

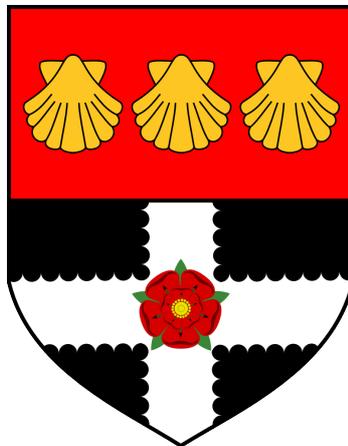
Caf1R-mediated regulation of the F1 surface antigen of *Yersinia pestis*

A thesis submitted for the degree of Doctor of Philosophy in
Molecular Microbiology

By

Dharmender Kumar

Division of Microbiology
School of Biological Sciences
University of Reading
Reading, UK



November 2015

Declaration

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

Signed...  .

Date... 13.11.2015

Abstract

This thesis investigates the fundamental aspects of regulation of expression of the surface antigen, Fraction 1 (F1), of the plague pathogen *Yersinia pestis*. F1 contains a single subunit, Caf1, polymerised into a flexible fibrillar structure. It is a key diagnostic tool and a primary component of older whole cell plague vaccines and newer subunit ones under development. Caf1 is encoded by the *caf* gene cluster, comprising *caf1R*, an AraC/XylS family regulator and the chaperone/usher *caf1MA1* operon. Despite considerable research on its assembly and role as a vaccine constituent there is virtually no information on how the expression of this key antigen is regulated.

Availability of a spontaneous Caf1R point mutation (E98G) that virtually abolished F1 assembly stimulated initial interest in regulation. Modelling of Caf1R revealed location of E98 within the DNA binding helix-6 and its involvement in a novel 'bridge-type' DNA-protein interaction linking DNA backbone bound R62 (helix-4) to helix-6. It is proposed that this interaction may be critical for the correct spatial orientation of the base-binding residues Q93 and R97 within the major DNA groove. Site-specific mutagenesis supported this model and defined the requirement of other Caf1R residues modeled as interacting with DNA backbone or specific nucleotides. Promoter-*lacZ* fusions identified a Caf1R dependent class II promoter (P_M) controlling expression of the *caf1MA1* operon and 2 potential promoters upstream of *caf1R* (P_{R^2} and P_{R^K}). P_{R^2} promoter, also a class II promoter, was autoactivated by Caf1R. Provision of Caf1R in *trans* from a P_{BAD} promoter, identified thermosensing within the *caf* locus is *via* 5' UTR of Caf1R. Potential RNA thermometer (RNAT) structures were predicted from the 5' UTR of transcripts from both promoters. The longer and more stable 5' UTR from P_{R^K} and the higher level of Caf1R independent transcription from this promoter would be consistent with an RNAT within this transcript controlling early expression of *caf1R* in response to temperature. Five potential Caf1R binding sites (repeat motifs) were identified within the *caf1R-caf1M* intergenic region, R1-3 upstream of *caf1R* and R3' and R4' upstream of *caf1M*, with the consensus sequence TGCRC^{BS1}RAMWAGCWARD^{BS2}. An electrophoretic mobility shift assay (EMSA) confirmed specific binding of Caf1R to R4'. Several tagged approaches were taken for purification of Caf1R. Low levels of soluble hCaf1R were recovered using the P_{BAD} promoter. Using the pET28a⁺ vector, hCaf1R^T was mainly recovered in inclusion bodies but low levels of soluble hCaf1R^T were recovered using codon optimisation and induction with glucose plus IPTG. Solubility of Caf1R was greatly enhanced with MBPCaf1R expressed in *E. coli*-K12 2508. While isolated Caf1R showed some activity in EMSA, comparison with activity in crude lysed cell fractions suggested either inactivation of Caf1R during isolation or loss of an activating factor.

Thus this study has identified the fundamental features controlling transcription within the *caf* locus for subsequent expression of F1. This is the first example of AraC/XylS type regulator, directly linked to a CU system of Y3-fimbriae/pili family. Working models are presented that can be used as a basis to further clarify autoregulation of Caf1R including activation, the possibility of repression and thermoregulation. This study has provided the fundamental information and tools for in-depth understanding of expression of F1 in *Y. pestis* and its applications in vaccine development.

Acknowledgements

First and foremost I would like to thank my supervisor Dr. S. MacIntyre for all her help and support over the last three years. I thank her for her guidance in my academic work and for her personal support. I have learned much from working with her, which will always remain in my memory. I would also like to thank my supervisory committee members, Prof. Simon. C. Andrews and Dr. Liam J. McGuffin who respectively helped me setting up β -galactosidase assays and modeling of DNA-Caf1R complex and were always available to give advice and feedback.

My colleague lab members, particularly Alma Lopez-Tolman and Mary Leonard who provided support, advice and also the pBADhCaf1R and pACYCF1_{SPM} plasmids for which I owe them thanks. In addition, the work of undergraduate project students, Athba Al Qahtani and Inês Garradas Domingues Alves Pereira who constructed the pRScaf1M'-R'-lacZ and pRScaf1R'-M'-lacZ plasmids. I also thank Alex Treadgold and in particular Gyles Ifill for their genuine help in optimisation trials of tagged Caf1R expression.

Mr. Kevan Hanson, senior lab technician in Knight Building helped me with general lab wares, often at short notice, and always with a good heart. I would also like to thank people in central lab facility of Knight building who provided me with glassware, media and other assistance whenever it was needed. I appreciate and would like to thank University of Reading, UK for a small financial help during the writing period of my PhD thesis.

My current lab colleague, Noratikah Othman has given me help when needed and has been a considerate colleague to work with. I would also like to thank Mr. David Johnstone who has provided friendship and advice, which has encouraged me on the PhD journey. A great thanks goes to my parents and family members for all their unconditional love and support, without it I could not have started let alone finished this stage of my studies.

My higher study in UK was made possible by a generous sponsorship from MoSJ & E Govt. of India, New Delhi and London High Commission, London (UK) for which I will always be grateful.

Last but not least; I bow my head before ALMIGHTY GOD for showering his blessing upon me always.

List of Abbreviations

Å	Angstrom
aa	Amino acid(s)
AGE	Agarose gel electrophoresis
Amp	Ampicillin
Amp ^R	Ampicillin resistance
APS	Ammonium persulfate
A _{xxx}	Absorbance at wavelength xxx nm
bp	Base pairs
BPROM	Bacterial promoter prediction program
BS	DNA binding site for protein
Caf1Rgs	Synthetic codon optimised Caf1R, purchased from GenScript, USA
Caf1R _N	N-terminal of Caf1R
CB	Coomassie Brilliant Blue R-250
Cm	Chloramphenicol
Cm ^R	Chloramphenicol resistance
CSE	Cell surface extract
CU	Chaperone Usher
CV	Column volume
DBD	DNA binding domain
dNTPs	Deoxynucleotide solution
ds	Double stranded
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
F1	Fraction 1, polymeric fibre of Caf1 subunit
fM	Formyl methionine
Fn	Fraction(s)
<i>g</i>	Relative centrifugal force
hCaf1R	His ₆ -tagged Caf1R
hCaf1R ^P	His ₆ -tagged Caf1R purchased from MyBiosource, USA
hCaf1R ^T	His ₆ -tagged Caf1R encoded in pET28a ⁺ vector
His ₆	Six histidine
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
HTH	Helix Turn Helix
IgG	Immunoglobulin G
IMAC	Immobilised-metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUPAC	International Union of Pure and Applied Chemistry
Kan	Kanamycin
Kan ^R	Kanamycin resistance
kb	Kilobase pairs
kDa	Kilo Daltons
LA	Luria-Bertani agar
LDF	Linear discriminant function
MBP	Maltose binding protein
MOPS	3-(N-morpholino) propanesulfonic acid buffer
MWCO	Molecular weight cut off
MWt	Molecular weight
NCM	Nitrocellulose membrane
NEB	New England Biolabs
nH ₂ O	Nanopure water
NI	Non-induced
nt	Nucleotide(s)
OD _{xxx}	Optical density at wavelength xxx nm
One OD unit	1 ml at an OD _{600nm} of 1

P	Pellet or insoluble fraction
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
P _M	<i>caf1M</i> promoter(s)
P _R	<i>caf1R</i> promoters
psi	Pound-force per square inch
PVDF	Polyvinylidene difluoride
QC-SDM	QuikChange site-directed mutagenesis
RBS	Ribosomal binding site
RNAP	RNA polymerase
RNAT	RNA thermometer(s)
rpm	Revolutions per minute
RT	Room temperature
S	Supernatant or soluble fraction
SD	Shine-Dalgarno
SDS	Sodium dodecyl sulphate
TAE	Tris-acetate EDTA buffer
TBST	Tris-buffered saline plus Tween-20
TEMED	Tetramethylethylenediamine
T _m	Melting temperature
TSS	Transcription start site
w.r.t.	With respect to
WB	Western immunoblot
WC	Whole cell
WT	Wild type

Amino acid	Three letter code	Single letter code	Amino acid	Three letter code	Single letter code
Alanine	Ala	A	Glutamate	Glu	E
Leucine	Leu	L	Serine	Ser	S
Arginine	Arg	R	Glutamine	Gln	Q
Lysine	Lys	K	Threonine	Thr	T
Asparagine	Asn	N	Glycine	Gly	G
Methionine	Met	M	Tryptophan	Trp	W
Aspartate	Asp	D	Histidine	His	H
Phenylalanine	Phe	F	Tyrosine	Tyr	Y
Cysteine	Cys	C	Isoleucine	Ile	I
Proline	Pro	P	Valine	Val	V

IUPAC ambiguity codes for DNA

Code	Description
M	AC
R	AG
W	AT
S	CG
Y	CT
K	GT
V	ACG
H	ACT
D	AGT
B	CGT
N	ACGT

Contents

Chapter 1: Introduction	1
1.1 An overview on <i>Yersinia pestis</i> and its F1 antigen	2
1.2 Historical perspective of plague its transmission and prevention	3
1.2.1 Plague history	3
1.2.2 Life cycle of <i>Y. pestis</i> and transmission of plague	4
1.2.3 Preventive measures.....	6
1.3 F1 capsular antigen	7
1.3.1 F1 encoding <i>caf</i> locus and regulation of F1 expression	7
1.3.2 F1 structure and assembly	9
1.4 Transcription regulation in bacteria	11
1.4.1 Regulation of expression of CU pathway assembled pili/fimbriae.....	14
1.4.2 Diversity of CU pathway assembled pili/fimbriae	15
1.4.3 Regulatory mechanisms of FGS group: Type 1 and Pap fimbriae/pili	16
1.4.4 Regulation of expression of FGL group CU systems	20
1.4.5 Regulation <i>via</i> AraC/XylS family protein-regulators	22
Aim and objectives:	32
Chapter 2: Materials and Methods	33
2.1 Introduction	34
2.1.1 Bacterial strains and culture conditions.....	34
2.1.2 Plasmids and their respective constructs.....	36
2.1.3 Chemicals and enzymes.....	42
2.1.4 Centrifuges	42
2.2 In vitro DNA techniques	42
2.2.1 Preparation of plasmid DNA.....	42
2.2.2 Quantification of plasmid DNA	42
2.2.3 Oligonucleotide primers.....	43
2.2.4 PCR amplification	50
2.2.5 Restriction digestion	53
2.2.6 Clean-up of restriction digestion reactions.....	53
2.2.7 Purification of DNA from agarose gel slice	53
2.2.8 Molecular cloning techniques	53
2.2.9 Transformation	55
2.2.10 DNA sequencing	56
2.3 Transcriptional fusions and β-galactosidase assay	56
2.3.1 Construction of promoter- <i>lacZ</i> fusion plasmids	56
2.3.2 Monitoring β -galactosidase activity	57
2.4 In vitro protein techniques	58
2.4.1 F1 analysis in whole cell.....	58
2.4.2 F1 extraction from cell surface.....	59
2.5 Expression and purification of tagged Caf1R	59
2.5.1 Expression, sampling and lysis of recombinant <i>caf1R</i> constructs in <i>E. coli</i>	59
2.5.2 Cell lysis	60
2.5.3 Clarification of cell lysates: preparation of soluble and insoluble fractions.....	60
2.5.4 Purification.....	60
2.5.5 Solubilisation of inclusion bodies by urea	61
2.6 Sodium dodecyl sulphate polyacrylamide gel electrophoresis	62
2.6.1 Preparation of discontinuous SDS-PAGE gel	62
2.7 Western Immunoblotting	63
2.8 Protein quantification	64
2.8.1 Quantification from solution	64
2.8.2 Quantification from a band on SDS-PAGE.....	64
2.9 Protein characterisation procedure	64
2.9.1 Concentration.....	64

2.9.2	Buffer exchange using viva spin concentrator.....	65
2.9.3	Buffer exchange using Mini D-tube dialyser.....	65
2.9.4	Size exclusion chromatography.....	65
2.10	Electrophoretic mobility shift assay (EMSA).....	65
2.10.1	Preparation of specific and non-specific <i>caf</i> DNA probes for use in EMSA.....	66
2.10.2	Preparation of cell lysate for EMSA.....	66
2.10.3	Purified, concentrated and dialysed protein samples used in EMSA.....	66
2.10.4	Setting up EMSA binding reaction.....	66
2.10.5	Polyacrylamide Gel Electrophoresis and Streptavidin-HRP based detection.....	67
2.11	Bioinformatic analysis.....	68
2.11.1	DNA sequence(s) analysis, primers and constructs design.....	68
2.11.2	Promoter prediction.....	68
2.11.3	Identifying potential Caf1R binding sites.....	68
2.11.4	Prediction of potential Shine-Dalgarno sequence for <i>caf1R</i> , <i>caf1M</i> and <i>caf1</i>	68
2.11.5	Prediction of potential transcriptional start site for <i>caf1R</i> , <i>caf1M</i> and <i>caf1</i>	69
2.11.6	Prediction of RNA thermometers (RNATs).....	69
2.11.7	Physiochemical properties prediction of Caf1R.....	69
2.11.8	Caf1R homologs search and their multiple sequence alignment.....	69
2.11.9	Caf1R secondary structure prediction.....	69
2.11.10	Modelling of Caf1R and its DNA-binding domain (DBD) mutants.....	69
Chapter 3: Demonstration of Caf1R as a positive regulator of the <i>caf</i> locus.....		71
3.1	Introduction.....	72
Results and Discussion.....		73
3.2	Generation of a stable construct expressing recombinant <i>caf</i> locus.....	73
3.2.1	Caf1R _{E98G} mutation is complemented by co-expression with wild type Caf1R.....	73
3.2.2	Caf1R regulator is an activator of F1 expression.....	74
3.2.3	Impact of N-terminal His ₆ -tag on Caf1R function.....	76
3.2.4	Repair of E98G mutation in pACYCF1 _{SPM}	78
3.2.5	Effect of Caf1R _{E98G} on transcriptional activity of <i>caf1R</i> and <i>caf1M</i> promoter(s).....	82
3.3	Bioinformatic analysis of Caf1R.....	84
3.3.1	Characterisation of Caf1R.....	84
3.3.2	Identification of Caf1R homologs.....	85
3.3.3	Identification of Caf1R homologs within PDB database.....	91
3.3.4	Modelling of Caf1R.....	95
3.4	Mutagenesis supports role of Glu98 in bridge interaction.....	96
3.4.1	Impact of E/G98 residues on Caf1R structure.....	96
3.4.2	E98 bridges critical DNA binding residues, R62, Q93 and R97.....	98
3.4.3	Effect of additional E98 substitutions on bridge interaction and F1 production....	101
3.5	Analysis of other key binding residues within N-terminal DBD of Caf1R.....	105
3.5.1	Localisation of other key residues modelled as interacting with DNA.....	105
3.5.2	Effect of predicted DNA binding residues of Caf1R-DBD on F1 expression.....	106
3.6	Conclusion.....	110
Chapter 4: Recombinant expression and purification of tagged Caf1R.....		113
4.1	Introduction.....	114
Results and Discussion.....		115
4.2	Expression and purification of tagged Caf1R from native gene.....	115
4.2.1	Expression and purification of hCaf1R from pBADhCaf1R.....	115
4.2.2	Expression of hCaf1R ^T and hCaf1R ^{T_N} from pET28a ⁺ based vector.....	121
4.2.3	Expression of MBPCaf1R and MBPCaf1R _N	127
4.3	Expression and purification of tagged Caf1R from the synthetic gene.....	133
4.3.1	Expression and purification of hCaf1R _{gs}	135
4.3.2	Expression and purification of MBPCaf1R _{gs}	143
4.4	Conclusion.....	146

Chapter 5: Identification and localisation of <i>caf</i> locus promoters and Caf1R binding site(s)	149
5.1 Introduction	150
Results and Discussion	151
5.2 Analysis of <i>caf1R-caf1M</i> and <i>caf1A-caf1</i> intergenic regions	151
5.2.1 Identification of potential promoter elements in <i>caf1R-caf1M</i> intergenic region... 153	
5.2.2 Identification of potential promoters elements in <i>caf1A-caf1</i> intergenic region 154	
5.2.3 Identification of potential Caf1R-binding sites in <i>caf1R-caf1M</i> intergenic region.. 155	
5.2.4 Repeat sequences within the <i>caf1A-caf1</i> intergenic region	157
5.2.5 Predicted binding sites for additional transcription factors within the <i>caf</i> locus... 157	
5.3 Promoter activity within the <i>caf</i> locus intergenic regions	158
5.3.1 Promoter activity upstream of <i>caf1R</i>	159
5.3.2 Promoter activity upstream of <i>caf1M</i>	160
5.3.3 Promoter activity upstream of <i>caf1</i>	162
5.4 Localisation of <i>caf</i> promoters and Caf1R-binding site(s)	163
5.4.1 Localisation of <i>caf1R</i> promoter and Caf1R-binding upstream of <i>caf1R</i>	163
5.4.2 Localisation of <i>caf1M</i> promoter and Caf1R-binding upstream of <i>caf1M</i>	166
5.4.3 Confirmation of P _M promoter by mutagenesis in pACYCF1	168
5.4.4 Localisation of promoter(s) upstream of <i>caf1</i>	168
5.5 <i>In vitro</i> confirmation of Caf1R binding to the R4' repeat, upstream of P_M	171
5.5.1 EMSAs with hCaf1R ^{Tgs}	171
5.5.2 EMSAs with MBPCaf1R and MBPCaf1R _N	174
5.5.3 Concentration-dependent binding of purchased hCaf1R ^P to F2-B probe	176
5.6 Conclusion	177
Chapter 6: Thermoregulation in the <i>caf</i> locus of <i>Y.pestis</i>	179
6.1 Introduction	180
Results and Discussion	182
6.2 Thermoinduction of F1 expression through Caf1R	182
6.3 Prediction of RNATs within the 5' UTR of <i>caf1R</i> and <i>caf1M</i>	185
6.3.1 Predicted RNA secondary structure within 5' UTR of <i>caf1R</i>	186
6.3.2 Predicted RNA secondary structure within 5' UTR of <i>caf1M</i>	189
6.4 Conclusion	191
Chapter 7: General discussion and future directions	193
7.1 Introduction	194
7.2 Caf1R-dependent and independent promoters of the <i>caf</i> locus	194
7.3 Caf1R residues involved in DNA binding	198
7.4 Future directions	201
Appendices	202
1. Transcriptional fusion vectors used to construct promoter- <i>lacZ</i> fusions.	203
2a. Similarity between RNAP holoenzyme of <i>Y. pestis</i> CO92 and <i>E. coli</i> K-12.....	204
2b. Promoters elements confirmed and predicted for some <i>Y. pestis</i> genes.	208
3. The <i>caf</i> locus promoters predicted upstream of <i>caf1R</i> , <i>caf1M</i> and <i>caf1</i>	209
4. Sequencing result of pRS <i>caf1R-caf1M'</i> - <i>lacZ</i>	210
5. ΔG of the predicted RNATs at 5' UTR of <i>caf1R</i> relative to transcript from P _R ²	211
6. ΔG of the predicted RNATs at 5' UTR of <i>caf1R</i> relative to transcript from P _R ^K	213
7. ΔG of the predicted RNATs at 5' UTR of <i>caf1M</i> relative to transcript from P _M	215
8. Comparison of Caf1R-DBD mutations with Ala scanning of MarA.....	219
References	220

