P G CCCGGCGGGA P S CCCGGCGGGA F W M S CCCGGCCGCCCT F S CCCCGCCGGAA CCCGGCCGCCCT F S CCCGTCGGCACCCT	ACCTTTTCGC TGGAAAAGCG G T A GGCACGGCTTAC CGTGGCGAA GGCACGGCTA CCGTGGCGGAT GCACGGCTA M V K AATGGTTAAA TTACCGATT Y G T TACGGGACCG ATGCCCTGGC ATGCCCTGGC ATGCCCTGGC ATGCCCTGGC ATGCCCTGGC ATGCCCTGGC ATGCCCTGCC ATGCCCTGC ATGCCCTGCC ATGCCCCTGCC ATGCCCCGCC ATGCCCTGCC ATGCCCCGCCC ATGCCCCCCCCCC	$\begin{array}{c} \text{CTTCCCGTTT}\\ \hline \text{GAAGGGCAAA}\\ R & L & A\\ \hline \text{CGTTGGGCG}\\ \text{GCAGACCGCG}\\ \text{GCAGACCGCG}\\ \text{GCAGACCGCG}\\ \hline \text{GCGCCCCAAT}\\ \text{GCGCGCCCAAT}\\ \text{GCGCGCCGAT}\\ \hline \text{T} & L & P\\ \hline \text{TACCCTGCCC}\\ \text{ATGGGACGGG}\\ \hline \text{R} & T & N\\ \hline \text{CGCACGAACA}\\ \text{GCGTCTTGT}\\ \hline \text{G} & E & K \\ \hline \text{GCGACGAACA}\\ \hline \text{GCGACGACCAC}\\ \hline \text{GCGACGACCAC}\\ \hline \text{GCGACGACCAC}\\ \hline \text{GCGACGACCAC}\\ \hline GCGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CGTTCAACTT GCAAGTTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGT CGTCTGGAC CGTACGAGCCTG CGTCGGACCAGTA TGCGGGCCATAT CTAGCTACAC A W Y GCCTGGTACAC CGGACCATAT Q N L T	AGTATAAAAA TCATATTTT F G V TTTTGGCCTC AAAACGGCA ATGAAGCGA ATGAAGCGA TTACTTCGCT D F V ACTATOTGGCT ACTATOTGGCT CATGCATCACCA I A I CATTGCTATT GTAACGATAA K I G K I G M AAAATGGGGA TTTTAACGATAA K I G AAAATGGGGA TTTTAACCCT T AAAATGGGA TTTTAACCCT I N ACAAGCTATAA AACAAGCTATAA ACAAGCTATAA ACAAGCTATAA
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2101 +1 2101 +1 2301	L Q R CTTAATTOTG GAATTAACAC GAATTAACAC AGCAGGCTTC TCGTCCGAAGC TCGTCCGACCGC TGCGCACTCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCCGCA CCCCGGCCGCA TACGCGGCG TACGGCCGCC TACGGCCCGCC ACCGCTAACT N C Q S CTGCCAAAGC	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TGCGGCCGCGCCG G T F GGTACGCCAG CCATCGGGTC D Y Q G G ACTATCAGG TGATAGTCCC D G E M TGGTGAAATCC K H A AAAACCGCCG	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGCGG G V F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y ACGCCACTGAT CTGGCACATA S V A ACGCTCCCTGAT CGCGCACGCACGC A T I G CGACGATCGC	T M L AGGTGTTTCC TCCACCAAAGG AAGTTCGGGC AAGTTCGGGC D Q I GATCAGATCC CTACTGCTAGG GGACTGCGG GGACTGCGG GGACTGCGG T S I CACCTCCATT GTGCAGGCGGC D W R R ACTGGCGCG D W R R ACGGCGCGGC V M C TGACCACGCGC	Q I S M GGCCTGAAAA CCGGACCTTG GAGCAACCTG CTCGTGGAC P S D Q CTTCCGATCA GAAGCTAGT R A A GAATACCGCC CTTATGGCGG GTCCGGGCCC A P G S A A CCGCCGGGAAG GGCCCCTCC A V N CCCCCGGGCAAT GCGCCACTTA S A A TCTGCCGCGT	GGAACTTTT GCATGATATA V N T T T T ACTTATGTAA T A P A ACTTATGTAA T A P A CCTGGGCGCA W S W S W S W A TCGATGGGAC A TCGATGGGAC A TCGATGGGAC A CCATGGGACCCTG A ACTTACGCCG A ACGATGGGAC C P G CCCGGGGGGAC C GGACCGGGGAC C F M Q CCCGGGGGGGACCCCCT C F W Q CCGAGGGGCCCCCCT C GGACCACCCCT C AGACCAGCGGT C	ACCTTTTCGC TGGANAAGCG TGGATAGCT TAGCCAAAGG G T A GGACCGCT CCGTGGCGCA GCACCGCT CCGTGGCGCA GCACCGGCT M V K AATGGTTAAA TTACCAATTT Y G T TACGGGCCG ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC TTATTACGG	$\begin{array}{c} \text{CTTCCCGTTT}\\ \textbf{GAAGGGCAAA}\\ \mathbb{R} L A \\ \hline \\ \text{CTTTGGCGC}\\ \textbf{GCAGACCGCG}\\ \textbf{GCAGACCGCGG}\\ \textbf{S} \mathbb{R} \mathbb{P} \textbf{L}\\ \textbf{GCGCCCCAAT}\\ \textbf{GCGCGCGGGTTA}\\ \hline \\ \textbf{T} L \mathbb{P} \\ \hline \\ \textbf{T} \textbf{C} \textbf{C} \\ \textbf{GCGCGCGATA}\\ \textbf{GCGGCCGAACA}\\ \textbf{GCGTCCTTGT}\\ \textbf{G} \mathbb{E} \textbf{K} \mathbb{P} \\ \textbf{T} \textbf{N}\\ \textbf{GCGCACGAACA}\\ \textbf{GCGTCCTTTTGG}\\ \hline \\ \textbf{G} \textbf{K} \textbf{K} \\ \textbf{T} \textbf{N}\\ \textbf{GCTGAAAAACC}\\ \textbf{CACTTTTGGG}\\ \hline \\ \textbf{L} \textbf{Q} \textbf{L} \\ \textbf{CTGCAGCTAK}\\ \end{array}$	CGTTCAACTT GGAAGTGAA L A L A TGGCACTGGC ACCGTGACCG AGACAGCCGC GGCACAGCCTG CGTCGAACG CGTCGAACG GGCACAGCCTG GGCACAGCCTG CGTCGACCG CGTCGAACTGTG GACCGCAGTA TTGCGGGCTAT I D Y ACGCCCAGTA I D Y GCTCGAACTGTG CACGCCAGTAT TGCGGTCATAT GGGCCGGTATA CGGACCATATTG Q N L T AAAATCTGAAC N L T AAAATCTGAAC	AGTATAAAAA TCATATTTT F G V TTTTGGCCTG AAAACGGCAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGT TGAAACACCA I A I CATTGCTATT GTAACGATAA K I G M AAAATTGGGA TTTTAACCCT M F D N TGTTCGAGAAAC ACCATCTATT I E N CATGGAGAAC GTAACTCTTG
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2101 +1 2101 +1 2301 +1	L Q R CTTAATTOTG GAATTAACAC GAATTAACAC AGCAGGCTTC TCGTCCGAAGC TCGTCCGACCGC CCCCACTCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA ATGCCGGCGCG TACGGCCCGCC TCGCCATGA ACCCCTAACT N C Q S CTGCCAAAGC TCCCCAAAGC	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TGCGGCCGCGCGC G T F GGTACGCCGCCGC G G T P GGTACGCCAG CCATCCGGTC D Y Q G G ACTATCAGG TGATAGTCCC K H A AAAACCGCCC D S V D	R K G CTGATGCTAC GACTACGATG GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGTGG Y F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y ACGCCACTGAT CTGGCACATA S V A ACGCCCCCC A T I G CGCACGCCC A G N	T M L AGGTGTTTCC TCCACCAAAGC AAGTTCGGCG AAGTTCGGCG D Q I GATCAGATCC CTACTGCTAGG GGACTCGGC GGGACTCGG GGACCCGCG CACCTCGATC GTTCCCGCCG CAAGGCGGCG D W R R ACTGGCGCCG D W R R ACTGGCGCCG T GTACCGCGCG V M C TGACCACACACA H P A	I S M GGCCTGAAAA CCGGACCTTG CCCGACTGACAC GACCACCTGACA GACCACCTGACA GACGACCTGACA GAACGCTAGTA GAACCCGCC N T GAATACCGCC CTTATGCGGG Q A CAGCCCCCGG GCCCCGCGGGACG GCCCCCGGGACG GCCCCCGGGCCCTC A P G A CCCCCGGGACC GCCCCGCGGACG GCGCCACTTC A S A TCTCCCCCGCTT S A CCTCCCCCGCT S A TCTCCCCCCGT A S A TCTCCCCCCGCT V A V A	GGAACTTTT GCATGATATA CCTTGATATAAAA V N T L TGAATACATT ACTTATGTAA T À P ACCGCCGCA T A P ACCGCCGCA GGCCCGCCA CCGTGGTCGC AGACTACGCCC AGACTACGCCC F W S Q TCTGGTCGCA AGACCACCGCT TCTGGTCGCA	ACCTTTTCGC TGGANAAGCG TGGATAGCT TAGCCAAAGG G T A GGACCGCT CCGTGGCGGA GCACCGCT CCGTGGCGGAT GCACCGGCT M V K AATGGTTAAA TTACCGATTT Y G T TACGGGCCG ATGCCCGGC ATGCCCGCC ATGCCCGCC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGC ATGCCCCGCCC ATGCCCCGC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCGC ATGCCCCGCC ATGCCCCGCC ATGCCCCGCC ATGCCCCCCCCCC	$\begin{array}{c} \text{CTTCCCGTTT}\\ \textbf{GAAGGGCAAA}\\ \mathbb{R} \ L \ A\\\\ \text{CTTTGGCGC}\\ \text{GCAGACCGCG}\\ \text{GCAGACCGCG}\\ \text{GCAGACCGCG}\\ \text{T} \ L \ P\\\\ \text{TACCTGCCC}\\ \text{ATGGGACGGG}\\ \mathbb{R} \ T \ N\\\\ \text{CGCACGAACA}\\ \text{GCGTGCTTGT}\\ \text{G} \ E \ K \ P\\\\ \text{TGGTAAAACC}\\ \text{CACTTTTGG}\\ \text{G} \ K \ P\\\\ \text{TGGTAAAACC}\\ \text{ACCATTTGG}\\ \text{ACCATTGGGA}\\ \text{ACCATTGGACAGCAAC}\\ \text{ACCATTGGACAACG}\\ \text{CACGACCAACA}\\ \text{CTGGACGACCACG}\\ \text{GACCTCATG}\\ \text{GACCTCATG}\\ GACTCACACACACACACACACACACACACACACACACACA$	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACCGG L S A TCTGTCGGCA AGACAGCCGG CGTCTGGAC CGCACGGCTG CGCACGGCTG CGTCGAAC ACCGCCGGTAT TGCGGGCATAT GCCCGGTATA CGGACCATAT Q N L T AAAATCTGAC V N I	AGTATAAAAA TCATATTTT F G V TTTTGGCCTG AAAACGGCAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTITGTGT TGAAACACCA I A I CATTGCTATT GTAACGATAA M F D N TGTTGGAGAAAC ACATGCTATT I E N CATGGAGAAC GTAACTCTTG I CATGCTATT I E N CATGGAGAAC
	-1 1601 +1 1701 +1 1801 +1 2001 +1 2101 +1 2301 +1 2401	L Q R CTTAATTOTG GAATTAACAC GAATTAACAC AGCAGGCTTC TCGTCCGAAGC T L S ACCCTGAGCG TGCGACTGGC K N F V AAAACTTGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA CCCCGGCCGCA ACCGCTAACT N C Q S CTGCCAAAGC TCGCCAAGC T L G ACGCTAGCCG ACGGTTCG T L G ACGCTAGCCG	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TGCGGCCGCGCCG G T F GGTACGCCGCCGC CCATCGGTC D Y G G ACTATCAGG TGATAGTCCC K H A AAAACCGCCG D S Y D ACTACCGTCGA	R K G CTGATGCTAC GACTACGATG GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGTGG Y F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y ACGCCACTGAT CTGGCACATA S V A ACGCCCCCTTA TCCCAGGAACGC A G N TGCCGGAAACCC A G N	T M L ASGTOTTICC TCCACCAAAGC AAGTICGGGC D Q I GATCAGACCC GGGACTCCGGC GGGACTCCGGC GGGACTCCGGC GGGACTCCGC GGGACCCGAC T S I CACCTCCACC CACCCCGACC CAAGGCGGCC D W R R ACTGGCCCG D W R R ACTGGCCCG T GTACCGCCGC T GACCACACACA H P A CATCCGCCGCG GTACGCCCCC	Q I S M GGCCTGAAAA CCGGAACCTTG GAGCAACCTG CTCGCGACCTG P S D Q CTTCCCATCA GAAGCTAGT N A A GAATACCGCC CTTATGCGG Q A A CAGGCCCGCG GTCCGGGCCC A F G S CCCCGGGCAAC GCGCCCCCGG GTCCGGCGCCA A F G S CCCCCGGCGCAA GCGCCACCTA S A A TCTGCCGCGTCC V A L R TGGCGTTCCC	GGAACTITTT GCATGATATTTT CCCTGAAAA V N T A T T ACTATACATT ACTATAGTAA T A P A CCGTGGTCGC GGAACCACCG GGACCACCACG V D ACTATAGTAA TCGACGCGA V D A TCGATGCGGA AGCCACACCG P G CCCGGCGGCACCCT P G CCCGGGCGGA GGACCACCCCT AGACCACGCGT T T GGACCACCCCT F S AGACCACGCGT T D TACGGATGGC TACGGATGCCCCCCCC	ACCTTTTCGC TGGANAAGCG G T A GGACGGTTCC TAGCCAAAGG G T A GGACCGCT CCGTGGCGGA GCACGGCT CCGTGGCGGAT GCACGGCT M V K AATGGTTAAA TTACGAGACCG ATGCCCGGC ATGCCCGCC CCTTATTACGG	CTTCCCGTTT GAAGGGCAAA R L CGTCTGGCGC GCAGGCCCAAT GCCCCCAAT GCCCCCAAT GCCCCCAAT GCCCCCAAT GCCCCCAAT GCCCCCAAT GCCCCCAAT GCCCCCAAT GCCCACGACCACC CCCATCGCC ATGCCTTCGCC ATGCCTTCGCC CCCATCGAACA GCCCACGAACA GCGACGAACA GCCATCGAAAACC CACATTGGC L P T N GCCACGAACA CACATTGGC L Q L CCCACGAGCTAC CTGCAGCTAC Q I N AGATTAACAA AGATTAACAA	CGTTCAACTT GGAAGTGAA L A L A TGGCACTGGC ACGGCACTGGC AGACAGCCGG AGACAGCCGG L S A TCTGTCGGAC AGACAGCCGG CGTACGTCGGAC CGTCGAACGCGG CGTCGAACGCGG CGTCGAACGCGG L S A Q P A Q P AGACGCGGCGG CGTCGAACGCGG CGTCGACGGTAT TGGGGGCATAT GGGCGGGTATA GGGCACATATCGAC A W Y GGCCGGTATAC CGGGACCATAT T AAAATCTGAC N L T A Q N L T AAAATCTGAC Q N L T GGTCGAATATC GGTCGAATATCGAC	AGTATAAAAA TCATATTTT F G V TTTTGGCCTG AAAACGCAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTITOTGGT TGAAACACCA I A I CATTGCTATT GTAACGATAA M F D N TGTTGGAGAAAC CATGCGAGAAC CGTACCTCTTG I E N CATGGAGAAC GTAACCTTTG
	-1 1601 +1 1701 +1 1801 +1 1901 +1 2101 +1 2301 +1 2401 +1	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAGC TCGTCCGACCGC TGCGCACTCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA CCCCGGCCGCA ATGCCGGCGCAACT N C Q S CTGCCAAACT CTGCCAAACT CCGCCAACC ACCGCTAACT N C Q S CTGCCAAACT CTGCCAAACC CCCCCAACC ACCGCTAACT	M G T A ATAAAACTAT TATTATGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TGCGGCGCGCGC G T F GGTACGCCGC G C T P GGTACGCCAG CCATCGGTG ACTATCAGG TGATAGTCCC CATGGTGAAATC K H A AAAACCGTCGA D S V D ACAGCGTGGA	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGCGG ATGCCGCGGG ATGCCGTGG G V T H GCGTTACGCA CGCAATGCGT T V Y ACGCCACGTATA CTGGCACATG CGCCACGCACGC A G N TGCCGGAACGC A G N TGCCGGAACGC A G N TGCCGGGAACGC A G N TGCCGGAACCC A G N TGCCGGAACCCC A C A G N TGCCGGAACCCCCC A C A C A C A C A C A C	T M L AGGTOTTICC TCCACCAAAGG AAGTICGGGG AAGTICGGGG D Q I GATCAGATCC CTACTGCTAGG GGACTCCGG GGACTCCGG GGACTCCGG T S I CACCTCCAT GTGCAGCCATA V P A GTTCCCGCGG CAAGGCGGGC D W R R ACTGGCGCG CAAGGCGGGC V M C TGACCACGCGGC V M C TGACCACGCGC CACTCGCCGC CACTCGCCGCG CACGCCGCC V M C TGACCACGCCGC CACTCGCCGCG CACGCCCCC V Q N R	Q I S M GGCCTGAAAA CCGGAACCTTG GAGCAACCTG CTCGCGACCTG P S D Q CTTCCGATCA GAAGCTAGT N L A GAATACCGCC CTTATGCGG Q A A CAGGCCGCGGAAG GGCCCCCGG GTCCGGGCCC A P G S A P G S CCCCCAGCGCGAAG GCGCCCCTTC S A A TCTCCCCGCGT S A A TCTCCCCCGCTTC S A A CCGCCAGCGCCA V A L R TGGCGTTCCC ACCGCAACG	GGAACTITTT GCATGATAAAA V N T L TGAATACATT ACTTATGTAA T À P ACCEGCEGCA T A P ACCEGCEGCA SECCEGEA V S CCETEGTCGC GGCACCACCEG AGCTACGCCG AGCTACGCCC CCEGEGEGCA AGACTACGCCCT F W S Q TCTGGTCGCA AGACCACCGT T D G TACGATGEGCCT T D G TACGATGEGC D R Q	ACCTTTTCGC TGGANAAGCG TGGANAGCG G T A GGACGGCTA CCGTGGCGAC GGCACCGCCT CCGTGGCGGAC GCACGGCTA M V K ATGGTTAAA TTACGAGACCG ATGCCCGGCA TACGGACG ATGCCCGGCA TACGGACG ATGCCCGGCA TACGGACG ATGCCCGGCA TACGGACG ATGCCCGGCA TACGGACG ATGCCCGGCA TACGACGACG ATGCCCGGCA CCAAAGTCC CAAAAGTCC CTTATTAACGG D K V GACAAAATCC CTGTTCAGG P R T L	CTTCCCGTTT GAAGGGCAAA R L CGTCTGGCGC GCCCCCAT GCCCCCAT GCCCCCAT GCCCCCAT GCCCCCAT GCCCCCAT GCCCCCAT GCCCCCAT GCCCCCAT GCCCCCAC ATGCCTCCCC ATGCCTCCCC ATGCCTCCCC CCCATCGAACA GCCGTCCTGT G K GCCACGAACA GCGTGAAAAACC CACTTTTGG L Q L CACATTGCAC ACCATTGCAC Q. I Y K AGATTAACAA TCTAATTGTT	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACTGGC ACCGGACTGGC AGACAGCCGG CTTCTGCGCA AGACAGCCGG CGTCTGGAC K R Q Y ACCGCCAGTA TGCGGGCATAT GCTCGGATCAT CTAGCTACACA A W Y GCCCGGTATA CGGACCATAT Q N L T AAAATCTGAC Q N L T AAAATCTGAC Q N L T AAAATCTGAC GGTCAATATC CCAGTAATATC CCAGTAATATC	AGTATAAAAA TCATATTTT F G V TTTTGGCCTG AAAACGCCAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGT TGAAACACCA I A I CATTGCTATT GTAACGATAA K I G M AAAATTGGGA TTTTAACCCT M F D N TGTTCGAGAAAC ACCAT CATGGCTATT I E N CATGGCATAT I E N CATGGAGAAC CATGGCATAT I G R Q CATGGCGCCC G GAACCGCCGC G GAACCGCCGCG E G D V
	-1 1601 +1 1701 +1 1801 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAGC TCGTCCGACCGC TCCGACTCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA CCCCGGCCGT ACCGCTAACT N C Q S CTGCCGATGGC TCCGATGGC ACCGCTAACT N C Q S CTGCCAAGC TCCGATGGC CACGGTTCG TCCGATGCCG CACGGTTCGG CCCCAAGCC C C C CACGCC C C C C C C C C C C C C C C C C C C	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TGCGGCGCGCGC G T F GGTACGCCGC CCATCCGGTC D Y Q G G T C GGTACGCCAG CCATCCGGTC D Y Q G G E M TGGTGGAAATC K H A AAAACCGTCGA D S Y D ACAACCGTCGA CCATCTTACCGC TGTGCGACATC	R K G CTGATGCTAC GACTACGATG GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGCGG G V F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y AGCGCTGTAT CTGGCACGATGGC A G N TGCGGGAACGC A G N TGCGGGAACGCCTTG N S G AACACCGCCC N S G AACACCGCCC	T M L AGGTOTTICC TCCACCAAAGC AAGTICGGGC D Q I GATCAGATCC GGGACTCGGC CTACTGCTAGG CCCTGACGCC GGGACTCGGC CT S I CACCTCCAT GTGCGGGCGC D W R R ACTGGCGCG D W R R ACTGGCGCG CAAGGCCGCC V M C TGCACCGCGGC V M C TGCACCTCGCGCG CACGCCCCCC V M C TGCACCGCGCC V M C TGCACCGCCGC V M C TGCACCGCCGC V M C TGCACCGCGCC V M C TGCACCGCCGC V M C TGCACCGCCCC V M C	Q I S M GGCCTGAAAA CCGGAACCTTG GAGCAACCTG CTCGCGACCTG P S D Q CTTCCGATCA GAATACCGCC CTTATGCGGC GAATACCGCC CTTATGCGGG GTCCGGGCCCCG GTCCGGGCCCCGG GTCCGGGCCCCTC A P G S A A CCGCCAGCGCGAAG GCGCCCCCGG GTCCCGGGCCC A V N S A A TCTCCCGCGTCCC A A L R TCGCGCAACCT CCGCAACCCC	GGAACTITTT GCATGATAAAA V N T L TGAATACATT ACTTATGTAA ACTTATGTAA T À P ACCEGCEGCA T A P W S CCETEGTCGC GGCACCACCG AGCTACGCGA AGACTACGCCA CCEGGCGCCCT F W S CCCGGCGCCCT TCTGACCCCG AGACCACCGG CCCCGTGGCGCA AGACCACCGGT TACGATGGCCCA CCACGCACGCC TACGCATCGCCACC D R Q GATCGCCACCC CTACGCCCACCG	ACCTTTTCGC TGGAAAGCG TGGAAAGCG G T A GGCCCCT CCGTGGCGAC GGCCCCGCT CCGTGGCGGAC GCACCGGCT M V K ATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCGGC ATGCCCGGC ATGCCCGGC ATGCCCGGC M N G GAATAACGGT TATATACGG N N G GAATAACGGT TATATACGG CTTATTACGA	CTTCCCGTTT GAAGGGCAAA R L CGTCTGGGCG GCACGCGCG GCCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAAT GCCCCCCAAT GCCCCCCAACA GCCCCCCAACA GCCCCCCAACA GCCCCCCAACA GCCCACGAACA GCCCACGAACA GCCACGCAACA Q L Q L GCCACTACCAACG Q L Q L Q L Q L Q L Q L Q L Q L Q L Q L GCCACTACCATG GCACTACCACTTG </th <th>CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACTGGC ACCGGACTGGC ACCGGCACTGGC CGTCTGGAC CGTCGGAC AGACCAGCCG CGTCGGAC K R Q Y ACCGCCAGTA TGGCGGTCAT TGGACGATATT GGTCGATATTGGAC Q N L T AAAATCTGAC Q N L T AAAATCTGAC GGTCGATATTGG GGTCGATATTGG CCAGTTATATGG S Y I AGCTCATATTG</th> <th>AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGT TGAACACCA I A I CATTGCTATT GTAACGATAA K I G M AAAATTGGGA TTTTTAACCCT M F D N TGTTCGGTAAC CATGGCATAT I E N CATGGCATAT I E N CATGGCATAT I CATGGCATAA ACAAGCTATT I E N CATGGCATAT CATGGCATAC TGTCGGCGCGC G G D V AAGGCGATGT TTGCGCGCC</th>	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACTGGC ACCGGACTGGC ACCGGCACTGGC CGTCTGGAC CGTCGGAC AGACCAGCCG CGTCGGAC K R Q Y ACCGCCAGTA TGGCGGTCAT TGGACGATATT GGTCGATATTGGAC Q N L T AAAATCTGAC Q N L T AAAATCTGAC GGTCGATATTGG GGTCGATATTGG CCAGTTATATGG S Y I AGCTCATATTG	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGT TGAACACCA I A I CATTGCTATT GTAACGATAA K I G M AAAATTGGGA TTTTTAACCCT M F D N TGTTCGGTAAC CATGGCATAT I E N CATGGCATAT I E N CATGGCATAT I CATGGCATAA ACAAGCTATT I E N CATGGCATAT CATGGCATAC TGTCGGCGCGC G G D V AAGGCGATGT TTGCGCGCC
	-1 1601 +1 1701 +1 1801 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2401 +1 +1 2501 +1 +1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACCCTGACCGC TGCGACTCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA CCCCGGCCGCA ATGCCGGCGCA ACCGCTAACT N C Q S CTGCCAACGC TACGGCCAAGC T L G ACGGTTAGCG ACGGTTAGCG ACGGTTAGCG CACGGTTCG T L G ACGGCTAGCG T L G ACGGCTAGCG T C Q N T T L G ACGGTAGCGGA T C Q N T T C TGCGGAAC	M G T A ATTAAAACTAT TATAAAACTAT TATAAAACTAT TATATATTGATAT AATGGATCAC TACCACTACAGE S S CCTGTAGCCC GGALACCAAC G G A A R TOCGGCGCCC A A R GTACCACCAGE G C A GTACCACTCAGG G T F GGTACGCCAGC CCATCCAGG TGGTGAAATG ACCACTTTAC K H A AAAACACGCCG T GGTGGAAACCGCCGA C T T GGTGGAATGAGE D G E M T ACCACTTTAC K H A AAACACCCCCG D S V D ACCACTTTAC S V T ACCACGTCGA T T TOTCCCACACCT S V T T T T CTTTOTGCACC S S R T <t< th=""><th>R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCCCCCC ATGCCCCCCCC ATGCCCCCCCC ATGCCCCCCCC ATGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC</th><th>T M L AGGTGTTTCC TCCACCAAAGC AAGTTCGGCC AAGTTCGGCC D Q I GATCAGATCC CTACTGCTAGG CCCTGACTCCACC GGGACTCGGC CACGTCCACT GTGCCGCGC D W R R ACTGGCGCGC D W R R ACTGGCGCGC V M C TGACCGCGCGC V M C TGACCGCGCGC V M C TGACCACCACA CACTCACACA H P A CATCGGCGCG V M C TGACCGCCGC V M C TGACCGCCGC C V M C TGACCGCCGC V M C TGACCGCCGCC V M C TGACCGCCCC C V M C TGACCGCCCC C V M C TGACCGCCCC C TGACGCCCCC C V M C TGACCGCCCCC V M C TGACCGCCCCC V M C TGACCGCCCCC V M C</th><th>Q I S M GGCCTGAAAA CCGGAACCTTG GAGCAACCTG CTCGTGGACC P S Q CTTCCGATCA GAAGCTAGC Q A A CAGCCCCCGG GTCCGGCGCC CTLATGCGGG GTCCGGCGCCC A P G S Q A A CAGCCCCCGG GTCCGGCGCCC A V N CCGCCAGCCCCG CCCCCGGGGAAG GCGCCCCTC A V N CCGCCCGCGGCAA GCGCCCCCC A C CCCCGGGCAA GCGCCCCCC A C CCCCCGGCGCA CCCCCCGCGGCAA CCCCCAACCT CCCCCCCCCC</th><th>GGAACTITTT GCATGAAAAA V N T L TGAATACATT ACTTATGTAA T À P ACCEGCEGCA T A P ACCEGCEGCA T A P ACCEGCEGCA GGCACCACCE GGCACCACCE AGCTACEGCEA AGACTACEGCA CCEGGEGCEA CCEGGEGCEA CCEGGEGCEA CCEGGEGCEA CCEGGEGCEA CCEGATEGC AGACCACCEGT TACCGATEGC AGACCACCEGT TACCGATEGC AGACCACCEGTACCE CTACCEGETCEC CTACCEGETCEC CTACCEGETCEC</th><th>ACCTTTTCGC TGGAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGC M Y K ATGGTGCGGAT GCACGGCT M Y K AATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCGGCA ATGCCCGGCA CCAAGCCC ATGCCCGGCA CCAACTCC AATGGTTAAA TTACCAATTC Y M P AATATTACGC ATGCCCGGCA CCAACTCC CCGCCTACGCT CTTATTACGG P R T L CGCCTACGCGA</th><th>$\begin{array}{c} \text{CTTCCCGTTT}\\ \textbf{GAAGGGCAAA}\\ \hline \textbf{R} \ \ \textbf{L} \ \ \textbf{A}\\ \hline \textbf{CCTCTGGCCC}\\ \text{GCLACACCGCG}\\ \text{GCLACACCGCG}\\ \textbf{S} \ \ \textbf{R} \ \ \textbf{P} \ \ \textbf{I}\\ \hline \textbf{CGCCCCGAT}\\ \text{GCGCGCCAAT}\\ \text{GCGGGGCTTA}\\ \hline \textbf{T} \ \ \textbf{L} \ \ \textbf{P}\\ \hline \textbf{T} \ \ \textbf{L} \ \ \textbf{T} \ \ \textbf{L} \ \ \textbf{T} \ \ \textbf{L} \ \ \textbf{T} \ \ \textbf$</th><th>CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACTGGC ACCGGACTGGC AGACGAGCCGG CGTCTGGAC CGTCTGGGCA AGACGCCG CGTCCGGACAT TGCGGGCATAT GCTCGGTATAT Q N L T AAAATCTGAC Q N L T AAAATCTGAC GGTCGATATA GGTCGATATATCG S Y I AGGTCAATATC CCAGTATATAG S Y I AGGTCATATATG</th><th>AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGATAA K I G M AAAATGGGA CATGGAGAAC TTTTAACCCT TTTTAACCCT TTTTAACCCT ACAACTATT CATGGAGAAC CATGGACAAC CATGGACAC CATGGACAC CATGGACAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGCCAC CATGGAC CATGGAC CATGAC CATGGAC CATGGAC CATGGAC CATGAC CATGGAC CATGGAC CATGGAC CATGAC CATGGAC CATGGAC CATGAC CATGAC CATGGAC CATGAC CATGGAC CATGAC CATGGAC CATG</th></t<>	R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCCCCCC ATGCCCCCCCC ATGCCCCCCCC ATGCCCCCCCC ATGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	T M L AGGTGTTTCC TCCACCAAAGC AAGTTCGGCC AAGTTCGGCC D Q I GATCAGATCC CTACTGCTAGG CCCTGACTCCACC GGGACTCGGC CACGTCCACT GTGCCGCGC D W R R ACTGGCGCGC D W R R ACTGGCGCGC V M C TGACCGCGCGC V M C TGACCGCGCGC V M C TGACCACCACA CACTCACACA H P A CATCGGCGCG V M C TGACCGCCGC V M C TGACCGCCGC C V M C TGACCGCCGC V M C TGACCGCCGCC V M C TGACCGCCCC C V M C TGACCGCCCC C V M C TGACCGCCCC C TGACGCCCCC C V M C TGACCGCCCCC V M C TGACCGCCCCC V M C TGACCGCCCCC V M C	Q I S M GGCCTGAAAA CCGGAACCTTG GAGCAACCTG CTCGTGGACC P S Q CTTCCGATCA GAAGCTAGC Q A A CAGCCCCCGG GTCCGGCGCC CTLATGCGGG GTCCGGCGCCC A P G S Q A A CAGCCCCCGG GTCCGGCGCCC A V N CCGCCAGCCCCG CCCCCGGGGAAG GCGCCCCTC A V N CCGCCCGCGGCAA GCGCCCCCC A C CCCCGGGCAA GCGCCCCCC A C CCCCCGGCGCA CCCCCCGCGGCAA CCCCCAACCT CCCCCCCCCC	GGAACTITTT GCATGAAAAA V N T L TGAATACATT ACTTATGTAA T À P ACCEGCEGCA T A P ACCEGCEGCA T A P ACCEGCEGCA GGCACCACCE GGCACCACCE AGCTACEGCEA AGACTACEGCA CCEGGEGCEA CCEGGEGCEA CCEGGEGCEA CCEGGEGCEA CCEGGEGCEA CCEGATEGC AGACCACCEGT TACCGATEGC AGACCACCEGT TACCGATEGC AGACCACCEGTACCE CTACCEGETCEC CTACCEGETCEC CTACCEGETCEC	ACCTTTTCGC TGGAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGC M Y K ATGGTGCGGAT GCACGGCT M Y K AATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCGGCA ATGCCCGGCA CCAAGCCC ATGCCCGGCA CCAACTCC AATGGTTAAA TTACCAATTC Y M P AATATTACGC ATGCCCGGCA CCAACTCC CCGCCTACGCT CTTATTACGG P R T L CGCCTACGCGA	$\begin{array}{c} \text{CTTCCCGTTT}\\ \textbf{GAAGGGCAAA}\\ \hline \textbf{R} \ \ \textbf{L} \ \ \textbf{A}\\ \hline \textbf{CCTCTGGCCC}\\ \text{GCLACACCGCG}\\ \text{GCLACACCGCG}\\ \textbf{S} \ \ \textbf{R} \ \ \textbf{P} \ \ \textbf{I}\\ \hline \textbf{CGCCCCGAT}\\ \text{GCGCGCCAAT}\\ \text{GCGGGGCTTA}\\ \hline \textbf{T} \ \ \textbf{L} \ \ \textbf{P}\\ \hline \textbf{T} \ \ \textbf{L} \ \ \textbf{T} \ \ \textbf{L} \ \ \textbf{T} \ \ \textbf{L} \ \ \textbf{T} \ \ \textbf$	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACTGGC ACCGGACTGGC AGACGAGCCGG CGTCTGGAC CGTCTGGGCA AGACGCCG CGTCCGGACAT TGCGGGCATAT GCTCGGTATAT Q N L T AAAATCTGAC Q N L T AAAATCTGAC GGTCGATATA GGTCGATATATCG S Y I AGGTCAATATC CCAGTATATAG S Y I AGGTCATATATG	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGATAA K I G M AAAATGGGA CATGGAGAAC TTTTAACCCT TTTTAACCCT TTTTAACCCT ACAACTATT CATGGAGAAC CATGGACAAC CATGGACAC CATGGACAC CATGGACAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGGAC CATGCCAC CATGGAC CATGGAC CATGAC CATGGAC CATGGAC CATGGAC CATGAC CATGGAC CATGGAC CATGGAC CATGAC CATGGAC CATGGAC CATGAC CATGAC CATGGAC CATGAC CATGGAC CATGAC CATGGAC CATG
	-1 1601 +1 1701 +1 1801 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2501 +1 2601	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAG T L S ACCCTGAGCG TGCGACTCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA CCCCGGCCGCA ATGCCGGCGCA ACCGCTAACT N C Q S CTGCCAAAGC T L G ACCGCTAACT N C Q S CTGCCAAGC CCCCAAGC CCCCCAACC CCCCAAGC CCCCCAACC CCCCCAACC C Q N T F ACACCACTAGCC Q N T F CTGTGGGAA	M G T A ATAAAACTAT TATTATGATAT TATTGGATGTC GACATCGAG G C S S CCTGTAGCTC GGACATCGAG G T P GGTACGCCGC CCATCGCGCCC G T P GGTACGCCAG CCATCGCGCCC CCATCGCGCCC CCATCGCGCCC G E M ACAACCACCG D S V D ACAACCGTCGA ACACCGTCGA CCACCTTAC K H A AAAACCGCCCG D S V D ACAACCGCCGCAC D S V D ACAACCGCCG GAAACACCGC GAAACACCGC GAAACACCGC GAAACACCGC CCACCGCCAC CCACCGCCGCAC D S V D ACAACCGCCGAC D S V D ACAACCGCCGAC CCACCGCCAC CCACCGCCGCCAC D S V D ACAACCGCCG GAAACACCGC GAAACACCGC GAAACACCCG CCACCGACCAC CCACCGCCAC CCACCGCCAC CCACCGCCAC CCACCGCCAC CCACCGCCAC CCACCCCCCCCCC	R K G CTGATGCTAC GACTACGATG GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGCGG G Y F X TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y AGCGCTGTAT CTGGCACGATGGC A G N TGCGGGAACGC A G N TGCGGGAACGC A G N TGCGGGAACGC CGACCTTGC N S G AACACCGCCC N S G AACACCGCCC N S G AACACCGCCC TTGCCCCCCCTTG	T M L AGGTGTTTCC TCCACCAAAGC AAGTTCGGGC AAGTTCGGGC D Q I GATCAGATCC CTACTGCTAGG CCCTGACGCC GGGACTCGGC CT S I CCCCTGACGCC GGGACTCGGC CACGCCCGCT CTCCCGCGC CAAGGGCGGC D W R R ACTGGCGCGC CAAGGGCGGC V M C TGACCACGCGCG V M C TGACCACGCGCC CACGCCCGCC V M C TGACCACGCGCC V M C TGACCACGCGCC C CACGCCCCCC TGTCACGCGCC C V M C TGACCACGCCCC C V M C TGACCGCCCC C TGTCCGCCGCC C CACGCCCCCC C CACGCCCCCCCCCC	Q I S M GGCCTGAAAAA CCGGAACCTTG GAGCAACCTG CTCGCGATGGAC P S D Q CTTCCGATCA GAATACGCC CTTATGCGG GAATACGCC CTTATGCGG GTCCGGGCCCC A P G S A A CCGCCGGGCAA GCGCCCCGG GTCCGGGCCCT A V N S A A TCTCCCGGTCCC A CGCCCGGGCCA S A A TCTCCCGCGTCC A CCGCAACCTTC ACCCGAACCGC ACCGCAACGCC ACCGCAACGCC ACCGCAACGCC ACCGCAACGCC ACCGCAACCTTC ACCGCAACCTC CCGCAACCTTCG ACCGCAACCTTCG ACCGCAACCTCC AGACGTTTGG	GGAACTITTI GGAACTITTI GCCTGAAAA V N T A T T ACTATACATT ACTATACATT ACTATACATT ACTATACATT ACCECCECT T T B V W S CCTGEGETCGC GGCACCACCG GGACTACGCCG A TCGATCGCGAC A TCTGACCCTG A ACCGCGGGGAC C P G CCCGGGGGGAC C AGACCACGCGT T T D AGACCACGCGT T T D TCTGGCCACGC T TACGGATCGCC A ATGCCTCACAGCGGT T D R GATCGCCACGGT G T D GGCCACGCCCC T T D GGCCCCCCCCC G T	ACCTTTTCGC TGGAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGC M Y K ATGGTGGCGGAT GCACGGCT M Y K AATGGTTAAA TTACCAATTT Y G T TACGGACGCA ATGCCCGGCA ATGCCCGGCA ATGCCCGGCA M N G GAATAACGGT TATATACGG N N G GAATAACGGT TATATACGG M N G GAATAACGGT CTTATTACGA	CTTCCCGTTT GAAGGGCAAA R L CGTCTGGCGC GCCCCCCCAT GCCCCCCCAT GCCCCCCCAT GCCCCCCCAT GCCCCCCCAT GCCCCCCCAT GCCCCCCCAT GCCCCCCAT GCCCCCCCAACCACCAC GCCCCCCCAACCACCACCACCACCACCACCCCACCAACCACCCC	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACCGG L S A TCTGTCGGCA AGACAGCCGG CGTCTGGAC CGCCTGCGGAC K R Q Y ACCGCCAGTA TGCGGCCAGTA TGCGGCCATAT GCTCGGATATA CGGACCATAT Q N L T AAAATCTGAC Q N L T AAAATCTGAC CCAGTATATCGGAC S Y I AGGTCAATATC CCAGTATATAG S Y I AGGTCAATATC CCAGTATATAG C S Y I AGGTCAATATC	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACCGCAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGATAA K I G M AAAATGGGA TTTTTAACCCT TGTTGGTATT GTAACGATAA K I G M AAAATGGGA TTTTTAACCCT I A I CATGGAGAAC CGTGGGGAGAC GTAGCGACTT I E N CATGGAGAAC CCTGGGGGAC GTAGCGACTT CATGGCGCCC GAACCGGCGG E G D V AAGGCGATGT TTCCGCCACA A P A CCCCCCCCCC
	-1 1601 +1 1701 +1 1801 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2501 +1 2501 +1 +1 2501 +1 +1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTAC TCGTCCGAAG T L S ACCCTGAGCG TGCAACTGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGGGCCGGCA CCCCGGCCGCA ACCCCTAACT N C Q S CTGCCGATGA ACCCGTAGCG ACCGCTAACT N C Q S CTGCCAAGCC CCCCAAGCC CCCCCAAGC T L G ACCGCTAGCC ACCGCTAGCC Q N T F AGAACACCTTT CCTGTGGGAA	M G T A ATAAAACTAT TATTATGATAT TATTAGATGTC TTACCTACAG A C S S CCTGTAGCTC GGACATCGAG G A A R TGCGGCGCGCGC G T F GGTACGCCAG CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGTC CCATCCGGCCCC CCATCCGGCCCC CCATCCGGCCCCC CCATCCGGCCCCC CCATCCGGCCCCCC CCATCCGGCCCCCC CAAACACTGG S C R TCCGGCCCCCCCC AGACCCTCGCA	R K G CTGATGCTAC GACTACGATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGCGG ATGCCGCGGG ATGCCGCGGGACG G V T H GCGTTACGCA CCCAATGCGT T V Y AGCGTCGCTGAT CTGGCACGATGGC A G N TGCGGGAACGC A G N TGCGGGACGCC A G N TGCGGGACGCC A G N TGCGGGCACCTTG N S G AACGCCTTGC N S G AACGCCCTTG N S G AACGCCCTTG N S G AACGCCCTTG N S G AACGCCCTTG N S G AACGCCCTTG N S G AACGCCCTTG N S G AACGCCCTTG CGCGGGACGC CGCGGCACCA G F L	T M L AGGTGTTTCC TCCACCAAAGC AAGTTCGGGC AAGTTCGGGC D Q I GATCAGATCC CTACTGCTAGG CCCTGACGCC GGGACTCGGC CT S I CCCCTGACGCC GGGACTCGGC CACGCCCAT GTTCCCGCGC CAAGGGCGGC D W R R ACTGGCGCGC CAAGGGCGGC V M C TGACCACGCGC TGACCACGCGC CAAGGCCGCC V M C TGACCACGCGCC CAAGGCCGCC V M C TGACCACGCGCC CACGCCCCC CAAGGCCCCC V Q N R TCCAGAACCG ACGTCTTGGC ACGTCTCGCCCC CAAGGCCCCC V Q N R TCCAGAACCG CAAGCCCTTGC CAAGCCCTCGC CAAGCCCTTGC CAAGCCCTTGC CAAGCCCTTGC	Q I S M GGCCTGAAAAA CCGGAACCTTG GAGCAACCTG CTCGCGACCTG P S D Q CTTCCGATCA GAAGCCTAGT A A CAGCCCCCGG GTCCGGGCCCC A P G S A A CAGCCCCCGGGGCC A P G S CCGCCGGGCCC CTATGCCGG GTCCGGGCCCC A P G S CCGCCGGGCAA GCGCCACCTC A A L R TCCGCCGCGCCA A C CCCCGGGCCA CCGCCGGGCCA CCGCCGGGCCA CCGCCGGGCCA C CCCCGGGCCA C CCCCGGGCCA C CCCCGGGCCA C CCCCGGGCCA C CCCCGGCGCA C CCCCGGGCCA C CCCCGGCGCA C CCCCGGCGCA C CCCCGGCGCA C CCCCGCGCCA C CCCCGCGCCA C CCCCGCGCCA C CCCCGCGCCA C CCCCGCGCCA C CCCCGCGCCA C CCCCGCGCCA C CCCCCGCGCA C CCCCCGCGCCA C CCCCCGCGCCA C CCCCCGCGCCA C CCCCCGCGCCC C CCCCGCGCCC C CCCCGCGCCC C CCCCGCGCCC C CCCCGCGCCC C CCCCGCGCCC C CCCCGCGCCC C CCCCGCGCCC C CCCCGCGCCC C C CCCCGCGCCC C C CCCCGCGCCC C C C C	GGAACTITTT GCATGAAAAA V N T L TGATACATT ACTTATGTAA ACTATAGTAA T À P ACCEGCEGCA T A P ACCEGCEGCA SECCEGEA V S CCETEGTCGC GGCACCACCE AGCTACGCCE AGACTAGGCA AGACTAGGGAC F W S CCEGEGEGCA AGACTAGGGAC AGACTAGGGAC TCTGGTCGCA AGACCACCGGT TA CGATGEGC CTAGCGATGEGC CTAGCGATCGC CTAGCGGCGCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGGCGCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCA CTAGCGCCCCA CTAGCCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCA CTAGCGCCCCCA CTAGCGCCCCA CTAGCGCCCCCA CTAGCGCCCCCA CTAGCGCCCCCA CTAGCCCCCCA CTAGCGCCCCCA CTAGCCCCCCA CTAGCGCCCCCA CTAGCGCCCCCA CTAGCGCCCCCA CTAGCGCCCCCA CTAGCGCCCCCA CTAGCCCCCCCA CTAGCGCCCCCCA CTAGCGCCCCCCA CTAGCGCCCCCCCCA CTAGCCCCCCCCCC	ACCTTTTCGC TGGAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGC GCGCGGCT GCACGGCT M V K ATGGTTAAA TTACCAATTT Y G T AATGGTTAAA TTACCAATTT Y G T TACGGGCACG ATGCCCGGCA ATGCCCGGCA CCTTATTACGG N N G GAATAACGGT TATTAACGG N N G GAATAACGGT TATTAACGG N N G GAATAACGGC M N G GAATAACGGC M N G GAATAACGGC TATTAACGG CTTATTACGC CTTATTACGC S R T CCCCCGCACA GAGGCCTACGCT S R T CCCCCGCACAC GAGGCCTACGCT	CTTCCCGTTT GAAGGGCAAA R L CGTCTGGGCG GCCCCCGC GCCCCCCCAT GCCCCCCCAT GCCCCCCCAT T L T L T L T L T L T L T L T L T L CGCCCCGATG GCCATCGACCAC GCCATCGAACA GCCGTCCATCT G K K CCATTTTGG K C ACCATTTGG C L Q L Q L C GGTGAACAACC CACTTTGG ACATTTGG Q I N K AGATTAACAA CCACTGCATG G Q Q L C Q Q I N GGTGAACAACC CCACTGCTG C Q Q E </th <th>CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACCGGC ACCGGCACGGC ACGCTGACCG CGACGAGCCG CGTCGGAC K R Q Y ACCGCCAGTA TGCGGCCATAT TGGATCATCTGAC CGGACCATAT CGGACCATAT CGGACCATATAC S Y I ACGGCCAATATC CCACTATATAG S Y I ACGGCCAATATC CCACTATATAG CCACTATATCGAC CCACTATATAG CCACTATATAG CCACTATATAG CCACTATATAG CCACTATATAG CCACTATATAC A Y V F CCTACCTCTT GGTACGACAA R S L</th> <th>AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACA I A I CATTGGTATT GTAACGATAA K I G M AAAATTGGA AAAATTGGA TTTTAACCCT I CATTGCTATT GTAACGATAA ACAAGCTATT I E N CATGGAGAAC CCTGGGGAGAC GTAGCGATGT I CATGCCATAA ACAAGCTATT I E N CATGGAGAAC CTTGGCGCCC GAACCGGCGGC E G D V AAGCCATGT TTCCGCTACA A P A CCCCCCCCCCCCC GCGGGCCCCC CCCGCCCCCCCC</th>	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACCGGC ACCGGCACGGC ACGCTGACCG CGACGAGCCG CGTCGGAC K R Q Y ACCGCCAGTA TGCGGCCATAT TGGATCATCTGAC CGGACCATAT CGGACCATAT CGGACCATATAC S Y I ACGGCCAATATC CCACTATATAG S Y I ACGGCCAATATC CCACTATATAG CCACTATATCGAC CCACTATATAG CCACTATATAG CCACTATATAG CCACTATATAG CCACTATATAG CCACTATATAC A Y V F CCTACCTCTT GGTACGACAA R S L	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACA I A I CATTGGTATT GTAACGATAA K I G M AAAATTGGA AAAATTGGA TTTTAACCCT I CATTGCTATT GTAACGATAA ACAAGCTATT I E N CATGGAGAAC CCTGGGGAGAC GTAGCGATGT I CATGCCATAA ACAAGCTATT I E N CATGGAGAAC CTTGGCGCCC GAACCGGCGGC E G D V AAGCCATGT TTCCGCTACA A P A CCCCCCCCCCCCC GCGGGCCCCC CCCGCCCCCCCC
	-1 1601 +1 1701 +1 1901 +1 2101 +1 2101 +1 2301 +1 2401 +1 2401 +1 2501 +1 2501 +1 2501	L Q R CTTAATTOTG GAATTAACAC AGCAGGCTTC TCGTCCGAAGC TCGTCCGACCGC TGCCAACCGC CCCCGACCGCC CCCCGCCGCT TTTTGAAGCA Y G P A ACCCCGGCCGCA CCCCGGCCGCA CCCCGGCCGCA ACCGCTAACT N C Q S CTGCCGATGGC ACCGCTAACT N C Q S CTGCCAAGC T L G ACCGCTAGCCG ACCGCTAGCCG C ACCGCTCAACT CTGCGCAAGCC C N T F ACCGCTAGCCG Q N T F ACCGCTAGCCG Q N T F ACCGCTAGCCG Q N T F ACCGCTAGCCG Q N T F ACCGCTAGCCG C N T F CCTATGCGAA C C C C AACCGCT C C C C AACC C C C C C C C C C C C C C C C C C C	M G T A ATTAAAACTAT TATAAAACTAT TATAAAACTAT TATATATTGATAT AATGGATCAC TACCACTACAGE S S CCTGTAGCCC GGALACCAAC G G A A R TOCGGCGCCCC A C S G T F GGTACGCCAC CCATCCGGTC G C M ACTATCAGG TGGTGAAATC ACCACTTTAC K H A AAAACACCCCC D G E M T ACTATCAGG TTTGTGGCAAC TTGGTGAAATC TTGGTGGAACC D S V D ACACCGTCGA TOTCCCACACCT F V T TTGTGGCACCCACC G R T TOTCCCACACCT F V T ACACCCTTACC G R T CTTTOTGCACC G R T GGAAACACCGCCG R <t< th=""><th>R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGCGG G Y F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y CGCCCTGTAT CTGGCACATG CGACCGTGTAT CTGGCACATGG A G N TGCGGGAACGC A G N TGCGGGACGCC A G N TGCGGGACGCC CGACCGTTGC CGCCCTTGC N S G AACACCGCCC N S G AACACCGCCC N S G AACACCGCCC N S G AACACCGCCC CTTGCCCGCCCTTG N S G AACACCGCCC CTTGCCCCCCCTTG N S G AACACCGCCC CTGCCCCCCTTG N S G AACACCGCCC CTTGCCCCCCCTTG CCCCCCCTTG CCCCCCCTTG CCCCCCCTTG CCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCCC</th><th>T M L AGGTATTCC TCCACCAAAGC AAGTTCGGGC AAGTTCGGGC D Q I GATCAGATCC CTACTGCTAGG CCCTGACGCC GGGACTCGGC CT S I CACCTCAT GTGGAGCTAA V F A CTTCCGCCGC CAAGGCCGCC V M C TGACCGCGCC V M C TGACCGCGCGC V M C TGACCGCCGCC V M C TGACCGCCCCC CAAGGCCCCC V M C TGACCGCCCCC C TGTCGACGCCCC C TGTCGACGCCCCC V M C TGACCGCCCCC V M C TGACCGCCCCC C TGTCGACGCCCCC V M C TGACCGCCCCC C TGTCGACGCCCCC C TGTCGACGCCCCC V M C TGACCGCCCCC C TGTCGACGCCCCC V M C TGACGCCCCC C TGTCGACGCCCC C TGTCGACGCCCCC C T TGACGCCCCCCCCC C TGTCGACGCCCCC C T T S I TGACCGCCCCC V M C TGACCGCCCCC C TGTCGCCCCCCCC C TGTCGCCCCCCCC</th><th>Q I S M GGCCTGAAAA CCGGAACCTTG CTCGTGGACC CTCGTGGACC CTCCGTGGAC CTTCCGATCA GAAGCCTAGT A A CAGCCCCCG GTCCGGCGCC CTTATGCCGG GTCCGGCGCCC A P G S A A CAGCCCCCGGGGAA GCGCCCCCG GTCCGGGGCC A V N CCGCCGCGGGAA GCGCCCCCG A C CCCCGGGGAA GCGCCCCTTA S A A TCTCCCCGCTTC ACCCCCACTTA S A A CCCCCACCTTA S A A TCTCCCCGCTTC ACCCCCACCTA</th><th>GGAACTITTI GCATGAAAAA V N T T GCATGAAAAA V N T T ACTATACATT ACTATAGTAA T A T A T A T A CCTTGATACATT ACTATAGTAA T A CCTGGCCCA GGCACCACCG ACCCACCCCG ACCCACCCCG CCCCGGCGCAC P G CCCCGGCGCACC F W AGACCACGCGT T D CCAGGACGCCC T D AGACCACGACGGT T D AGACCACGACGCGT T D CTACGGATCGCC D R CTACGGACGGCC T D AGACCACACCGG AGACCACGACGACCACCACG D R GATGCCACACCACGACG</th><th>ACCTTTTCGC TGGAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGA GCACCGGCT GCACGGCTA M V K AATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCGGCA ATGCCCGGCA ATGCCCGGCA CCAAGCCC ATGCCCGGCA CCAAGCCCGCA CCTTATTACGG N N G GAATAACGGT N N G GAATAACGGT CTTATTACGA</th><th>CTTCCCGTTT GAAGGGCAAA R L CCTCTCGCCC GCCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAAT GCCCCCCAAT GCCCCCCAACCACCAC GCCTCTCTT G K T L CCCATCGACCAAC GCCGTCCTTGT G K CCATTTTTGG G K GCTGAAAAACC CCACTTTTGG ACCATTTGG C Q L Q L CCACTCGACTAC Q L Q L Q L CCACTGCATG Q I N K GGTGAACGACGAC CCACTGCATG CCACTGCATG Q I N K GGTGAACGAAC CCACTGCATG CCACTGCATG Q Q E CCACCACGAAC <</th><th>CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACCGG L S A TCTGTCGGCA AGACAGCCGG CGTCTGGAC CGCCTGCGGAC K R Q Y ACCGCCAGTA TGCGGCCAGTA TGCGGCCAGTA TGCGGCCATAT GCTCGGTATATC Q N L T AAAATCTGAC CGGACCATAT GGTCGATATAC S Y I AGGTCAATATC CCAGTATATCG CACTGATATAC CCAGTATATAC S Y I AGGTCAATATC CCAGTATATAC CGATCATATAC CGATCATATAC</th><th>AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACAA K I G M CATTGGTATT GTAACGATAA K I G M AAAATTGGGA TTTTTAACCCT TTTTTAACCCT TTTTTAACCCT I A I CATGGAGAAC CGTGGGGAGCAC GTAGCGACACTT I E N CATGGAGAAC CATGGAGAAC CATGGCACTT I E N CATGGAGAAC CATGGCACTT CATGGCACCA CTTGGCGCCC GAACCGCGCGC E G D V AAGCGATGT TTCCGCCACA C D A CATCGCACAC</th></t<>	R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACCCCACCC ATGCCGCGG G Y F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y CGCCCTGTAT CTGGCACATG CGACCGTGTAT CTGGCACATGG A G N TGCGGGAACGC