List of Figures

Chapter 1: Introduction.....	1
Figure 1.1 India ink stained <i>Y. pestis</i> smear, indicating F1 capsular antigen on the cell surface.....	2
Figure 1.2 Life cycle and transmission of <i>Y. pestis</i> along with three types of plagues.....	6
Figure 1.3 Organisation of the <i>caf</i> locus and <i>caf1R</i> gene.....	8
Figure 1.4 F1 capsular and molecular architecture.....	9
Figure 1.5 F1 Assembly.....	11
Figure 1.6 Gene clusters/operons of some FGL and FGS group fimbriae.....	15
Figure 1.7 Regulation of Type 1 fimbriae expression.....	18
Figure 1.8 Regulation of Pap pilus expression.....	20
Figure 1.9 Caf1R-DBD closely related to metabolic regulator, XylR of <i>E. coli</i>	23
Figure 1.10 Light-switch mechanism of AraC.....	24
Figure 1.11 Transcription activation by MarA.....	26
Figure 1.12 Transcription activation by Rob.....	28
Chapter 2: Materials and Methods.....	33
Figure 2.1 DNA HyperLadder I™ (Bioline).....	43
Figure 2.2 An example of plasmid construct designing by Infusion cloning method.....	55
Chapter 3: Demonstration of Caf1R as a positive regulator of the <i>caf</i> locus.....	71
Figure 3.1 Plasmid map of pACYCF1.....	73
Figure 3.2 Caf1R _{E98G} mutation is complemented by co-expression with wild type Caf1R.....	74
Figure 3.3 Construction of pACYC-MA1.....	75
Figure 3.4 F1 expression is abolished in pACYC-MA1.....	75
Figure 3.5 Construction of pBADΔhCaf1R.....	76
Figure 3.6 N-terminal His ₆ -tag plus enterokinase cleavage site does not affect Caf1R function.....	77
Figure 3.7 Site-directed mutagenesis of pACYCF1 _{S_{PM}}	78
Figure 3.8 Comparative analysis of F1 from repaired pACYCF1 to spontaneous mutant pACYCF1 _{S_{PM}}	79
Figure 3.9 Quantification of F1 in pACYCF1 (WT) and pACYCF1 _{S_{PM}} (Caf1R _{E98G} mutant).....	81
Figure 3.10 Strategy to compare impact of wild type and mutant (E98G) Caf1R on transcription.....	83
Figure 3.11 Caf1R _{E98G} lower the transcription activation at both <i>caf1R</i> and <i>caf1M</i> promoter(s).....	84
Figure 3.12 Closest homologue of Caf1R is not associated with CU system.....	87
Figure 3.13 Identified Caf1R homologs associated with CU-system.....	88
Figure 3.14 Phylogenetic tree of Caf1R homologs from AraC/XylS family and CU-type regulators.....	89
Figure 3.15 Multiple sequence alignment of selected Caf1R homologs.....	90
Figure 3.16 Caf1R secondary structure prediction by PSIPRED.....	93
Figure 3.17 Multiple sequence alignment of Caf1R DBD with MarA, Rob and SoxS DBD.....	94
Figure 3.18 Caf1R modelling.....	96
Figure 3.19 Impact of E/G98 residues on Caf1R structure.....	98
Figure 3.20 Caf1R- R62, Q93 and R97 are essential for bridge-type interaction and F1 production.....	100
Figure 3.21 Caf1R- 98 th residue is essential for bridge-type interaction and F1 production.....	103
Figure 3.22 Caf1R _{E98K} causes reduced transcription activation at both P _M and P _R promoter(s).....	104
Figure 3.23 Caf1R- DBD mutagenesis model in complex with DNA.....	106
Figure 3.24 Quantification of F1 recovery from DH5α/pACYCF1-R mutants.....	108
Chapter 4: Recombinant expression and purification of tagged Caf1R.....	113
Figure 4.1 The pBADhCaf1R construct.....	116
Figure 4.2 Expression of native hCaf1R from <i>E. coli</i> Top10/pBADhCaf1R.....	117
Figure 4.3 Solubility analysis of hCaf1R by ultracentrifugation and 2 different cell lysis buffers.....	118
Figure 4.4 Ni ²⁺ affinity chromatography of soluble hCaf1R.....	119
Figure 4.5 Molecular size determination of hCaf1R by gel filtration chromatography.....	120
Figure 4.6 Construction of pEThCaf1R.....	122
Figure 4.7 Expression of hCaf1R ^T from <i>E. coli</i> BL21(DE3)/pEThCaf1R at two different temperatures and IPTG concentrations.....	124
Figure 4.8 Construction of pEThCaf1R _N	125
Figure 4.9 The hCaf1R ^{T_N} is insoluble following expression at both 27 and 37°C.....	126
Figure 4.10 Constructs, pMALc2-Caf1R and pMALc2-Caf1R _N	128
Figure 4.11 MBPCaf1R expression at 37°C and 25°C from <i>E. coli</i> BL21(DE3)/pMALc2-Caf1R.....	130
Figure 4.12 MBPCaf1R _N expression from <i>E. coli</i> BL21(DE3)/pMALc2-Caf1R _N	131
Figure 4.13 MBPCaf1R and MBPCaf1R _N expression from <i>E. coli</i> K12-ER2508.....	132
Figure 4.14 Pairwise sequence alignment of native <i>caf1R</i> with synthetic gene (<i>caf1Rgs</i>).....	134
Figure 4.15 The pEThCaf1Rgs construct.....	135

Figure 4.16 Expression of hCaf1R ^T gs from <i>E. coli</i> BL21(DE3)/pEThCaf1Rgs.....	136
Figure 4.17 Expression of hCaf1R ^T gs from <i>E. coli</i> BL21(DE3)/pEThCaf1Rgs at lower temperatures.....	137
Figure 4.18 Expression of hCaf1R ^T gs from <i>E. coli</i> BL21 Star TM (DE3) pLysS and SHuffle [®] T7 carrying pEThCaf1Rgs plasmid.....	138
Figure 4.19 The pBADhCaf1Rgs construct.....	140
Figure 4.20 Expression of hCaf1Rgs from <i>E. coli</i> Top10/pBADhCaf1Rgs.....	141
Figure 4.21 IMAC purification of soluble hCaf1R ^T gs from <i>E. coli</i> BL21(DE3)/pEThCaf1Rgs.....	142
Figure 4.22 Solubilisation of hCaf1R ^T gs from inclusion bodies. of <i>E. coli</i> BL21(DE3)/pEThCaf1Rgs by urea treatment.....	143
Figure 4.23 The pMALc2-Caf1R _{gs} construct.....	144
Figure 4.24 Recovery of soluble MBPCaf1Rgs from <i>E. coli</i> K12-ER2508/pMALc2-Caf1R _{gs}	145
Figure 4.25 Purification of soluble MBPCaf1R _{gs} by maltose affinity chromatography.....	146
Chapter 5: Identification and localisation of <i>caf</i> locus promoters and Caf1R binding site(s).....	149
Figure 5.1 The <i>caf</i> locus of <i>Yersinia pestis</i>	151
Figure 5.2 Sequence analysis of <i>caf1R-caf1M</i> and <i>caf1A-caf1</i> intergenic regions of <i>caf</i> locus.....	152
Figure 5.3 Additional predicted ORF within <i>caf1</i>	155
Figure 5.4 Potential Caf1R-binding motifs, within the <i>caf1R-caf1M</i> intergenic region of the <i>caf</i> locus.....	157
Figure 5.5 Promoter- <i>lacZ</i> fusion constructs of the <i>caf</i> locus intergenic regions.....	158
Figure 5.6 Promoter activity upstream of <i>caf1R</i> with and without Caf1R.....	160
Figure 5.7 The <i>cis</i> and <i>trans</i> constructs used to test influence of Caf1R on P _M promoter(s) activity.....	160
Figure 5.8 Promoter activity upstream of <i>caf1M</i> with and without Caf1R.....	162
Figure 5.9 Promoter activity upstream of <i>caf1</i> with and without Caf1R.....	163
Figure 5.10 Localisation of <i>caf1R</i> promoter and Caf1R-binding site(s) upstream of <i>caf1R</i>	165
Figure 5.11 Localisation of <i>caf1M</i> promoter and Caf1R binding upstream of <i>caf1M</i>	167
Figure 5.12 Loss of F1 production on mutation of P _M promoter.....	168
Figure 5.13 Localisation of <i>caf1</i> promoter and lack of effect of Caf1R.....	170
Figure 5.14 Biotin tagged specific (F2-B) and non-specific (F1-B) EMSA probes.....	171
Figure 5.15 EMSA with cell lysates of wild type and mutated hCaf1R ^T gs.....	173
Figure 5.16 <i>In vitro</i> binding of IMAC purified hCaf1R ^T gs to Caf1R specific F2-B probe.....	174
Figure 5.17 EMSA with cell lysates of MBPCaf1R and MBPCaf1R _N	175
Figure 5.18 <i>In vitro</i> binding of purified MBPCaf1R _{gs} to Caf1R specific F2-B probe.....	176
Figure 5.19 Concentration-dependent binding of purchased hCaf1R ^P to Caf1R specific F2-B probe.....	177
Chapter 6: Thermoregulation in the <i>caf</i> locus of <i>Y.pestis</i>.....	179
Figure 6.1 General principle of RNAT mediated control of protein expression.....	181
Figure 6.2 Dependency of thermoregulation of F1 on <i>caf1R</i> 5' UTR.....	184
Figure 6.3 Caf1R-mediated thermoinduction at P _{R²} , P _{R^K} and P _M promoters.....	185
Figure 6.4 DNA fragment used to predict RNATs at 5' UTR of <i>caf1R</i> and <i>caf1M</i>	186
Figure 6.5 Predicted RNATs at 5' UTR of <i>caf1R</i>	188
Figure 6.6 Predicted RNATs at 5' UTR of <i>caf1M</i>	190
Chapter 7: General discussion and future directions.....	193
Figure 7.1 Caf1R-mediated transcription activation at the class II P _M promoter.....	195
Figure 7.2 Working model of Caf1R- independent and dependent transcription of <i>caf1R</i>	197
Figure 7.3 Modelled interactions of Caf1R residues at R4' repeat.....	199
Appendices.....	202
Figure A.1 Transcriptional fusion vectors used to construct promoter- <i>lacZ</i> fusions.....	203
Figure A.2a Similarity between RNAP holoenzyme of <i>Y. pestis</i> CO92 and <i>E. coli</i> K-12.....	207
Figure A.2b Promoter elements confirmed (light) and predicted (dark) for some <i>Y. pestis</i> genes.....	208
Figure A.3 The <i>caf</i> locus promoters predicted upstream of <i>caf1R</i> , <i>caf1M</i> and <i>caf1</i>	209
Figure A.4 Sequencing result of pRS <i>caf1R-caf1M'</i> - <i>lacZ</i>	210
Figure A.5 ΔG of the predicted RNATs at 5' UTR of <i>caf1R</i> relative to transcript from P _{R²}	212
Figure A.6 ΔG of the predicted RNATs at 5' UTR of <i>caf1R</i> relative to transcript from P _{R^K}	214
Figure A.7 ΔG of the predicted RNATs at 5' UTR of <i>caf1M</i> relative to transcript from P _M	218
Figure A.8 Site-specific mutagenesis of Caf1R-DBD with respect to MarA.....	219

List of Tables

Chapter 2: Materials and Methods	33
Table 2.1 <i>E. coli</i> strains used in this study.....	34
Table 2.2 Plasmid constructs used and designed during this study	36
Table 2.3 Oligonucleotide primers used in PCR reactions	45
Chapter 3: Demonstration of Caf1R as a positive regulator of the <i>caf</i> locus	71
Table 3.1 Predicted structural homologs of Caf1R identified in PDB database.	92
Table 3.2 Impact of Caf1R-DBD site-specific mutations on the expression of F1.	109

Chapter 1

Introduction

1.1 An overview on *Yersinia pestis* and its F1 antigen

Yersinia pestis, the causative agent of plague, has been one of the most devastating diseases in human history. *Y. pestis* belongs to the *Enterobacteriaceae* family of Gram-negative bacteria and possesses a rod like structure of 1-3 μm in length and 0.5-0.8 μm in diameter, which is surrounded by a unique capsular material known as surface antigen, F1 (**Fig. 1.1**) (MacIntyre, 2004; Miller et al., 1998; Runco et al., 2008). F1 is encoded by the *caf* locus (consist of *caf1R* regulator and *caf1MA1* operon) of the *Y. pestis* specific virulence plasmids, pFra (also known as pMT1). The typical size of this plasmid is ≈ 100 kb; 96.21 kb for pFra from *Y. pestis* CO92 (Hu et al., 1998) and 100.98 kb for pMT1 from *Y. pestis* KIM5 (Filippov et al., 1990; Lindler et al., 1998; Prentice et al., 2001). *Y. pestis* evolved from its progenitor species *Y. pseudotuberculosis*, around 15,00-20,000 years ago *via* several distinct genetic gain and losses, resulting in a very different mode of pathogenicity and transmission (Achtman et al., 1999; Morelli et al., 2010; Reuter et al., 2014; Skurnik et al., 2000). *Y. pseudotuberculosis* is transmitted by the faecal-oral route and causes mild and self-limiting gastroenteritis; in contrast, *Y. pestis* is transmitted by flea bites directly into subcutaneous tissues causing bubonic or septicemic plague or by aerosols resulting to pneumonic plague (Perry and Fetherston, 1997). Genomic analysis indicates that *Y. pestis* shares $\geq 97\%$ nucleotide identity in about 75% of their genes with *Y. pseudotuberculosis* (Achtman et al., 1999; Reuter et al., 2014; Skurnik et al., 2000).

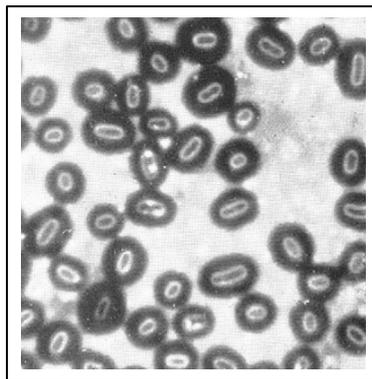


Figure 1.1| **India ink stained *Y. pestis* smear, indicating F1 capsular antigen on the cell surface.** Picture taken from (Amies, 1951).

F1 is a key detection tool (WHO manual) and extensive research has been carried out over the last 20-25 years using F1 as a primary component of both subunit and whole cell (attenuated and killed) vaccines to combat plague (Derbise et al., 2012; Hart et al., 2012; Rocke et al., 2008; Titball et al., 1997). F1 is unique to *Y. pestis*, likely to be required for transmission from the flea vector to the mammalian hosts and thus establishment of infection in a mouse model of bubonic pelage (Sebbane et al., 2009). Because of its uniqueness to *Y. pestis* it is being used to identify bacterium and monitoring anti-F1 antibodies from the recovering patients of plagues or the ancient human remains of plague (Bianucci et al., 2007; Chanteau et al., 2000). F1 in concert with Type three secretion system (T3SS) protects *Y. pestis* from the host immune by its antiphagocytic property, as shown by comparison of native EV76 strain-mediated resistance to phagocytosis in the macrophage-like cell line J774 to the reduced level of phagocytosis by mutant strain carrying in-frame deletion in the *caf1M* chaperone gene (Du et al., 2002). It has been always arguable that F1 role in virulence is not critical as naturally occurring F1 negative

'virulent' strains are known and these strains along with strains carrying genetically defined nonpolar mutations in the *caf1* structural gene retained their virulence for mice and nonhuman primates (Friedlander et al., 1995). African green monkeys following 'aerosol' exposure to F1 positive and negative strain (CO92 and CO91-C12, respectively) all died between 4 to 10 days of exposure and had lesions consistent with primary pneumonic plague, suggesting F1 is not an essential virulence factor (Davis et al., 1996). A recombinant subunit vaccine composed of F1 and V antigens (rF1+V) had been shown to protect mice against bubonic and pneumonic plague, and also shown to give a better protection than F1 or V antigen alone carrying vaccines against the F1 positive strains only (Heath et al., 1998). These results indicate a requirement of designing new vaccines to protect from the F1 negative virulent strains and suggest F1 alone can't be an efficient target to boost immune response albeit antibodies to F1 contributes to protection. Likewise, in one another study, F1 negative strains were found fully virulent in animal models (BALB/c mice) of bubonic and pneumonic plague and broke the immune responses generated with live-attenuated strains or F1 subunit vaccines (Quenee et al., 2008). In a relatively new study, intranasal instillation of guinea pigs by fully virulent strain of *Y. pestis* CO92 showed that expression of the F1 was not required for development of the pneumonic plague; however, it appeared to be essential for establishment of the bubonic plague (Quenee et al., 2011). Immunisation of guinea pigs with recombinant subunit vaccine, carrying F1 and V antigens (rF1+V), generated robust humoral immune responses in contrast to vaccines carrying either V or F1 antigen alone, partial protection by V antigen carrying vaccine against pneumonic plague and no protection by recombinant F1 carrying vaccine (Quenee et al., 2011). Again these results suggest F1 is not a required virulence factor; however, it should be considered that virulence may also be dependent on the animal models used in these studies hence, it is not so clear that the F1 could fulfil the requirement of molecular Koch's postulates (Falkow, 1988).