A G N TGCGGGACGCC A G N TGCGGGACGCC CGACCGTTGC CGCCCTTGC N S G AACACCGCCC N S G AACACCGCCC N S G AACACCGCCC N S G AACACCGCCC CTTGCCCGCCCTTG N S G AACACCGCCC CTTGCCCCCCCTTG N S G AACACCGCCC CTGCCCCCCTTG N S G AACACCGCCC CTTGCCCCCCCTTG CCCCCCCTTG CCCCCCCTTG CCCCCCCTTG CCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCTTG CCCCCCCCCTTG CCCCCCCCCC	T M L AGGTATTCC TCCACCAAAGC AAGTTCGGGC AAGTTCGGGC D Q I GATCAGATCC CTACTGCTAGG CCCTGACGCC GGGACTCGGC CT S I CACCTCAT GTGGAGCTAA V F A CTTCCGCCGC CAAGGCCGCC V M C TGACCGCGCC V M C TGACCGCGCGC V M C TGACCGCCGCC V M C TGACCGCCCCC CAAGGCCCCC V M C TGACCGCCCCC C TGTCGACGCCCC C TGTCGACGCCCCC V M C TGACCGCCCCC V M C TGACCGCCCCC C TGTCGACGCCCCC V M C TGACCGCCCCC C TGTCGACGCCCCC C TGTCGACGCCCCC V M C TGACCGCCCCC C TGTCGACGCCCCC V M C TGACGCCCCC C TGTCGACGCCCC C TGTCGACGCCCCC C T TGACGCCCCCCCCC C TGTCGACGCCCCC C T T S I TGACCGCCCCC V M C TGACCGCCCCC C TGTCGCCCCCCCC C TGTCGCCCCCCCC	Q I S M GGCCTGAAAA CCGGAACCTTG CTCGTGGACC CTCGTGGACC CTCCGTGGAC CTTCCGATCA GAAGCCTAGT A A CAGCCCCCG GTCCGGCGCC CTTATGCCGG GTCCGGCGCCC A P G S A A CAGCCCCCGGGGAA GCGCCCCCG GTCCGGGGCC A V N CCGCCGCGGGAA GCGCCCCCG A C CCCCGGGGAA GCGCCCCTTA S A A TCTCCCCGCTTC ACCCCCACTTA S A A CCCCCACCTTA S A A TCTCCCCGCTTC ACCCCCACCTA	GGAACTITTI GCATGAAAAA V N T T GCATGAAAAA V N T T ACTATACATT ACTATAGTAA T A T A T A T A CCTTGATACATT ACTATAGTAA T A CCTGGCCCA GGCACCACCG ACCCACCCCG ACCCACCCCG CCCCGGCGCAC P G CCCCGGCGCACC F W AGACCACGCGT T D CCAGGACGCCC T D AGACCACGACGGT T D AGACCACGACGCGT T D CTACGGATCGCC D R CTACGGACGGCC T D AGACCACACCGG AGACCACGACGACCACCACG D R GATGCCACACCACGACG	ACCTTTTCGC TGGAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGA GCACCGGCT GCACGGCTA M V K AATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCGGCA ATGCCCGGCA ATGCCCGGCA CCAAGCCC ATGCCCGGCA CCAAGCCCGCA CCTTATTACGG N N G GAATAACGGT N N G GAATAACGGT CTTATTACGA	CTTCCCGTTT GAAGGGCAAA R L CCTCTCGCCC GCCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAAT GCCCCCCAAT GCCCCCCAACCACCAC GCCTCTCTT G K T L CCCATCGACCAAC GCCGTCCTTGT G K CCATTTTTGG G K GCTGAAAAACC CCACTTTTGG ACCATTTGG C Q L Q L CCACTCGACTAC Q L Q L Q L CCACTGCATG Q I N K GGTGAACGACGAC CCACTGCATG CCACTGCATG Q I N K GGTGAACGAAC CCACTGCATG CCACTGCATG Q Q E CCACCACGAAC <	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGGACCGG L S A TCTGTCGGCA AGACAGCCGG CGTCTGGAC CGCCTGCGGAC K R Q Y ACCGCCAGTA TGCGGCCAGTA TGCGGCCAGTA TGCGGCCATAT GCTCGGTATATC Q N L T AAAATCTGAC CGGACCATAT GGTCGATATAC S Y I AGGTCAATATC CCAGTATATCG CACTGATATAC CCAGTATATAC S Y I AGGTCAATATC CCAGTATATAC CGATCATATAC CGATCATATAC	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACAA K I G M CATTGGTATT GTAACGATAA K I G M AAAATTGGGA TTTTTAACCCT TTTTTAACCCT TTTTTAACCCT I A I CATGGAGAAC CGTGGGGAGCAC GTAGCGACACTT I E N CATGGAGAAC CATGGAGAAC CATGGCACTT I E N CATGGAGAAC CATGGCACTT CATGGCACCA CTTGGCGCCC GAACCGCGCGC E G D V AAGCGATGT TTCCGCCACA C D A CATCGCACAC
	-1 1601 +1 1701 +1 1901 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2401 +1 2501 +1 2501 +1 2701 -1 2701 -1 -1 -1 -1 -1 -1 -1 -1 -1 -	L Q R CTTAATTOTG GAATTAACAC ACCAGCTTC TCGTCCGAAGC TCGTCCGACCGC TCCCACTCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A GGCGCCGCC CCCCGCCGCC TCCGCCGCC M F G ATGCCGGCG TGCGATGC ACCGCTAACT N C Q S CTGCCAACGCC TCCGACGCC CACGGTTCG T L G ACCGCTAACT N C Q S CTGCCAACGCC CACGGTTCG T L G ACCGCTAACT T L G ACCGCTAACT TCTTGTGGAA V D M V GAATATGTT CCTATACCAA ACCGTACGCC Q N T F ACCGCTAGCG CCCCAACGCC C N T S CTGCCAACGCC C N T S CTGCCAACCCC C N T S CCCCCAACCCCC C N T S CCCCCAACCCCCC C N T S CCCCCAACCCCCC C N T S CCCCCAACCCCCC C N T S CCCCCAACCCCCC C N T S CCCCCAACCCCCCC C N T S CCCCCAACCCCCC C N T S CCCCCAACCCCCCC C N T S CCCCCAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	M G T A ATTAAAACTAT TATAAAACTAT TATAAAACTAT TATATAACTAT TATAAAACTAT A AATGGATGTC TACACTACAGE G TTACCTACAGE A A TOCGGCGCCCGC G A A TOCGGCGCCCGC G T F GGTACGCCAG CCATCCAGG GAACACCACGC G D G E M T ACTATCAGG TGGTGGAATG ACCACTTTAC K H AACACCGTCGA ACCACTTTAC K H AAAACACGCCG D S V D ACAGCGTCGACGC S S R T T TOTGCGCAACCTGG AAACACGCCG ACACACTGG AAACACGCCG S S R T T TATTTTGCAACCGCGACGCCGACGCCG AGACACTGGG AAACCACTGG AAACACGCCGGAAGCGCG MACACCCCACG S R T Y ATATTTACTA TATTTAATTAACTA <th>R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACCCCCCCC ATGCCCCCCCC ATGCCCCCCCC ATGCCCCCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCC</th> <th>T M L AGGTACTACC TCCACCAAAGC AAGTICGGGC AAGTICGGGC D Q I GATCAGATCC CTACTACTAGG CCCTGACGCC GGGACTCGGC CACCTCATT GTGGAGCTAA V P A CACCTCACGCCG CAAGGCCGCC V M C TGACCGCGCGC V M C TGACCGCGCGC V M C TGACCGCCGCC V M C TGACCGCCGCC V M C TGACCGCCGCC V M C TGACCACCACTACACA ACATACACACACACA H P A CATCGGCGCCC V M C TGACCGCCGCC V M C TGACCGCCGCC V M C TGACCGCCGCC V M C TGACCGCCCCC V M C TGACCGCCCCC C ACGTACTCACA ACATACACACACACACACACACACACACACACACACACA</th> <th>Q I S M GGCCTGAAAA CCGGAACTTT GAGCAACCTG CTCGTGGACC CTCGCGACC CTCCGATCA GAAGCCTAGA A A CAGCCCCCG GTCCGGCCCC A V N CCGCCGGGCAAC CCCCCGGGGCAAC CCCCCGGGGCAAC CCCCCGGGGCAAC CCCCCGGGGCAAC CCCCCGCGGGCAAC CCCCCGCGGCCCA A A CCGCCCGCGCCAAC CCCCCGCGCCCAAC CCCCCGCGCCCCAAC CCCCCCGCGCCCAAC CCCCCCGCCCCCC A C V A L R TCCCCCCCCTTG ACCCCCAACCT CCCCCAACCTCC CCCCCGCGCCCCACC ACCCCCAACCTCC CCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCC</th> <th>GGAACTITTI GCATGAAAAA V N T T GCATGAAAAA V N T A T A T A T A T A T A T A T B W S W S W S CCCTGGETCGC GGCACCACCG A TCGATGCGCA A TCTGACCCCTG AGACCACGCGT T AGACCACGCCC T CAGACCACCCG T AGACCACGCCC T T G G T A GCCCACCACCG T T GCACGACCCCC T CCACGATGCCC T CAGACCACCACCGT</th> <th>ACCTTTTCGC TGGAAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGA GCACCGGCT GCACGGCTA M V K AATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCGGCA GAATGCCCGGCA M N G GAATAACGGT TATATACGG N N G GAATAACGGC TTATTACGG CTACTGCCAAAGTCC CTGTTCAGG P R T L CGCCTACGCTA CCTTATTACGG S R T CTCCCGCAAC GAAGGCCTG GAAGGCCTG GAAGGCCTG GAAGGCCTG GAAGGCCTG CTACCGCAAC</th> <th>CTTCCCGTTT GAAGGGCAAA R L CGTCTGGCGC GCCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAACCACCAC CCCTTCTGT G K T L CCCATCGACCAC GCCCCCGAACA GCCCACGAACA CCCATTGTG G K ACATTTGG L Q L CCACTTTTGG GCCACGACGACC CCACTTACAA ACATTACAA CCACTGCATG Q Q E CCACTGCTGCT N GGTGACGAACA CCCACTGCTGCTC N GCCACTGCACGACA Q Q E CACACTACGACAA CCACTGCTTCC N CCACTGCTTC Q Q E CCACTGCACGACAAC QCCACTGCTTGC A A C <tr< th=""><th>CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGG CGTCTGGAC CGTCTGGAC K R Q Y ACCGCCAGTA TTGCGGTCAT TGCTGGGTCAT AACCGCAGTA TGCGGCCATAT GCTCGGTATAAC Q N L T AAAATCTGAC CGGACCATAT TTTTTAGACTG CCAGTATATCG S Y I AGGTCAATATC CCAGTATATAG S Y I AGGTCAATATC CCAGTATATAG C S Y I AGGTCAATATC CCAGTATATAC A Y Y F CCTACGTCTT GGATCAGCAG C Y T C</th><th>AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACAA K I G M CATGGATAAT GTAACGATAA K I G M AAAATTGGGA TTTTTAACCCT TTTTTAACCCT TTTTTAACCCT I A I CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAC CATGGAGAC CATGGCGACG CGACCGGCGCG CGGCCCGGCG CGCGCGCGC CTTCGACGAC CTTCGACGAC CTCGACGAC CTCCGACGAC CTCCGCGCC CGCCCGCGC CGCCCGCGC CTCCACGCGCC CTCCGACGAC CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCCGCC CTCCGCACG CTCCGCCGCC CTCCGCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCCCCCCC CTCCCCCCCC CTCCCCCCCC</th></tr<></th>	R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACCCCCCCC ATGCCCCCCCC ATGCCCCCCCC ATGCCCCCCCC CCCCCCCCCCCCCCCCCCCCCCCCCCCC	T M L AGGTACTACC TCCACCAAAGC AAGTICGGGC AAGTICGGGC D Q I GATCAGATCC CTACTACTAGG CCCTGACGCC GGGACTCGGC CACCTCATT GTGGAGCTAA V P A CACCTCACGCCG CAAGGCCGCC V M C TGACCGCGCGC V M C TGACCGCGCGC V M C TGACCGCCGCC V M C TGACCGCCGCC V M C TGACCGCCGCC V M C TGACCACCACTACACA ACATACACACACACA H P A CATCGGCGCCC V M C TGACCGCCGCC V M C TGACCGCCGCC V M C TGACCGCCGCC V M C TGACCGCCCCC V M C TGACCGCCCCC C ACGTACTCACA ACATACACACACACACACACACACACACACACACACACA	Q I S M GGCCTGAAAA CCGGAACTTT GAGCAACCTG CTCGTGGACC CTCGCGACC CTCCGATCA GAAGCCTAGA A A CAGCCCCCG GTCCGGCCCC A V N CCGCCGGGCAAC CCCCCGGGGCAAC CCCCCGGGGCAAC CCCCCGGGGCAAC CCCCCGGGGCAAC CCCCCGCGGGCAAC CCCCCGCGGCCCA A A CCGCCCGCGCCAAC CCCCCGCGCCCAAC CCCCCGCGCCCCAAC CCCCCCGCGCCCAAC CCCCCCGCCCCCC A C V A L R TCCCCCCCCTTG ACCCCCAACCT CCCCCAACCTCC CCCCCGCGCCCCACC ACCCCCAACCTCC CCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCC	GGAACTITTI GCATGAAAAA V N T T GCATGAAAAA V N T A T A T A T A T A T A T A T B W S W S W S CCCTGGETCGC GGCACCACCG A TCGATGCGCA A TCTGACCCCTG AGACCACGCGT T AGACCACGCCC T CAGACCACCCG T AGACCACGCCC T T G G T A GCCCACCACCG T T GCACGACCCCC T CCACGATGCCC T CAGACCACCACCGT	ACCTTTTCGC TGGAAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACCGCT CCGTGGCGGA GCACCGGCT GCACGGCTA M V K AATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCGGCA GAATGCCCGGCA M N G GAATAACGGT TATATACGG N N G GAATAACGGC TTATTACGG CTACTGCCAAAGTCC CTGTTCAGG P R T L CGCCTACGCTA CCTTATTACGG S R T CTCCCGCAAC GAAGGCCTG GAAGGCCTG GAAGGCCTG GAAGGCCTG GAAGGCCTG CTACCGCAAC	CTTCCCGTTT GAAGGGCAAA R L CGTCTGGCGC GCCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAT GCCCCCCAACCACCAC CCCTTCTGT G K T L CCCATCGACCAC GCCCCCGAACA GCCCACGAACA CCCATTGTG G K ACATTTGG L Q L CCACTTTTGG GCCACGACGACC CCACTTACAA ACATTACAA CCACTGCATG Q Q E CCACTGCTGCT N GGTGACGAACA CCCACTGCTGCTC N GCCACTGCACGACA Q Q E CACACTACGACAA CCACTGCTTCC N CCACTGCTTC Q Q E CCACTGCACGACAAC QCCACTGCTTGC A A C <tr< th=""><th>CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGG CGTCTGGAC CGTCTGGAC K R Q Y ACCGCCAGTA TTGCGGTCAT TGCTGGGTCAT AACCGCAGTA TGCGGCCATAT GCTCGGTATAAC Q N L T AAAATCTGAC CGGACCATAT TTTTTAGACTG CCAGTATATCG S Y I AGGTCAATATC CCAGTATATAG S Y I AGGTCAATATC CCAGTATATAG C S Y I AGGTCAATATC CCAGTATATAC A Y Y F CCTACGTCTT GGATCAGCAG C Y T C</th><th>AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACAA K I G M CATGGATAAT GTAACGATAA K I G M AAAATTGGGA TTTTTAACCCT TTTTTAACCCT TTTTTAACCCT I A I CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAC CATGGAGAC CATGGCGACG CGACCGGCGCG CGGCCCGGCG CGCGCGCGC CTTCGACGAC CTTCGACGAC CTCGACGAC CTCCGACGAC CTCCGCGCC CGCCCGCGC CGCCCGCGC CTCCACGCGCC CTCCGACGAC CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCCGCC CTCCGCACG CTCCGCCGCC CTCCGCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCCCCCCC CTCCCCCCCC CTCCCCCCCC</th></tr<>	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGG CGTCTGGAC CGTCTGGAC K R Q Y ACCGCCAGTA TTGCGGTCAT TGCTGGGTCAT AACCGCAGTA TGCGGCCATAT GCTCGGTATAAC Q N L T AAAATCTGAC CGGACCATAT TTTTTAGACTG CCAGTATATCG S Y I AGGTCAATATC CCAGTATATAG S Y I AGGTCAATATC CCAGTATATAG C S Y I AGGTCAATATC CCAGTATATAC A Y Y F CCTACGTCTT GGATCAGCAG C Y T C	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACAA K I G M CATGGATAAT GTAACGATAA K I G M AAAATTGGGA TTTTTAACCCT TTTTTAACCCT TTTTTAACCCT I A I CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAAC CATGGAGAC CATGGAGAC CATGGCGACG CGACCGGCGCG CGGCCCGGCG CGCGCGCGC CTTCGACGAC CTTCGACGAC CTCGACGAC CTCCGACGAC CTCCGCGCC CGCCCGCGC CGCCCGCGC CTCCACGCGCC CTCCGACGAC CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCACG CTCCGCCGCC CTCCGCACG CTCCGCCGCC CTCCGCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCCCCCCC CTCCCCCCCC CTCCCCCCCC
	-1 1601 +1 1701 +1 1901 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2501 +1 2601 +1 2601 +1 2701	L Q R CTTAATTOTG GAATTAACAC ACCAGGCTTC TCGTCCGAAC TCGTCCGACCGAC CCCCGACTCGC M P G ACCCTGACTCGC M P G ACCCGGCGCA ATCCCGGCCGCA CCCCGGCCGACCGC ACCGCTAACT N C Q S CTGCCAACGCC ACCGCTAACT N C Q S CTGCCAACGCC CCACCGCC Q N T F ACCGCTAGCGA ACCGCTAGCGA CCCGCAACGCC Q N T F CTGCGATCGCA CCCACACGCCA ACCGCTACCAAC CCCACACGCC ACCGCTACCAAC CCCACACGCC ACCGCTACCAAC CCCACACGCC ACCGCTACCAAC CCCACACGCCAAC CCCACACGCCAAC CCCACACGCCAAC CCCACACGCCAAC CCCACACCAAC CCCACACCAAC ACCGCTACCAAC CCCACACCAAC CCCACACCAC ACCGCCAACCCAAC	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG GCTGTAGCTC GGACATCGAG G A A R TOCGGCCGCCGC G A A R TOCGGCCGCGCGC G T F GGTACGCCGCCGC CCATGCGCCG G T F GGTACGCCAG CCATGCGCCGCCCC CCATGCGCCC CCATGCGCCCC CCATGCGCCCC CCATGCGCCCCCCCCC CCATGCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACGCCACCC ATGCGCGCGG G Y F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y CGCGCACGT CGCCACTGTAT CTGGCACGTAC CGCCACGGCACCA AGCGCTCGTG CGCCGCACGG A G N TCGCAGCACGG CGACCGTCGCC CGCCCTTGC CGCGCACCAC CGCCCTTGC CGCGCCCCG A G N TGCCGGAAAC ACGCCCTTG TGCCGGGACCA CGCCCTTGC CGCGCCCCG TTGCCCGCGC TGCCGCGCCCC CGCCCTTTTC CCGCCGCCCC CGCCCTTTTC CCGCCGCCCC CGCCCCTTG TGCCCGCCCCC TTGCCCCCCC TGCCCCCCCC CGCCCTTTTC CCGCCCCCCC CGCCCCTTG CCCCCCCCCC	T M L AGGTATTTCC TCCACCAAAGC AAGTTCGGCG AAGTTCGGCG D Q I GATCAGATCC CTACTGCTAGG CCCTGACGCC GGGACTCCGG T S I CACCTCCATT GTGGAGCTAA V P A ACTGGCGCCC D W R R ACTGGCCCCC D W R R ACTGGCCCCC T GTCCCGCCG CAAGGCCGCCC V M C TGACCGCGCGC V M C TGACCGCGCGC V M C TGACCGCGCGC V M C TGACCGCGCGC V M C TGACCGCGCCC V M C TGACCGCCGCC C T GTCCCGCCGC C CAAGGCCCCCC V M C TGACCGCCCCC V M C TGACCGCCGCC V M C TGACCGCCCCC V M C TGACCGCCCCC V M C TGACCGCCCCC V M C TGACCGCCCCC V M C TGACCGCCCCC C T TGACCGCCCC C C C D V R R C TGCCACTATGC C ACGTCCGCCCC C C D N R C TGCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	Q I S M GGCCTGAAAA CCGGAACTTT GAGCAACCTG CTCGTACGAC GAAGCCAACTT R A A GAATACGCC CTTAGGCG GAATACGCC CTTAGGCG GCGCCCGGAAC CAGCCCCGG GCGCCCCGGAAC CAGCCCCCGG GCGCCCCGCAAC A V N CGCGGTGAAC CGCCGCGCCCTA S A A TCTCCCCCGTT ACCACCTTCG ACCCCCCCCCCCCCCCC A C V N CGCCCTTCGAACC CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GGAACTTTT GCATGAAAA V N T T CCTTGAATACATT ACTATACATT ACTATACATT ACTATACATT ACTATACATT ACTATACATT ACTATACATT ACTATACATT ACTATACATT ACTATAGTAA Y A T A P Y W S CCTGGETCGC GGCACCACCG L T ACGTACGGCA ACTACCGCGC AGACTACGCGC CCCGGGCGGA AGACCACCGCT G T D AGACCACGCGT T AGCCACGACCGC TACGATACCG D R Q AGTCGCACGCGT T AGGCTACCGATGGC AGGCTACCGG D R Q AGTCACACGATGG AGCCACGCGT AGGCTACCGCGC AGGCTACCGGC AGGCTACCGCGCC A AGGCTACCGCAGCC AGGCCACAC	ACCTTTTCGC TGGAAAAGCG G T A GGCACCGCTT CCGTGGCGAA GG T A GGCACCGCT CCGTGGCGGAT GCACCGGCT CCGTGGCGGAT GCACCGGCT ATGCCTGGCGAC ATGCCCTGGC K Y M P AATGTTAA TTACGAATTT Y G T TACGGACCC ATGCCCTGGC K Y M P AATGTTAA ATGCTTAA TACGGACCC M N G GAATAACGGC TTATTACGG CAAAGTCC CTGCTGCCGCAC GGCATGCGA A GGCATGCCAAAGTCC CTGCCGGCAC GGCGCCTGCCT GCGCACCCTTC CTCCCGGCAC A K P W	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCCC GCACACCGCC S R P I CGCCCCCCAT GCCGCCCCAT GCCGCCCAT GCCGCCCAT GCCCCCAT GCCCCCAACA CCCTTGT G E K P TGGTAAACCC CACTTTTGG G K P TGGTAAACCC CACTTTTGG G K P TGGTAAACCC CACTTTTGG G K P TGGTAAACCC CACTTTTGG Q I N K ACATTACAA CTAATACAA CTAATACAA CCACTGCTG Q I N K AGATTAACAA CCACTGCTG GACGTCGACCAC CCCCCGCTGC GCCCCCCTTG CCCCCCCTTG CCCCCCCTTG CCCCCCCTTG CCCCCCCTTGC CCCCCCCTTGC CCCCCCCTTGC CCCCCCCTTGC CCCCCCCTTGC CCCCCCCTTGC CCCCCCCTTGC CCCCCCCC	CGTTCAACTT GCAACGTGAAC L A L A TCGCCACTGGC ACCGTGACCGC L S A TCTCTCGGCA AGACAGCCGG CGTCTGGAC CGTCTCGGAC K R Q Y ACCCCCAGTA TGCGGCCAGTA TGCGGCCATAT GCTCGGTATAAC AACGCCAGTA TTGGACAATATC CGACCAATATC CCACTGATATAC CCACTGATATAC CCACTGATATAC A Y Y F CCTCACGTCT TGGATCAATATC CGACCATATAAC A Y V F CCTCACGTCT GGGTCCAATATC CGACCATATAAC A Y V F CCTCACGTCT GGGTCCACAA R S L CCGCTCCGCTG GGCCACCCCC GGCCACCCCCC GGCCACCCCCCCG GGCCACCCCCCCG GGCCACCCCCCCC	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACCGCAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACAA K I G M CATGGATAAT GTAACGATAA K I G M AAAATTGGA K I G M F D N TGTTCGATAAT CATGGAGAAC CTTGGAGAAC CTTGGAGAAC CTTGGCGCGC GTAACCTTG L G R Q CTTGGCGACAC CTTGGCGCGCC GAACCGCGCG CGCGCCGGCC CTACGCGCGC CTACGCGTC CTCGGCGCCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCCGCCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCCGCCCC CTCCCGCCC CTCCGCCCCC CTCCCGCCC CTCCGCCCCC CTCCGCCCCC CTCCCGCCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCGCCC CTCCCCCCCC CTCCCCCCCC CTCCCCCCCC CTCCCCCCCC
	-1 1601 +1 1701 +1 1901 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2501 +1 2601	L Q R CTTAATTOTG GAATTAACAC ACCAGCTTC TCGTCCGAAC TCGTCCGACCGC TCCCACCGCC CCCACCGCC TTTTTTTAACAC M P G ACCCTGACTCGC M P G ACCCCGCCGCC CCCCGCCGCC M A I E GGGCCCGCAACCGC TCCGCCAACCGC TCCGCCAACCGC TCCGCCAACCGC TCCGCCAACCGC TCCGCCAACCGC TCCGCCAACCGC TCCGCCAACCGC CCCCCAACCGCC Q N T E ACCGCTAGCCG ACCGCTAGCCG CCCACACGCC Q N T E ACCGCTAGCCG CCCACACCGC Q N T E ACCGCTAGCCG CCCACACCGC CCCACACCCA CCCACACCCA CCCACACCCA CCCACACCCA CCCACACCCA CCCACACCCA CCCACACCCA CCCACACCCA CCCACCCA	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG GACATCGAG A C S S CCTGTAGCTC GGACATCGAG G A A R TOCGGCCGCCCC G C T F GGTACGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCGCAG TGGTGAAATG ACCACTTAC K H A AAAACCCCCC D S V D ACACCGTCGA CCACTTACC K H A AAAACCCCCC D S V D ACACCCTCAG TTGTGCGCAGCT CTTGTGCGCACT CCCACCCCCCCCCCC S G R TCCCGCAGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACGCCACCC ATGCGCGCGG G Y F A TATTTTGCCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y CGCCGTACGCA CGCACTGCTG CGCCGCTGTAT CTGGCACGATCGC A G N TCGCAGCATCGC CGCCACTGGC CGCCGTCACC A G N TCGCAGCATCGC CGCCGTCACC A G N TCGCAGCATCGC CGCCGCTCA CGCCCTTAGCCCC CGCCCTTACC CGCCCCTGCT CCCGGAAAC CGCCCTTGCCCCGC A G N TGCCGGCACCA CGCCCTTGCCCCGC CCCCCGCACA CGCCCTTGCCCCCC CCCCCCTTACC CCCCCCTCACC A G N TGCCGGCACCAC CGCCCTTGCCCCCC CCCCCCTTACC CCCCCCTCACC CCCCCCTCACC CCCCCCCCCC	T M L AGGTGTTTCC TCCACCAAAGC AAGTTCGGCG AAGTTCGGCG D Q I GATCAGATCC CTACTGCTAGG CCCTGACGCC GGGACTCGGG T S I CACCTCGATT GTGGAGCTAA V F A CATCGCCGC D W R R ACTGGCGCCG D W R R ACTGGCGCCG D W R R ACTGGCCGCG CAAGGCCGCC V M C TGTACCGCGCG CAAGGCCGCC V M C TGTACCGCGCG CAAGGCCGCC V M C TGTACCGCGCG CAAGGCCCCC CAAGGCCCCC CAAGGCCCCC CAAGGCCCCC CAAGGCCCCC CAAGGCCCCC CAAGGCCCCC C T D N C TGTACAGAACCG CAACGCCTTGGC ACGTCTGGATAAC CCAAGCTATG D S V I C S V I C S V I C S V I C S V I C CACCCCACTGTTG C CAACGCTACTGC CAAGCCTCGCC CAAGCCCCCC C CAAGCCCCCC C C N C C C C N C C C C C N C C C C	Q I S M GGCCTGAAAA CCGGAACTTT GAGCAACCTG CTCGTTGGAC CTCGTTGGAC CTTCGGATCA GAATACGGCC CTTATGGCGG GCCCCGCGCCCGG GTCCGCCGCGCCC A P G S CAGCCCCCGG GCCCCGCGCCCCG A P G S CCCCCGCGCCCCG A P G CCCCCGCGCCCC A V N CCCCGCGCCCCC A CCCCCGCCCCC A CCCCCGCCCCCC A CCCCCCCCCC	GGAACTTTT GCATGATAAA V N T ACATTACATA ACTTATACTAA T A T A T A T A T A T A T A T B W S CCGTGGTCGC GGCACCACCG L T T T ACGATGGGAC A TCGATGCGCA A ACGATGGGAC C AGACTACGCGC A AGACCACCGT T AGACCACGCGT T T D AGACCACGACGGT C D R CALGCATGCCACCGT T AGGTCACCACGGT T A A AGGTCACCACCGT A AGGTCACCACCGT A AGGTCACCACCGT A AGGTCACCACCGT A AGG	ACCTTTTCGC TGGAAAGCG G T A GGCACGGTTCC TAGCCAAAGG G T A GGCACGGCT CCGTGGCGGAC GCACGGCT AATGGTTAAA TTACCAATTT Y G T TACGGACGCA ATGCCTGGC K Y M P AATGTTAAA TTACCGACGCT ATGCCCGGCC K Y M P AATGTTAAA TTACGGACG N N G GAATAACGGC TTATTACGG CAAAGTCC CTTATTGCAA	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCCG GCAGACCGCG S R P I CGCCCCCGAT GCCGCCCGAT GCCGCGGGTTA T L P T L P TACCTTGCCC ATGGGACGGG R T N CGCCGACGAACA GCGTGCTTGT G E K P TGGTAAACCC CACTTTTTGG G K P TGGTAAACCC CACTTTTTGG CACTCGATG ACCATTACAAA CCATTACCAA CCACTTACGAGTAA CCACTGCTG Q I N K AGATTAACAA CCACTGCTG GCCCCCCTGCTG CCCCCCTGCTGC CCCCCCTGGCGCCCCCCCTGC GCCCCCCTTG CCCCCCTTGG CCCCCCCTTG CCCCCCCTTGC CCCCCCTTGC CCCCCCTGCACCC CCCCCCCTGC	CGTTCAACTT GCAAGTGAA L A L A TGGCACTGGC ACCGTGACCG L S A TCTGTCGGCA AGACAGCCGG CGTCTGGAC CGTCTGGAC K R Q Y ACCCCCAGTA TTGCGGCCATAT TGCGGCCATAT GCTCGGTATAAC AACGCCACATAT CGGACCATAT GCTCGGTATATAC S Y I ACGGCCATATATC CCACTATATCG S Y I ACGGCCATATAC CACTCAATATC CCACTATATAC S Y I ACGTCCAATATC CCACTATATAC S Y I ACGTCCAATATC CCACTATATAC CA Y Y F CCTCGCTCGAC R S L CCGCTCCGCTG GGCCAGCGAC Y I S	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGGT TGAACGACAA K I G M CATGGATAAT GTAACGATAA K I G M AAAATGGGA TTTTTAACCCT M F D N TGTTCGATAAT AAAATGGGA ACTTCGTTAT GTAACGATAA AAAATGGGA CATGGAGAAC CTTGGCGCGCC GTAGCGACACTT I E N CATGGAGAAC CATGGCATT I E N CATGGAGAAC CATGGCATAT CATGGCATAT CATGGCATAT CATGGCATAT CATGGAGAAC CTTTCACCTTTG CATGGCATAT CATGGAGAAC CTTCGCGCGCC GCACCCGCGCC GCGCCCGGCC CTACCGCGCC CTTCCGCCGCC CTTCCGCCGCC CTCCGCCCC CTCCGCCGCC CTCCGCCGCC CTCCGCCCCC CTCCCGCCC CTCCCGCCCC CTCCCGCCCC CTCCCGCCCC CTCCCGCCCC CTCCCGCCCC CTCCCGCCCC CTCCCGCCCC CTCCCCCCCC CTCCCCCCCC CTCCCCCCCC
	-1 1601 +1 1701 +1 1901 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2501 +1 2601	L Q R CTTAATTOTG GAATTAACAC ACCAGGCTTC TCGTCCGAAG TCGTCCGACGGC TCCGACTCGC K N F V AAAACTTCGT TTTTGAAGCA V G P A ACCCGGCGCA CCCCGGCCGCA CCCCGGCCGCA CCCGGCCGACTGCC M P G ATGCCGGCGCA ACCGCTAACT N C Q S CTGCCAACGCC CACCGCTAACT N C Q S CTGCCAACGCC TCTGTGGGAT CCGACTAGCCA ACCGCTAACCA CCGCCAACGCC Q N T F ACCGCTAGCGA CCGCAACGCCT CCTGCGATCGCA CCCACACGCC Q N T F CCTACCACACGCC Q N T S CTGCCAACGCCT CCTACCCACGCA CCCACACGCC Q N T S CCGCCAACGCCA CCCACACGCCA CCCACACGCCA CCCACCACCAC CCCACCACCAA CCCACCCAC	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG GACATCGAG G S S CCTGTAGCTC GGALACCGAG G T F GGTACGCCGC CCATGCGCCG G T F GGTACGCCAG CCATGCGCCG CCATGCGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCCCAG CCATGCCCAG CCATGCCCAG CCATGCCCAG CCATGCCCAG CCATGCCCAG CCAGCCCAG S G R TCCGGCAGCCCAG S G R TCCGGCAGCCCAG CCATGCCCAG CCATGCCCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCATGCAG CCACGCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCCAG CCACCAG CCACCCAG CCACCAG CCACCCAG CCACCAG CCACCAG CCACCCAG CCACCAG CCACCAG CCACCAG CCACCAG CCACCACCCAG CCACCACCAG CCACCACCAG CCACCACCAG CCACCACCACCAG CCACCACCACCACCACCACCACCACCACCACCACCACCA	R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACGCCACCC ATGCGCGCGC G Y F A TATTTTGCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y CGCGACATGGC CGCACTGTAT CTGGCACGATGGC A G N TCGCAGGCACCA A G N TCGCAGCATCGC CGACGTCGCTG CGCACCATCGC CGCCCTTGC ACGGCCCTGG TGGCCGCGCCC A G N S G AACACCGCCC CGACACCGCC CGCCGTAACC CGCCGTCGTC CGCGCACCAAC CGCCCTTGC CGCCGCCCCC A G N S G AACACCGCCC TGGCCGCCCC TGGCCGCCCC CGCCGCAAAAC CGCCCTTGC CCCCGGAAAC CCCCCGCAACAC CGCCCTTGC CCCCCGCAACAC CCCCCCCCC CCCCGCAACAC CCCCCCCCCC	T M L AGGTACTACC TCCACCAAAGC AAGTTCGGCC AAGTTCGGCC D Q I GATCAGACC CTACTACTAGG CCCTGACACCC GGGACTCCGG T S I CCCCTGACCC GGGACTCGGC CACCTCCATT GTGGAGCCTAA V P A CATCCGCCGC CAAGGCCGCC V M C TGACCGCGCGC CAAGGCCGCC V M C TGACCGCGCGC V M C TGACCGCGCGC CAAGGCCGCC V M C TGACCGCGCGC CAAGGCCGCC V M C TGACCGCGCCC V M C TGACCGCGCGC C TGTCCGCCGCC CAAGGCCGCC C V M C TGACCGCCGCC C TGTCCGCCGCC C C TGTCCGCCGCC C C D N C TGACCGCCCC C T D N C TGACCGCCCC C C D N C TGTCGATAGC C CACCTCCACCCCC C C C C C C C C C C C	Q I S M GGCCTGAAAAA CCGGAACTTT GAGCAACCTG CTCGTAGACC CTCGTAGACC GAAGCCTAGA P S Q CTTCGGATCA GAATACGGC CTTATGGCGG GCGCCCGGAAC GCGCCCGGAAC CAGCCCCGG GCGCCCGCGAAC CCCCCGGCGCC A V N CGCGGCAACTTC S A A CCCCCGGCGCC A C C CCCCGCGCGAAC CCCCCGCGCCCA CCCCCGCCCC A C C CCCCCGCCCC A C C CCCCCGCCCC A C CCCCCGCCCCC A C CCCCCGCCCCC A C CCCCCGCCCCC A C CCCCCCCCCCC A C CCCCCCCCCC	GGAACTTTT GCATGATAAAA V N T T CCTTGAATACATT ACTTACATA T A T A T A T A T A T A T B W S W S CCTGGETCGC GGCACCACCG L T TCGATGCGCA A TCGATGCGCA GGCCACCACCG AGACTACGGCGAC C F G G CCCGGGCGGAA AGACCACGCGT T D G AGACCACGACGGT T AGACCACGCGT T D G AGGCTCACCACGGT C AGACCACGCGT T D G AGACCACGACGGT A AGCCACGACGGT CAGGCTCCCACCGT A A AGGTCACCACGGT A A AGGTCTCAC	ACCTTTTCGC TGGAAAAGCG G T A GGCACCGCTT CCGTGGCGAA GG T A GGCACCGCT CCGTGGCGGAT GCACCGGCT CCGTGGCGGAT GCACCGGCT M V K ATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCTGGC K Y M P AATGTTAAA TTACCGAACT TACGGCACG M N G GAATAACGGC TTATTACGG CAAAGTCC CTTATTGCAA CTTATTACGG CAAAGTCC CTGCTTCCGGAC GAGGCCTTG CTCCCGGACA A K P W GCAAACGTCG GGTTGGCAC A K P W CCAACCGTA GGTTGGCACC	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCCC GCACACCGCC S R P I CGCCCCCCAT GCCCGCGGGTTA T L P T L P TACCTTGCCC ATGGGACGGG R T N CGCCGACGAACA GCGTGCTTGT G E K P TGGTAAACCC CACTTTTGG G K P TGGTAAACCC CACTTTTGG G K P TGGTAAACCC CACTTTTGG Q I N K ACATTACAA CTAATACAAT CTAATACAAT CTAATACAAT CCACTGCTG Q I N K AGATTAACAA CCACTGCTG GCCCCCCCTGC GCCCCCCTGCTGC CCCCCTGGCACCC CCCCCTGGCACCC CCCCCTGGCACCC CCCCCTGCGCCCCCCC CCCCCTTGG CCCCCCTTGG CCCCCCCTGCACCCC CCCCCCTGCGCCCCCCCCCC	CGTTCAACTT GCAACGTGAAC L A L A TCGCCACTGGC ACCGTGACCG L S A TCTCTCGGCA AGACAGCCGT CGTCTTCGGAC AGACCAGCCG CGTCTCGGAC K R Q Y ACCCCCAGTA TTGCGGTCAT TGCGTCGGAC ACCGCCAGTA TTGCGGTCGGAC ACCGCCACTAT CGGACCATATC CGGACCATATC CCAGTATATCG CCAGTATATCG CCAGTATATAC CAA Y F CCCTCGGTCGTAT AAAATCCGAC S Y I AGGTCCAATATC CCAGTATATAC CAA Y F CCCTCGGCTG CCAGTATATAC CGGCCCGCTG GGCCAGCCAC CCGCTCCGCT	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGT TGAACGACAA K I G M AAAATGGGA TTACTATT GTAACGATAA K I G M AAAATGGGA TTTTTAACCT M F D N TGTTCGATAA AAAATGGGA AAAATGGGA TTTTTAACCT I E N CATGGAGAAC CATGGAGAAC CATGGCATAT I E N CATGGAGAAC CATGGCATAT I E N CATGGAGAAC CATGGCATAT CATGGCATAT CATGGCATAT CATGGAGAC CTTTCACACT CATGCCACA CATGCCACCGC CTACCCCGCC CTACCCCGCC C CACCGCGCC C CACCGCGCC C CTACCGCGCC C TTCCACGCGCC C TTCCACGCCC C TTCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCCC C TTCCCCCCCCCC
	-1 1601 +1 1701 +1 1901 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2501 +1 2601 +1 2801 +1 2801 +1 2901	L Q R CTTAATTOTG GAATTAACAC ACCAGGCTTC TCGTCCGAAG T L S ACCCTGACTGCC TCCACCGCC K N F V AAAACTTGT TTTTGAAGCA V G P A ACCCGGCGCA CCCCGGCCGCA CCCCGGCCGCA M F G ATGCCGGCGCA ACCGCTAACT N C Q S CTGCCAACGCC TCTGGGAAC ACCGCTAACT N C Q S CTGCCAACGCC TCTGTGGGAA CCCACAACGCCTTCG T L G ACCGCTAGCCG ACCGCTAACT CTGCGATCGGA CCCACACGCC Q N T F ACCGCTAGCGCA CCCACACGCC Q N T S CTGCCAACGCCT CCTATACCAA CCCACACGCCCAAGCT CCCACACCCA CCCACACCCAA CCCACACCCAA CCCACACCCAA CCCACACCCAA CCCACACCCAA CCCACACCCAA CCCACACCCAA CCCACCCCCCAACT CCCACCCCCCCCCC	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG GACATCGAG G T S S CCTGTAGCTC GGACATCGAG G A A R TOCGGCCGCCGC G C A A C S CCATGCGCCG G T F GGTACGCCAG CCATGCGCCG CCATGCGCCG G E M TGGTGAAATG ACCACTTAC K H A AAAACCCCCG D S V D ACAACCGTCGA ACCACTTAC S G R TCCGGACGACCG G C V CCGGTCAGGCT G C V ACGGCCGGCCCG G C C CCGGCCAGGCC CCACCCCGC CCACCCCGC CCACCCCGC CCACCCCGCCCC CCACCCCGCCCCC CCACCCCCGCC T G T C CCGCCACCCCGCCCC T G T C CCGCCACCCCGCCCCCCCCCCCCCCCCCCCCCCCCCC	R K G CTGATGCTAC GACTACCATG GACTACCATG T P P TACGCCACCC ATGCGCGCGG G V F A CGGTTACGCA CGCATGCGC CGCATGCGCACATA S V A AGCGTCGCTG CGCCGTGTAT CTGGCACATA CGCCACGGCACCA A G N TGCCGGGAACA AGCGTCGCGG CGACGCTGCGCG CGCCGTCACC CGCCACGGCACCA A G N TGCCGGGACCA A G N TGCCGGGACCA CGCCCTTGCCCGC CGCCCTTACC CGCCCTTACC CGCCCCTGCT CGCCGCACCA CGCCCTTACC A G N TGCCGGCACCA CGCCCTTGCCCGC CCCCCGGAACA CGCCCTTGCCCCCC CCCCCGGAACA CGCCCTTTC CCCCGCACACA CGCCCCTGCT TGCCCGCCCCC A G N TGCCCGGCACCA CCCCCCTGC CCCCCCCCCC CCCCCCCCCC A G N TGCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCC	T M L AGGTGTTTCC TCCACCAAAGC AAGTTCGGCC AAGTTCGGCC D Q I GATCAGATCC CTACTGCTAGG CCCTGACGCC GGGACTCGGC CT S I CCCTGACGCC GGGACTCGGC D W R R ACTGGCGCCC D W R R ACTGGCGCCC CAAGGCCGCC CAAGGCCGCC V M C TGTCCCGCCG CAAGGCCGCC V M C TGTCCGCCGCC CAAGGCCGCC V M C TGTCCGCCGCC CAAGGCCGCC V M C TGTCCGCCGCC CAAGGCCGCC V M C TGTCCGCCGCC CAAGGCCGCC V M C TGTCCGCCGCC CAAGGCCGCC V M C TGTCCGCCGCC CAAGGCCGCC V M C TGTCGATAAC CACCTACTACA CACCTACACAC CAAGCTTGC D S V I D S V I C S V C D S V I C S V C C CACGCCGCC CAAGGCTCAC CAACGCTCACAC CAAGGCTCACAC CACGCCTACACAC CACGCCTACACAC CCACGCCCCC CACGCCTCGCCCC CACGCCCCCCCCCC	Q I S M GGCCTGAAAA CCGGAACTTT GAGCAACCTG CTCGTTGGAC P S Q CTTCCGATCA GAAGCCTAGT N L A GAATACCGCC CTTATGCGG GCCCCGCGCCCG GCCCGCGCAAC CCCCCGCGCCCC A V N CCCCCGCGCCCC A V N CCCCCGCGCCCC A V N CCCCCGCGCCCC A CCCCCGCCCCC A V N CCCCCGCGCCCC A CCCCCCCCCCCCC A CCCCCCCCCC	GGAACTTTT GGAACTTTT CCTTGAATACATT ACTTACATA ACTTACATA T ACTACATT ACTTACATA T ACTACATT ACTACATT ACTATACATT ACTATACATT ACTATAGTAA T ACTACATACATT ACCECCEGE T GGAACACACG GGACCACACCG ACTACGCGC AGACTACGCGC AGACCACCCGT AGACCACCGCGT T AGACCACACCG T AGACCACACCGT AGACCACCGTC AGACCACCGCT T AGACCACCACCGT AGACCACCACCGT AGACCACCACCGT AGGTCACCACCGT AGGTCACCACCGT AGGTCACCACCAC CTAGCGAGCCC CAGGCAGCCCT A GGTCACCACCAC A GGTCACCACCAC A GGTCACCACCAC CTAGCGAGCCGC A GGTCACCACCAC	ACCTTTTCGC TGGAAAGCG G T A GGCACGGCTTCC CCGTGGCGAC GGCACGGCT CCGTGGCGGAC GCACCGGCT CCGTGGCGGAC GCACCGGCT ATGCCTGGCGAC ATGCCCTGGC M V K ATGGTTAAA TTACCAATTT Y G T TACGGACG ATGCCCTGGC K Y M P AATATATCCC CAAACTCCC GCAAACTCCC GCGCATGCGA D K V GACAAATCCCT GCGCATGCGA CTTATTACGGCAC P R T CTCCCGGACC GATGGCCTTC CGCCTTCCCG GATGGCCTTC CCCCGCACC A K P W GCCAACCGTG GGTTGGCAC A K P W CCCAACCGTG GGTTGGCAC A K P W CCCAACCGTG	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCCC GCAAACGCCG S R P I CGCCCCCAAT GCGCGGGGTTA T L P T L P TACCTTGCCC ATGGGACGGG R T N CGCCACGAACA GCGTGCTTGT G E K P TGGTAAACCC CACTTTTTGG G K P TGGTAAACCC CACTTTTTGG CACTCAGGTAA CCATTACAAA CCATTACAAA CCATTACAAA CTTAAATGTT N GGCGACGCAGC CCACGCCAGCAAC CCACTGCTGG Q I N K AGATTAACAA CCATTGCTGATG CCACTGCTTGC CCACGCCGCACC CCCACTGCTGC CCCACGCTGGC CCCACGCTGGG CCCCCCTTGG GCCCCCCTTGG CCCACGCTGGG CCCCCCTTGG GCCCCCCTTGG CCCCCCTTGG CCCCCCCTGG CCCCCCTTGG CCCCCCTGGG CCCCCCTGG CCCCCCTGGG CCCCCCTGGG CCCCCCTGGG CCCCCCTGGG CCCCCCTGGG CCCCCCTGGG CCCCCCTGGG CCCCCCTGGG CCCCCCTGGG CCCCCCTGGG CCCCCCTTGG CCCCCCTTGG CCCCCCTTGG CCCCCCCC	CGTTCAACTT GCAACGTGAAC L A L A TCGCCACTGCC ACCGTGACCG L S A TCTCTCGCCA AGACAGCCGT CGTCTTCGGAC AGACCAGCCG CGTCTCCGAC K R Q Y ACCCCCAGTA TTGCGGTCAT TG ACCCCCAGTA TGCGGCCATAT GCTCGGTATAAC A AACCCCACATAT CGGACCATAT CGGACCATAT CGGACCATAT CGGACCATAT CGGACCATAT CGGACCATAT TTTTTAGACTG Y N I GGTCAATATC CCACTTATATG CCACTATATAC A Y Y F CCTCACGTCT TGGATATAAC A Y Y F CCCTCGCTCGAC CACTATATAC CGGCCCCTC GGCCAGCGAC Y I S CTCATCTCGA CAGTAGACAA TATCTGAT	AGTATAAAAA TCATATTTT F G V TTTTGGCGTG AAAACGCGAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGT TGAACGACAA K I G M CATGGATAAT GTAACGATAA K I G M AAAATGGGA TTTTTAACCCT M F D N TGTTCGATAAT CATGGAGAAC K I G M CATGGAGACACTT I E N CATGGAGAAC CTTGGCGCGC GTAACCTTG CATGGCATAT AAAATTGGCGA CTTGGCGCGCC GTAACCTTG CATGGCACCGC GTAACCCTTG CATGCCACCAC CTTGCGCGCCC GAACCGCGCG C D V D A GATGCCACGC CTACGCGTC CTCGCGCGCC CTCCGCGCCC CTCCGCGCCC CTCCGCGCCC CTCCGCCGCC CTCCGCGCCC CTCCGCGCCC CTCCGCCGCC CTCCGCCGCC CTCCGCGCCC CTCCGCGCCC CTCCGCGCCC CTCCGCGCCC CTCCGCGCCC CTCCGCCGCC CTCCGCGCCC CTCCGCGCCC CTCCGCGCCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCGCCC CTCCGCCGCC CTCCGCCGCC CTCCGCGCCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCGCCC CTCCGCGCCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCGCC CTCCGCCCCC CTCCCGCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCGCCCCC CTCCCCCCCC CTCCCCCCCC CTCCCCCCCC
	-1 1601 +1 1701 +1 1901 +1 2001 +1 2101 +1 2301 +1 2401 +1 2401 +1 2501 +1 2601 +1 2601 +1 2701 +1 2801 +1 2801 +1 2801 +1 2801 +1 2801 +1 2801 +1 2801 +1 2801 +1 2801 +1 2901 +1 2001 +1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -	L Q R CTTAATTOTG GAATTAACAC ACCAGGCTTC TCGTCCGAAG T L S ACCCTGACTGGC TTTTTTTAACAC K N F V AAAACTTCGT TTTTTGAAGCA V G P A ACCCGGCGCA CCCCGGCCGCA CCCCGGCCGCA M F G ATGCCGGCGCA ACCGCTAACT N C Q S CTGCCAACGCC TCTGTGGGAT ACCGCTAACT N C Q S CTGCCAACGCC CCCCACAGCC CCCCACAGCC ACCGCTACCA CCCCCACAGCC CCCCACAGCC A N T E S ACCGCTACCAC CCCCACAGCC A N T S CCCCACAGCCCA ACCGCTACCAC CCCCACAGCC A N T S CCCCACAGCCCA ACCGCTACCAC CCCCACAGCCCA CCCCACAGCCCA CCCCACAGCCCAC CCCCACAGCCCA CCCCACAGCCA CCCCACCCACCAC CCCCACCCACCACCACCAC CCCCCACCCA	M G T A ATAAAACTAT TATTAGATGTC TTACCTACAG GACATCGAG G S S CCTGTAGCCC GGALACCGAG G T F GGTACGCCGCCCG G T F GGTACGCCGCCCG CCATGCGCCC CCATGCGCCC CCATGCGCCC CCATGCGCCAG CCATGCGCCAG CCATGCGCCAG CCATGCGCCCCCCC D S V D ACAACCCCCC CAACCCCCCC CAACCCCCCCCCC	R K G CTGATGCTAC GACTACCATG ATTTAACTTT TAAATTGAAA T P P TACGCCACCC ATGCGCGGG G Y F A TATTTTGCT ATAAAACGGA G V T H GCGTTACGCA CGCAATGCGT T V Y CGCGTACGCA CGCAATGCGCT CGCGCACGATGGC CGCCGCTGTAT CGGGCACCATG CGCCGCCGCACGA AGCGTCGCTG CGCCGCGGAAAA ACGCCCTTG TGGCCGGCACCA A G N TGCCGGAAAAC CGCCGCTGT TGGCCGGCGCC A G N S G AACGCCCTG TGGCCGGCACCA CGCCGCTGT TGGCCGCGCCC CGCCGCAAAAC CGCCCTTGCCCGC CGCCGCCGCC CGCCGCCCC A G N TGCCGGAAAAC CGCCCTGC TGGCCGCCCC TGGCCGCCCC TGGCCGCCCCC CCCCGAAAAC CGCCCTTGC CCCCGAAAAC CGCCCCTGC TGGCCCCCCC TGGCCCCCC CCCCAAAACCGCCC C CCCCCCCCC CCCCCCCCCC	T M L AGGTOTTICC TCCACCAAGCCC AAGTCGGCC AAGTCGGCC D Q I GATCAGATCC CTACTGCTAGG GGACTCGGC GGGACTCGGC CTACTGCTAG GTGCGCCCCAT GTGCGCGCCC D W R R ACTGGCGCCC D W R R ACTGGCGCCC D W R R ACTGGCCCCC D W R R ACTGGCCCCC D W R R ACTGCCCCCC T GTCCCGCCG CAAGGCCGCCC V M C TGTCCGCCGCC CAAGGCCGCCC V M C TGTCCGCCGCC CAAGGCCGCCC V M C TGTCCGCCGCC CAAGGCCGCCC V M C TGTCCGCCCCC C T TGTCCGCCCCC C C D N R C TCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	Q I S M GGCCTGAAAA CCGGAACTTT GAGCAACCTG CTCGTAGAC CTCGTGGAC P S D Q CTTCGGAC GAATACCGCC CTTATGCGG GCCCCCGGAAC GCCCCCGCAAC CAGCCCCCG CCCCCGCGCCCC A V N CGCCGCGCCCCC A V N CGCCGCGCCCCC A V N CGCCGCCCCCCCC A C CCCCCCCCCCCCCCC A C CCCCCCCCCC	GGAACTTTT GCATGAAAAA V N T L TCAATACATT ACTTATGTAA T À P ACCECEGECT TI GECECEGA GGCACCACCE GGCACCACCE AGCTACCGCE AGCTACCGCE AGCTACCGCE AGCTACCGCE CCTGECECA AGACCACCCCT TCTCGACCGCCCC TCTCGCCCAC AGACCACCCCC D R Q CCACCACCEC AGCCCACCCCC TCCGCCCCCC TCCGCCCCCC TCCGCCCCCC AGCCCCCCCCCC	ACCTTTTCGC TGGAAAAGCG S V S ATCGCTTACC TAGCCAAAGC G T A GCACCGCT CCGTGGCGGAT GCACCGGCT CCGTGGCGGAT GCACCGGCT M V K ATGGTTAAA TTACCAATTT Y G T TACGGACCC ATGCCCTGGC K Y M P AATGTTAAA TTACCAATTT TACGGACCC ATGCCCTGGC M N G GAATAACGGC TTATTACGG M N G GAATAACGCT GCAAAGTCC CTGTTCCGGAAA S R T CTCCCGGACAA S R T CTCCCGGACAA S R T CTCCCGGACAAA S R T CTCCCGGACAC A K P W A K P W CAAAACTCC CCAACCGTG GGTTGGCAC A K P W T ACCAACGTG GGTTGGCAC A K P W	CTTCCCGTTT GAAGGGCAAA R L A CGTCTGGCCC GCACACCGCC S R P I CGCCCCCAT GCCGCCCAT GCCGCCCAT GCCGCCCAT GCCGCCCAT GCCCCCACACA CCCTTGT G E K P TACCTGCCC ATGGAACACC CCCTTTTGG G K P TGGTAAACCC CACTTTTGG G K P TGGTAAACCC CACTTTTGG G K P TGGTAAACCC CACTTTTGG G C CACGCATGA CCACTGCTG Q I N K AGATTAACAA CCACTGCTG GCCCCCCTGCTGC CCCCCCTGCTGC CCCCCCTGCACCC CCCCCCTGCACCC CCCCCCTGCACCC CCCCCCCTGCACCC CCCCCCTGCACCCC CCCCCCCTGCACCCC CCCCCCTGCGCCCCCCCC CCCCCCTGCGCCCCCCCC	CGTTCAACTT GCAAGTTGAA L A L A TCGCACTGGC ACCGTGACCG L S A TCTCTCGGCA AGACAGCCGG CGTCTGGAC CGTCTCGGAC K R Q Y ACCCCCAGTA TTGCGGTCGAC ACCGCCAGTA TTGCGGTCGAC ACCGCCAGTA TTGCGGTCGAC ACCGCCAGTA TTTTTAGACTG V N L GGTCAATATC CCAGTATATAC CAATCAATATC CCAGTATATAC CAA Y V F CCTCGCTCGAC S Y I AGCTCCAATATC CCAGTATATAC CAA Y V F CCTCGCTCGAC S Y I ACCGCCCCTGG GGCCAGCCAC V I S CCTCCGCTCGAC CAGTAGACCAT	AGTATAAAAA TCATATTTT F G V TTTTGCCGTG AAAACCGCAC N E A K AATGAAGCGA TTACTTGGCT D F V V ACTTTGTGT TGAACGATAA K I G M AAAATGGGA X G M K I G M AAAATGGGA X G M K I G M AAAATGGGA X G M K I G M AAAATGGGA TTTTTAACCCT M F D N TGTTCGATAA CATGGAGAAC CTTGGAGGACAC GTAACCTTG L G R Q CTTGGAGGACAC CTTGGCGCGCC GAACCGGCGCG E G D V AAGGCCATGT TTCCGCGCGC CGCGCGGCGC CTACGCGCGC CTTGCAGGCAA R G V CCCCGGTGTA GGCGCCGCG CTTGCAGGCAA R G V CTTGCAGCAA CCCCGGTGTA CGCCCGGCG CTTGCAGGCAA R G V CTTGCAGCAA CCCCGGTGTA CCCCGGTGTA

```
1-ybhC >tr|B5QX57|B5QX57_SALEP Possible pectinesterase OS=Salmonella enteritidis PT4 (strain P125109) GN=ybhC PE=3 SV=1
```

```
MNTLSVSRLALALAFGVTLSACSSTPPDQIPSDQTAPGTASRPILSANEAKNFVAARYFA
SLTPNTAPWSPSPITLPAQPDFVVGPAGTPGVTHTSIQAAVDAAMVKRTNKRQYIAIMPG
DYQGTVYVPAAPGSLTLYGTGEKPIDVKIGMAIDGEMSVADWRRAVNPGGKYMPGKPAWY
MFDNCQSKHAATIGVMCSAAFWSQNNGLQLQNLTIENTLGDSVDAGNHPAVALRTDGDKV
QINKVNILGRQNTFFVTNSGVQNRLQTDRQPRTLVTNSYIEGDVDMVSGRGAVVFDNTNF
QVVNSRTQQEAYVFAPATLSNIYYGFLAINSRFNASGDGVAQLGRSLDVDANTNGQVVIR
DSVINEGFNVAKPWADAVISKRPFAGNTGTVDDKDEVQRNLNDTNYNRMWEYNNRGVGSK
VVAEPKQ
```