1.2 Historical perspective of plague its transmission and prevention

1.2.1 Plague history

Plague has been responsible for millions of deaths over the last 15 centuries and today approximately 1000-3000 cases of plague arise each year (Ninh, 2011). Three main pandemics of plague are documented, Justinian plague (6th century A. D.), Black Death (14th century A. D.) and bubonic plague (1665-till date), causing an estimated 137 million deaths worldwide reviewed in (Butler, 2009; MacIntyre, 2004; Morelli et al., 2010). The 19th century pandemic (originating in 1855 in Yunnan province, China) killed over 12 million people in India and China alone (Cohn, 2003), together with continued 20th century plague outbreaks reported from India, Madagascar and Africa, suggest that plague is not a disease of the past and will remain a threat in the future, hence requiring continued study (Butler, 2009). Despite attempts to eradicate plague it is still endemic in certain regions of the world. Australia and European countries are currently plague-free but regions in Africa, Asia and the America have experienced epidemics in recent decades (Ninh, 2011). The most recent presumptive positive case of plague was reported by California Department of Public Health (USA), involving a person from Georgia who visited Yosemite National Park, California, USA in August, 2015 (Botelho, 2015). According to the World Health Organisation (WHO), the most recent outbreak of plague was reported in China (July 2014) where a man died of bubonic plague and 30,000 residents of Yumen City were

quarantined from the outside world. In 2013, a plague outbreak was reported in Madagascar with 84 reported plague cases and 32 deaths (Associated Press in Antananarivo, 2013). In 2012, there were 256 cases of plague and 60 deaths reported in Madagascar (BBC, 2013). In the same year (July 2012), The Guardian newspaper reported an Oregon State citizen recovering from plague after contracting the disease while trying to take a mouse from the jaw of a choking cat (Quinn, 2012). In 2010, the Ministry of Health, Peru reported 17 cases of plague of which 4 were pneumonic, 12 were bubonic and one was septicemic. During this investigation, 10 distinct clones of *Y. pestis* were isolated from rodents and domestic cats as well as from humans (Haensch et al., 2010). Albeit recent plague cases are relatively low in numbers, *Y. pestis* is an organism of high concern due to the fact it is so virulent, the emergence of multi-drug resistant strains (Welch et al., 2007) and potential misuse as a bioterrorist weapon combined with the continued fear of this bacterial disease (MacIntyre, 2004; Rollins et al., 2003). Taking these facts into account, the Center for Disease Control and prevention (CDC) classified *Y. pestis* as an 'A' class pathogen requiring high attention (Inglesby et al., 2000).

1.2.2 Life cycle of *Y. pestis* and transmission of plague

Y. pestis is a zoonotic pathogen, primarily transmitted to humans and animals by an infected flea-vector bites (for example bite from the rat flea, *Xenopsylla cheopis*) (Chouikha and Hinnebusch, 2012; Hinnebusch, 2003, 2005; Sebbane et al., 2009; Stenseth et al., 2008) as depicted in **Fig. 1.2a**. Briefly, once fleas acquire *Y. pestis* through their blood meal from the infected mammalian host, the bacterium establishes in the flea midgut where it expresses selected colonisation factors such as the biofilm generating Hemin storage locus (Hms) and a phospholipase D (Ymt) (Chouikha and Hinnebusch, 2012; Hinnebusch, 2003, 2005; Zhou et al., 2012). The biofilm extends into the proventriculus and oesophagus, and blocks the digestive tract. This blockage forces the flea to regurgitate the contaminated blood meal when feeding on the next host, thus transmitting *Y. pestis* to a new host (Chouikha and Hinnebusch, 2012; Hinnebusch, 2003, 2005). With every attempt to feed about 25,000-100,000 bacteria are regurgitated into the host skin (Reed et al., 1970). During plague epizootics, many rodents including rats die, causing *Y. pestis* loaded hungry fleas to seek an alternate source of blood meal. As a result of this, humans or animals in close proximity are at high risk of being infected from the flea bites (CDC, 2012). Following *Y. pestis* loaded flea bite(s), bacteria are transmitted from cutaneous lymphatics to the regional lymph nodes and multiply quickly. Once established in lymphoid tissues bacteria replicate primarily extracellularly within necrotic lesions or abscesses and resist host macrophages and neutrophil mediated phagocytosis (Pujol and Bliska, 2005). During the very early stage of infection, *Y. pestis* enters into macrophages and neutrophils by active invasion or by passive phagocytosis (Lukaszewski et al., 2005). Neutrophils normally kill the bacterial cells but *Y. pestis* can survive within macrophages by subverting some macrophage functions for example inhibition of production of Nitric oxide (Pujol and Bliska, 2005). Invasion of *Y. pestis* into nonprofessional phagocytes or epithelial cells is mediated by expression and binding of adhesive factors, reviewed in (Ke et al., 2013). These include, the chromosomally encoded outer membrane protein, Ail (attachment-invasion locus) (Miller et al., 2001), the 9.5 kb pPCP1 plasmid encoded Plasminogen activator (Pla, a 17.5 kDa outer membrane protein of the omptin family of proteases) (Kukkonen and Korhonen, 2004; Lahteenmaki et al., 2001), the chromosomally encoded pH6 antigen, Psa (Bartra et al., 2008; Huang and Lindler, 2004) and a

chromosomally encoded adherence factor, yadBC (Forman et al., 2008; Sha et al., 2008). Upon entering into macrophages, *Y. pestis* express an alternate set of virulence factors. These include the pCD1 plasmid (≈ 70 kb) encoded type 3-secretion machinery (T3SS), *Yersinia* effector proteins (Yops) and the low-calcium response V antigen, LcrV (Pettersson et al., 1999; Plano and Schesser, 2013). Expression of T3SS genes is enhanced by low calcium (Bi et al., 2009) and controlled by the AraC/XylS-like regulator LcrF (Plano and Schesser, 2013). During the intracellular phase, the pFra plasmid encoded F1 is initially expressed (Fukuto et al., 2010). After multiplying within macrophages, *Y. pestis* escapes from the cells and develops resistance to phagocytosis. T3SS injects Yops through a needle like injectisome into neutrophils, macrophages and dendritic cells leading to cytokine activation and apoptosis (Marketon et al., 2005). The antiphagocytic ability of *Y. pestis* has been attributed to the expression of F1, pH6 antigen and Yops of T3SS (Benedek et al., 2004; Du et al., 2002; Ke et al., 2013). This allows the bacteria to exist and multiply rapidly in the extracellular environment.

Following infection, the host immune response destroys most bacterial cells; however survival of some *Y. pestis* cells within macrophage and migration to the regional lymph nodes, causes a remarkably tender inflammation in the lymphatic tissues. These inflammations are round oval-shaped structures, known as buboes (**Fig. 1.2b (i)**), and the plague caused from these buboes is known as bubonic plague, the most common primary manifestation (Inglesby et al., 2000). Bursting of buboes causes bleeding under the skin and organs and release of *Y. pestis* into the blood where it multiplies. This form of plague is known as septicemic plague. The symptoms of septicemic plague are similar to those of the bubonic form with an additional necrosis leading to blackness on the fingers (**Fig. 1.2b (ii)**). Septicemic plague has an even higher rate of mortality. When bacteria from septicemic stage infect lungs and cause pneumonia then this form of plague is known as pneumonic plague (**Fig. 1.2b (iii)**). Pneumonic plague is the most contagious and devastating form of plague as it can be communicated *via Y. pestis* charged respiratory droplets from person to person (**Fig. 1.2a**). Pneumonic plague is thought to lead to 100% mortality rate if left untreated. The pneumonic route of plague is the route that has been of concern with respect to its potential use as a bioweapon (**Fig. 1.2a**) (Morelli et al., 2010).

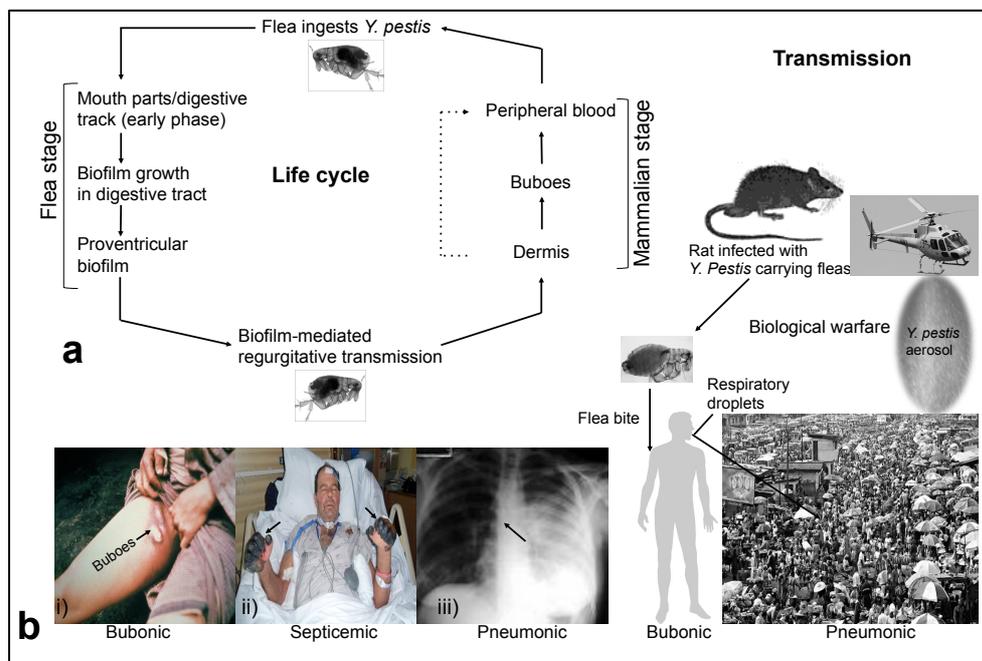


Figure 1.2| **Life cycle and transmission of *Y. pestis* along with three types of plagues.**

a) Arrows indicate flea and mammalian stages of life cycle and transmission of *Y. pestis*. Pictures adapted from (Chouikha and Hinnebusch, 2012; Cole and Buchrieser, 2001). **b)** Three forms of plague caused by *Y. pestis* in human. i). Bubonic plague, indicated by buboes on thigh. ii) Septicemic plague, an Oregon patient recovering from plague. iii). Pneumonic plague, an X-ray picture of pneumonic plague infected lungs. Pictures, i) and iii) taken from CDC image library ii) is taken from The Guardian newspaper (Quinn, 2012).

1.2.3 Preventive measures

In the early stage of infection, plague can be controlled by antibiotics. Streptomycin has been the most commonly used and effective antibiotic against *Y. pestis*. Gentamycin, tetracycline, chloramphenicol, or combinations of antibiotics have alternatively been used. More recently, levofloxacin, ciprofloxin and doxycycline have also been recommended to treat plague cases (CDC, 2015). However, once infection is established antibiotics are less successful. In addition, isolation of multi-drug resistant strains of *Y. pestis*, there is concern over excessive use of antibiotics associated with re-emergence of plague (Galimand et al., 1997; Guiyoule et al., 2001). To control plague, live attenuated and killed whole cells vaccines were used in the past (Feodorova and Motin, 2012; MacIntyre, 2004; Titball and Williamson, 2001). But a number of local and systemic side effects were observed with both types of vaccine, for example, strong pain at the site of injection, swelling, regional lymphadenopathy, malaise, headache, giddiness, anorexia, weakness and mild fever (Feodorova and Motin, 2012; MacIntyre, 2004; Titball and Williamson, 2001). In addition, killed whole cell vaccine is unable to protect against the pneumonic form of plague and both live attenuated and killed whole cell vaccines require an annual booster. Hence, there has been a strong focus on development of less reactogenic, safe, stable and efficient vaccines to combat plague (Feodorova and Motin, 2012). There has been particular focus on subunit vaccines and safe designed attenuated vaccines. Subunit vaccines have been primarily focused on 2 major antigens, F1 capsular antigen and LcrV, produced as recombinant proteins. Both have been shown to induce a protective immune response when used individually and a synergistic response when combined (Titball and Williamson, 2001). The recombinant F1-V fusion (rF1-V) provides stronger protection than killed whole cells and live attenuated vaccines, when used in mice (Rocke et al., 2008). Additionally, the rF1-V fusion

vaccine is being evaluated in Phase 2b human clinical trials (Hart et al., 2012). Several other new strategies are being developed in order to produce a successful vaccine, such as using heterologous bacterial, viral and plant delivery systems (Del Prete et al., 2009; Feodorova and Corbel, 2009). Among these *Salmonella* is the most characterised bacteria, used as a heterologous live carrier platform. Attenuated *Salmonella enteritica* Serovars *Minnesota*, *Typhimurium*, and *Typhi* have been used to express F1, LcrV and a combination of both (Andrews et al., 1996; Motin et al., 1994). A study from Carniel and coworkers suggests that a single dose of live attenuated *Y. pseudotuberculosis* V674F1, expressing recombinant *Y. pestis* F1, gives protection against pneumonic plague (Derbise et al., 2012). As plague is still endemic in some parts of the world and can be used as a bioweapon, there is continued interest in the development of next-generation vaccines including live attenuated and subunit vaccines.

1.3 F1 capsular antigen

1.3.1 F1 encoding *caf* locus and regulation of F1 expression

Initial sequencing identified that the pFra plasmid encoded *caf* locus (5.128 kb) (**Fig. 1.3a**) has three genes, *caf1M* (777 bp), *caf1A* (2502 bp) and *caf1* (513 bp), organised into two possible transcriptional units (in the same orientation), one for *caf1M* and *caf1A*, and the other for *caf1* alone (Galyov et al., 1990). These genes code for a periplasmic chaperone, Caf1M (26.3 kDa), an outer membrane usher protein, Caf1A (90.5 kDa), both required for F1 assembly, and the F1 structural subunit, Caf1 (15.5 kDa) (Chapman et al., 1999). Further an additional divergently orientated transcriptional unit of 924 bp was identified at 327 bp upstream from the *caf1M* start codon (Karlyshev et al., 1992). This encodes a protein of 301 amino acids with homology to the AraC/XylS family of DNA-binding transcriptional regulators and was shown to be required for F1 assembly. Hence, it was named *caf1R*, encoding Caf1R regulator (36.04 kDa). Bioinformatic analysis of Caf1R identified two domains; a DNA-binding domain (DBD, 1-112 amino acids) and a C-terminal domain (113-301 amino acids). Unlike many AraC/XylS family proteins, Caf1R-DBD constitutes the N-terminal half rather than the C-terminal half of the protein (Schuller et al., 2012). Two different start sites have been annotated for Caf1R, one start from ATG (Met) codon and codes for a 301 amino acid long product (**Fig. 1.3b**) of 36.05 kDa (Karlyshev et al., 1992). The second start site was annotated, six codons upstream, starting at the rare start codon TTG (Leu or ^fMet) (**Fig. 1.3b**). This would encode a 307 amino acid long product of 36.82 kDa (Parkhill et al., 2001; Prentice et al., 2001; Song et al., 2004; Zhou et al., 2004). Three possible promoters and two inverted repeat motifs (IR1 and IR2) were identified (Karlyshev et al., 1992) within the *caf1R-caf1M* intergenic region (**Fig. 1.3b**). The -10 and -35 elements of two predicted promoters were located upstream of *caf1M*, suggesting that either of these could be the active promoter for *caf1M* or *caf1MA1* transcription. The third promoter was a predicted promoter for *caf1R* (**Fig. 1.3b**). A distinct transcriptional unit for *caf1* alone was predicted (Galyov et al., 1990) and it was speculated that the small intergenic region (80 bp) between *caf1A* and *caf1* might be an additional regulatory site of the *caf* locus controlling the high level of F1 production (Galyov et al., 1990; MacIntyre, 2004). There has been little specific focus on the regulation of F1 expression. However, it has been well documented that production of F1 is temperature regulated, a high level at 35-37°C and none at lower temperature of 26°C, correlating with production in mammalian hosts but not in the flea vector (Simpson et al., 1990). More recent

transcriptomic studies have confirmed upregulation of the *caf* genes at 37°C and the fold increase upon temperature transition was highest for *caf1* in the entire genome (Chauvaux et al., 2007; Han et al., 2007; Li et al., 2011; Motin et al., 2004; Simpson et al., 1990). A transcriptome analysis of *Y. pestis* grown under different conditions found upregulation of *caf* genes in the presence of low Mg²⁺ concentration, nutrient deficiency and low pH, conditions mimicking life inside a macrophage (Han et al., 2007) and analysis of *Y. pestis* mRNA expression inside a macrophage, revealed a time dependant induction of *caf* mRNA (Fukuto et al., 2010).

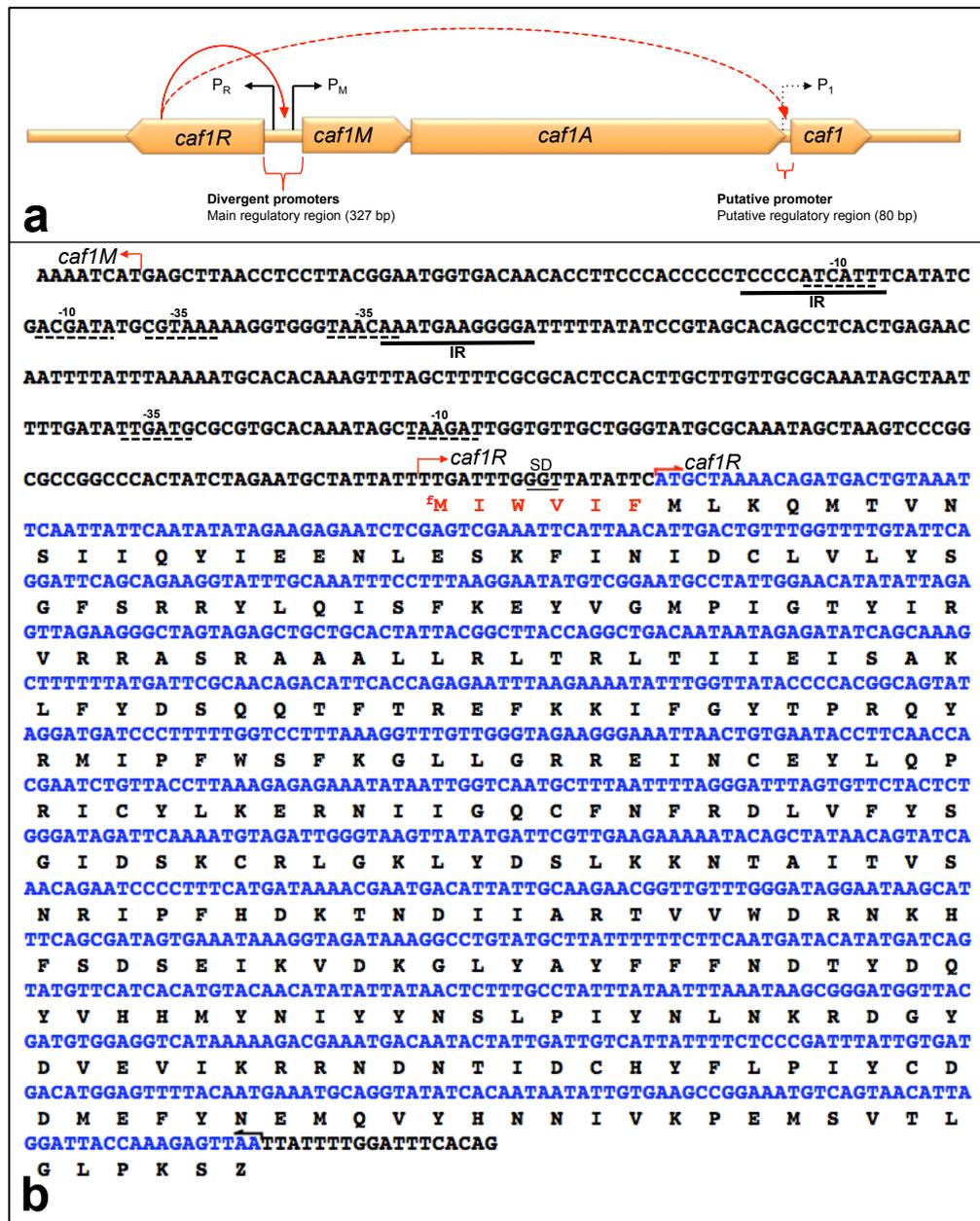


Figure 1.3| Organisation of the *caf* locus and *caf1R* gene.

a) *Y. pestis* pFra plasmid (≈100 kb) encoded *caf* locus (5.128 kb), solid red arrow indicates proposed activation of *caf* genes by Caf1R and dotted red arrow indicates speculation that Caf1R might also influence expression from the *caf1A-caf1* intergenic region. Proposed and putative promoter for *caf1R*, *caf1M* and *caf1* are indicated by P_R, P_M and P₁, respectively. **b)** The intergenic region between *caf1R-caf1M* and the coding sequence of *caf1R* with deduced amino acids is depicted. Thick lines indicate inverted repeat (IR) motifs and dashed lines represent the predicted -10 and -35 elements of putative *caf1R* and *caf1M* promoters. Red arrows indicate the assigned start sites for *caf1R* and *caf1M*. Two different start sites, ATG (Met or M) (PubMed ID 1633857) together with SD sequence (underlined) (Karlyshev et al., 1992) and TTG (Leu or ^fM) (PubMed ID 11586360) (Parkhill et al., 2001) are indicated with red bent arrows, which encode 301 and 307 amino acid long product with a calculated molecular weight of 36.05 and 36.82 kDa, respectively.