 $\begin{array}{l} \textbf{2-hutI} > sp|B5QX58|HUTI_SALEP\ Imidazolone propionase\ OS=Salmonella\ enteritidis\ PT4\ (strain\ P125109)\ GN=hutI\ PE=3\ SV=1 \end{array}$

```
MRQLLPGDTVWRNIRLATMDPQRQAPYGLVDNQALIVREGHICDIVPETQLPVSGDNIHD
MQGRLVTPGLIDCHTHLVFAGNRAAEWEQRLNGASYQHISAQGGGINATVSATRACAEET
LYLLARERMMRLASEGVTLLEIKSGYGLELATEEKLLRVAAKLAAENAIDISPTLLAAHA
TPAEYRDDPDGYITLVCETMIPQLWQKGLFDAVDLFCESVGFNVAQSERVLQTAKALGIP
VKGHVEQLSLLGGAQLVSRYQGLSADHIEYLDEAGVAAMRDGGTVGVLLPGAFYFLRETQ
RPPVELLRRYQVPVAVASDFNPGTSPFCSLHLAMNMACVQFGLTSEEAWAGVTRHAARAL
GRQATHGQLRADYRADFVVWDAEQPVEVVYEPGRNPLYQRVYRGQIS
```

Appendix 5: Sequencing of the potential promoter fragments

Green colour for restriction sites; yellow for genes sequences; Ns ambiguous nucleotides Pink colour for mismatching nicleotides

1- pJET-RP2-ybhC

Salmonella enterica subsp. Enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

putativ	ve acyl-C	oA thioester hydrolase ybhCimidazolonepropionase	
Query	44	ATCAGCGCCTGGTTATCCACCAGCCCGTACGGGGCTTGCCGCTGCGGGTCCATTGTCGCC	103
Sbjct	3794408	ATCAGCGCCTGGTTATCCACCAGCCCGTACGGGGCTTGCCGCTGCGGGTCCATTGTCGCC	3794349
Query	104	AGCCTGATGTTTCGCCAGACAGTATCGCCCGGTAAAAGTTGCCGCATTCCTGTCGCTCTC	163
Sbjct	3794348	AGCCTGATGTTTCGCCAGACAGTATCGCCCGGTAAAAGTTGCCGCATTCCTGTCGCTCTC	3794289
Query	164	TTGCCTGTCATGAGTTGTATAGACATTTATTTTCTTTCTGCTCCGGATTGTCAACTCAAA	223
Sbjct	3794288	TTGCCTGTCATGAGTTGTATAGACATTTATTTTCTTTCTGCTCCGGATTGTCAACTCAAA	3794229
Query	224	GCGCGAAAGTTGTTGCTTAATTGTGATAAAACTATCTGATGCTACAGGTGTTTCCCGGCCT	283
Sbjct	3794228	GCGCGAAAGTTGTTGCTTAATTGTGATAAAACTATCTGATGCTACAGGTGTTTCCGGCCT	3794169

PhD T	PhD Thesis		
Query	284	GAAAAGGAACTTTTTACCTTTTCGCCTTCCCGTTTCGTTCAACTTAGTATAAAAAAGCAG	343
Sbjct	3794168	GAAAAGGAACTTTTTACCTTTTCGCCTTCCCGTTTCGTTCAACTTAGTATAAAAAAGCAG	3794109
Query	344	GCTTCAATGGATGTCATTTAACTTTTTCAAGCCCGGAGCAACCTGTGAATACATTATCGG	403
Sbjct	3794108	GCTTCAATGGATGTCATTTAACTTTTTCAAGCCCGGAGCAACCTGTGAATACATTATCGG	3794049
Query	404	TTTCCCGTCTGGCGCTGGCACTGGCTTTTGGCGTGACGCTGAGCGCCTGTAGCTCTACGC	463
Sbjct	3794048	TTTCCCGTCTGGCGCTGGCACTGGCTTTTGGCGTGACGCTGAGCGCCTGTAGCTCTACGC	3793989
Query	464	CACCCGATCAGATCCCTTCCGATCAA 489	
Sbjct	3793988	CACCCGATCAGATCCCTTCCGATCAA 3793963	

T7F-ybhC

Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

putativ	ve acyl-Co	A thioester hydrolase ybhCimidazolonepropionase	
Query	51	TTGATCGGAAGGGATCTGATCGGGTGGCGTAGAGCTACAGGCGCTCAGCGTCACGCCAAA	110
Sbjct	3793963	TTGATCGGAAGGGATCTGATCGGGTGGCGTAGAGCTACAGGCGCTCAGCGTCACGCCAAA	3794022
Query	111	AGCCAGTGCCAGCGCCAGACGGGAAACCGATAATGTATTCACAGGTTGCTCCGGGCTTGA	170
Sbjct	3794023	AGCCAGTGCCAGCGCCAGACGGGAAACCGATAATGTATTCACAGGTTGCTCCGGGCTTGA	3794082
Query	171	AAAAGTTAAATGACATCCATTGAAGCCTGCTTTTTTATACTAAGTTGAACGAAACGGGAA	230
Sbjct	3794083	AAAAGTTAAATGACATCCATTGAAGCCTGCTTTTTTATACTAAGTTGAACGAAACGGGAA	3794142
Query	231	GGCGAAAAGGTAAAAAGTTCCTTTTCAGGCCGGAAACACCTGTAGCATCAGATAGTTTTA	290
Sbjct	3794143	GGCGAAAAGGTAAAAAGTTCCTTTTCAGGCCGGAAACACCTGTAGCATCAGATAGTTTTA	3794202
Query	291	TCACAATTAAGCAACAACTTTCGCGCTTTGAGTTGACAATCCGGAGCAGAAAGAA	350
Sbjct	3794203	TCACAATTAAGCAACAACTTTCGCGCTTTGAGTTGACAATCCGGAGCAGAAAGAA	3794262
Query	351	ATGTCTATACAACTCATGACAGGCAAGAGAGCGACAGGAATGCGGCAACTTTTACCGGGC	410
Sbjct	3794263	ATGTCTATACAACTCATGACAGGCAAGAGAGCGACAGGAATGCGGCAACTTTTACCGGGC	3794322
Query	411	GATACTGTCTGGCGAAACATCAGGCTGGCGACAATGGACCCCGCAGCGGCAAGCCCCGTAC	470
Sbjct	3794323	GATACTGTCTGGCGAAACATCAGGCTGGCGACAATGGACCCCGCAGCGGCAAGCCCCGTAC	3794382
Query	471	GGGCTGGTGGATAACCAGGCGCTGAT 496	
Sbict	3794383	GGGCTGGTGGATAACCAGGCGCTGAT 3794408	

2-pJET-RP2-SEN1435

NNNNNNNNNNNNNNNGGAGANCTTCTAGNNNATCAC <mark>GAATTC</mark> CAAAGCCCAGTCCTCGTGCAGAACCCGTTACCAGCGCCGTTTTCCCAGT
TAAATCAAATAAAGCGGTCATGTTGTTTCCTCACTTGTTTAATTTGTATGACGACTATCCTTTTTTAGGTTGAATTTTCGCCCTGATAAAATCA
ACAGTTCACCCATGAATTTGCAACAAGGATCACAAACAGCTCCACATGCCGACCGCGTAATTAAT
TGGTTTAAATTTAACGCAGTTTGATCGCTGTCACAGAATGGCACTCGCAGCGATCCGCTGTAAAAGAAGCGTGATATAACAGCATAAAGTTGTA
GGACAACTTACGTATATCTGTTGTATCATCCACAACGGTATGACATGCGGTAAATTCGCTGAGTTAAGGAGTGAAAGTGAGTAACCTGAAAATT
<mark>ACCAACGTGAAAACGATTCTGACGGCGCCGGGCGGCATTGATTG</mark>
${\tt CTGAAAAACTCGAGCCATCCGGAAGATCTGGCGGCCGCTCTCCCTATAGTGAGTCGTATTACGCCGGATGGAT$
TAAAGCAGTTGATTTATTCACTATGATGAAAAAAACAATGAATG
AGTAAGATTAGAGATAATACAACAATAAAAAAATGGTTTAGAACTTACTCACAGCGTGATGCTACTAATTGGGACAATTTTCCCAGATGAAGTA
TCATCTAAGAATTTAAATGAAGAAGACTTCAGAGCTTTTGTTAAAATTATTTGGCAAAATAATAATTCGGCTGCAGGGGCGGCCTCNTGATA
${\tt CGCCTATTTTTATNGGTTAATGTCATGATAANANGGTTTCTTAGACGTCAGGNGNNCTTTTNGGGAANGNGCNNNGAACCCTANTTNNNNTTTT$
${\tt CNAANNCNTCANTATNNNTCNNCTCATGNNNNAAACCNNATAANGCTTCANANNTNNANNNNNNNNNN$
TTTGCNNNTTNCNNGNNNNNNNNNNNNNNNNNNNNNAGANTCNTNAANNANTNNNNGNNN

Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE6-00219-16, complete genome Sequence ID: <u>gb|CP009088.1|</u>Length: 4677619Number of Matches: 1

Thr operon leader peptidegluconate 5-dehydrogenase

1111 0	JOI 011 104	aer peptraegraeonate z'aen jarogenase	
Query	46	CAAAGCCCAGTCCTCGTGCAGAACCCGTTACCAGCGCCGTTTTCCCAGTTAAATCAAATA	105
Sbjct	1522936	CAAAGCCCAGTCCTCGTGCAGAACCCGTTACCAGCGCCGTTTTCCCAGTTAAATCAAATA	1522995
Query	106	AAGCGGTCATGTTGCTTCCCCCACTTGTTTAATTTGTATGACGACTATCCTTTTTTAGGTT	165
Sbjct	1522996	AAGCGGTCATGTTGTTTCCTCACTTGTTTAATTTGTATGACGACTATCCTTTTTAGGTT	1523055
Query	166	GAATTTTCGCCCTGATAAAATCAACAGTTCACCCATGAATTTGCAACAAGGATCACAAAC	225
Sbjct	1523056	GAATTTTCGCCCTGATAAAATCAACAGTTCACCCATGAATTTGCAACAAGGATCACAAAC	1523115
Query	226	AGCTCCACATGCCGACCGCGTAATTAATATTAATTAATTGAATTATATGTATATATTTGG	285
Sbjct	1523116	AGCTCCACATGCCGACCGCGTAATTAATATTAATTAATTGAATTATATGTATATATA	1523175
Query	286	TTTAAATTTAACGCAGTTTGATCGCTGTCACAGAATGGCACTCGCAGCGATCCGCTGTAA	345
Sbjct	1523176	TTTAAATTTAACGCAGTTTGATCGCTGTCACAGAATGGCACTCGCAGCGATCCGCTGTAA	1523235
Query	346	AAGAAGCGTGATATAACAGCATAAAGTTGTAGGACAACTTACGTATATCTGTTGTATCAT	405
Sbjct	1523236	AGAAGCGTGATATAACAGCATAAAGTTGTAGGACAACTTACGTATATCTGTTGTATCAT	1523295
Query	406	CCACAACGGTATGACATGCGGTAAATTCGCTGAGTTAAGGAGTGAAAGTGAGTAACCTGA	465
Sbjct	1523296	CCACAACGGTATGACATGCGGTAAATTCGCTGAGTTAAGGAGTGAAAGTGAGTAACCTGA	1523355
Query	466	AAATTACCAACGTGAAAACGATTCTGACGGCGCCGGGCGGCATTGATTTGGCAGTCGTTA	525
Sbjct	1523356	AAATTACCAACGTGAAAACGATTCTGACGGCGCCGGGCGGCATTGATTTGGCAGTCGTTA	1523415
Query	526	AGATAGAAACCAACGAGCCAGGGC 549	
Sbjct	1523416		

T7F- SEN1435

 Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE6-00219-16, complete genome Sequence ID: <u>gb|CP009088.1|</u>Length: 4677619Number of Matches: 1

Thr operon leader peptidegluconate 5-dehydrogenase				
Query	52	GCCCTGGCTCGTTGGTTTCTATCTTAACGACTGCCAAATCAATGCCGCCCGGCGCCGTCA	111	
Sbjct	1523439	GCCCTGGCTCGTTGGTTTCTATCTTAACGACTGCCAAATCAATGCCGCCCGGCGCCGTCA	1523380	
Query	112	GAATCGTTTTCACGTTGGTAATTTTCAGGTTACTCACTTTCACTCCTTAACTCAGCGAAT	171	
Sbjct	1523379	GAATCGTTTTCACGTTGGTAATTTTCAGGTTACTCACTTTCACTCCTTAACTCAGCGAAT	1523320	
Query	172	TTACCGCATGTCATACCGTTGTGGATGATACAACAGATATACGTAAGTTGTCCTACAACT	231	
Sbjct	1523319	TTACCGCATGTCATACCGTTGTGGATGATACAACAGATATACGTAAGTTGTCCTACAACT	1523260	
Query	232	TTATGCTGTTATATCACGCTTCTTTTACAGCGGATCGCTGCGAGTGCCATTCTGTGACAG	291	
Sbjct	1523259	TTATGCTGTTATATCACGCTTCTTTTACAGCGGATCGCTGCGAGTGCCATTCTGTGACAG	1523200	
Query	292	CGATCAAACTGCGTTAAATTTAAACCAAATATATACATATAATTCAATTAATTAATATA	351	
Sbjct	1523199	CGATCAAACTGCGTTAAATTTAAACCAAATATATACATATAATTCAATTAATAA	1523140	
Query	352	ATTACGCGGTCGGCATGTGGAGCTGTTTGTGATCCTTGTTGCAAATTCATGGGTGAACTG	411	
Sbjct	1523139	ATTACGCGGTCGGCATGTGGAGCTGTTTGTGATCCTTGTTGCAAATTCATGGGTGAACTG	1523080	
Query	412	TTGATTTTATCAGGGCGAAAATTCAACCTAAAAAAGGATAGTCGTCATACAAATTAAACA	471	
Sbjct	1523079	TTGATTTTATCAGGGCGAAAATTCAACCTAAAAAAGGATAGTCGTCATACAAATTAAACA	1523020	
Query	472	AGTGAGGAAACAACATGACCGCTTTATTTGATTTAACTGGGAAAACGGCGCTGGTAACGG	531	
Sbjct	1523019	AGTGAGGAAACAACATGACCGCTTTATTTGATTTAACTGGGAAAACGGCGCTGGTAACGG	1522960	
Query	532	GTTCTGCACGAGGACTGGGCTTTG 555		
Sbjct	1522959	GTTCTGCACGAGGACTGGGCTTTG 1522936		

3-pJET-RP2- SEN1436

Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE6-00219-16, complete genome Sequence ID: <u>gb|CP009088.1|</u>Length: 4677619Number of Matches: 1

Thr operon leader peptidegluconate 5-dehydrogenase				
Query	47	GCCCTGGCTCGTTGGTTTCTATCTTAACGACTGCCAAATCAATGCCGCCCGGCGCCGTCA	106	
Sbjct	1523439	GCCCTGGCTCGTTGGTTTCTATCTTAACGACTGCCAAATCAATGCCGCCCGGCGCCGTCA	1523380	
Query	107	GAATCGTTTTCACGTTGGTAATTTTCAGGTTACTCACTTTCACTCCTTAACTCAGCGAAT	166	
Sbjct	1523379	GAATCGTTTTCACGTTGGTAATTTTCAGGTTACTCACTTTCACTCCTTAACTCAGCGAAT	1523320	

PhD T	hesis		Appendix
Query	167	TTACCGCATGTCATACCGTTGTGGATGATACAACAGATATACGTAAGTTGTCCTACAACT	226
Sbjct	1523319	TTACCGCATGTCATACCGTTGTGGATGATGATACAACAGATATACGTAAGTTGTCCTACAACT	1523260
Query	227	TTATGCTGTTATATCACGCTTCTTTTACAGCGGATCGCTGCGAGTGCCATTCTGTGACAG	286
Sbjct	1523259	TTATGCTGTTATATCACGCTTCTTTTACAGCGGATCGCTGCGAGTGCCATTCTGTGACAG	1523200
Query	287	CGATCAAACTGCGTTAAATTTAAACCAAATATATACATATAATTCAATTAATTAATATTA	346
Sbjct	1523199	CGATCAAACTGCGTTAAATTTAAACCAAATATATACATATAATTCAATTAATTAATATAA	1523140
Query	347	ATTACGCGGTCGGCATGTGGAGCTGTTTGTGATCCTTGTTGCAAATTCATGGGTGAACTG	406
Sbjct	1523139	ATTACGCGGTCGGCATGTGGAGCTGTTTGTGATCCTTGTTGCAAATTCATGGGTGAACTG	1523080
Query	407	TTGATTTTATCAGGGCGAAAATTCAACCTAAAAAAGGATAGTCGTCATACAAATTAAACA	466
Sbjct	1523079	TTGATTTTATCAGGGCGAAAATTCAACCTAAAAAAGGATAGTCGTCATACAAATTAAACA	1523020
Query	467	AGTGAGGAAACAACATGACCGCTTTATTTGATTTAACTGGGAAAACGGCGCTGGTAACGG	526
Sbjct	1523019	AGTGAGGAAACAACATGACCGCTTTATTTGATTTAACTGGGAAAACGGCGCTGGTAACGG	1522960
Query	527	GTTCTGCACGAGGACTGGGCTTTG 550	
Sbjct	1522959	GTTCTGCACGAGGACTGGGCTTTG 1522936	

T7F- SEN1436

Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE6-00219-16, complete genome Sequence ID: <u>gb|CP009088.1|</u>Length: 4677619Number of Matches: 1

I III O	Joi on Iou	der peptidegraeonate 5 den varogenase	
Query	51	CAAAGCCCAGTCCTCGTGCAGAACCCGTTACCAGCGCCGTTTTCCCAGTTAAATCAAATA	110
Sbjct	1522936	CAAAGCCCAGTCCTCGTGCAGAACCCGTTACCAGCGCCGTTTTCCCAGTTAAATCAAATA	1522995
Query	111	AAGCGGTCATGTTGTTTCCTCACTTGTTTAATTTGTATGACGACTATCCTTTTTTAGGTT	170
Sbjct	1522996	AAGCGGTCATGTTGTTTCCTCACTTGTTTAATTTGTATGACGACTATCCTTTTTTAGGTT	1523055
Query	171	GAATTTTCGCCCTGATAAAATCAACAGTTCACCATGAATTTGCAACAAGGATCACAAAC	230
Sbjct	1523056	GAATTTTCGCCCTGATAAAATCAACAGTTCACCCATGAATTTGCAACAAGGATCACAAAC	1523115
Query	231	AGCTCCACATGCCGACCGCGTAATTAATATTAATTGAATTATATGTATATATTGG	290
Sbjct	1523116	AGCTCCACATGCCGACCGCGTAATTAATTAATTAATTGAATTATATGTATATATTTGG	1523175
Query	291	TTTAAATTTAACGCAGTTTGATCGCTGTCACAGAATGGCACTCGCAGCGATCCGCTGTAA	350
Sbjct	1523176	TTTAAATTTAACGCAGTTTGATCGCTGTCACAGAATGGCACTCGCAGCGATCCGCTGTAA	1523235
Query	351	AAGAAGCGTGATATAACAGCATAAAGTTGTAGGACAACTTACGTATATCTGTTGTATCAT	410
Sbjct	1523236	AAGAAGCGTGATATAACAGCATAAAGTTGTAGGACAACTTACGTATATCTGTTGTATCAT	1523295

Thr operon leader peptidegluconate 5-dehydrogenase

Query	411	CCACAACGGTATGACATGCGGTAAATTCGCTGAGTTAAGGAGTGAAAGTGAGTAACCTGA	470
Sbjct	1523296	CCACAACGGTATGACATGCGGTAAATTCGCTGAGTTAAGGAGTGAAAGTGAGTAACCTGA	1523355
Query	471	AAATTACCAACGTGAAAACGATTCTGACGGCGCCGGGCGGCATTGATTTGGCAGTCGTTA	530
Sbjct	1523356	AAATTACCAACGTGAAAAACGATTCTGACGGCGCCGGGCGGCATTGATTTGGCAGTCGTTA	1523415
Query	531	AGATAGAAACCAACGAGCCAGGGC 554	
Sbjct	1523416	AGATAGAAACCAACGAGCCAGGGC 1523439	

4-pJET-RP2- SEN1432

Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

Query	47	TCAGTTCCACTTCTGAGGGCAAACGGCTACCCGGCGCCCAACTGCCGTCCAGCAGTTTGC	106
Sbjct	4504119	TCAGTTCCACTTCTGAGGGCAAACGGCTACCCGGCGCCCAACTGCCGTCCAGCAGTTTGC	4504178
Query	107	TACTTATCTGATCATAAATTTCATTTACAACATTCTGTTTTTGAATGGATTTTATGCTCA	166
Sbjct	4504179	TACTTATCTGATCATAAATTTCATTTACAACATTCTGTTTTTGAATGGATTTTATGCTCA	4504238
Query	167	AGATGGGTATCCGTTAAGATGTCGCTGAAGTGCTTTATTATAACAATTCTCTTTTAAG	226
Sbjct	4504239	AGATGGGTATCCGTTAAGATGTCGCTGAAGTGCTTTATTATAACAATTCTCTTTTAAG	4504298
Query	227	AACAAAGCGCCCGGCCGGCGCTTTTCACCGTTAATCGAAACGAATGAGTACCTTAGCAG	286
Sbjct	4504299	AACAAAGCGCCCGGCCGGCGCGCTTTTCACCGTTAATCGAAACGAATGAGTACCTTAGCAG	4504358
Query	287	AGACATTTTTGTCTGTGGCGGTAATCAGCGCGTCTTCAATTTGCTGGGGCGGGAACTCGG	346
Sbjct	4504359	AGACATTTTTGTCTGTGGCGGTAATCAGCGCGTCTTCAATTTGCTGGGGCGGGAACTCGG	4504418
Query	347	CGCTGATAAGCGGGCGAGGATCGACGCCCCATCTTCCAGCCAG	406
Sbjct	4504419	CGCTGATAAGCGGGCGAGGATCGACGCGCCCATCTTCCAGCCAG	4504478
Query	407	ACTCACCGATAAAACGGAATGAGCCGACCCAGTTGAGTTCTTTAACCAGCATCGTTGACA	466
Sbjct	4504479	ACTCACCGATAAAACGGAATGAGCCGACCCAGTTGAGTTCTTTAACCAGCATCGTTGACA	4504538
Query	467	CCGG 470	
Sbjct	4504539	CCGG 4504542	

bacterial regulatory s, gntR family proteinL-idonate 5-dehydrogenase

T7F- SEN1432

Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

bacterial regulatory s, gntR family proteinL-idonate 5-dehydrogenase				
Query	50	CCGGTGTCAACGATGCTGGTTAAAGAACTCAACTGGGTCGGCTCATTCCGTTTTATCGGT	109	
Sbjct	4504542	CCGGTGTCAACGATGCTGGTTAAAGAACTCAACTGGGTCGGCTCATTCCGTTTTATCGGT	4504483	
Query	110	GAGTTCATCACCGCGGTACGCTGGCTGGAAGATGGGCGCGTCGATCCTCGCCCGCTTATC	169	
Sbjct	4504482	GAGTTCATCACCGCGGTACGCTGGCTGGAAGATGGGCGCGTCGATCCTCGCCCGCTTATC	4504423	
Query	170	AGCGCCGAGTTCCCGCCCCAGCAAATTGAAGACGCGCTGATTACCGCCACAGACAAAAAT	229	
Sbjct	4504422	AGCGCCGAGTTCCCGCCCCAGCAAATTGAAGACGCGCCTGATTACCGCCACAGACAAAAAT	4504363	
Query	230	GTCTCTGCTAAGGTACTCATTCGTTTCGATTAACGGTGAAAAGCGCCCGGCCGG	289	
Sbjct	4504362	GTCTCTGCTAAGGTACTCATTCGTTTCGATTAACGGTGAAAAGCGCCCGGCCGG	4504303	
Query	290	TGTTCTTAAAAGAGAATTGTTATATAAAAGCACTTCAGCGACATCTTAACGGATACCC	349	
Sbjct	4504302	TGTTCTTAAAAGAGAATTGTTATAATAAAAGCACTTCAGCGACATCTTAACGGATACCC	4504243	
Query	350	ATCTTGAGCATAAAATCCATTCAAAAACAGAATGTTGTAAATGAAATTTATGATCAGATA	409	
Sbjct	4504242	ATCTTGAGCATAAAATCCATTCAAAAACAGAATGTTGTAAATGAAATTTATGATCAGATA	4504183	
Query	410	AGTAGCAAACTGCTGGACGGCAGTTGGGCGCCGGGTAGCCGTTTGCCCTCAGAAGTGGAA	469	
Sbjct	4504182	AGTAGCAAACTGCTGGACGGCAGTTGGGCGCCGGGTAGCCGTTTGCCCTCAGAAGTGGAA	4504123	
Query	470	CTGA 473		
Sbict	4504122	CTGA 4504119		

5-pJET-RP2- dgoR

Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE1-1019-1, complete genome Sequence ID: gb|CP009083.1|Length: 4678914Number of Matches: 2

Thr or	peron lead	der peptidegalactonate operon transcriptional repressor	
Query	42	CAGCGCCGAACCGGGTACGTATTTACCGCTGACAATCTGTTTGCCCAGCGTGATAACGAT	101
Sbjct	3901708	CAGCGCCGAACCGGGTACGTATTTACCGCTGACAATCTGTTTGCCCAGCGTGATAACGAT	3901767
Query	102	GCGATCGGTTTTATTGAGAGTCATAGAGAGTCCTTGTGCTCGATGTGAACTCTCTTACTT	161
Sbjct	3901768	GCGATCGGTTTTATTGAGAGTCATAGAGAGTCCTTGTGCTCGATGTGAACTCTCTTACTT	3901827
Query	162	TACCGCGATAGCTGAATTACGCCGCAATTTTGTAGTACTAGCGTGATATCAACCGTCGTT	221
Sbjct	3901828	TACCGCGATAGCTGAATTACGCCGCAATTTTGTAGTACTAGCGTGATATCAACCGTCGTT	3901887
Query	222	ATCATGCCATTAATGTAGTACAACATAATTATGTTGTACTACAATTTAGATCACAAAAAAC	281
Sbjct	3901888	ATCATGCCATTAATGTAGTAGTACAACATAATTATGTTGTACTACAATTTAGATCACAAAAAAC	3901947
Query	282	AACAATTGGTTATGGGAACGTTATAAGACGTAAACGAAAGACATAAAAAAACCCGCAGCA	341
Sbjct	3901948	AACAATTGGTTATGGGAACGTTATAAGACGTAAACGAAAGACATAAAAAAAA	3902007
Query	342	AGTGCGGGTCGTTAAGCGCGTATTTGCCCGATGGCGGCTATGCCTTATCGGGCGGG	401
Sbjct	3902008	AGTGCGGGTCGTTAAGCGCGTATTTGCCCGATGGCGGCTATGCCTTATCGGGCGGG	3902067
Query	402	TGGCCGGATGAGTGATCGGGGTTCAGCACAAATTTTTCAATCGCCCAGGCAACACCATCT	461
Sbjct	3902068	TGGCCGGATGAGTGATCGGGGTTCAGCACAAATTTTTCAATCGCCCAGGCAACACCATCT	3902127
Query	462	TCAAGGTTCGATTTAGTCACAAAGTTAGCCACCTCTTTGACCGACGGAATGGCGTTGTCC	521
Sbjct	3902128	TCAAGGTTCGATTTAGTCACAAAGTTAGCCACCTCTTTGACCGACGGAATGGCGTTGTCC	3902187
Query	522	ATTGCCACGCCCATACCGGCGTATTCGATCATCGCAATGTCGTTTTCCTGATCGCCAATC	581
Sbjct	3902188	ATTGCCACGCCCATACCGGCGTATTCGATCATCGCAATGTCGTTTTCCTGATCGCCAATC	3902247
Query	582	GCCATCACCTC 592	
Sbjct	3902248	GCCATCACCTC 3902258	

T7F-dgoR

NNNNNNNNNNNNNNNGNTGGCTCGAGTTTTTCNGCAAGATGAG<mark>GGATCC</mark>GAGGTGATGGCGATTGGCGATCAGGAAAACGACATTGCGATGAT CGAATACGCCGGTATGGGCGTGGCAATGGACAACGCCATTCCGTCGGTCAAAGAGGTGGCTAACTTTGTGACTAAATCGAACCTTGAAGATGGT **GCGCTTAACGACCCGCACTTGCTGCGGGGTTTTTTTTATGTCTTTCGTTTACGTCTTATAACGTTCCCATAACCAATTGTTGTTTTTGTGATCTAA** ATTGTAGTACAACATAATTATGTTGTACTACATTAATGGCATGATAACGACGGTTGATATCACGCTAGTACTACAAAATTGCGGGCGTAATTCAG CTATCGCGGTAAAGTAAGAGAGTTCACATCGAGCACAAGGACTCTCTATGACTCTCAATAAAACCGATCGCATCGTTATCACGCTGGGCAAACA GATTGTCAGCGGTAAATACGTACCCGGTTCGGCGCTGGAATTC ${\tt ATGTTCTTTTATTCTCTCAAGATTTTCAGGCTGTATATTAAAAACTTATTAAGAACTATGCTAACCACCTCATCAGGAACCGTTGTAGGT$ ${\tt AACGAGGTTAGAGCAAGCTTCAGGAAACTGANACAGGAATTTTATTAAAAATTTTAATTTTGAAGAAGNTCNGGNNAATAGCNTCCATTTTTTTGC$

Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE1-1019-1, complete genome Sequence ID: gb|CP009083.1|Length: 4678914Number of Matches: 4

Thr operon leader peptidegalactonate operon transcriptional repressor				
Query	51	GAGGTGATGGCGATTGGCGATCAGGAAAACGACATTGCGATGATCGAATACGCCGGTATG	110	
Sbjct	3902258	GAGGTGATGGCGATTGGCGATCAGGAAAACGACATTGCGATGATCGAATACGCCGGTATG	3902199	
Query	111	GGCGTGGCAATGGACAACGCCATTCCGTCGGTCAAAGAGGTGGCTAACTTTGTGACTAAA	170	
Sbjct	3902198	GGCGTGGCAATGGACAACGCCATTCCGTCGGTCAAAGAGGTGGCTAACTTTGTGACTAAA	3902139	

PhD Thesis

Appendix

Query	171	TCGAACCTTGAAGATGGTGTTGCCTGGGCGATTGAAAAATTTGTGCTGAACCCCGATCAC	230
Sbjct	3902138	TCGAACCTTGAAGATGGTGTTGCCTGGGCGGATTGAAAAATTTGTGCTGAACCCCGATCAC	3902079
Query	231	TCATCCGGCCATTTCCCCGCCCGATAAGGCATAGCCGCCATCGGGCAAATACGCGCTTAA	290
Sbjct	3902078	TCATCCGGCCATTTCCCCGCCCGATAAGGCATAGCCGCCATCGGGCAAATACGCGCTTAA	3902019
Query	291	CGACCCGCACTTGCTGCGGGTTTTTTTATGTCTTTCGTTTACGTCTTATAACGTTCCCAT	350
Sbjct	3902018	CGACCCGCACTTGCTGCGGGGTTTTTTTTTTTTTTTTCGTTTTCGTTTACGTCTTATAACGTTCCCAT	3901959
Query	351	AACCAATTGTTGTTTTTGTGATCTAAATTGTAGTACAACATAATTATGTTGTACTACATT	410
Sbjct	3901958	AACCAATTGTTGTTGTTGTGATCTAAATTGTAGTACAACATAATTATGTTGTACTACATT	3901899
Query	411	AATGGCATGATAACGACGGTTGATATCACGCTAGTACTACAAAATTGCGGCGTAATTCAG	470
Sbjct	3901898	AATGGCATGATAACGACGGTTGATATCACGCTAGTACTACAAAATTGCGGCGTAATTCAG	3901839
Query	471	CTATCGCGGTAAAGTAAGAGAGTTCACATCGAGCACAAGGACTCTCTATGACTCTCAATA	530
Sbjct	3901838	CTATCGCGGTAAAGTAAGAGAGTTCACATCGAGCACAAGGACTCTCTATGACTCTCAATA	3901779
Query	531	AAACCGATCGCATCGTTATCACGCTGGGCAAACAGATTGTCAGCGGTAAATACGTACCCG	590
Sbjct	3901778	AAACCGATCGCATCGTTATCACGCTGGGCAAACAGATTGTCAGCGGTAAATACGTACCCG	3901719
Query	591	GTTCGGCGCTG 601	
Sbjct	3901718	GTTCGGCGCTG 3901708	

6-pJET-RP2-dgoT

NNNNNNNNNNNNNAGGAGATCTTCTAGAAAGATCACGAATTCGTTGGCGCGATCGACGTAGCAAATCACCACGGTAATAAAGATCATCAC CAGCGTCAGATAGCGGCGACGCCCCGGCTGTGCTGCTGTAACTGAAATATCCATCGTCATCTGTCTCCAGATTCTGGGCATAGCGAGGCCGCCC ACCATGCCCTGTAAATTACAGAGGGTGTGTTTTTATATTTAAATTGGGTTGCCCGGAGGGCGACGTTTGTTGAGCCTACAGCGTGGCGAGCACCC ACTCGGCTACCGATCCGTCAGCGTGCCGCCACAACGGATTACGCCAGTCCGGCGCGCTTTTGTTAAGTTCAATCACCCTGGCCTCGTCAATGTC TACGCCAAGACCCGGTTTGGTTAAGGGGTTTAAAGAAGCCGCCGTCCATGCTGAAGTCTTCTTTGTTTTTCACAAAGTCGAGCAGCTCCGCGCCC TGCTATAGT GGATAGCCCCCGCCACAAGGGTTTAAAGAAGCCGCCGCCGTCCATGCTGAAGTCTCTTTGTTTTTCACAAAGTCGAGCAGCTCCGCGCC GGATANGGTGTTCNNGCACAAGTGTTANNGCNGTTGATTTTATTNACTATNANGAANAAACAATGAATGGAACCTGCTCNANGTTNAANANNN AGATAANANNNAAANNCTNANCNANT

Salmonella enterica subsp. enterica serovars Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

Query	47	GTTGGCGCGATCGACGTAGCAAATCACCACGGTAATAAAGATCATCACCAGCGTCAGATA	106
Sbjct	2202171	GTTGGCGCGATCGACGTAGCAAATCACCACGGTAATAAAGATCATCACCAGCGTCAGATA	2202230
Query	107	GCGGCGACGCCCCGGCTGTGCTGCTGTAACTGAAATATCCATCGTCATCTGTCTCCAGAT	166
Sbjct	2202231	GCGGCGACGCCCCGGCTGTGCTGCTGCTGTAACTGAAATATCCATCGTCATCTGTCTCCAGAT	2202290
Query	167	TCTGGGCATAGCGAGGCCGCTCACCATGCCCTGTAAATTACAGAGGGTGTGTTTTTATAT	226
Sbjct	2202291	TCTGGGCATAGCGAGGCCGCTCACCATGCCCTGTAAATTACAGAGGGTGTGTTTTTATAT	2202350
Query	227	TTAAATTGGGTTGCCCGGAGGGCGACGTTTGTTGAGCCTACAGCGTGGCGATCACCACTC	286
Sbjct	2202351	TTAAATTGGGTTGCCCGGAGGGCGACGTTTGTTGAGCCTACAGCGTGGCGATCACCACTC	2202410
Query	287	GGCTACCGATCCGTCAGCGTGCCGCCACAACGGATTACGCCAGTCCGGCGCGCTTTTGCT	346
Sbjct	2202411	GGCTACCGATCCGTCAGCGTGCCGCCCCCACAACGGATTACGCCAGTCCGGCGCGCTTTTGCT	2202470
Query	347	AAGTTCAATCACCCTGGCCTCGTCAATGTCTACGCCAAGACCCGGTTTGGTTAAGGGTTT	406
Sbjct	2202471	AGTTCAATCACCCTGGCCTCGTCAATGTCTACGCCAAGACCCGGTTTGGTTAAGGGTTT	2202530

D-galactonatetransporterD-galactonate dehydratase

PhD Thesis			Appendix	
Query	407	AAAGAAGCCGCCGTCCATG0		466
Sbjct	2202531	AAAGAAGCCGCCGTCCATG	TGAAGTCTTCTTTGTTTTTCACAAAGTCGAGCAGCTCCGC	2202590
Query	467	GCCCTTGTTATAGTG	487	
Sbjct	2202591	GCCCTTGTTATAGTG	2202611	

T7F-dgoT

NNNNNNNNNNNNNNNNNNGGANNGGCTCGAGTTTTTNGCANATGAG

Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

Ouerr		anspond D-ganacionale a conversional a babaabaabaabaabaabaabaabaabaabaabaaba	1.0.1
Query	42		101
Sbjct	2202611	ACTATAACAAGGGCGCGGGGGGGGGCGCTGCTCGACTTTGTGAAAAAACAAAGAAGAAGACTTC	2202552
Query	102	AGCATGGACGGCGGCTTCTTTAAACCCTTAACCAAACCGGGTCTTGGCGTAGACATTGAC	161
Sbjct	2202551	AGCATGGACGGCGGCTTCTTTAAACCCTTAACCAAACCGGGTCTTGGCGTAGACATTGAC	2202492
Query	162	GAGGCCAGGGTGATTGAACTTAGCAAAAGCGCGCGGACTGGCGTAATCCGTTGTGGCGG	221
Sbjct	2202491	GAGGCCAGGGTGATTGAACTTAGCAAAAGCGCGCCGGACTGGCGTAATCCGTTGTGGCGG	2202432
Query	222	CACGCTGACGGATCGGTAGCCGAGTGGTGGTCGCCACGCTGTAGGCTCAACAAACGTCGC	281
Sbjct	2202431	CACGCTGACGGATCGGTAGCCGAGTGGTGATCGCCACGCTGTAGGCTCAACAAACGTCGC	2202372
Query	282	CCTCCGGGCAACCCAATTTAAATATAAAAACACACCCCCTCGTAATTTACAGGGCATGGTG	341
Sbjct	2202371	CCTCCGGGCAACCCAATTTAAATATAAAAACACACCCCCTCTGTAATTTACAGGGCATGGTG	2202312
Query	342	AGCGGCCTCGCTATGCCCAGAATCTGGAGACAGATGACGATGGATATTTCAGTTACAGCA	401
Sbjct	2202311	AGCGGCCTCGCTATGCCCAGAATCTGGAGACAGATGACGATGGATATTTCAGTTACAGCA	2202252
Query	402	GCACAGCCGGGGCGTCGCCGCTATCTGACGCTGGTGATGATCTTTATTACCGTGGTGATT	461
Sbjct	2202251	GCACAGCCGGGGCGTCGCCGCTATCTGACGCTGGTGATGATCTTTATTACCGTGGTGATT	2202192
Query	462	TGCTACGTCGATCGCGCCAAC 482	
Sbjct	2202191	TGCTACGTCGATCGCGCCAAC 2202171	

D-galactonatetransporterD-galactonate dehydratase

7-pJET-RP2- SEN2978

NNNNNNNNNNNNGNNNGAGATCTTCTAGANNATCAC <mark>GAATTC</mark> GATATGGTGTAACGCCGTTACCACGCCGGTTGCGCCAGCCTGGCGTACATC
TGACAGCGTTACCGGGTCGTTAGGTCCGTACCAGCGCCAGGTTTGTTT
ATTGCCAGCCAACATCGAAACGTGCTTTGTAAACCCGTTCTGACCCCTAAATTCAACCAAAATTTTTCTCATGTCAACCTTATTGTCTAAATTG
GCTAACCAAATCACAAATATCATCATCACGGTCTGCCAATTTTATTTA
AACTTGAAATGAGTTGATGTATTTATTTCAAGAATATTAAGGGCGGGAGTTGCCGCCAGATTTTGACCGGTCCGGATGAGAAAATATTGATTG
TCAACCAATTTTTGTGATTTCAGTTTTCCCGCTACAGGTCAGACGGCGCGGGGGCTAATGTTTTTAACGAGGCTTTATCATGAAGATGACAAAA
TTAAGATGGTGGATTATCGGCCTGGTCTGCGTAGGG <mark>GGATCC</mark> CTCATCTTGCTGAAAAACTCGAGCCATCCGGAAGATCTGGCGGCCGCTCTCC
${\tt CTATAGTGAGTCGTATTACGCCGGATGGATATGGTGTTCAGGCACAAGTGTTAAAGCAGTTGATTTATTCACTATGATGAAAAAAAA$
${\tt TGGAACCTGCTCCAAGTTAAAAATAGAGATAATACCGAAAACTCATCGAGTAAGATTAGAGATAATACAACAATAAAAAAATGGTTTAGAA$
${\tt CTTACTCACAGCGNGATGCTACTAATTGGGACAATTTTCCAGATGAAGTATCATCTAANAATTTAAATGAAGAAGACTTCAGAGCTTTTGTTAAATGAAGAAGACTTCAGAGCTTTTGTTAAATGAAGAAGAAGAAGAAGAAGAAGAAGAAGA$
AAATTATTTGGCAAAAATAATATAATTCGGCTGCNGGGGCGGCCTCGTGATACNCCTATTTTNTNGGTAATGNCATGATAATAATGNNNTCNTA
NACGTCNGNGNNNTTTTCGGGAANNNNNNNNCCCNANTGTTTNTTTTNCNAANNANNTCAANNTNNTCCNCTCATNNNNNNNTANCCTGANNA
NNNTNNNNNNTNAAAGNNNNNNNNNNCCNNNTNNNNNNNNTCNTTTNNNNNTNTNCNNCGNTTNNNCNNNC

Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

major	Facilitate	or Superfamily proteinmannonate dehydratase	
Query	44	GATATGGTGTAACGCCGTTACCACGCCGGTTGCGCCAGCCTGGCGTACATCTGACAGCGT	103
Sbjct	1481450	GATATGGTGTAACGCCGTTACCACGCCGGTTGCGCCAGCCTGGCGTACATCTGACAGCGT	1481391
Query	104	TACCGGGTCGTTAGGTCCGTACCAGCGCCAGGTTTGTTTCATATCTCGTTTCCTCTTCTT	163
Sbjct	1481390	TACCGGGTCGTTAGGTCCGTACCAGCGCCAGGTTTGTTTCATATCTCGTTTCCTCTTCTT	1481331
Query	164	GCGATAACGTCTTCGTGGTTGACCCATTGCCAGCCAACATCGAAACGTGCTTTGTAAACC	223
Sbjct	1481330	GCGATAACGTCTTCGTGGTTGACCCATTGCCAGCCAACATCGAAACGTGCTTTGTAAACC	1481271
Query	224	CGTTCTGACCCCTAAATTCAACCAAAATTTTTCTCATGTCAACCTTATTGTCTAAATTGG	283
Sbjct	1481270	CGTTCTGACCCCTAAATTCAACCAAAATTTTTCTCATGTCAACCTTATTGTCTAAATTGG	1481211
Query	284	CTAACCAAATCACAAATATCATCATCACGGTCTGCCAATTTTATTTA	343
Sbjct	1481210	CTAACCAAATCACAAATATCATCATCACCGGTCTGCCAATTTTATTTA	1481151
Query	344	CAATTTTTGCTGGGTGAAAAGCATTCACCATTCAACTTGAAATGAGTTGATGTATTTATT	403
Sbjct	1481150	CAATTTTTGCTGGGTGAAAAGCATTCACCATTCAACTTGAAATGAGTTGATGTATTTATT	1481091
Query	404	TCAAGAATATTAAGGGCGGGAGTTGCCGCCAGATTTTGACCGGTCCGGATGAGAAAATAT	463
Sbjct	1481090	TCAAGAATATTAAGGGCGGGGGGTTGCCGCCAGATTTTGACCGGTCCGGATGAGAAAATAT	1481031
Query	464	TGATTGGTCAACCAATTTTTGTGATTTCAGTTTTCCCGCTACAGGTCAGACGGCGCGGAG	523
Sbjct	1481030	TGATTGGTCAACCAATTTTTGTGATTTCAGTTTTCCCGCTACAGGTCAGACGGCGCGGAG	1480971
Query	524	CTAATGTTTTTTAACGAGGCTTTATCATGAAGATGACAAAATTAAGATGGTGGATTATCG	583
Sbjct	1480970	CTAATGTTTTTTAACGAGGCTTTATCATGAAGATGACAAAATTAAGATGGTGGATTATCG	1480911
Query	584	GCCTGGTCTGCGTAGGG 600	
Sbjct	1480910	GCCTGGTCTGCGTAGGG 1480894	

Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE6-00219-16, complete genome Sequence ID: <u>gb|CP009088.1|</u>Length: 4677619Number of Matches: 1

Thr operon leader peptidehexuronate transporter			
Query	44	GATATGGTGTAACGCCGTTACCACGCCGGTTGCGCCAGCCTGGCGTACATCTGACAGCGT	103
Sbjct	3176576	GATATGGTGTAACGCCGTTACCACGCCGGTTGCGCCAGCCTGGCGTACATCTGACAGCGT	3176517
Query	104	TACCGGGTCGTTAGGTCCGTACCAGCGCCAGGTTTGTTTCATATCTCGTTTCCTCTTCTT	163
Sbjct	3176516	TACCGGGTCGTTAGGTCCGTACCAGCGCCAGGTTTGTTTCATATCTCGTTTCCTCTTCTT	3176457

PhD Thesis			Appendix
Query	164	GCGATAACGTCTTCGTGGTTGACCCATTGCCAGCCAACATCGAAACGTGCTTTGTAAACC	223
Sbjct	3176456	GCGATAACGTCTTCGTGGTTGACCCATTGCCAGCCAACATCGAAACGTGCTTTGTAAACC	3176397
Query	224	CGTTCTGACCCCTAAATTCAACCAAAATTTTTCTCATGTCAACCTTATTGTCTAAATTGG	283
Sbjct	3176396	CGTTCTGACCCCTAAATTCAACCAAAATTTTTCTCATGTCAACCTTATTGTCTAAATTGG	3176337
Query	284	CTAACCAAATCACAAATATCATCATCACGGTCTGCCAATTTTATTTA	343
Sbjct	3176336	CTAACCAAATCACAAATATCATCATCACGGTCTGCCAATTTTATTTA	3176277
Query	344	CAATTTTTGCTGGGTGAAAAGCATTCACCATTCAACTTGAAATGAGTTGATGTATTTATT	403
Sbjct	3176276	CAATTTTTGCTGGGTGAAAAGCATTCACCATTCAACTTGAAATGAGTTGATGTATTTATT	3176217
Query	404	TCAAGAATATTAAGGGCGGGAGTTGCCGCCAGATTTTGACCGGTCCGGATGAGAAAATAT	463
Sbjct	3176216	TCAAGAATATTAAGGGCGGGAGTTGCCGCCAGATTTGACCGGTCCGGATGAGAAAATAT	3176157
Query	464	TGATTGGTCAACCAATTTTTGTGATTTCAGTTTTCCCGCTACAGGTCAGACGGCGCGGGAG	523
Sbjct	3176156	TGATTGGTCAACCAATTTTGTGATTTCAGTTTTCCCGCTACAGGTCAGACGGCGCGGAG	3176097
Query	524	CTAATGTTTTTTAACGAGGCTTTATCATGAAGATGACAAAATTAAGATGGTGGATTATCG	583
Sbjct	3176096	CTAATGTTTTTTAACGAGGCTTTATCATGAAGATGACAAAATTAAGATGGTGGATTATCG	3176037
Query	584	GCCTGGTCTGCGTAGGG 600	
Sbjct	3176036	GCCTGGTCTGCGTAGGG 3176020	

T7F- SEN2978

Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

major	Facilitate	or Superfamily protein mannonate dehydratase	
Query	49	CCCTACGCAGACCAGGCCGATAATCCACCATCTTAATTTTGTCATCTTCATGATAAAGCC	108
Sbjct	1480894	CCCTACGCAGACCAGGCCGATAATCCACCATCTTAATTTTGTCATCTTCATGATAAAGCC	1480953
Query	109	TCGTTAAAAAACATTAGCTCCGCCGCCGTCTGACCTGTAGCGGGAAAACTGAAATCACAAA	168
Sbjct	1480954	TCGTTAAAAAACATTAGCTCCGCGCCGTCTGACCTGTAGCGGGAAAACTGAAATCACAAA	1481013
Query	169	AATTGGTTGACCAATCAATATTTTCTCATCCGGACCGGTCAAAATCTGGCGGCAACTCCC	228
Sbjct	1481014	AATTGGTTGACCAATCAATATTTTCTCATCCGGACCGGTCAAAATCTGGCGGCAACTCCC	1481073
Query	229	GCCCTTAATATTCTTGAAATAAATACATCAACTCAATTTCAAGTTGAATGGTGAATGCTTT	288
Sbjct	1481074	GCCCTTAATATTCTTGAAATAAATACATCAACTCATTTCAAGTTGAATGGTGAATGCTTT	1481133
Query	289	TCACCCAGCAAAAATTGACACAGATCAAATAAAATAAAA	348
Sbjct	1481134	TCACCCAGCAAAAATTGACACAGATCAAATAAAATTAAAATTGGCAGACCGTGAATGATGAT	1481193

PhD T	PhD Thesis			
Query	349	ATTTGTGATTTGGTTAGCCAATTTAGACAATAAGGTTGACATGAGAAAAATTTTGGTTGA	408	
Sbjct	1481194	ATTTGTGATTTGGTTAGCCAATTTAGACAATAAGGTTGACATGAGAAAAATTTTGGTTGA	1481253	
Query	409	ATTTAGGGGTCAGAACGGGTTTACAAAGCACGTTTCGATGTTGGCTGGC	468	
Sbjct	1481254	ATTTAGGGGTCAGAACGGGTTTACAAAGCACGTTTCGATGTTGGCTGGC	1481313	
Query	469	CACGAAGACGTTATCGCAAGAAGAGGAAACGAGATATGAAACAAAC	528	
Sbjct	1481314	CACGAAGACGTTATCGCAAGAAGAGGAAACGAGATATGAAACAAAC	1481373	
Query	529	GACCTAACGACCCGGTAACGCTGTCAGATGTACGCCAGGCTGGCGCAACCGGCGTGGTAA	588	
Sbjct	1481374	GACCTAACGACCCGGTAACGCTGTCAGATGTACGCCAGGCTGGCGCAACCGGCGTGGTAA	1481433	
Query	589	CGGCGTTACACCATATC 605		
Sbjct	1481434			

8-pJET-RP2-SEN2977

Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

major	Facilitate	or Superfamily protein mannonate dehydratase	
Query	44	GATATGGTGTAACGCGGTTACCACGCCGGTTGCGCCAGCCTGGCGTACATCTGACAGCGT	102
Sbjct	1481450	GATATGGTGTAACGCCGTTACCACGCCGGTTGCGCCAGCCTGGCGTACATCTGACAGCGT	1481391
Query	103	TACCGGGTCGTTAGGTCCGTACCAGCGCCAGGTTTGTTTCATATCTCGTTTCCTCTTCTT	162
Sbjct	1481390	TACCGGGTCGTTAGGTCCGTACCAGCGCCAGGTTTGTTTCATATCTCGTTTCCTCTTCTT	1481331
Query	163	GCGATAACGTCTTCGTGGTTGACCCATTGCCAGCCAACATCGAAACGTGCTTTGTAAACC	222
Sbjct	1481330	GCGATAACGTCTTCGTGGTTGACCCATTGCCAGCCAACATCGAAACGTGCTTTGTAAACC	1481271
Query	223	CGTTCTGACCCCTAAATTCAACCAAAATTTTTCTCATGTCAACCTTATTGTCTAAATTGG	282
Sbjct	1481270	CGTTCTGACCCCTAAATTCAACCAAAATTTTTCTCATGTCAACCTTATTGTCTAAATTGG	1481211
Query	283	CTAACCAAATCACAAATATCATCATCACGGTCTGCCAATTTTATTTA	342
Sbjct	1481210	CTAACCAAATCACAAATATCATCATCACGGTCTGCCAATTTTATTTA	1481151
Query	343	CAATTTTTGCTGGGTGAAAAGCATTCACCATTCAACTTGAAATGAGTTGATGTATTTATT	402
Sbjct	1481150	CAATTTTTGCTGGGTGAAAAGCATTCACCATTCAACTTGAAATGAGTTGATGTATTTATT	1481091
Query	403	TCAAGAATATTAAGGGCGGGAGTTGCCGCCAGATTTTGACCGGTCCGGATGAGAAAATAT	462
Sbjct	1481090	TCAAGAATATTAAGGGCGGGGGGTTGCCGCCAGATTTGACCGGTCCGGATGAGAAAATAT	1481031
Query	463	TGATTGGTCAACCAATTTTTGTGATTTCAGTTTTCCCGCTACAGGTCAGACGGCGCGGAG	522
Sbjct	1481030	TGATTGGTCAACCAATTTTTGTGATTTCAGTTTTCCCGCTACAGGTCAGACGGCGCGGAG	1480971
Query	523	CTAATGTTTTTTAACGAGGCTTTATCATGAAGATGACAAAATTAAGATGGTGGATTATCG	582
Sbjct	1480970	LIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	1480911

T7F- SEN2977

NNNNNNNNNNNNNNNNNNNNGGCTCGAGTTNNNNGCAAGATGAG<mark>GAATTC</mark>CCCTACGCAGACCAGGCCGATAATCCACCATCTTAATTTTG TCATCTTCATGATAAAAGCCTCGTTAAAAAACATTAGCTCCGCGCCGTCTGACCTGTAGCGGGAAAACTGAAAATCACAAAAATTGGTTGACCAAT <mark>ACGCCAGGCTGGCGCAACCGGCGTGGTAACCGCGTTACACCATATC<mark>GGATCC</mark>GATCTTTCTAGAAGATCTCCCTACAATATTCTCAGCTGCCATG</mark> ${\tt G} {\tt A} {\tt A} {\tt A} {\tt A} {\tt C} {\tt G} {\tt A} {\tt G} {\tt A} {\tt C} {\tt A} {\tt C} {\tt C$ NGTTGNAGGTGGCGTGGGTTTTCTTGGCAATCGACTCTCATGAAAACTACNAGCTAAATATTCAATATGTTCCTCTTTGACCANCTTNATTCTG CNTTTTTTTGAANNAGGTTTANNGCANGNTTCANGAANNTT

Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome Sequence ID: <u>gb|CP008928.1|</u>Length: 4678927Number of Matches: 1

major Facilitator Super family protein manifoliate denyuratase			
Query	54	CCCTACGCAGACCAGGCCGATAATCCACCATCTTAATTTTGTCATCTTCATGATAAAGCC	113
Sbjct	1480894	CCCTACGCAGACCAGGCCGATAATCCACCATCTTAATTTTGTCATCTTCATGATAAAGCC	1480953
Query	114	TCGTTAAAAAACATTAGCTCCGCGCCGTCTGACCTGTAGCGGGAAAACTGAAATCACAAA	173
Sbjct	1480954	TCGTTAAAAAACATTAGCTCCGCGCCGTCTGACCTGTAGCGGGAAAACTGAAATCACAAA	1481013
Query	174	AATTGGTTGACCAATCAATATTTTCTCATCCGGACCGGTCAAAATCTGGCGGCAACTCCC	233
Sbjct	1481014	AATTGGTTGACCAATCAATATTTTCTCATCCGGACCGGTCAAAATCTGGCGGCGACCTCCC	1481073
Query	234	GCCCTTAATATTCTTGAAATAAATACATCAACTCATTTCAAGTTGAATGGTGAATGCTTT	293
Sbjct	1481074	GCCCTTAATATTCTTGAAATAAATACATCAACTCATTTCAAGTTGAATGGTGAATGCTTT	1481133
Query	294	TCACCCAGCAAAAATTGACACAGATCAAATAAAAAATTGGCAGACCGTGAATGATGAT	353
Sbjct	1481134	TCACCCAGCAAAAATTGACACAGATCAAATAAAATAAAA	1481193
Query	354	ATTTGTGATTTGGTTAGCCAATTTAGACAATAAGGTTGACATGAGAAAAATTTTGGTTGA	413
Sbjct	1481194	ATTTGTGATTTGGTTAGCCAATTAGACAATAAGGTTGACATGAGAAAAATTTTGGTTGA	1481253
Query	414	ATTTAGGGGTCAGAACGGGTTTACAAAGCACGTTTCGATGTTGGCTGGC	473
Sbjct	1481254	ATTTAGGGGTCAGAACGGGTTTACAAAGCACGTTTCGATGTTGGCTGGC	1481313
Query	474	CACGAAGACGTTATCGCAAGAAGAGGAAACGAGATATGAAACAAAC	533
Sbjct	1481314	CACGAAGACGTTATCGCAAGAAGAGGAAACGAGATATGAAACAAAC	1481373
Query	534	GACCTAACGACCCGGTAACGCTGTCAGATGTACGCCAGGCTGGCGCAACCGGCGTGGTAA	593
Sbjct	1481374	GACCTAACGACCCGGTAACGCTGTCAGATGTACGCCAGGCTGGCGCAACCGGCGTGGTAA	1481433
Query	594	CCGCGTTACACCATATC 609	
Sbjct	1481434	CGGCGTTACACCATATC 1481450	

major Facilitator Super family protein mannonate debudrates
9-pJET-RP2- SEN2979

NNNNNNNNNNNNNNNNNNGANNNCTTCNNNANNGNGAG <mark>GGATCC</mark> GAAGAAGAGCACCGTCGTAAAGCCGAAGGTAGCGACGATCTGATCCC
AATGCGCCCGGACCACGGTCATCAGATGCTGGACGATCTGAAGAAGAAAACGAATCCGGGTTATTCCGCCATTGGCCGTCTGAAAGGGCTTGCG
GAAGTCCGCGGCGTCGAACTGGCTATCCAGCGCGCTTTCTTT
CAATGCAATAGCAACATGCCTCGCCCCGGAGATCGCGGGGGGGAAGACGTCGAATGACAGGAGTTTGCAATGGAACAGAATATCGCCACCGCCCAG
<mark>GTTTCCGTCGCCCGCCCAAACTGGGACAAATCACGTCTGGTATCCCGTATTGTGCATCTGGGCTGCGGGGG<mark>GAATTC</mark>GTGATCTTGCTGAAAAAC</mark>
${\tt TCGAGCCATCCGGAAGATCTGGCGGCCGCTCTCCCTATAGTGAGTCGTATTACGCCGGATGGAT$
TGATTTTATTCACTATGATGAAAAAAAACAATGAATGGAACCTGCTCCAAGTTNAAAATAGAGATAATACCGAAAACTCATCGAGTAGTAAGATT
AGAGATAATACAACAATAAAAAAATGGTTTAGAACTTACTCACAGCGTGATGCTACTAATTGGGACAATTTTCCAGATGAAGTATCATCTNNGA
ATTTAAATGAAGAAGACTTTCAGAGCTTTTGTTAAAAATTATTTGGCAAAAATAATATATTCNGCTGCAGGGGCGGCCTCNNGATACGCCTAT
${\tt TTTTATAGGNTAATGTCATGATAATAANGGNTTNTTANACNTCAGGNGGCACTTTTTCNGGGAAATGTGCGNNGGAACCCCTATTGTTATTTTT$
${\tt CTAANACATTCAAATATGTATCCGNNNNNGANACATANCCTGANNAATGNTTCATAATATNAAAANNANNATNATGAGNNTTCNACATTCCNNG$
NNNNCCTNATNNCCNNTTTTGNGNNNTTTNNCNTCCNNNTTTGCTCNCCANAANNNCTGGTNAAGTAAAGANNCTGANNATCNNTNGGNGNNAN
CNANNNNTNNNNNNNNNNNNNNNNNGNANANCNNGNNNNNNCCCNNNNNNTNCNNNNNNCACTTTNTANNNNNNNATNNNNNNN

Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE6-00219-16, complete genome Sequence ID: <u>gb|CP009088.1|</u>Length: 4677619Number of Matches: 1

Thr operon leader peptidemannonate dehydratase

Query	48	GAAGAAGAGCACCGTCGTAAAGCCGAAGGTAGCGACGATCTGATCCCAATGCGCCCGGAC	107
Sbjct	3177468	GAAGAAGAGCACCGTCGTAAAGCCGAAGGTAGCGACGATCTGATCCCAATGCGCCCGGAC	3177527
Query	108	CACGGTCATCAGATGCTGGACGATCTGAAGAAGAAAACGAATCCGGGTTATTCCGCCATT	167
Sbjct	3177528	CACGGTCATCAGATGCTGGACGATCTGAAGAAGAAAACGAATCCGGGTTATTCCGCCATT	3177587
Query	168	GGCCGTCTGAAAGGGCTTGCGGAAGTCCGCGGCGTCGAACTGGCTATCCAGCGCGCTTTC	227
Sbjct	3177588	GCCGTCTGAAAGGGCTTGCGGAAGTCCGCGGCGTCGAACTGGCTATCCAGCGCGCTTTC	3177647
Query	228	TTTAGCAAATAACCTTCTTTCGCATGGCGCGACGCGTCATGCGATTTCCCCTACTCAATG	287
Sbjct	3177648	TTTAGCAAATAACCTTCTTTCGCATGGCGCGACGCGTCATGCGATTTCCCCTACTCAATG	3177707
Query	288	CAATAGCAACATGCCTCGCCCCGGAGATCGCGGGGGGAAGACGTCGAATGACAGGAGTTTG	347
Sbjct	3177708	CAATAGCAACATGCCTCGCCCCGGAGATCGCGGGCGAAGACGTCGAATGACAGGAGTTTG	3177767
Query	348	CAATGGAACAGAATATCGCCACCGCCCAGGTTTCCGTCGCCCGCC	407
Sbjct	3177768	CAATGGAACAGAATATCGCCACCGCCCAGGTTTCCGTCGCCCGCC	3177827
Query	408	CACGTCTGGTATCCCGTATTGTGCATCTGGGCTGCGGGG 446	
Sbjct	3177828	CACGTCTGGTATCCCGTATTGTGCATCTGGGCTGCGGGG 3177866	

T7F- SEN2979

Salmonella enterica subsp. enterica serovar Enteritidis strain OLF-SE6-00219-16, complete genome Sequence ID: <u>gb|CP009088.1|</u>Length: 4677619Number of Matches: 1

Thr or	beron lead	der peptidemannonate dehydratase	
Query	48	GAAGAAGAGCACCGTCGTAAAGCCGAAGGTAGCGACGATCTGATCCCAATGCGCCCGGAC	107
Sbjct	3177468	GAAGAAGAGCACCGTCGTAAAGCCGAAGGTAGCGACGATCTGATCCCAATGCGCCCGGAC	3177527
Query	108	CACGGTCATCAGATGCTGGACGATCTGAAGAAGAAAACGAATCCGGGTTATTCCGCCATT	167
Sbjct	3177528	CACGGTCATCAGATGCTGGACGATCTGAAGAAGAAAACGAATCCGGGTTATTCCGCCATT	3177587
Query	168	GGCCGTCTGAAAGGGCTTGCGGAAGTCCGCGGCGTCGAACTGGCTATCCAGCGCGCTTTC	227
Sbjct	3177588	GGCCGTCTGAAAGGGCTTGCGGAAGTCCGCGGCGTCGAACTGGCTATCCAGCGCGCTTTC	3177647
Query	228	TTTAGCAAATAACCTTCTTTCGCATGGCGCGACGCGTCATGCGATTTCCCCTACTCAATG	287
Sbjct	3177648	TTTAGCAAATAACCTTCTTTCGCATGGCGCGACGCGTCATGCGATTTCCCCTACTCAATG	3177707
Query	288	CAATAGCAACATGCCTCGCCCCGGAGATCGCGGGCGAAGACGTCGAATGACAGGAGTTTG	347
Sbjct	3177708	CAATAGCAACATGCCTCGCCCCGGAGATCGCGGGCGAAGACGTCGAATGACAGGAGTTTG	3177767
Query	348	CAATGGAACAGAATATCGCCACCGCCCAGGTTTCCGTCGCCCGCC	407
Sbjct	3177768	CAATGGAACAGAATATCGCCACCGCCCAGGTTTCCGTCGCCCGCC	3177827
Query	408	CACGTCTGGTATCCCGTATTGTGCATCTGGGCTGCGGGGG 446	
Sbjct	3177828	CACGTCTGGTATCCCGTATTGTGCATCTGGGCTGCGGGGG 3177866	

Appendix 6: Beta-galactosidase rate with error bar.



Appendix



Appendix 7: Physical maps of the transcriptional *lacZ* fusion vectors generated during this study. All plasmids included pRS1274 as the vector. Inserts are in red and proximal region of fused genes are indicated by small green arrows just upstream of *lacZ*. Maps have been drawn using the Vector NTI program.



Appendix 8: Sequencing of PCR fragments for knock out. Grey for primers sequences; Ns ambiguous nucleotides; Pink colour for mismatching nucleotides.

1432_pJET12-For

emplate plasmid pKD3, complete sequence Sequence ID: AY048742.1Length: 2804Number of Matches: 3

Query	149	CATATGAATATCCTCCTTAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAA	208
Sbjct	1044	CATATGAATATCCTCCTTAGTTCCTATTCCCGAAGTTCCTATTCTCTAGAAAGTATAGGAA	985
Query	209	CTTCGGCGCGCCTACCTGTGACGGAAGATCACTTCGCAGAATAAATA	268
Sbjct	984	CTTCGGCGCGCCTACCTGTGACGGAAGATCACTTCGCAGAATAAATA	925
Query	269	CTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACG	328
Sbjct	924	CTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACG	865
Query	329	TAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGT	388
Sbjct	864	TAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTGAGT	805
Query	389	TGTCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCA	448
Sbjct	804	TGTCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCA	745
Query	449	CCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAG	508
Sbjct	744	CCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAG	685
Query	509	AATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGA	568
Sbjct	684	AATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGA	625
Query	569	AAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTC	628
Sbjct	624	AAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTC	565

Appendix

Query	629	ATCCGGAATTACGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACC	688
Sbjct	564	ATCCGGAATTACGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACC	505
Query	689	CTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACC	748
Sbjct	504	CTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACC	445
Query	749	ACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAA	808
Sbjct	444	ACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAA	385
Query	809	ACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCT	868
Sbjct	384	ACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCT	325
Query	69	GGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCG	928
Sbjct	324	GGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCG	265
Query	929	TTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTC	988
Sbjct	264	TTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTC	205
Query	989	AGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGATGCTTAATGAATACAACAG	1045
Sbjct	204	AGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGATGCTTAATGAATACAACAG	145

1432-pJET12-REV

GTTCTTAAAAGAGAATTGTTATATA<mark>ATAAAGCACTTCAGCGACATCTTAACGGATACCCATCTTGAGCATA</mark> AATGTGTAGGCTGGAGCTGCTTCGAAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGAACTTC ATTTAAATGGCGCGCCTTACGCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTG CCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTG CGTATAATATTTGCCCATGGTGAAAAACGGGGGGGGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAAC TGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCC AGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTC ACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATA TCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGA ATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAAC GGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATA TATCAACGGTGGTATATCCAGTGATTTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAAATCTCGACAAC TCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAAC GAATGATCTTCCGTCCAGGTAGGCCCGCCAAATTCCTAACTTTCTAAAGATAGGACTTCGGATAGGACTAG GAGGAATTCAATGAACCGTTGGTACAGCCAAGTTAAAGCATTTACCTTTGGATC

Template plasmid pKD3,	complete sequence.	Sequence ID:	AY048742.1.