1.3.2 F1 structure and assembly

F1 has a capsular appearance as it excludes India ink, but it actually consists of a mesh of very thin fibres, collapsing together to act as a protective coat surround the cell (**Fig. 1.4a**). F1 is a high molecular weight ($0.5-1.0 \times 10^3$ kDa) thin fibrillar polymer, made of thousands of repeating Caf1 subunits (MacIntyre, 2004; MacIntyre et al., 2001; Zavialov et al., 2003). A high-resolution crystal structure of two Caf1 subunits bound to Caf1M chaperone (Caf1M-Caf1'-Caf1'') was solved (Zavialov et al., 2003) and this together with earlier mutagenic studies (MacIntyre et al., 2001; Zavialov et al., 2002) demonstrated the structural basis of these long polymers. Each subunit within the polymer is linked together by a non-covalent donor strand complementation interaction. Based on the assumption that the subunit-subunit orientation is retained throughout the F1 polymer, the polymer was modelled as a thin extended linear fibre of nearly 20 Å diameter, wound in an open right-handed helical structure (**Fig. 1.4b**). The modelled fibre has a wavy appearance with a rise of ~47 Å per subunit (roughly the length of a single Caf1 subunit) and ~2.7 subunits/turn.

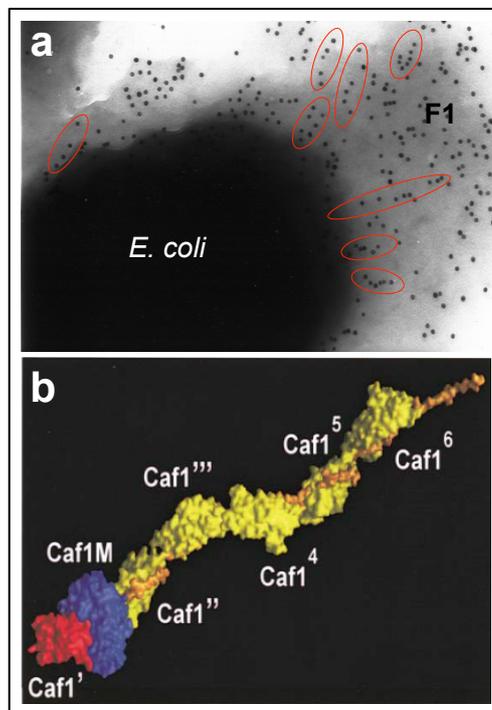


Figure 1.4| **F1 capsular and molecular architecture.**

a) Transmission electron micrograph of recombinant *E. coli* expressing F1 capsular antigen, made up of thin fibers (indicated in red circles). F1-labeled with anti-F1 antibody and gold labeled anti IgG (MacIntyre, S., unpublished data). **b)** F1 fibre, modelled on the crystal structure of periplasmic chaperone (Caf1M, blue)-subunit (Caf1', red)-subunit (Caf1'', yellow) complex, generated by assuming same orientation between successive subunits (Caf1''' to Caf1⁶ in picture). An orange colored string indicates donor β -strand. Image taken from (Zavialov et al., 2003).

F1 is assembled by a bacterial chaperone-usher (CU) pathway (Di Yu et al., 2012; MacIntyre et al., 2001; Zavialov et al., 2003; Zavialov et al., 2002), one of the most common assembly mechanisms of pili in bacteria belonging to the *Enterobacteriaceae* family (Busch and Waksman, 2012; Choudhury et al., 1999; Drummelsmith and Whitfield, 2000; Phan et al., 2011). Briefly, following synthesis of the F1 subunit (Caf1) in the bacterial cytoplasm, the subunit is transported across the inner membrane *via* the Sec pathway and enters the periplasm (**Fig. 1.5a, iia**). The newly synthesised subunits would be predicted to have an incomplete immunoglobulin (Ig)-like fold of only six β -strands instead of the seven required to complete the fold. This would lead to an exposed hydrophobic groove or cleft in the subunit. On this account the subunits on their own are very unstable, tend to aggregate (misfold) and are rapidly degraded (**Fig. 1.5a, iib**). To avoid degradation and assists in stabilisation of Caf1 subunit, Caf1M chaperone interacts selectively with Caf1 and donates its G_1 β -strand *via* a mechanism known as donor strand complementation (DSC) (Choudhury et al., 1999; Zavialov et al., 2003) (**Fig. 1.5a, iiii**). This capture of subunit by Caf1M prevents aggregation of nascent Caf1, helps in correct folding of Caf1 and is involved in targeting the correctly folded subunits to the outer membrane usher, Caf1A (**Fig. 1.5a, iii-iv**) (Chapman et al., 1999; MacIntyre et al., 2001; Yu et al., 2012).

Caf1A usher forms a transmembrane β -barrel with four additional functional domains, an N-terminal domain (NTD), a middle or plug domain (Huang et al., 2009; Remaut et al., 2008; Yu et al., 2009) and two C-terminal domains (CTD1 and CTD2) (Dubnovitsky et al., 2010; Ford et al., 2010; Phan et al., 2011). The Caf1A_{NTD} selectively interacts with Caf1 loaded chaperone (**Fig. 1.5a, iiii**) (Yu et al., 2012), positioning the incoming subunit correctly for donor strand exchange with a C-terminal bound Caf1:Caf1M complex and fibre elongation (polymerisation). During donor strand exchange, the N-terminal extension (G_D strand) of the incoming Caf1 subunit displaces the Caf1M chaperone G_1 strand and occupies the hydrophobic cleft of the preceding Caf1 subunit in the polymer. Thus by exchanging the chaperone G_1 strand for a subunit G_D strand, the Caf1-Caf1 subunits are sequentially interlocked and form F1 capsule at the cell surface (**Fig. 1.5a, iv**). Selective binding of loaded chaperone (Caf1M-Caf1) to the Caf1A usher is a consequence of allosteric changes in the surface of the chaperone. In free or unloaded Caf1M, two invariant Proline residues (Pro103 and Pro104) at the end of the F_1 strand create a sharp kink at the beginning of the subunit-binding loop, causing a condensed conformation of the Caf1A-binding surface (**Fig. 1.5b, i**). During DSC the interaction of Caf1M with Caf1 stretches the F_1 - G_1 loop altering the position of Pro103 and Pro104 thus opening this 'proline lock' and generating a hydrophobic pocket into which an absolutely conserved F4 residue of Caf1A can fit. This interaction is key to stabilising the binding of Caf1M-Caf1 binary complex to Caf1A_{NTD} (**Fig. 1.5b, iiii**). Prior to elongation of Caf1 polymer through Caf1A usher, a central plug of the usher must be displaced to permit docking of a subunit in the central usher pore (Eidam et al., 2008; Yu et al., 2009). A high-resolution crystal structure of the Fim translocon complex revealed details of the chaperone: subunit: adhesin complex docked on the C-terminus of FimD usher (Phan et al., 2011). Based on the facts that chaperone: subunit complex binds both N-terminus (Di Yu et al., 2012) and C-terminus (Dubnovitsky et al., 2010) of Caf1A, Mycroft, Z. proposed possible mechanisms of Caf1A opening (Mycroft, 2011). In one model, binding of Caf1M-Caf1 binary complex or Caf1M-Caf1-Caf1 ternary complex to Caf1A_{NTD} was sufficient to open the pore

and in the case of ternary complex docking of one subunit within the pore. This would be accompanied by transfer of the chaperone subunit complex to the C-terminus, positioning it for polymerisation with an incoming Caf1A_{NTD} bound binary subunit: chaperone complex. The other model involved opening of the pore upon direct binding of the ternary complex to the C-terminus. This would then also be appropriate for polymerisation with an incoming binary complex. However, it is still elusive what activates opening of the Caf1A pore.

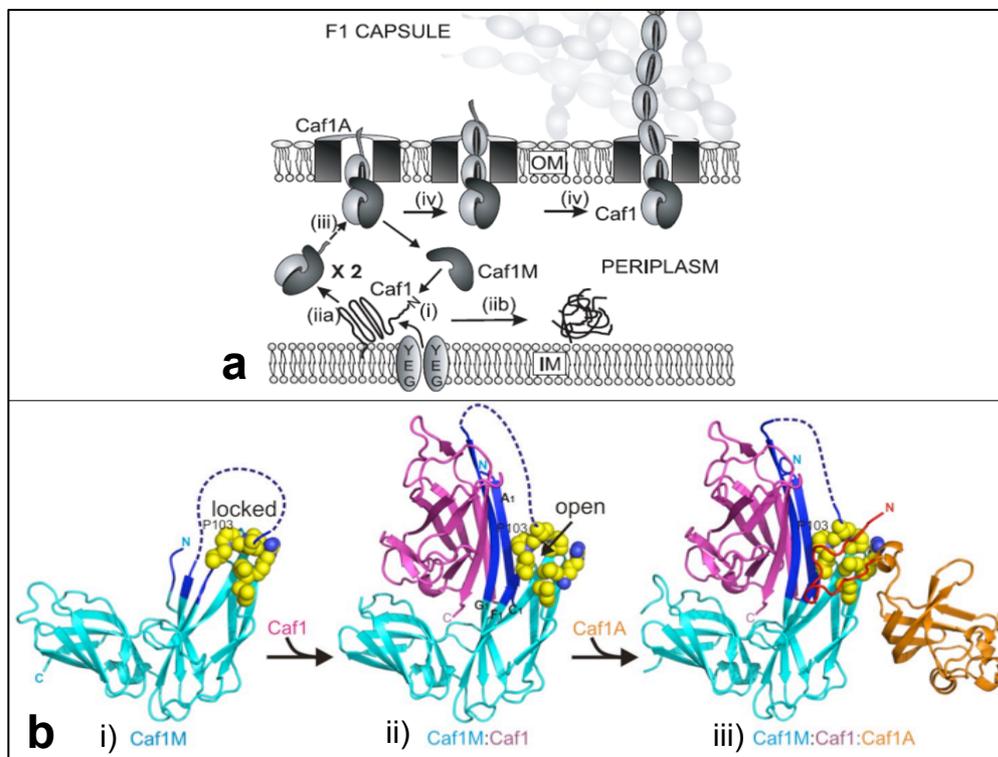


Figure 1.5| **F1 Assembly.**

a) Schematic presentation of F1 assembly. F1 subunit (Caf1) is transported into the periplasm *via* the sec pathway (iia). In the periplasm, chaperone (Caf1M) aids folding of subunit by donating its G₁ strand (i) and edge strand interactions by a process known as donor strand complementation (DSC). The chaperone:subunit complex is then targeted to the outer membrane usher (iii) where the subunit is incorporated into the growing fiber as chaperone: subunit interaction is exchanged for subunit: subunit interaction in a process known as donor strand exchange. This leads to translocation through the Caf1A outer membrane usher (iv), forming a capsular like structure of F1 on the cell surface. **b**) Selective targeting of Caf1M-Caf1 binary complex to Caf1A_{NTD} *via* opening of a proline lock at Caf1M-P103/P104. Caf1M-Caf1 interaction during DSC leads to shortening of the chaperone F₁-G₁ binding loop. Repositioning of Pro104 results in opening of a hydrophobic pocket into which the Phe4 residue of Caf1A can fit. This stabilises the Caf1A_{NTD} interaction specifically with loaded chaperone. Turquoise, Caf1M chaperone; royal blue, Caf1M G₁ and F₁ strands or dashed line F₁-G₁ loop; yellow spheres, Pro103 and Pro104 magenta, Caf1 subunit; orange, Caf1A_N; red line Caf1A unstructured N-terminal 1-24 residues. Picture **(a)** is taken from (MacIntyre, 2004) and **(b)** from (Di Yu et al., 2012).

1.4 Transcription regulation in bacteria

To protect from the adverse environment and ensure energy economy, bacterial cells respond quickly and regulate gene expression by making the right products in the correct amount at the appropriate time. *E. coli* has served as model organism to understand bacterial gene regulation. The key steps to regulate gene(s) lie in the initiation of transcription or RNA-transcript formation. The central component of transcription initiation is DNA-dependent RNA polymerase (RNAP), a multi-subunit holoenzyme, which is composed of 2 α , β , β' , σ and ω subunits (Ebright, 2000). Active site for RNAP, the determinant for the binding of both template DNA and the RNA

product, is made up from the large β and β' subunits. Each of the two identical α subunits contains two independently folded domains, a larger amino-terminal domain (α NTD) and a smaller carboxy-terminal domain (α CTD), which are joined by flexible linker (Korzheva et al., 2000). The α NTD dimerises and is responsible for the assembly of the β and β' subunits while α CTD is a DNA-binding module has an important role at certain promoters (Gourse et al., 2000). No direct role in transcription has been assigned for the ω subunit but it seems to function as a chaperone to assist the folding of the β' subunit (Mathew and Chatterji, 2006). To begin transcription at a particular promoter, the core RNAP (made up of 2α , β , β' and ω subunits) must first interact with σ subunit to form the holoenzyme. The σ subunit has 3 main functions, 1) to ensure the recognition of specific promoter elements, 2) to position the RNAP holoenzyme at a target promoter and 3) to facilitate unwinding of the DNA duplex near the transcript start site (Wosten, 1998). Most bacteria contain multiple σ factors, enabling the recognition of different set of promoters. With an exception of σ 54 (Merrick, 1993), all σ factors share common features, which are multi-domain proteins, consist of up to 4 different domains joined by linkers (Campbell et al., 2002). Domains (regions) 2, 3 and 4 are known to be involved in promoter recognition, whereas the function of domain (region) 1 is not understood (Browning and Busby, 2004).

Transcription initiation requires the interaction of RNAP holoenzyme with promoter DNA to form an open complex, in which unwinding of duplex DNA begins near the transcript start-site (deHaseth et al., 1998). Following this, synthesis of DNA template-directed RNA starts, with the formation of the first phosphodiester bond between the initiating and adjacent nucleoside triphosphate. After the initiation stage, RNAP is moved into the elongation stage and thus extension of RNA-chain occurs. A crucial step in transcription initiation is the promoter recognition by RNAP and the different DNA elements recognised by RNAP have been studied extensively (Busby and Ebright, 1994). Four DNA elements have been identified. The two principal elements are -10 and -35, with a consensus of TATAAT (-10) and TTGACA (-35), located at 10 and 35 bp upstream from the transcription start site (TSS). The -10 elements are recognised by domain 2 and -35 elements by domain 4 of the σ subunit (Murakami et al., 2002). The remaining two DNA elements are the extended -10 (by 3-4 bp) and the UP-element (\approx 20 bp), located upstream of the -35 element. The extended -10 elements are recognised by domain 3 of the σ subunit whereas UP-elements are recognised by α CTD (Murakami et al., 2002; Ross et al., 2001). Following initial binding of RNAP with promoter DNA (closed complex), the DNA strands from approximately -10 to +2 (relative to TSS) are unwound and form the open complex (Tsujikawa et al., 2002) and thus permitting transcription.

To ensure transcription at the correct promoter when required, productive distribution of RNAP, promoter sequence, σ factors, small ligands, transcription factors and the folded chromosome structure all have critical roles, as elegantly described by Busby and co-workers (Browning and Busby, 2004). A brief overview on how different transcriptional factors (regulators) control transcription is described here. In *E. coli* K-12, 304 genes are predicted to encode regulators to regulate (up and down) transcription of almost half of the total genes in this bacterium (Perez-Rueda et al., 2015). About 60% of these regulators contain two-domains, 30% have single domain, and 10% contain 3 and 4 structural domains. The most frequent DNA-

binding domains correspond to the winged helix-turn-helix motifs. Most of these regulators are predicted to bind specifically to the promoter DNA. Some of these regulators are global and hence control expression of large number of genes whereas some are dedicated regulators controlling expression of one or two genes or an operon. On the basis of sequence analysis and predicted structures, these regulators are grouped in different families correspond to well-studied families, LacI, AraC, LysR, CRP and OmpR (Perez-Rueda et al., 2015).

Transcriptional regulators controlling the expression of genes in response to environmental signals must be regulated either by controlling their activity or their expression. Several mechanisms are utilised to achieve this. First, the DNA binding affinity of the regulators can be modulated by small ligands, nutrient availability or stress signal. A well-characterised example of this is LacI repressor whose DNA-binding affinity is modulated by allolactose, in the presence of lactose (Daber et al., 2007). Second, the activity of some regulators is modulated by covalent modification and such regulators are known as response regulators. For example, NarL, which binds to its target DNA only once it is phosphorylated by the cognate sensor kinase such as NarX and NarQ, which are activated by binding to extracellular nitrite or nitrate ions (Stock et al., 2000). Third, the concentration of regulators inside the cell controls their activity. In such cases concentration is determined either by the rate of expression or by proteolysis. For examples, SoxS (AraC-type regulator) whose transcription is controlled by SoxR, which in its turn is regulated by interaction with oxidizing ligands (Dempse, 1996) and enhanced proteolysis of RovA global regulator (SlyA/Hor family) of *Y. pestis* at 37°C (Quade et al., 2012). Fourth, a less common mechanism for regulating the effective concentration of transcriptional regulator is sequestration by a regulatory protein to which it binds (Plumbridge, 2002).

When a transcriptional regulator binds to its cognate promoter, it can either activate or repress transcription. Although, some regulators function solely as activators or repressors others can function as both activator and repressor according to the target promoter (Perez-Rueda and Collado-Vides, 2000). Activators enhance the performance of the target promoter by improving its affinity to RNAP. Hence, it is quite likely that most activators first bind to target promoter before binding to RNAP. Although, some regulators first interact with free RNAP and then scan the sequence of target promoter DNA, examples include, MarA and SoxS (Martin et al., 2002; Shah and Wolf, 2004). Different mechanisms have been documented to express genes in response to changing environment. Although each mechanism has variation in the level of gene expression which is known as fine-tuning but there are some exceptions which do not require fine-tuning such as expression of the fimbrial gene. DNA segment containing the promoter for the fimbrial genes can be switched from one orientation in which transcription of fimbrial gene occurs, to the opposite orientation in which promoter is directed away and thus creating an effective ON/OFF switch (Blomfield, 2001). Transcription activation at many promoters is simple and thus involves contribution of a single activator. Three basic mechanisms are utilised for simple activation, class I, class II and activator-mediated conformational change of the target promoter, enabling interactions of RNAP with -10 and -35 recognition elements. The class I activation requires regulator binding to the target promoter sequence located upstream of the -35 element and thus recruits RNAP by interacting with α CTD of RNAP. Example of this kind of activation is the action of cAMP receptor protein (CRP) at *lac* promoter (Ebright, 1993).