Query	126	TACACATCTTGAGC	185
Sbjct	13	TACACGTCTTGAGC ATTGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAG	70
Query	186	${\tt A}{\tt G}{\tt A}{\tt A}{\tt G}{\tt A}{\tt A}{\tt C}{\tt A}{\tt C}{\tt C}{\tt C}{\tt C}{\tt C}{\tt C}{\tt C}{\tt C$	245
Sbjct	71	AGAATAGGAACTTCGGAATAGGAACTTCATTTAAATGGCGCGCCCTTACGCCCCGCCCTGC	130
Query	246	CACTCATCGCAGTACTGTTGTATTCATTAAGCATCTGCCGACATGGAAGCCATCACAAA	305
Sbjct	131	CACTCATCGCAGTACTGTTGTATTCATTAAGCATCTGCCGACATGGAAGCCATCACAAA	188
Query	306	ACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATAT	365
Sbjct	189	ACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATAT	248
Query	366	TTGCCCATGGTGAAAACGGGGGGGGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAA	425
Sbjct	249	TTGCCCATGGTGAAAAACGGGGGGGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAA	308
Query	426	CTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTA	485
Sbjct	309	CTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTA	368
Query	486	GGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAAC	545
Sbjct	369	GGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAAC	428
Query	546	TGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGG	605
Sbjct	429	${\tt TGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGG$	488
Query	606	AAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCC	665
Sbjct	489	AAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCC	548
Query	666	ATACGTAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAA	725
Sbjct	549	ATACGTAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAA	608
Query	726	AACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTC	785
Sbjct	609	${\tt AACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTC}$	668
Query	786	TGGTTATAGGTACATTGAGCAACTGACTGAAAATGCCTCAAAATGTTCTTTACGATGCCAT	845
Sbjct	669	${\tt TGGTTATAGGTACATTGAGCAACTGACTGAAAATGCCTCAAAAATGTTCTTTACGATGCCAT$	728
Query	846	TGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCCATTTTAGCTTCCTTAGCT	905
_			
Sbjct	729	TGGGATATATCAACGGTGGTATATCCAGTGATTTTTTCTCCCATTTTAGCTTCCTTAGCT	788

PhD Th	nesis		
Query	906	CCTGAAAATCTCGACAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGA	965
Sbjct	789	CCTGAAAATCTCGACAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGA	848
Query	966	AAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTCGCCAAAAGTTGGCCCAGGGC	1025
Sbjct	849	AAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGC	908

dgoR-pJET12-FOR

CTCACGGAATTTCGGGAAGGGTCGAGTTTTTCAGCAAGATGTGTAACGCCTGCTTCTGATTGCCTGCTCTC CAGGCATTTGTCGCCCTGGTAAAGCCAGGCG<mark>CGCAGATTGGTCGATCCCCAGTCAATTGCGATGTAGCGAG</mark> CTGTCACATATGAATATCCTCCTTAGTTCCTATTCCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCG GCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATG AAATAAGATCACTACCGGGCGTATTTTTTGAGTTGTCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAG AAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCA GTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGA AAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTA CGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGA GCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATT CGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTC GTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGC CCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTC ATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGATGCTTAATGAATACAACAGTACTGCGATGAGTGGCA GGGCGGGGCGTAAGGCGCGCCATTTA

Template plasmid pKD3, complete sequence. Sequence ID: AY048742.1

Query	149	CATATGAATATCCTCCTTAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAA	208
Sbjct	1044	CATATGAATATCCTCCTTAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAA	985
Query	209	CTTCGGCGCGCCTACCTGTGACGGAAGATCACTTCGCAGAATAAATA	268
Sbjct	984	CTTCGGCGCGCCTACCTGTGACGGAAGATCACTTCGCAGAATAAATA	925
0	260		220
Query	269	CIGIIGAIACCGGGAAGCCCIGGGCCAACIIIIGGCGAAAAIGAGACGIIGAICGGCACG	328
Sbjct	924	CTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGCGAAAATGAGACGTTGATCGGCACG	865
Query	329	TAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTGAGT	388
Sbjct	864	TAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGT	805
-	200		4.4.0
Query	389	TGTCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCA	448
Sbjct	804	TGTCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCA	745
Ouerv	449	CCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAG	508
2			
Sbjct	744	CCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAG	685

Appendix

Appendix

Query	509	AATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGA	568
Sbjct	684	AATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGA	625
Query	569	AAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTC	628
Sbjct	624	AAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTC	565
Query	629	ATCCGGAATTACGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACC	688
Sbjct	564	ATCCGGAATTACGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACC	505
Query	689	CTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACC	748
Sbjct	504	CTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACC	445
Query	749	ACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAA	808
Sbjct	444	ACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAA	385
Query	809	ACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCT	868
Sbjct	384	ACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCT	325
Query	869	GGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCG	928
Sbjct	324	GGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCG	265
Query	929	TTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTC	988
Sbjct	264	TTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTC	205
Query	989	AGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGATGCTTAATGAATACAACAG	1048
Sbjct	204	AGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGATGCTTAATGAATACAACAG	145
Query	1049	TACTGCGATGAGTGGCAGGGCGGGGGGGGGGGGGGGGGG	
Sbjct	144	TACTGCGATGAGTGGCAGGGCGGGGGGGGGGGGGGGGGG	

dgoR-pJET12-REV

TCTCTTTATTGTAGGAGATCTTCTAGAAGATTGGCATGATAACGACGGTTGATATCACGCTAGTACTACAA AATTGCGGCGTAATTCAGCTATCGCGGTAAA<mark>GTAAGAGAGTTCACATCGAGCACAAGGACTCTCTATGACT</mark> CTCAAT</mark>TGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGAA CTTCATTTAAATGGCGCGCCCTACGCCCGCCCTGCCACTCATCGCAGTACTGTTGTATTCATTAAGCATC TGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGCGCATCAGCACCTTGTCGCCT TGCGTATAATATTTGCCCATGGTGAAAACGGCGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAA ACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGG CCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATGTGTAGAAACTGCCGGAAATCGTCGTGGTAT TCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGGAAAACGGTGTAACAAGGGTGAACACTATCCCA TATCACCGGCACGTCTTTCATTGCCCATACGTAATTCCCGGATGAGCATTCATCAGGCGGGCAAGAATGT GAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAGGCCCGTAATATCCAGCTGA

Appendix

Template plasmid pKD3, complete sequence Sequence ID: <u>AY048742.1</u>Length: 2804Number of Matches: 3

Query	147	ATTGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGA	206
Sbjct	28	ATTGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGA	87
Query	207	ATAGGAACTTCATTTAAATGGCGCGCCCTTACGCCCCGCCCTGCCACTCATCGCAGTACTG	266
Sbjct	88	ATAGGAACTTCATTTAAATGGCGCGCCCTACGCCCCGCCCTGCCACTCATCGCAGTACTG	147
010000	267	۳۳٬۹۳۵ ۳۳٬۵ ۳٬۹۶ ۵٬۹۵۵ ۳٬۹۶ ۵٬۹۹۵ ۵٬۹۹۵ ۵٬۹۹۵ ۳٬۹۶ ۵٬۹۹۹ ۳٬۹۶ ۵٬۹۹۹ ۳٬۹۶ ۵٬۹۹۹ ۵٬	226
Query	207		520
Shiat	1/0		207
SDJCL	140	I I I I I I I I I I I I I I I I I I I	207
Query	327	TCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGG	386
Sbjct	208	TCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGG	267
Query	387	GGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGG	446
Sbjct	268	GGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGG	327
010000	447		EOG
Query	447		500
Shiat	328		387
50 JCC	520		507
Query	507	CACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGT	566
Sbjct	388	CACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGT	447
Query	567	ATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGT	626
Sbjct	448	ATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGT	507
Ouerv	627	Саасастатсссататсассасстсассстсттсаттсосатасстаатсссатсас	686
guor j	01,		000
Sbjct	508	GAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAG	567
Query	687	CATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTCT	746
Sbjct	568	CATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTCT	627

Appendix

Query	747	TTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAG	806
Sbjct	628	TTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAG	687
Query	807	CAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGG	866
Sbjct	688	CAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGG	747
Query	867	TATATCCAGTGATTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACT	926
Sbjct	748	TATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACT	807
Query	927	CAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGT	986
Sbjct	808	CAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGT	867
Query	987	GCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGA	1046
Sbjct	868	GCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGA	927
Query	1047	CACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTAGGCGCGCCGAAGTTC	1106
Sbjct	928	CACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTAGGCGCGCCGAAGTTC	987
Query	1107	CTATACTTTCTAGAGAATAGGAAC	
Sbjct	988	CTATACTTTCTAGAGAATAGGAACTTCGGAATA 1020	

Appendix 9: Sequencing of pmrAB operon. Grey for primers sequences; Ns ambiguous nucleotides; Pink colour for mismatching nicleotides. Green colour for restriction sites.

pmrAB-pJET12-FOR

CGGTTTATCTTCAGATGGCTCGAGTTTTTCAGCAGATGAG<mark>GGATCC</mark>CAACATCCGCGTATCGATGAATAAA TTTCGCGCTTAAGGTTCGCTTAATCTCTCGCGGGCATACTCTCCTCCATACCTTTGGAGGAGAGCGTCATG AAAAGCTATATTTATAAAAGTTTGACGACCCTGTGTAGTGTGCTGATTGTCAGCAGTTTTATCTATGTGTG GGTCACGACGTATTAAACGCCTGTTATGCCTTTTTCAACAGCACCCAGGCACGGGTGCCTGTTCTTTCCGT ACGGTTTTGCAGGAAAAACTGTCCCTGATGTAGCTGGGTGATGCGGCTGACGATACTCAGCCCCAGGCCAA TTCCGCCATAACGGCTGTCCATCCGCACGAACGCTTCGCTTAGCTTCCCGCATTTGCTTTCATCAATACCC GGCCCCTCGTCTTCGACCGCCATAATAGCGTCGGGGTCGGCGCTAATGTGGATAGTGATATGGGTTCCTTC AGGGCTATAGCGATGCGCGTTTTCCACCAGATTTCGCAGCAGCATACGCAGTAACGTCGCGTCACCGCGCA CTACCACGTCCGCCGCACTTTCCGGCAGCAACAGAGTTTGCTGGCGCGCGTTTCCAGCATGGTGTTCAGCTCA TCGTAGGAGGGGGGGATCACATCTTCCAGCAGTTTTACTTCCTGATAATTCCCCGGAAGAGAATGACTGGCC CACGCGCGCCAGTTGCAGAAGCTGGGAGACGCCTATCCATCTCGGTCAAGACGGGCGATAAGCGGCGCGA CGCCGAGACGACGGACTCAATCTCAAGCGTGGAGCTGTGAATGGCGATCGGCGCCAGATTATCCGCCGTCC GCGCTTCCAGCTCTTTTTGCAGTTCGGCGAGCGGACGGGTAATACGCCGTACCGCCTGGTAACAAATTAGC AGCGTCAGGCTAACCATAAATACGCCGGGGACGATCAGGCTGGCGACCGCCTCCCGAATTCCGGGCATGAA TGGG

Salmonella enterica subsp. enterica serovar Enteritidis strain SE86 chromosome, complete genome. Sequence ID: <u>CP019681.1</u> Length: 4685718Number of Matches: 1			
Query	47	CAACATCCGCGTATCGATGAATAAATTTCGCGCTTAAGGTTCGCTTAATCTCTCGCGGGC	106
Sbjct	4389306	CAACATCCGCGTATCGATGAATAAATTTCGCGCTTAAGGTTCGCTTAATCTCTCGCGGGC	4389365
Query	107	ATACTCTCCCATACCTTTGGAGGAGAGCGTCATGAAAAGCTATATTTATAAAAGTTTG	166
Sbjct	4389366	ATACTCTCCTCCATACCTTTGGAGGAGAGCGTCATGAAAAGCTATATTTATAAAAGTTTG	4389425
Query	167	ACGACCCTGTGTAGTGTGCTGATTGTCAGCAGTTTTATCTATGTGTGGGTCACGACGTAT	226
Sbjct	4389426	ACGACCCTGTGTAGTGTGCTGATTGTCAGCAGTTTTATCTATGTGTGGGTCACGACGTAT	4389485
Query	227	TAAACGCCTGTTATGCCTTTTTCAACAGCACCCAGGCACGGGTGCCTGTTCTTTCCGTAC	286
Sbjct	4389486	TAAACGCCTGTTATGCCTTTTTCAACAGCACCCAGGCACGGGTGCCTGTTCTTTCCGTAC	4389545
Query	287	GGTTTTGCAGGAAAAACTGTCCCTGATGTAGCTGGGTGATGCGGCTGACGATACTCAGCC	346
Sbict	4389546	GGTTTTGCAGGAAAAACTGTCCCTGATGTAGCTGGGTGATGCGGCTGACGATACTCAGCC	4389605
Query	347	CCAGGCCAATTCCGCCATAACGGCTGTCCATCCGCACGAACGCTTCGCTTAGCTTCCCGC	406
Chiat	4280606		4290665
Ouerv	4389606		4565
Quor 1	107		
Sbjct	4389666		4389725
Query	467		526
Sbjct	4389726	CGCTAATGTGGATAGTGATATGGGTTCCTTCAGGGCTATAGCGATGCGCGTTTTCCACCA	4389785
Query	527	GATTTCGCAGCAGCATACGCAGTAACGTCGCGTCACCGCGCACTACCACGTCCGCCGCAC	586
Sbjct	4389786	GATTTCGCAGCAGCATACGCAGTAACGTCGCGTCACCGCGCACTACCACGTCCGCCGCAC	4389845
Query	587	TTTCCGGCAGCAACAGAGTTTGCTGGCGCGTTTCCAGCATGGTGTTCAGCTCATCGTAGG	646
Sbjct	4389846	TTTCCGGCAGCAACAGAGTTTGCTGGCGCGTTTCCAGCATGGTGTTCAGCTCATCGTAGG	4389905
Query	647	AGGGGAGGATCACATCTTCCAGCAGTTTTACTTCCTGATAATTCCCGGAAGAGAATGACT	706
Sbjct	4389906	AGGGGAGGATCACATCTTCCAGCAGTTTTACTTCCTGATAATTCCCGGAAGAGAATGACT	4389965
Query	707	GGCCCACGCGCCAGTTGCAGAAGCTGGGAGACGCTATCCATCATCTGGTCAAGACGGG	766
Sbjct	4389966	GGCCCACGCGCCCAGTTGCAGAAGCTGGGAGACGCTATCCATCTTGGTCAAGACGGG	4390025

PhD Thesis		Appendix	
Query	767	CGATAAGCGGCGCGACATCAACATTGTGGGTTTTTGACAATAATTCCAGATGCAAACGCA	826
Sbjct	4390026	CGATAAGCGGCGCGACATCAACATTGTGGGTTTTTGACAATAATTCCAGATGCAAACGCA	4390085
Query	827	CCCCCGACAGCGGCGTGCGTAGCTCATGGGCCACATCGGCGGTAAAAAGGCGTTCATTGT	886
Sbjct	4390086	$\tt CCCCCGACAGCGGCGTGCGTAGCTCATGGGCCACATCGGCGGTAAAAAGGCGTTCATTGT$	4390145
Query	887	CGAGCGTGGTGGTCAAACGCGTAACCAGTTGATTGATCGCCGAGACGACGGACTCAATCT	946
Sbjct	4390146	CGAGCGTGGTGGTCAAACGCGTAACCAGTTGATTGATCGCCGAGACGACGGACTCAATCT	4390205
Query	947	CAAGCGTGGAGCTGTGAATGGCGATCGGCGCCAGATTATCCGCCGTCCGCGCTTCCAGCT	1006
Sbjct	4390206	CAAGCGTGGAGCTGTGAATGGCGATCGGCGCCAGATTATCCGCCGTCCGCGCTTCCAGCT	4390265
Query	1007	CTTTTTGCAGTTCGGCGAGCGGACGGGTAATACGCCGTACCGCCTGGTAACAAATTAGCA	1066
Sbjct	4390266	CTTTTTGCAGTTCGGCGAGCGGACGGGTAATACGCCGTACCGCCTGGTAACAAATTAGCA	4390325

pmrAB-pJET12-REV

CCTTTTATTGTACGAGATCTTCTAGAGATCACGAATTCCCCACGTGTAGTTAATGTTATCGCAACAGGCCGG ATAGCGCAGGTTATCCGGCCGCCACCAACATTAAGTTCTTAAGGTTCACTTAATTTTACTTTGTCACGATT AGCGTCACCGAATCGATGGACGCATCAACATGTTAAAGCGCTTTCTTAAAAGACCTGTTCTTGGGCAAATC GCCTGGCTTCTGCTTTTTTCCTTTTATATTGCCGTCTGCCTGAACATTGCGTTCTACAAGCAGGTACTACA AGACCTACCGTTAAACTCGCTGCGCAATGTACTGGTGTTTATTTCCATGCCGGTCGTCGCGTTTAGCGTGG GGATACCACGCCCGCGGAAACCTTTGCGCTGATGACGCCGCAAATGGTGCTGACGCTGGGATTAAGCGGCG CGCCTGGTCAGCGTCCTGATCTCTATTTTATTAGTTATTCTGGTCGCCGCCTTTTTCTATAAAGATTACGC CTCGCTATTTCGAAATAATAAACAGTTGATCAAAGCGTTAAGCCCATCGAACAGTATTGTCGCCAGTTGGT CATGGTATTCGCATCAACGGCTGGCGAATTTGCCGCTGGTACGCATTGGCGAGGATGCCCATCGCAATCCA TTAATGCTGAAAAGCGATCGCAAAAACCTGACGATTCTCATCGTTGGCGAAACCTCGCGCGGCGATGATTT CTCTCTTGGCGGCTATCCGCGCGACACCAATCCGCGGCTGGCGAAAGACGATGTGATCTATTTCCCGCATA CCACCTCTTGCGGTACGGCGACCGCGATCTCCGTTCCCTGCATGTTTTCTGAAATGCCGCGCAAACTCACG AATTC CCACGTGTAGTTAATGTTATCGCAA

Salmonella enterica subsp. enterica serovar Enteritidis strain SE86 chromosome, complete genome. Sequence ID: <u>CP019681.1</u>Length: 4685718Number of Matches: 1

Query	39	${\tt CCACGTGTAGTTAATGTTATCGCAACAGGCCGGATAGCGCAGGTTATCCGGCCGCCACCA}$	98
Sbjct	4393017	${\tt CCACGTGTAGTTAATGTTATCGCAACAGGCCGGATAGCGCAGGTTATCCGGCCGCCACCA}$	4392958
Ouerv	99	ຉ຺ຨຉຠຠຉຉຌຌຠຠຒຠຠຉຉຌຌຒຠຌຉຉຠຠຠຉຎຠຠຌຌຒຨຌຌຎຨຌຏຎຨຏຌຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏຏ	158
Query			190
Sbjct	4392957	ACATTAAGTTCTTAAGGTTCACTTAATTTTACTTTGTCACGATTAGCGTCACCGAATCGA	4392898

PhD T	hesis		Appendix
Query	159	TGGACGCATCAACATGTTAAAGCGCTTTCTTAAAAGACCTGTTCTTGGGCAAATCGCCTG	218
Sbjct	4392897	TGGACGCATCAACATGTTAAAGCGCTTTCTTAAAAGACCTGTTCTTGGGCAAATCGCCTG	4392838
Query	219	GCTTCTGCTTTTTTCCTTTTATATTGCCGTCTGCCTGAACATTGCGTTCTACAAGCAGGT	278
Sbjct	4392837	GCTTCTGCTTTTTTCCTTTTATATTGCCGTCTGCCTGAACATTGCGTTCTACAAGCAGGT	4392778
Query	279	ACTACAAGACCTACCGTTAAACTCGCTGCGCAATGTACTGGTGTTTATTTCCATGCCGGT	338
Sbjct	4392777	ACTACAAGACCTACCGTTAAACTCGCTGCGCAATGTACTGGTGTTTATTTCCATGCCGGT	4392718
Query	339	CGTCGCGTTTAGCGTGGTCAATAGTGTGCTGACGCTGGCCTCATTTATTT	398
Sbjct	4392717	CGTCGCGTTTAGCGTGGTCAATAGTGTGCTGACGCTGGCCTCATTTATTT	4392658
Query	399	ACCGCTGGCCTGCGTTTTTATTCTGGTCGGCGCTGCCGCGCAGTATTTTATTTTGACTTA	458
Sbjct	4392657	ACCGCTGGCCTGCGTTTTTATTCTGGTCGGCGCTGCCGCGCAGTATTTTATTTTGACTTA	4392598
Query	459	CGGCATCATCATCGATCGTTCCATGATCGCCAATATGATGGATACCACGCCCGCGGAAAC	518
Sbjct	4392597	CGGCATCATCATCGATCGTTCCATGATCGCCAATATGATGGATACCACGCCCGCGGAAAC	4392538
Query	519	CTTTGCGCTGATGACGCCGCAAATGGTGCTGACGCTGGGATTAAGCGGCGTTCTGGCAGC	578
Sbjct	4392537	CTTTGCGCTGATGACGCCGCAAATGGTGCTGACGCTGGGATTAAGCGGCGTTCTGGCAGC	4392478
Query	579	CGTGATTGCCTTCTGGGTCAAAATCCGTCCGGCGACGCCGCGCTTACGTAGCGGGGCTTTA	638
Sbjct	4392477	CGTGATTGCCTTCTGGGTCAAAATCCGTCCGGCGACGCCGCGCTTACGTAGCGGGCTTTA	4392418
Query	639	CCGCCTGGTCAGCGTCCTGATCTCTATTTATTAGTTATTCTGGTCGCCGCCTTTTTCTA	698
Sbjct	4392417	CCGCCTGGTCAGCGTCCTGATCTCTATTTATTAGTTATTCTGGTCGCCGCCTTTTTCTA	4392358
Query	699	TAAAGATTACGCCTCGCTATTTCGAAATAATAAACAGTTGATCAAAGCGTTAAGCCCATC	758
Sbjct	4392357	TAAAGATTACGCCTCGCTATTTCGAAATAATAAACAGTTGATCAAAGCGTTAAGCCCATC	4392298
Ouerv	759	GAACAGTATTGTCGCCAGTTGGTCATGGTATTCGCATCAACGGCTGGCGAATTTGCCGCT	818
guor j			010
Sbjct	4392297	GAACAGTATTGTCGCCAGTTGGTCATGGTATTCGCATCAACGGCTGGCGAATTTGCCGCT	4392238
Querv	819	GGTACGCATTGGCGAGGATGCCCATCGCAATCCATTAATGCTGAAAAAGCGATCGCAAAAA	878
- 4			
Sbjct	4392237	GGTACGCATTGGCGAGGATGCCCATCGCAATCCATTAATGCTGAAAAAGCGATCGCAAAAA	4392178
Query	879	CCTGACGATTCTCATCGTTGGCGAAACCTCGCGCGGCGATGATTTCTCTCTTGGCGGCTA	938
-			
Sbjct	4392177	CCTGACGATTCTCATCGTTGGCGAAACCTCGCGCGGCGATGATTTCTCTCTTGGCGGCTA	4392118

PhD Thesis		Appendix	
Query	939	TCCGCGCGACACCAATCCGCGGCTGGCGAAAGACGATGTGATCTATTTCCCGCATACCAC	998
Sbjct	4392117	${\tt TCCGCGCGACACCAATCCGCGGCTGGCGAAAGACGATGTGATCTATTTCCCGCATACCAC}$	4392058
Query	999	CTCTTGCGGTACGGCGACCGCGATCTCCGTTCCCTGCATGTTTTCTGA	1058
Sbjct	4392057	CTCTTGCGGTACGGCGACCGCGATCTCCGTTCCCTGCATGTTTTCTGA	4391998
Query	1059	AC 1060	
Sbjct	4391997	AC 4391996	

Appendix 10: Overexpression constructs maps.







Appendix 11: Sequencing of PCR fragments for overexpression. Grey for primers sequences; Green colour for restriction sites; Ns ambiguous nucleotides; Pink colour for mismatching nicleotides. *Nde*I: CATATG, *Hind*III: AAGCTT. Nucleotides added to

SEN1432-21R

SEN1432-28R

dgoA-21R

GAG<mark>CATATG</mark>AAAATAACTCACATCACCACGTACCGTTTAC<mark>C</mark>TCCACGTTGGATGTTCCTGAAAATCGAAAC GGATGAAGGCGTGGTTGGCTGGGGAGAGCCGGTCATTGAAGGTCGGGCACGTACTGTAGAGGCGGCAGTAC ATGAGTTTGCCGACTACCTGATAGGGAAAGATCCGGCGCGTATCAACGACCTATGGCAGGTAATGTACCGG GCCGGTTTTTATCGCGGCGGCCCGATTATGATGAGCGCCATCGCCGGTATTGACCAGGCATTGTGGGATAT CAAAGGCAAGGTGTTGAATGCGCCGGTCTGGCAGCTCATGGGCGGCCTAGTGCGCGACAAAATCAAGGCCT ATAGCTGGGTGGGTGGCGATCGTCCGGCAGACGTCATTGACGGTATTGAAAAATTGCGCGGTATTGGTTTT CAATACCGTGGCGCAAATCCGCG<mark>AAGCTT</mark>TCGGCAGTGAAATTGAGTTTGGGCTCGACTTCCACGGTCGCG TTAGCGCGCCGATGGCGAAGGTGCTGATTAAAGAACTGGAACCCTATCGCCCGCTGTTTATTGAAGAGCCG GTGCTGGCGGAACAGGCGGAATATTATCCGCGCCTGGCAGCGCAAACGCATATTCCGATTGCCGCAGGCGA TATCCCACGCGGGCGGCATTACCGAATGCTATAAAATCGCCGGAATGGCGGAAG<mark>CATATG</mark>ATGTGGCGCTG GGTATTCCAGGAGCAGAGCATGGGCATTCACTATAACAAGGGCGCGGAGCTGCTCGACTTTGTGAAAAACA AAGAAGACTTCAGCATGGACGGCGGCTTCTTTAAACCCTTAACCAAACCGGGTCTTGGCGTAGACATTGAC GAGGCCAGGGTGATTGAACTTAGCAAAAGCGCGCGGGACTGGCGTAATCCGTTGTGGCGGCACGCTGACGG ATCGGTAGCCGAGTGG<mark>AAGCTT</mark>CAC

dgoA-28R

GAG<mark>CATATG</mark>AAAATAACTCACCACCACGTACCGTTTACCTCCACGTTGGATGTTCCTGAAAATCGAAAC GGATGAAGGCGTGGTTGGCTGGGGAGAGCCGGTCATTGAAGGTCGGGCACGTACTGTAGAGGCGGCAGTAC ATGAGTTTGCCGACTACCTGATAGGGAAAGATCCGGCGCGTATCAACGACCTATGGCAGGTAATGTACCGG GCCGGTTTTTATCGCGGCGGCCCGATTATGATGAGCGCCATCGCCGGTATTGACCAGGCATTGTGGGATAT CAAAGGCAAGGTGTTGAATGCGCCGGTCTGGCAGCTCATGGGCGGCCTAGTGCGCGACAAAATCAAGGCCT ATAGCTGGGTGGGTGGCGATCGTCCGGCAGACGTCATTGACGGTATTGAAAAATTGCGCGGTATTGGTTTT CAATACCGTGGCGCAAATCCGCG<mark>AAGCTT</mark>TCGGCAGTGAAATTGAGTTTGGGCTCGACTTCCACGGTCGCG TTAGCGCGCCGATGGCGAAGGTGCTGATTAAAGAACTGGAACCCTATCGCCCGCTGTTTATTGAAGAGCCG GTGCTGGCGGAACAGGCGGAATATTATCCGCGCCTGGCAGCGCAAACGCATATTCCGATTGCCGCAGGCGA TATCCCACGCGGGCGGCATTACCGAATGCTATAAAATCGCCGGAATGGCGGAAG<mark>CATATG</mark>ATGTGGCGCTG GGTATTCCAGGAGCAGAGCATGGGCATTCACTATAACAAGGGCGCGGAGCTGCTCGACTTTGTGAAAAACA AAGAAGACTTCAGCATGGACGGCGGCTTCTTTAAACCCTTAACCAAACCGGGTCTTGGCGTAGACATTGAC GAGGCCAGGGTGATTGAACTTAGCAAAAGCGCGCGGGACTGGCGTAATCCGTTGTGGCGGCACGCTGACGG ATCGGTAGCCGAGTGG**TAATAA</mark>AAGCTT</mark>CAC**

Appendix 12: Codon analysis report for SEN1432 generated using GenScript. The G + C content of the gene, G+C content <30% or >70% will negatively affect transcription and translation efficiency.