Regulators function in class I manner can bind at several locations upstream of promoters, which is accompanied by flexible linker joining the α CTD and α NTD of RNAP. The class II activation requires activator binding to the target sequence, which overlaps -35 element and thus contacts α CTD and region 4 of σ subunit of the RNAP (Dove et al., 2003). The third mechanism of activation, which is by conformational change of the target promoter by activator, requires activators binding at or close to the promoter elements (Sheridan et al., 2001). Promoters activated by MerR family regulators are examples of this kind of activation, where spacing between -10 and -35 is not optimum for the binding of RNAP. Activators belonging to MerR family bind to the -10 and -35 spacer sequence and twist the DNA to reorientate -10 and -35 elements in such a way that they can be bound by RNAP σ subunit (Brown et al., 2003).

Like transcription activation, repression of transcription by repressor proteins, which decreases the rate of transcription by binding to the target promoter is also simple and thus involves a single repressor. Three mechanisms have been illustrated for transcription repression, first, steric hindrance of RNAP binding to the target promoter. In this case, repressor-binding site are located in or close to -10 and -35 elements, example include repression by LacI repressor by binding to *lac* promoter (Daber et al., 2007). Second, repressor might not prevent RNAP binding but might interfere with post recruitment step of transcription initiation (Muller-Hill, 1998). Third mechanism of repression involves DNA looping, where multiple repressor molecules binds to promoter distal sites, causing loop formation and thus shuts off transcription initiation. The best-characterised example of transcription repression by DNA-looping is regulation of *araBAD* operon by AraC, discussed in later sub section. Unlike activation, complex cases of repression also have been characterised where repressor acts as an anti-activator. For example, CytR-repressed promoters those are dependent on activation by CRP, which in turn is mediated by CytR direct-interaction with CRP (Shin et al., 2001). CytR recognises tandem-bound CRP and the bound CRP confers the specificity for CytR-mediated repression (Valentin-Hansen et al., 1996). It has been documented that CRP, FNR, IHF, Fis, ArcA, NarL and Lrp transcription regulators control 50% of all regulated genes, whereas about 60 other transcription regulators control only a single promoter (Martinez-Antonio and Collado-Vides, 2003). A descriptive survey of CU pathway assembled fimbriae/pili regulation mechanisms and more specifically regulation by AraC/XylS family of transcriptional regulators is discussed in the following subsections.

1.4.1 Regulation of expression of CU pathway assembled pili/fimbriae

To establish an infection, attachment of bacterial pathogens to specific host cells is often a crucial step. This is often mediated by surface-exposed adhesive structures known as pili or fimbriae, assembled by the CU pathway (Sauer et al., 2004; Soto and Hultgren, 1999; Thanassi et al., 1998; Zavialov et al., 2007). Genes involved in the biogenesis of the pili or fimbriae are clustered into operons. These operons always contain a gene for the major structural subunit or pilin, a periplasmic chaperone and an outer membrane usher protein. Often, these operons contain additional genes encoding structural proteins such as genes for minor subunits and adhesins located at the tip of the pilus and sometimes the locus includes additional chaperones and regulatory protein(s) (**Fig. 1.6**).

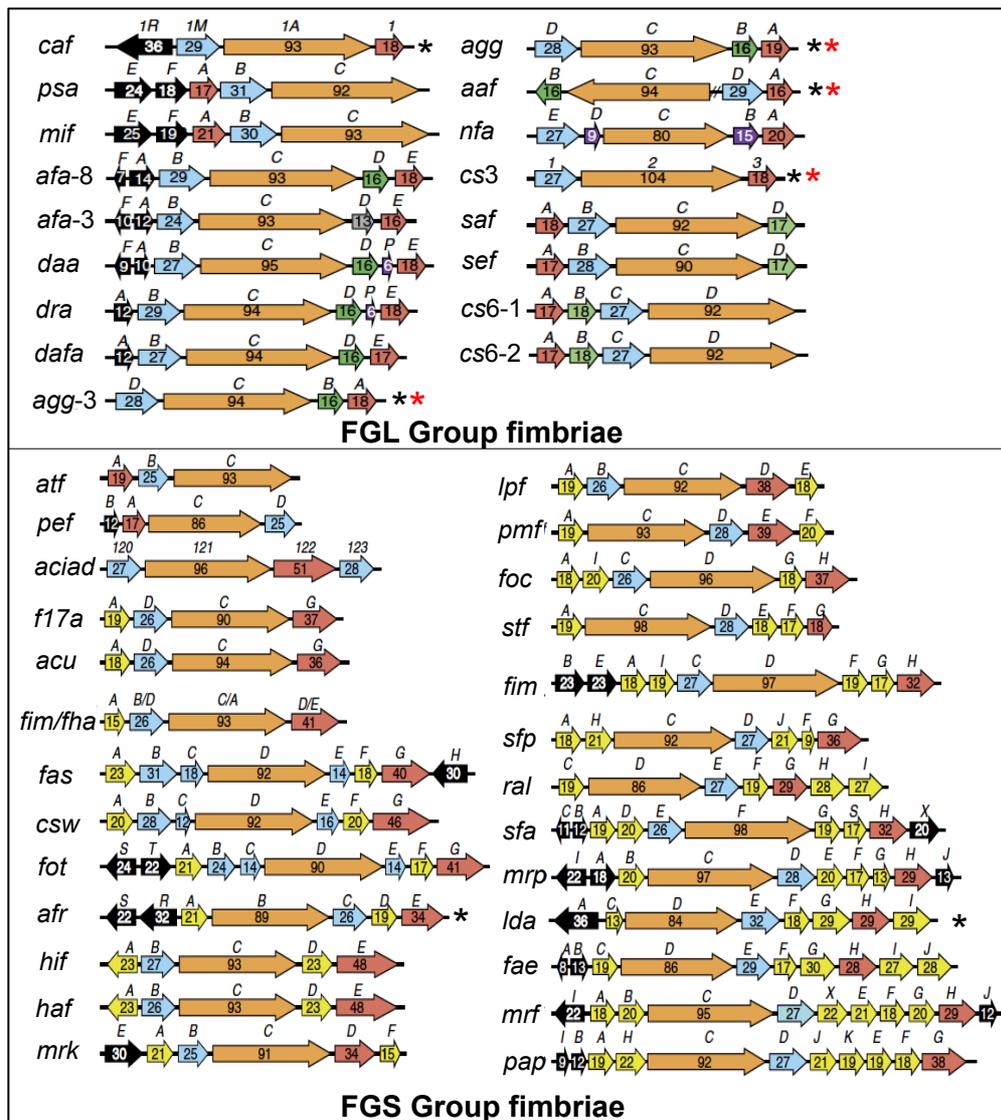


Figure 1.6| Gene clusters/operons of some FGL and FGS group fimbriae.

Colored arrows indicate the orientation of each gene in the cluster/operon; black (regulator/s), blue (periplasmic chaperone), orange (outer membrane usher), red (adhesin and structural subunit/s), yellow (subunit/s with unknown function), green (protein/s with putative function) and pink (protein/s of unknown function). Numbers inside the colored arrows designate the molecular weight in kDa of the corresponding protein. Black asterisks (*) indicate CU loci known to be controlled by an AraC/XylS family regulator, AggR (29.4 kDa), AfrR (32 kDa), Caf1R (36 kDa), LdaA (36 kDa) and Rns (32 kDa). AggR and Rns (*, not indicated) are global regulators. AggR controls expression of *agg*, *agg-3* and *aaf* fimbriae (Morin et al., 2013; Nataro et al., 1994). Rns controls the expression of *cs1-4*, *cs14*, *cs17* and *cs19* fimbriae (Bodero et al., 2008; Caron et al., 1989). Picture modified from (Zav'yalov et al., 2010).

1.4.2 Diversity of CU pathway assembled pili/fimbriae

The CU class was initially divided into two structurally and functionally distinct groups, FGS and FGL (**Fig 1.6**), where the assembly of pili or fibrillar structures is assisted either by a short (S) or a long (L) F₁-G₁ loop of the periplasmic chaperone (Hung et al., 1996). A well-studied example of CU pathway from the FGL group is the assembly of *Y. pestis* F1 capsular antigen. Most of the FGL group, like the Caf system assemble simple, thin and flexible fibres of only one or two subunits. The pili belonging to the very large FGS group are generally thick and rigid pili also known as adhesive fimbriae. These fimbriae/pili have a more complex structure with a major structural subunit (fimbrin/pilin) making up the shaft of the pilus, several additional subunits forming a thin flexible tip fibrillum and a single specialised adhesin with both pilin and adhesin domains

on the tip of pilus (Hung et al., 1996) unlike F1 (MacIntyre, 2004; Zavialov et al., 2003). Phylogenetically, FGS group is much more diverse and is subdivided into several clades including, β -, γ 1-, γ 2-, γ 4-, κ - and π -fimbriae (Nuccio and Baumler, 2007). Unlike FSG, FGL is a small monophyletic group having only one clade, γ 3.

Expression of pili/fimbriae from both groups is often tightly controlled to avoid expression when not needed. Different mechanisms used to regulate expression of CU systems, either belonging to FGS or FGL group. Well-studied examples of both the assembly and the gene regulation of fimbriae/pili from the FGS group are Type 1 and Pap fimbriae of *E. coli* (Blomfield, 2001; Busch and Waksman, 2012; Choudhury et al., 1999; Drummelsmith and Whitfield, 2000; El-Labany et al., 2003; Fernandez and Berenguer, 2000; Hernday et al., 2002; Phan et al., 2011; Schwan, 2011; Sohanpal et al., 2004; Sohanpal et al., 2007). An overview of regulatory mechanisms controlling expression of the Type 1 fimbriae and Pap pilus (FGS group), and regionally well-studied examples from the FGL group are described below.

1.4.3 Regulatory mechanisms of FGS group: Type 1 and Pap fimbriae/pili

i) Regulation of Type 1 fimbriae expression: promoter inversion

Type 1 fimbriae of uropathogenic *E. coli* are synthesised from a chromosomally encoded gene cluster known as the *fim* locus. This locus is composed of 9 genes, *fimB*, *fimE*, *fimA*, *fimI*, *fimC*, *fimD*, *fimF*, *fimG* and *fimH* (**Fig. 1.7a**), encoding two regulatory proteins, FimB and FimE; a major subunit pilin, FimA; two assembly proteins, a periplasmic chaperone, FimC and an outer membrane usher, FimD; two tip associated fibrillum subunits, FimF and FimG; the mannose-binding terminal adhesin, FimH; and FimI, a subunit responsible for anchoring the pilus to the cell surface. Regulation of expression of the *fim* locus is mediated by Phase variation, which is accompanied by an invertible DNA fragment (314 bp), known as *fimS*, flanked by 9 bp inverted repeats, IRL and IRR (⁵TTGGGGCCA³) (**Fig. 1.7b**). A promoter located on *fimS* is responsible for transcription of major fimbrial subunit (FimA) encoding gene, *fimA*. Switching the orientation of *fimS* is responsible for transcription of *fimA*, forward orientation lead to transcription (Phase-ON, pilliated state) whereas in opposite orientation of *fimS*, no transcription is possible (Phase-OFF, non-pilliated state). The regulatory proteins, FimB and FimE influence the orientation of *fimS*. They are site-specific recombinases, share 48% amino acid identity and both contain a tetrad of conserved amino acids (Arg47, His141, Arg144 and Tyr176; in FimB and Arg41, His136, Arg139 and Tyr171; in FimE), which has been shown to be essential for recombinase activity (Blomfield, 2001; Burns et al., 2000; Dorman and Higgins, 1987; Schwan, 2011; Smith and Dorman, 1999). FimB can bind to *fimS* in both states, Phase-ON and Phase-OFF with a slight bias toward the Phase-OFF state (Schwan, 2011). In contrast, FimE primarily mediates the inversion of *fimS* from ON to OFF state (**Fig. 1.7b**). The mechanism and factors influencing orientation and recombination of the *fimS* element are clearly complicated and are influenced by the concentrations of both FimB and FimE, as well as other factors. Three promoters have been identified for *fimB* expression (Donato et al., 1997; Schwan et al., 1994) while for *fimE* a single promoter has been identified (Olsen and Klemm, 1994).

At least 20 additional auxiliary genes are known to influence the expression of Type 1 fimbriae (**Fig. 1.7c**). The product of *pilG*, an allele of the *hns* gene, was reported to influence inversion. A mutation in the *pilG* was shown to increase inversion of *fimS* up to 100-fold using a *fimA-lacZ* fusion (Spears et al., 1986). The global regulatory protein, H-NS, represses transcription of both *fimB* and *fimE* genes by binding to promoters of both with high specificity (Olsen and Klemm, 1994; Olsen et al., 1998; Schwan et al., 2002). H-NS may also directly affect inversion by binding to sequences adjacent to the *fimS* DNA fragment (Donato and Kawula, 1999; Schwan et al., 2002). In addition to H-NS, integration host factor (IHF) and leucine-responsive protein (Lrp) also affect Phase variation (Brinkman et al., 2003; Dorman and Higgins, 1987; Eisenstein et al., 1987). Both IHF and Lrp cause sharp bends in the DNA structure, introducing hairpin loops that facilitate recombination. Other site-specific recombinases (HbiF, IpuA, IpuB and IpbA) have been identified as possibly influencing inversion of *fimS* (Xie et al., 2006). There is also cross talk between regulators of different pili systems in *E. coli*. Both PapB, a regulatory protein of the Pap fimbriae locus, and SfaB from the S pili locus, can influence the orientation of *fimS* by inhibiting the Phase-OFF to the Phase-ON switch (Holden et al., 2006; Holden et al., 2001; Xia et al., 2000). The ppGpp alarmone also appears to be required for optimum expression of Type 1 fimbriae and the ppGpp alarmone along with RNA polymerase-binding protein, DksA was shown to stimulate transcription of *fimB* from the P2 promoter (Paul et al., 2005; Schwan, 2011). Environmental factors such as temperature and pH also play a crucial role in the expression of type 1 fimbriae. Several studies have shown that the Phase switching from OFF to ON state increases at lower temperature (Dorman and Ni Bhriain, 1992; Gally et al., 1993; Olsen et al., 1998). Under conditions of low pH, proteins such as SlyA or RcsB may activate *fimB* and prevent H-NS from binding thus promoting Type 1 fimbriae synthesis (Schwan, 2011).

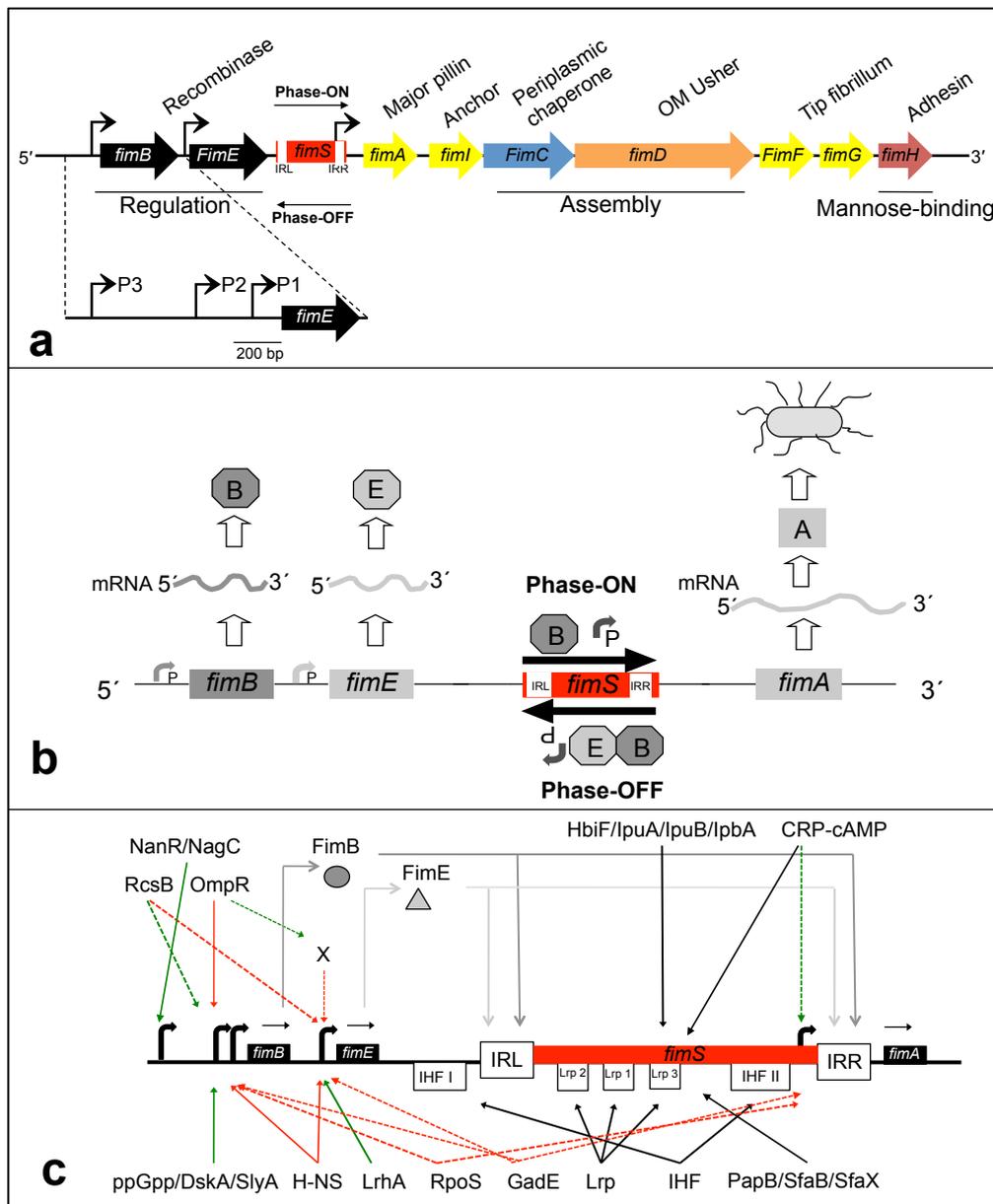


Figure 1.7| Regulation of Type 1 fimbriae expression.

a) The *fim* locus indicating key features, role of each gene, promoters and invertible *fimS* DNA element flanked by inverted repeats IRL and IRR, responsible for Phase ON/OFF states. **b)** FimB and FimE mediated inversion of *fimS* DNA element. Fim B influences ON state and pilated bacteria. FimB and FimE both bind to influence switch to non-piliated bacteria. **c)** Role of 20 auxiliary proteins in regulation of Type 1 fimbriae. Invertible motifs, IRL and IRR are indicated in open boxes. IHFs (I-II) and Lrp (1-3) binding sites are also represented as open boxes. The *fimB*, *fimE* and *fimA* are displayed in black boxes. All identified promoters are shown as bent black arrows. Dark and light gray arrows correspond to FimB and FimE, respectively. Black arrows signify effect on *fimS*. Confirmed and presumed binding associated with stimulatory effects is presented by solid and dashed green arrows, respectively; whereas confirmed and presumed binding associated with repressing effects are shown by solid and dashed red arrows, respectively. Picture (a) modified from (Fernandez and Berenguer, 2000; Schwan, 2011) and pictures, (b) and (c) were modified and provided by Bill Schwan (Schwan, 2011).

ii) Regulation of Pap pili expression: methylation/epigenetic switch

Genes encoding Pap pilus are clustered on a chromosomally encoded *pap* locus which contains 11 genes, *papI*, *papB*, *papA*, *papH*, *papC*, *papD*, *papJ*, *papK*, *papE*, *papF* and *papG*. The role of each gene product is depicted in **Fig. 1.8a**. PapB and PapI, are regulatory proteins essential for the expression Pap pilus. PapA is the major pilin, while PapK, PapE and PapF code for the tip fibrillum, PapG is a galactose-binding adhesion and PapH is the pilus anchor. The other two gene

products PapD and PapC are the two assembly proteins, a periplasmic chaperone, and outer membrane usher, respectively (Fernandez and Berenguer, 2000; Sauer et al., 2004). Regulation of expression of the *pap* locus is also mediated by Phase variation but in a different way to that of Type 1 fimbriae. Phase variation in Pap is determined by binding of Lrp to 2 different Dam methylation sites within the promoter region. This prevents methylation of the site leading to either activate (Phase-ON) or repress (Phase-OFF) transcription of the *pap* gene cluster (Blomfield, 2001), as depicted in **Fig. 1.8b-c**. The regulatory proteins, PapI (8 kDa) and PapB (12 kDa) positively control expression of the *pap* gene cluster in association with the global regulatory proteins, Lrp, CAP and H-NS (Hernday et al., 2002; White-Ziegler et al., 1998; White-Ziegler et al., 2000). The intergenic regulatory region (416 bp) between *papI* and *papB* contains 6 Lrp-binding sites (1-6), which control transcription of both *papA* and *papB* (**Fig. 1.8b**). Within this there are two DNA methylase (Dam) sites, GATC^{prox} and GATC^{dist} located within Lrp-binding sites 2 and 5, respectively (**Fig. 1.8b-c**). When accessible these sites are methylated at the A base of the GATC motif, but when Lrp is bound methylation of the sites is blocked. Lrp-binding proximal to *papBA* promoter inhibits transcription whereas binding at the distal site permits methylation at site 2 and activates transcription of *papBA* (ON state). Lrp binds with higher affinity to 1-3 sites than 4-6 (Braaten et al., 1994; Nou et al., 1995; Nou et al., 1993). The regulator PapI promotes binding of LRP to sites 3-5 rather than 1-3, promoting the ON state. Transcription is also under control of catabolite repression thus requiring binding of cAMP-CAP complex 60 bp upstream of the Lrp-binding site (Baga et al., 1985; Goransson et al., 1989; Hung et al., 2001; Weyand et al., 2001).

As with regulation of expression of Type 1 fimbriae, expression of Pap pili is also controlled by a number of environmental stimuli (Baga et al., 1985; White-Ziegler et al., 2000). Pap pili transcription is significantly repressed during growth at lower temperatures (<26°C) in Luria broth (Goransson et al., 1990; Goransson and Uhlin, 1984; White-Ziegler et al., 2000). In addition to environmental factors, proteins not belonging to the *pap* locus have been shown to contribute to positive regulation of Pap pilus transcription. For example, the two-component sensor-regulator CpxAR appears to be activated by misfolded Pap subunits in the periplasm as well as on binding of Pap pilus to solid surfaces. This initiates a phosphotransfer relay from CpxA to CpxR, which in turn upregulates transcription of a number of genes, including *degP* and *pap*. It has also been shown that when CpxAR system is activated, the transcription of *papI* and *papBA* is enhanced 2-fold (Otto and Silhavy, 2002).

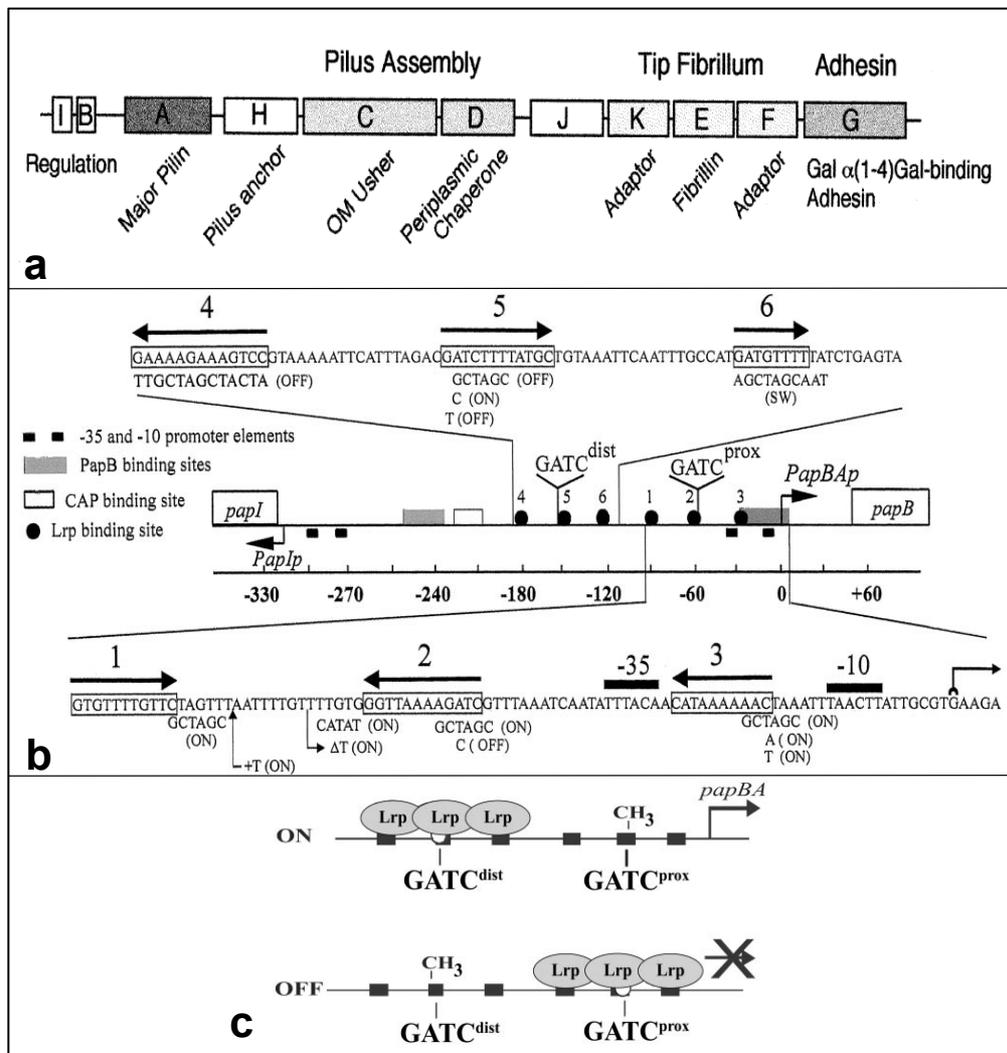


Figure 1.8| **Regulation of Pap pilus expression.**

a) The *pap* locus, indicating key features. Genes encoding regulatory proteins, PapI and PapB; pilus assembly proteins, PapH, PapC and PapD; tip fibrillum proteins, PapJ, PapK, PapE and PapK; and adhesin protein, PapG are indicated by different shaded boxes and labeled accordingly. **b**) Regulatory network of *pap* locus. Regulatory region of divergently transcribed *papBA* and *papI* promoters is depicted. Dam methylation sites, $GATC^{prox}$ and $GATC^{dist}$ are located within Lrp binding sites 2 and 5 as indicated. Legends for PapB, CAP and Lrp binding sites along with -10 and -35 elements are indicated on left side of the figure. Orientation of Lrp sites with consensus sequence $5'Gnn(n)TTTt^3'$ are indicated with arrows above the sequence. Mutations (substitution, deletion and insertion) shown below the wild type sequence switch the phenotype of mutants is indicated in parentheses. **c**) DNA-methylation states of Phase-ON and Phase-OFF cells. DNA methylation of the GATC sites is blocked by Lrp-binding at 4-6 sites in Phase-ON cells and at 1-3 sites in Phase-OFF cells. Picture (a) is taken from (Fernandez and Berenguer, 2000) and (b) and (c) are taken from (Hernday et al., 2002) with permission.

1.4.4 Regulation of expression of FGL group CU systems

Within the FGL (Y3) group of CU loci, the regulatory mechanisms of expression of only few loci have been studied and understanding is very limited when compared to the detailed knowledge of regulation of expression of Type 1 and Pap pili of *E. coli*. One system which has been studied in some detail is expression of the chromosomally located *psa* locus which encodes the pH6 fibrillar antigen of *Y. pestis* and is closely related to the *myf* locus of *Y. pseudotuberculosis*) (Iriarte and Cornelis, 1995; Zavialov et al., 2007; Zhang et al., 2013). The pH6 antigen is expressed at low pH and inside macrophage (Huang and Lindler, 2004) hence its name. The *psa* locus is composed of five genes, *psaE*, *psaF*, *psaA*, *psaB* and *psaC* (Fig. 1.6). PsaE (24 kDa) and

PsaF (18 kDa) are regulators both located within the inner membrane and required for expression of pH6 fibrillum. PsaE contains a cytosolic N-terminal DNA-binding domain with homology to PhoB and OmpR of *E. coli*, ToxR of *Vibrio* and HilA of *S. typhimurium* (Yang and Isberg, 1997). The C-terminal domain of PsaE lies in the periplasm and interacts with the periplasmic domain of PsaF (Yang and Isberg, 1997). The central regulator, RovA (18 kDa) has been shown to interact with promoter regions of *psaE* and *psaA* (Cathelyn et al., 2006). A recent study from Zhang et al., demonstrated that PhoP and RovA recognise the promoter-proximal regions of *psaEF* and *psaABC* (Zhang et al., 2013). RovA activated *psaEF* and *psaABC*, whereas PhoP repressed both *psaEF* and *psaABC* by an involvement of direct association between RovA/PhoP and the target promoter regions. It was suggested that this reciprocal regulation of *psa* genes by PhoP and RovA could contribute to the tight regulation of pH6 antigen expression during infection (Zhang et al., 2013). Expression of the Afa-3 and -8 fimbriae of *E. coli*, controlled by two regulatory proteins, AfaF (10 kDa) and AfaA (12 kDa). AfaF and AfaA are homologous to the regulatory proteins, PapI and PapB of the Pap pilus, which suggests that these fimbriae might have a similar kind of Phase switching mechanism to regulate their expression (Servin, 2005). Gene cluster of Sef fimbriae of *S. enteritidis* contains, *sefABC* genes, which encode a structural subunit (SefA fimbrin), a periplasmic chaperone (SefB) and an outer membrane usher (SefC) to assemble Sef fimbriae (Clouthier et al., 1993). The regulatory mechanism of Sef fimbriae is mediated by two stem loop mRNA secondary structure, a class of rho-independent transcription terminator, located in between *sefA* and *sefB* genes (Clouthier et al., 1993).

Expression of Agg and Aaf fimbriae of enteroaggregative *E. coli* is controlled by an AraC/XylS family regulator, AggR (29.4 kDa), which is encoded on a 60 MDa plasmid and located 9 kbp upstream from the *agg* locus (Gallegos et al., 1993; Nataro et al., 1994). AggR was recently shown to act as a global regulator (Morin et al., 2013). Forty-four additional AggR targets, including genes encoding dispersin surface protein (Aap), dispersin translocator (Aat) and Aai type-VI secretion system were identified. AggR was found to activate the expression of Agg fimbriae in response to temperature, oxygen tension, and osmolality as well as the medium composition (Nataro et al., 1994). Similarly, gene cluster of chromosomally encoded CS1-3 pili of enterotoxigenic *E. coli* is also regulated by an AraC/XylS family regulator, Rns (32 kDa) (Munson and Scott, 1999, 2000). Rns is encoded on a plasmid, unlinked to any CU system, and is flanked by transposases on one side and pseudogenes on the other side (Caron et al., 1989; Munson and Scott, 1999, 2000). A detailed description of gene regulation by Rns is discussed in section 1.5.5.

Many pathogens undergo a temperature transition during their life-cycle to establish an infection in the host. In such cases, expression of virulence factors is frequently under thermo-control. Temperature-dependent expression of a number of virulence factors including the *caf* locus (Chauvaux et al., 2007; Han et al., 2007; Li et al., 2011; Motin et al., 2004; Simpson et al., 1990), has been identified in *Y. pestis*, but the basis of temperature regulation of F1 expression has never been investigated. As mentioned earlier (section 1.3.1), Caf1R is also a regulator of the AraC/XylS family, but despite the fact that this regulator controls expression of a major virulence component of *Y. pestis*, there has been no further characterisation of this regulator or its action on the *caf* locus.

In conclusion, many very different mechanisms of regulation have been co-opted by CU systems to control the expression of fimbriae. Despite the common nature of CU assembly pathways of both FGS and FGL groups of fimbriae, even closely related systems such as *psa* and *caf* loci may use very different methods of regulation.

1.4.5 Regulation via AraC/XylS family protein-regulators

Regulators belonging to AraC/XylS family are widespread among Gram-negative and Gram positive bacteria, controlling expression of genes involved in varied biological processes including carbon metabolism, stress responses and pathogenesis (Gallegos et al., 1997; Ibarra et al., 2008; Martin and Rosner, 2001; Schuller et al., 2012). Proteins of this family are characterised by having a conserved DNA-binding domain (DBD, ≈ 100 amino acids long), most frequently located at the C-terminus and a highly variable sensing or oligomerisation N-terminal domain (100-200 amino acids long) (Gallegos et al., 1997; Ibarra et al., 2008; Martin and Rosner, 2001; Schuller et al., 2012). A few examples of this family of regulators contain only the DBD, including MarA and SoxS, which have been characterised (Griffith et al., 2004; Griffith and Wolf, 2002). DBD binds to the promoter/operator region of target operon(s) and often activates expression although some may act as both activator and repressor (Gallegos et al., 1997; Ibarra et al., 2008; Martin and Rosner, 2001; Schuller et al., 2012). DBD contains a conserved tertiary structure composed of 7 α -helices, including two helix-turn-helix (HTH) motifs, which are connected by a central α -helix or turn (Ibarra et al., 2008; Schuller et al., 2012).

A recent bioinformatic analysis of putative and characterised DBDs of 62 AraC/XylS-type regulators grouped Caf1R-DBD with regulators controlling metabolic processes (**Fig. 1.9**). Unexpectedly, it was aligned very distantly to AggR and VirF, regulators of virulence factors including fimbriae (Schuller et al., 2012). Caf1R was most closely aligned with XylR regulator of *E. coli* K-12, which is essential for transcription of genes required for the utilisation of D-xylose and L-arabinose (Ni et al., 2013). Whether this relationship to XylR and other regulators of metabolic processes is an indication of metabolic processe(s) regulated by Caf1R or is an indication of small molecule sensing by Caf1R remains to be seen.

Despite the very large size of the AraC/XylS family (about 1900 members), only a very small number of regulators belonging to this family have been characterised experimentally and among these so far only two regulators, MarA and Rob of *E. coli* have been cocrystallized with their cognate promoter DNA (Kwon et al., 2000; Rhee et al., 1998). An overview of transcription activation by some well-characterised regulators of this family is described below.

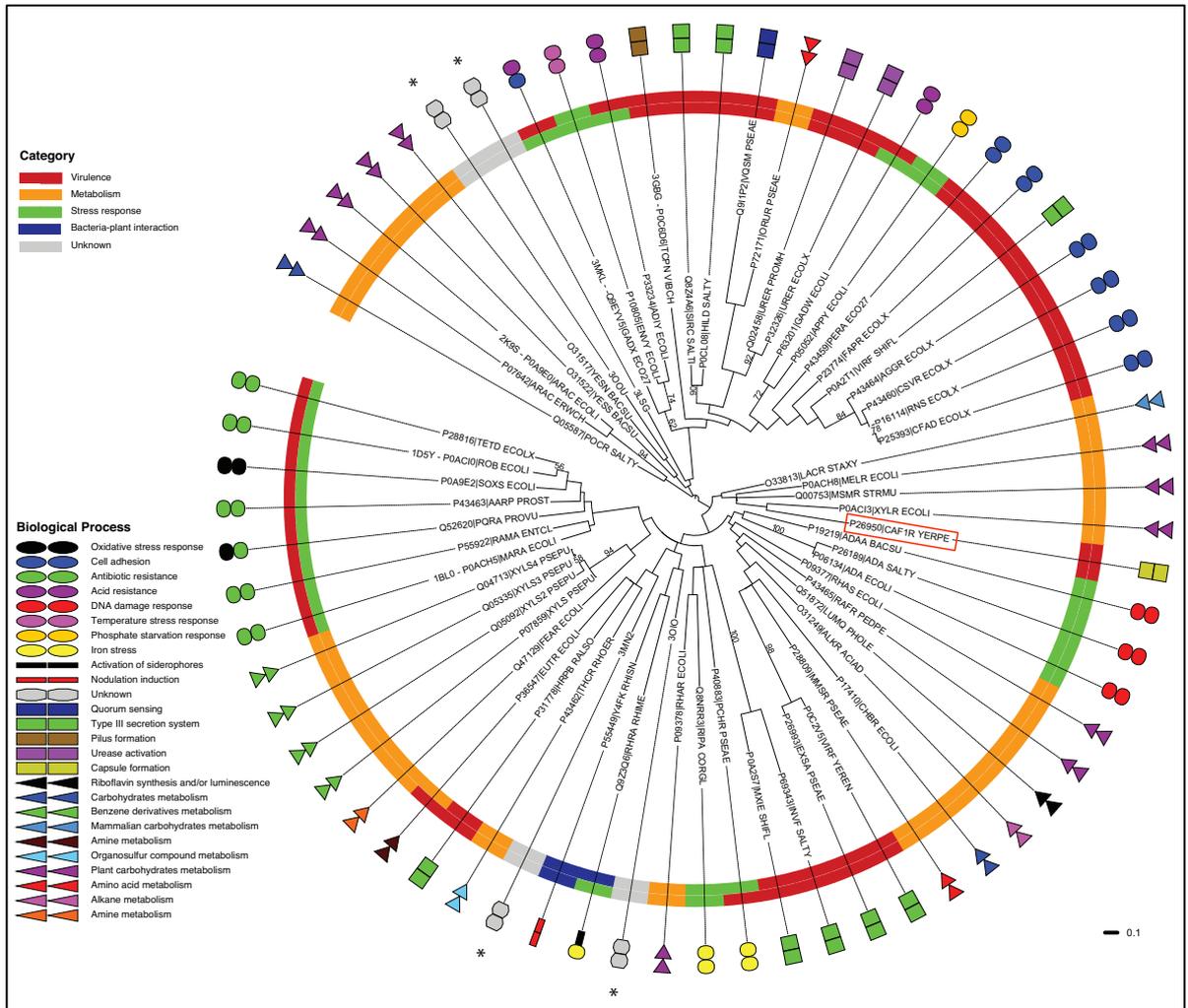


Figure 1.9| Caf1R-DBD closely related to metabolic regulator, XylR of *E. coli*.

A dendrogram of DNA binding domains (DBDs) from 62 AraC/XylS family regulators. Functional category and biological process controlled by each each regulator is indicated by different colors schemes, on left. Caf1R is indicated in red box. Picture kindly provided by Francisco Melo (Schuller et al., 2012).

i) Transcription activation by AraC: a sugar responsive regulator

AraC regulator is one of the best-characterised regulators of AraC/XylS family. It regulates expression of the *araBAD* operon of *E. coli* for the catabolism of arabinose (Bustos and Schleif, 1993; Schleif, 2010). It acts as both activator and repressor in the presence and absence of arabinose inducer, respectively (Bustos and Schleif, 1993; Schleif, 2010). AraC is a homodimeric protein (60 kDa), where each monomer binds to a single molecule of L-arabinose prior to activating transcription of the *araBAD* operon (Bustos and Schleif, 1993; Schleif, 2010). Genetic and biochemical studies showed that a short N-terminal arm (residues 7-17) of AraC plays a critical role in both activation and repression (Rodgers et al., 2009; Saviola et al., 1998). It does this *via* differential contact with either DBD or L-arabinose, a process termed the 'Light-switch' mechanism (Schleif, 2000, 2010) (**Fig. 1.10**). According to this mechanism, in the absence of L-arabinose (-), the NTD-arm region binds the DBD limiting flexibility of the two domains. The N-terminal domains of 2 monomers interact and the DBDs of these monomer bind at two distantly spaced sites, *araO2* and *araI1*, repressing the transcription of *araBAD* operon. Induction occurs in the presence of L-arabinose, which binds the N-terminal domain and the arm releasing the

constraint on the DBD. During this state (L-arabinose bound), the NTD and DBD are flexibly connected and adopt a conformation where both DBDs bind to closely spaced half sites, *araI1* and *araI2* on the *araBAD* promoter (P_{BAD}) region activating transcription.

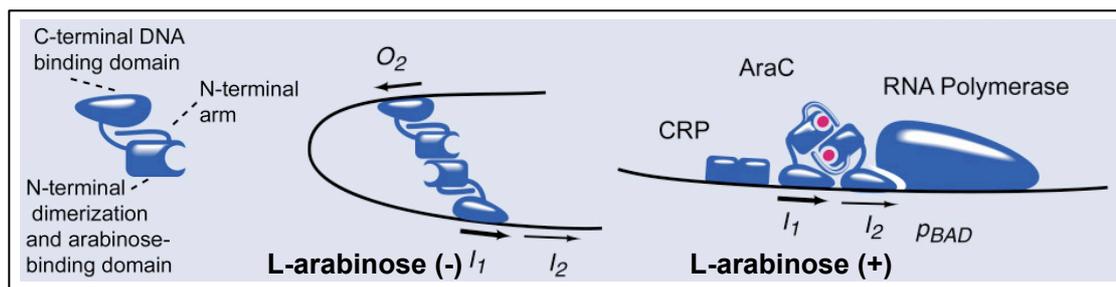


Figure 1.10| **Light-switch mechanism of AraC.**

In the absence (-) of L-arabinose, the arm region of AraC_{NTD} interacts with CTD_{DBDs} and the DBDs binds to *araO2* and *araI1* sites. In the presence (+) of L-arabinose, the arm region of AraC_{NTD} bends over L-arabinose and thus releases the DBDs. DBDs from both monomers bind. Picture modified from (Schleif, 2000).

ii) Transcription activation by MarA: conferring multiple antibiotic resistance

MarA (15 kDa) of *E. coli* is a global regulator within the *Enterobacteriaceae* family and regulates transcription of more than 40 genes, involved in conferring resistance to multiple antibiotics, toxic compounds, acidic pH and antimicrobial peptides (Duval and Lister, 2013). In particular, MarA activates expression of the multidrug efflux pump, AcrAB-TolC and the porin OmpF leading to decreased permeability of the outer membrane and a multiple antibiotic resistance phenotype (Okusu et al., 1996). MarA is encoded within the chromosomally located *marRAB* operon (Cohen et al., 1993). The function of *marB* is largely unknown. It was shown to somehow increase the level of *marA* message and also contains a predicted periplasmic signal sequence, suggesting a role at the posttranscriptional level (Vinue et al., 2013). Another report suggested a role in transcriptional regulation of *inaA*, one of the MarA regulated genes (Lee et al., 2012). The first gene of the *marRAB* operon, *marR* encodes a transcriptional repressor, MarR that functions as an autorepressor of the *marRAB* operon. It forms a dimer and contains a winged-type HTH motif (Alekhshun et al., 2001) that promotes highly specific binding to two palindromic sequence motifs (P1 and P2), located within promoter region (*marO*) upstream of the *marRAB* operon (Alekhshun et al., 2001), repressing transcription (**Fig. 1.11a**). Binding of MarR is abolished on binding of antibiotics such as chloramphenicol and tetracycline (Hachler et al., 1991) or phenolic compounds such as sodium salicylate, 2, 4-dinitrophenol or 2, 3-dihydroxybenzoate (Martin and Rosner, 1995). In the absence of MarR binding, MarA activates transcription of the *marRAB* operon and other genes within the *mar* regulon. (Barbosa and Levy, 2000).

MarA activates transcription at both class I and class II promoters (Martin et al., 1999; Martin and Rosner, 2002). These promoter classes are defined according to the orientation and location of DNA-binding site, relative to -35 and -10 promoter elements, recognised by RNAP (Martin et al., 1999; Martin and Rosner, 2002). In majority of the class I promoters, transcriptional regulators bind around 7-27 bp upstream of -35 element, to a DNA motif oriented in the backward (←) direction. In class II promoters, the DNA-binding site of the regulators is oriented in the forward (→) direction and often overlaps with -35 element. Examples of class I promoters

include, MarA/Rob/Sox activated *mar*, *zwf* and *map* promoters (Duval and Lister, 2013), although *zwf* and *map* are exceptions with forward (→) oriented binding sites. Examples of MarA/Rob/SoxS activated class II promoters include *fumC* and *micF* (Duval and Lister, 2013). The consensus DNA binding motif for MarA/Rob/SoxS regulators is AGRGCACRWWNNRYAAAGN (R=A/G, Y=C/T, W=A/T and N=A/T/C/G) and contains two binding sites, BS1 (RGCAC) and BS2 (NRYAAAA), identified by DNase I foot printing of MarA-DNA binding (Martin et al., 1996) and analysis of the cocrystal structure of *mar*-MarA complex (Rhee et al., 1998). BS1 is recognised by MarA helix-3 and BS2 by helix-6. Other class II promoters include, *pur* and *hdeA* promoters, but both are repressed by MarA/Rob/SoxS (Duval and Lister, 2013). Two different orientations and locations of MarA binding site imply two configurations of ternary complex (MarA-RNAP-DNA) formation to activate transcription. NMR studies of MarA-DNA-RNAP complexes (Dangi et al., 2004) showed that class I promoter, transcriptional activation requires MarA interaction with the CTD of the α -subunit from RNAP whereas class II promoters, transcription activation requires MarA interaction with both α -CTD and region 4 of σ^{70} subunit of the RNAP (Dangi et al., 2004; Martin et al., 1999).

MarA activates its own transcription by binding (as a monomer) to a 20 bp motif (*mar* box), located upstream of the -35 promoter element in the promoter region of *marRAB* operon (Martin et al., 1996). MarA is rapidly degraded by Lon protease and has an extremely short half-life (about 3 min) that ensures rapid removal of the response cascade as the stress signal is removed. Thus in the absence of a stress signal the level of MarA is drastically low (Griffith et al., 2004). Fis (Factor for inversion stimulation), a small DNA-binding and bending protein that has been shown to activate the transcription of the *marRAB* operon by binding to the *marO* promoter region (Martin and Rosner, 1997). Fis-mediated activation of *marRAB* operon requires the presence of MarA or the auxiliary proteins, SoxS or Rob (Martin and Rosner, 1997). MarA contains only DBD (127 amino acids long). It lacks the effector domain, although MarR effectively act as a sensor for MarA (Duval and Lister, 2013).

The 3D cocrystal structure of *mar*-MarA complex (PDB-1bl0) showed that MarA is composed of seven α -helices (1-7) and a long C-terminal loop (**Fig. 1.11b**). The α -helices are folded into two structurally identical subdomains (SD1 and SD2), encompassing residues, 10-61 (SD1) and 62-110 (SD2). Residues from, 31-52 and 79-102 form two HTH motifs, HTH1 and HTH2, comprised of helices, 2-3 and 5-6. The two HTH motifs are connected by a flexible linker, helix-4 (Rhee et al., 1998). Each HTH motif contains a recognition helix, helix-3 in HTH1 and helix-6 in HTH2, both of which bind to adjacent major grooves of the *mar* DNA. Binding involves bending of the DNA ($\approx 35^\circ$) and makes base specific contacts in the major grooves (Rhee et al., 1998). Alanine-scanning mutagenesis (Gillette et al., 2000) and the solved cocrystal structure of *mar*-MarA complex (Rhee et al., 1998) has identified the amino acid residues of both S1 and S2 subdomains of MarA, which are involved in the *mar* DNA-binding. A schematic representation of interactions of SD1 and SD2 with cognate *mar* DNA base(s) is shown in (**Fig. 1.11c**). In the recognition helix-3 and 6, amino acid residues R46 and R96 protrude into the adjacent major groove and make hydrogen bonds (interatomic distance $< 3.5\text{\AA}$) with G-20, G-30, C-31 bases (with R46 of helix-3) and A-9, G-10, G-40 bases (with R96 of helix-6). In addition, T93 of recognition helix-6 makes water mediated sequence-specific hydrogen bonds with C-41 and T-42 bases of *mar* DNA (Rhee

et al., 1998). Moreover, W42 of recognition helix-3 and Q92 of recognition helix-6 may make sequence-specific hydrogen bonds with corresponding C-32 and T-7 bases of *mar* DNA (Rhee et al., 1998). It was suggested that the overall specificity of protein-DNA interactions in *mar*-MarA complex is provided by hydrogen bonds and the shape complementarity of the MarA residues within the recognition helices and the *mar* DNA bases of the adjacent major grooves (Rhee et al., 1998).

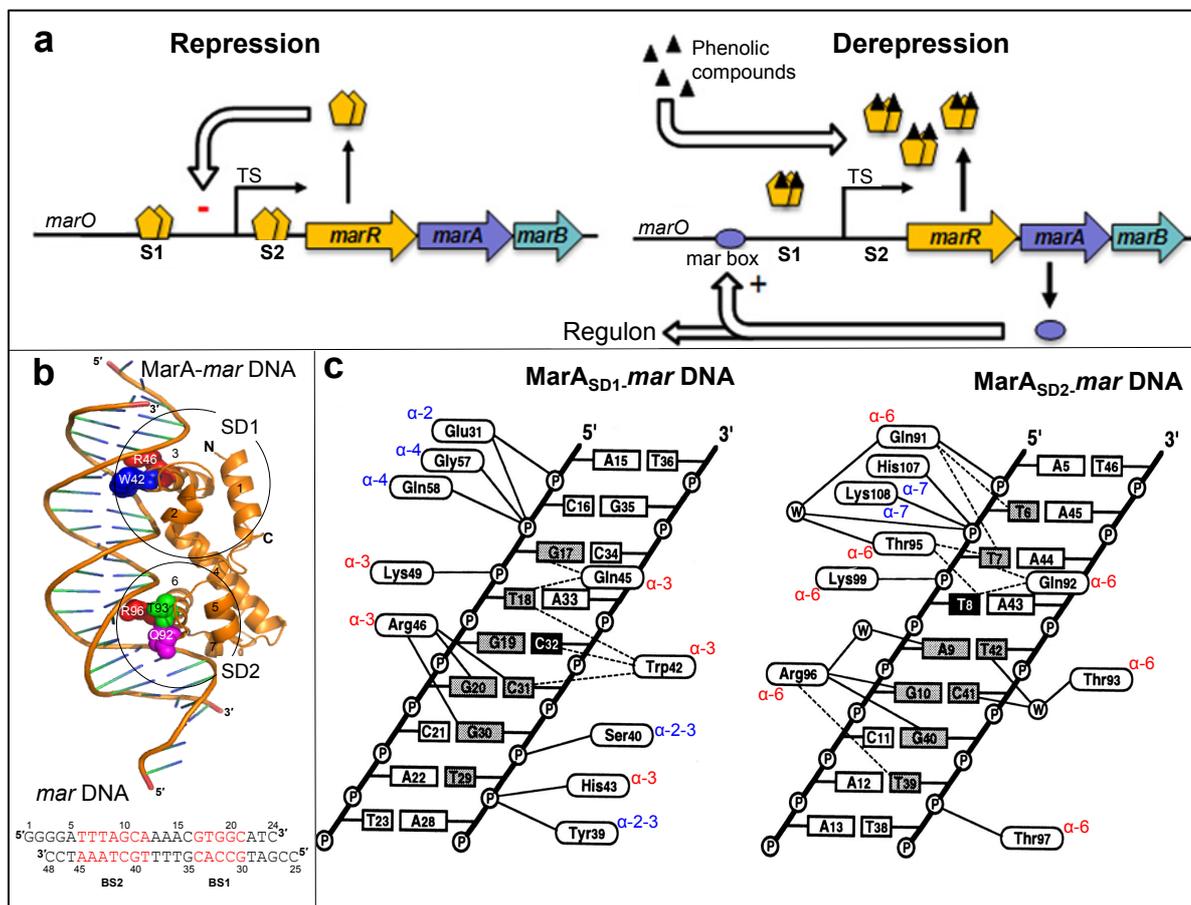


Figure 1.11| **Transcription activation by MarA.**

a) MarA regulon, in the absence of sensing signal (phenolic compounds/antibiotics), MarR binds (as a dimer) to S1 and S2 sites on the *marO* promoter region, repressing the *mar* regulon. In presence of a sensing signal, the signaling compound binds MarR and inactivates it. Release of MarR leads to derepression of the operon, production of MarA, which binds to the *mar* box and thus activating transcription of genes in the *mar* regulon. **b)** *mar*-MarA cocrystal structure complex (PDB-1bl0). MarA subdomain 1 and 2 are encircled. Specificity determining residues, R46 (in recognition helix-3) and R96 (in recognition helix-6) are shown as red balls. W42 (in helix-3) and Q92 and T93 (in helix-6) were also suggested to provide sequence specific interactions. The *mar* DNA used in the cocrystal structure is indicated underneath. Binding site (BS) 1 and 2 for corresponding helix-3 and 6 interactions are shown in red text. **c)** DNA-protein interactions mediated by residues from SD1 and SD2 with *mar* DNA. Amino acid residues in the corresponding helix (α) are encircled while DNA bases are boxed. Recognition helices, 3 and 6 are indicated as α -3 and α -6. Van der Waal interactions (interatomic distance $< 4\text{\AA}$) are indicated by solid lines whereas hydrogen bonds (interatomic distance $< 3.5\text{\AA}$) are indicated by dashed lines. Water molecule (W) involved in SD2 binding are encircled. DNA bases buried by MarA binding are shaded in grey. DNA bases (T-8 and C-32), which may contribute sequence specificity are indicated in black boxes. Picture **(a)** is adapted from (Duval and Lister, 2013), pictures **(b)** and **(c)** are adapted from (Rhee et al., 1998).

iii) Transcription activation by Rob: conferring resistance to heavy metals

Rob (33.1 kDa) of *E. coli* is also a global regulator within the *Enterobacteriaceae* family and regulate transcription of more than 40 genes, conferring resistance to antibiotics, organic solvent and heavy metals (Duval and Lister, 2013). The homologous proteins, MarA and SoxS can also activate the same set of genes albeit the level of activation, differs among these regulators (Duval and Lister, 2013). Transcription activation of these genes occurs by Rob binding (as a monomer) to a highly degenerate sequence motif, located either upstream of or within the promoter of the target gene (Jair et al., 1996b; Tanaka et al., 1997). Rob transcription has been shown to be repressed by SoxS, MarA and by itself (Chubiz et al., 2012; McMurry and Levy, 2010; Michan et al., 2002; Schneiders and Levy, 2006). Rob is synthesised constitutively unlike MarA and SoxS whose expression requires inducing signals (Rosenberg et al., 2003). However, being abundant inside the cell, a major fraction of Rob remains inactive due to its sequestration in inclusion bodies (Azam et al., 2000).

Unlike MarA and SoxS, Rob is a longer protein (289 amino acids) and contains a C-terminal sensing domain (residues, 121-289), which interacts with bile salts or dipyrindyl to activate Rob (Griffith et al., 2009; Rosenberg et al., 2003; Rosner et al., 2002) as depicted in **Fig. 1.12a**. Like MarA, the N-terminal DBD (residues, 1-120) contains seven α -helices with two HTH motifs (HTH1; residues, 26-46, (α 2-T- α 3) and HTH2; residues, 73-96 (α 5-T- α 6)), joined by a linker helix (helix-4), which fixes their relative orientation. The C-terminal sensing domain of Rob contains a pair of α -helices and 10 antiparallel β -strands, organised in such a way that eight β -strands are sandwiched between this pair of α -helices. The remaining two β -strands are present at the C-terminus (Kwon et al., 2000). The interface between DBD and CTD is composed of a triad of hydrophobic residues, V64, L68, W109 (from DBD) and L149, I152, W244 (from CTD), and an electrostatic interaction between E108 and R288 of DBD and CTD, respectively. The N-terminal DBD of Rob shares 51% identity and 71% similarity with MarA (Duval and Lister, 2013). In the 3D cocrystal structure of *micF*-Rob complex (PDB-1d5y), unlike *mar*-MarA complex (PDB-1bl0), only one HTH motif (HTH1) of DBD protrudes into the adjacent major groove of the *micF* DNA (**Fig. 1.12b**) and makes sequence specific contacts (Kwon et al., 2000). The second HTH motif (HTH2) interacts with the DNA backbone instead of protruding into the adjacent major groove of DNA and as a result of this the *micF* DNA appears unbent (**Fig. 1.12b**) (Kwon et al., 2000). In this structure, the amino acid residues within HTH1 motif, interact with conserved A-box (relative to recognition α -helix 3) of the *micF* DNA major groove includes, W36, Q39 and R40. W36 and Q39 make van der Waals interactions with C-7 base whereas R40 makes hydrogen bond with G-6 base within the A-box of the *micF* DNA. Unlike *mar*-MarA cocrystal structure, no specific interactions were observed within HTH2 motif and the adjacent major groove of the *micF* DNA (Kwon et al., 2000; Rhee et al., 1998). Although some non-specific interactions were observed, formed by R90 and K94 with phosphate group of the *micF* DNA backbone (Kwon et al., 2000). The *micF*-Rob interactions at the A-box are analogous to the *mar*-MarA complex interactions at the recognition helix-3. Recently, *micF*-Rob interactions from HTH2 motif with the B-box of *micF* DNA (relative to recognition helix-6) have been confirmed by *in vivo* epigenetic and mutagenesis studies (Taliaferro et al., 2012). This suggests that absence of HTH2 motif interaction with B-box of *micF* DNA in the cocrystal structure of *micF*-Rob complex

could be an artifact. Recently it has also been shown that class II promoters directed transcription, activated by Rob-DBD interaction with region 4 of σ^{70} subunit of RNAP and the CTD does not substantially alter this mode of interaction (Taliaferro et al., 2012).

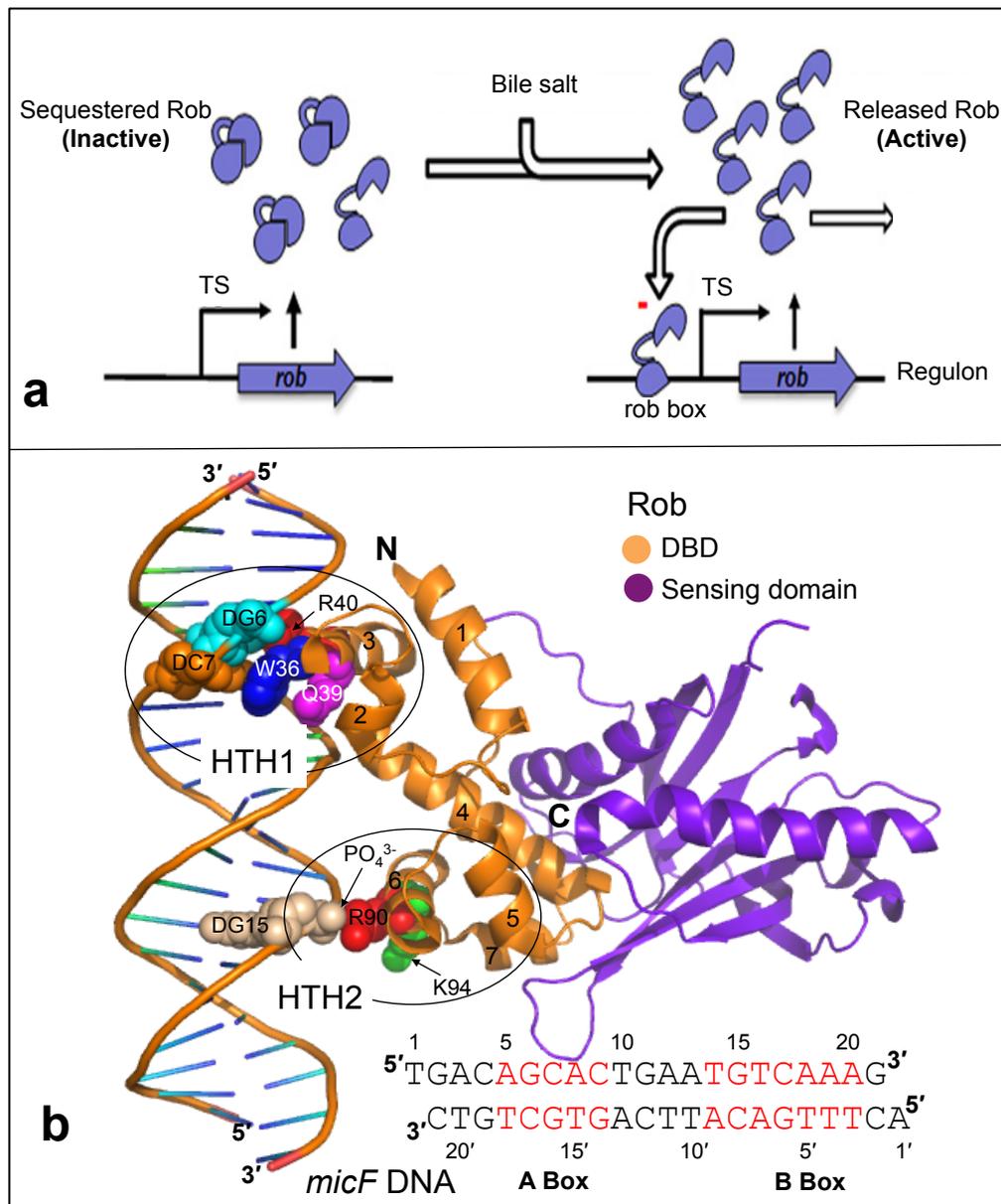


Figure 1.12| **Transcription activation by Rob.**

a) Rob regulon, in the absence of sensing signal (bile salt), Rob is sequestered in the intracellular foci and remains in the inactive form. Once the cell comes in contact with sensing signals then the sequestered form of Rob is released from the intracellular foci and Rob becomes active and binds (as monomer) to the Rob box, located at the upstream promoter region. The activated Rob then activates the expression of other genes in its regulon. **b)** Rob-*micF* DNA cocrystal structure complex (PDB-1d5y). N-terminal DNA binding domain (DBD) and C-terminal sensing domain are indicated in different colors. Specific interactions from HTH1 residues, W36, Q39 and R40 with G-6 and C-7 bases of *micF* DNA and non-specific interaction from HTH2 residues, R90 and K94 with phosphate group of *micF* DNA backbone are indicated in different color balls. The *micF* DNA bound in the cocrystal structure is indicated underneath with corresponding helix-3 and 6 binding sites as box A and box B. Picture **(a)** is adapted from (Duval and Lister, 2013) whereas pictures **(b)** is adapted from (Kwon et al., 2000).

iv) Transcription activation by SoxS: a stress response regulator

SoxS (12.9 kDa) is a global regulator that regulates more than 40 genes involved in oxidative stress and protection of bacterial cells from the oxidative stresses. This is governed by binding to a highly degenerate 'soxbox' DNA motif, (ANVGCACWWWNKRHCAAHN consensus, where N: A/T/G/C, V: A/C/G, W: A/T, K: G/T, R: A/G, H: A/C/T and bold letters represent two binding sites), located upstream of promoter region in the chromosomally encoded *soxR-soxS* regulon (Duval and Lister, 2013). SoxS is the smallest member (107 amino acids) of the AraC/XylS family and closely related to MarA, having 41% identity and 67% similarity (Duval and Lister, 2013). Like MarA, SoxS activates transcription at both class I and class II promoters by binding as a monomer (Duval and Lister, 2013). In *soxR-soxS* regulon, *soxR* encodes SoxR regulator (a member of MerR family) whose expression is divergently transcribed from a promoter embedded in the *soxS* structural gene (Amabile-Cuevas and Demple, 1991). In the absence of an oxidative stress signal, SoxR binds to the *soxS* promoter region as a homodimer and prevents transcription (Hidalgo et al., 1998). Due to the inactivity of Lon and FtsH proteases on SoxR a very low level of SoxS is synthesised, in the absence of inducing stress signal (Griffith et al., 2004). Once SoxR is oxidized, it becomes an activator of *soxS* transcription and synthesises SoxS (Nunoshiba et al., 1992). When SoxS is synthesised optimally it is able to repress its own transcription, in addition to the repression mediated by MarA and Rob (Chubiz et al., 2012; Nunoshiba et al., 1993). SoxS regulon shows a remarkable degree of overlap with MarA regulon albeit different degree of affinities for different promoters and the methods of activation (Duval and Lister, 2013; Pomposiello et al., 2001). Therefore, SoxS is the first to respond to any oxidative stress and MarA is the first to deal with antibiotic assaults (Duval and Lister, 2013).

A comprehensive alanine scanning mutagenesis of SoxS identified amino acid residues, required for transcription activation at both class I (*zwf* and *fpr*) and class II (*fumC* and *micF*) promoters and DNA-binding (Griffith and Wolf, 2002). It was demonstrated that the predicted N-terminal HTH motif (HTH1, residues 25-46) of SoxS interacts with highly conserved region of the 'soxbox' assigned as recognition element 1 (RE1, GCAC consensus) and the predicted C-terminal HTH motif (HTH2, residues 73-96) interacts with less conserved region of the 'soxbox' assigned as recognition element 2 (RE2, CAAA consensus). It has been shown that Ala substitution of predicted surface exposed residues of recognition helices, 3 and 6 of SoxS were more detrimental to DNA-binding than the substitutions of residues predicted to make base specific contacts (Griffith and Wolf, 2002). Substitution of many residues within predicted recognition helix-6, which is assumed to make base specific contact with RE2 had a relatively small effect on DNA-binding, suggesting the possibility of alternate protein-DNA interactions than those inferred from the cocrystal structure of the *mar*-MarA complex (PDB-1bl0). Furthermore, mutant SoxS_{K30A} failed to activate transcription at both class of the promoters. A similar phenotype was observed upon substitution of putatively surface-exposed surrounding residues, H3A, K5A, D9A, S31A and V45A mutants, suggesting surface generated by H3, K5, D9, K30, S31 and V45 residues is essential for the interactions with RNAP and transcription activation at both classes of SoxS-dependent promoters. In contrast, the Ala substitutions of F74, D75, M78, D79 and Q85 residues appeared to create a surface for RNAP interactions and transcription activation only at class II promoters (Griffith and Wolf, 2002).

v) Transcription activation by Rns: a fimbrial regulator

Rns (32 kDa) or its closest homolog CfaD/CfaR (95% amino acid identity) activates the expression of several pili/fimbriae from enterotoxigenic *E. coli* (ETEC) (Bodero et al., 2008; Caron et al., 1989; Caron and Scott, 1990). About 25 different fimbriae have been identified thus far in various strains of ETEC, although most strains express only two or three types (Gaastra and Svennerholm, 1996; Grewal et al., 1997; Peruski et al., 1999; Pichel et al., 2000). Most commonly expressed fimbriae are, CS1-CS6 and CFA/I (Evans et al., 1975), out of which the regulation mechanisms of CS1, CS2 and CFA/I are reasonably well studied (Caron et al., 1989; Caron and Scott, 1990). Fimbriae whose expression is activated by Rns include, CS1-CS4, CS14, CS17 and CS19 (Bodero et al., 2008; Caron et al., 1989). Expression of CFA/I fimbriae is activated by Rns homologs; CfaD/CfaR, which are functionally interchangeable with Rns and recognises the same binding sites as Rns (Caron and Scott, 1990). In addition to activation of the expression of fimbrial genes, Rns positively regulates its own gene (*rns*) expression by binding (direct) at three sites, centered at -227, +43 and +82 (relative to transcription start site (TSS)) (Froehlich et al., 1994). Position of uncommon downstream binding sites (at +43 and +82) is often associated with transcription repression. However, it has been shown that activation of *rns* requires at least one of the two downstream binding sites along with the upstream binding site (Froehlich et al., 1994). Rns activates the expression of several other putative virulence genes (Munson et al., 2002; Pilonieta et al., 2007).

A DNaseI footprinting analysis revealed that Rns binding site is located immediately upstream of -35 element of *P_{coo}* promoter (required for the expression of CfaA, CS1 fimbriae) and other promoters of CS17, CS19 and PCF071 fimbriae (Bodero et al., 2008; Munson and Scott, 1999), suggesting Rns may activate transcription by direct contact with RNAP. In addition to binding site near -35 promoter element, a second binding site was identified at all the promoter regions of these fimbriae (at -88 in CfaA, at -144 in CS1, at -109.5 in CS17 and PCF071 and at -108.5 in CS19) (Bodero et al., 2008; Munson and Scott, 1999). Each site showed an additive effect on Rns-dependent transcription activation and thus required for Rns full activation (Bodero et al., 2008; Munson and Scott, 1999). Rns homologs have been identified in several strains of ETEC and other enteric bacterial pathogens. Proteins having strong homology to Rns include CsvR and FapR from ETEC (De Haan et al., 1991; Klaasen and de Graaf, 1990), PerA of enteropathogenic *E. coli* (Tobe et al., 1996), AggR of enteroaggregative *E. coli* (Nataro et al., 1994), ToxT of *Vibrio cholerae* (Lowden et al., 2010) and VirF of *Shigella* (Sakai et al., 1986). Among these, AggR, CsvR, PerA and ToxT regulate the expression of fimbrial genes (De Haan et al., 1991; Klaasen and de Graaf, 1990; Lowden et al., 2010; Nataro et al., 1994; Tobe et al., 1996). VirF of *Shigella* activates the expression of *icsA* and *virB* genes, having roles in invasion and cell-to-cell spread of bacteria in the host epithelial cells, respectively (Sakai et al., 1988).

Transcription activation by Rns and its homologs is thermo-sensitive, which is repressed by H-NS at low temperatures (Munson and Scott, 2000). Like most AraC/XylS family regulators, Rns is also a two domain protein, a NTD (largely unknown function) and a conserved C-terminal DBD with two predicted HTH motifs (Mahon et al., 2010). A mutagenesis study (pentapeptide insertion) demonstrated that Rns uses both HTH motifs to make DNA contacts and thereby

activates the expression of CS1 fimbrial genes (Mahon et al., 2010). In support of this a uracil interference study by Munson et. al., showed that Rns contacts two major grooves of the DNA in order to activate transcription (Munson and Scott, 1999, 2000). Like DBD of several AraC/XylS family regulators, Rns DBD is believed to be involved in making RNAP contacts, although the region involved in making such contacts has not been identified yet (Munson and Scott, 1999, 2000). Comparatively, the role of Rns-NTD is largely unknown. It is not clearly known whether Rns acts primarily as a monomer or dimer as previous studies regarding this are contradictory to each other (Basturea et al., 2008; Mahon et al., 2012). Moreover, there is no evidence that Rns responds to any effector molecules (Basturea et al., 2008). The N-terminal deletion mutagenesis showed that Rns-NTD is essential for transcription activation both *in vivo* and *in vitro* (Basturea et al., 2008). A truncated version of Rns (first 61 amino acids deleted) has also shown the complete loss of DNA-binding and transcription activation or repression at the corresponding promoter regions (Basturea et al., 2008). Rns-NTD motif from I12-M18 is highly conserved among the closest homologs, sharing about 74% amino acid identity compared to 26% identity in the overall NTDs, suggesting a crucial role of this motif in the overall function of Rns. In support of this, Munson et. al., isolated two random mutants (I14T and N16D) of Rns-NTD and found their activities were decreased dramatically at the *rns* promoter, indicating the importance of I14 and N16 residues in transcription activation (Basturea et al., 2008). Additionally, it was found that the I14T and N16D mutations have disparate effects at other Rns regulated promoters and suggested that these residues may interact with DBD in a manner to activate transcription (Basturea et al., 2008).

Aim and objectives:

Fraction 1 (F1) is a key antigen of *Y. pestis*, encoded by virulence plasmid (pFra or pMT1) borne *caf* locus and is used in both plague detection and vaccine design. A wealth of knowledge is available on its structure and assembly by the chaperone usher (CU) system. Despite that, virtually no information is available on regulation of the *caf* locus. Understanding the regulation of expression of key antigens from a bacterial pathogen is one strategy to design antibacterial therapeutics. With respect to F1, this might be ensuring optimum production of F1 in attenuated vaccines or heterologous vaccines expressing native F1. Initial sequencing identified Caf1R is a transcriptional activator of the *caf* locus, belonging to an AraC/XylS family of bacterial transcriptional regulators that contain a conserved DNA binding domain (DBD). Additionally, production of F1 is known to be temperature regulated and transcriptomic studies have provided some relevant data, but there is no further study on Caf1R itself or on the mode of regulation within the *caf* locus. Regulation of CU systems is very diverse and, unlike the wealth of knowledge available on regulation of *E. coli fim* and *pap* operons, AraC/XylS-type regulators controlling expression of CU systems remain poorly characterised. A detailed molecular characterisation of Caf1R would provide useful information of relevance to other CU systems using the same mechanism of regulation. F1 is naturally expressed from a low copy number plasmid. Expression of the *caf* locus from the native promoter on high copy number recombinant plasmids can be a metabolic burden on the cell, leading to deletions and mutations that reduce expression. Previous studies in the laboratory identified a single spontaneous point mutation (E98G) in Caf1R-DBD that lead to a dramatic reduction in F1 expression (Lopez-Tolman, A.; unpublished). Understanding the impact of this mutation on Caf1R function was the starting point of this study, which focused on a detailed study of Caf1R-DBD and its function as an activator of *caf* transcription. The main objectives were as follows:

- To characterise the N-terminal DBD of Caf1R, identify key DNA binding residues and explain the impact of the Caf1R-E98G mutation on Caf1R function and assembly of F1.
- To create and compare tagged Caf1R constructs, optimise the expression and purification of tagged Caf1R for use in a DNA-protein interaction study.
- To localise active promoters within the *caf* locus and monitor the impact of Caf1R on their activity, in order to define Caf1R binding.
- Analysis of thermoregulation in the *caf* locus.

Chapter 2

Materials and Methods

2.1 Introduction

This chapter covers all the materials, methods, tools and software used routinely. In the case where additional explanation of detail or method is required this is dealt with in the appropriate chapter.

2.1.1 Bacterial strains and culture conditions

Details of *E. coli* bacterial strains used in this research are listed in Table 2.1. Bacteria were grown at 26°C or 37°C in Luria Bertani broth (LB) (1% (w/v) trypton, 0.5% (w/v) yeast extract, 0.09 M NaCl; pH 7.0)). For overnight (15-24 h) growth 0.6% (w/v) glucose was added in LB along with appropriate antibiotic(s), when required. All bacterial cultures (in LB) were grown with shaking at 225 rpm. The additional details of culture conditions are discussed in the respective section. Bacterial strains and transformants carrying confirmed plasmid constructs were stocked in LB supplemented with 15% (w/v) glycerol at -80°C. Luria Bertani agar (LA) (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.09 M NaCl, 1.5% (w/v) agar; pH 7.0)), with appropriate antibiotic(s) was used throughout for culture of transformants. Antibiotic stocks (mg/ml) used in this study were Ampicillin (Amp-100 mg/ml), Chloramphenicol (Cm-50 mg/ml) and Kanamycin (Kan-50 mg/ml) with final working concentrations, 100, 10-34 and 30 µg/ml, respectively. All antibiotic stocks were kindly provided by Hanson, K. (a senior lab technician for Knight building).

Table 2.1| *E. coli* strains used in this study

Strain	Description	Genotype	Source
TOP10 (*R1120)	General cloning and expression studies of pBADHisA, pRS415 and pRS550 plasmid based constructs. Feature: Keeps constant level of L-arabinose during growth.	<i>F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (Str^R) endA1 nupG λ-</i>	Invitrogen Cat No. C4040-10
DH5α (R1023)	General cloning and expression studies of pACYCDuet-I based constructs.	<i>F⁻ endA1 hsdR17(r_k⁻ m_k⁺) supE44 thi -1 recA1 gyrA (Nal^r) relA1 Δ(lacIZYA- argF)U169 deoR (φ80dlacΔ(lacZ)M15)</i>	(Sambrook and Russell, 2001); Lab stock.
Stellar™	Culturing of Infusion mixtures. Feature: Enhanced ability to repair nicks in the infused DNA fragment.	<i>F⁻ endA1 supE44 thi-1 recA1 relA1 gyrA96 phoA Φ80d_lac ZΔM15 Δ(lacZYA- argF)U169 Δ(mrr-hsdRMS- mcrBC) ΔmcrA λ-</i>	Clontech Cat No. 636763
BL21(DE3) (R837)	His ₆ and MBP-tagged Caf1R overexpression from pET28a ⁺ and pMALc2x plasmids based constructs. Feature: General expression strain of λDE3 lysogen contains a chromosomal copy of T7	<i>F⁻ ompT gal dcm lon hsdS_B (r_B⁻ m_B⁻) λ(DE3)</i>	(Studier and Moffatt, 1986); Lab stock.

	RNAP gene under the control of <i>lacUV5</i> promoter.		
BL21 Star™ (DE3)pLysS	His ₆ -tagged Caf1R overexpression from pET28a ⁺ plasmid based constructs Feature: BL21(DE3) derivative, provides enhanced mRNA stability with low proteins background expression.	F ⁻ <i>ompT hsdSB</i> (r _B ⁻ , m _B ⁻) <i>gal dcm rne131</i> (DE3) pLysS (Cm ^R)	Life Technologies Cat. No. C6020-03
Rosetta-gami™ 2 (DE3)	His ₆ -tagged Caf1R overexpression from pET28a ⁺ plasmid based constructs. Feature: i) BL21 (DE3) derivative with combined properties of Rosetta 2 and Origami 2 strains. ii) Alleviate codon bias by supplying 7 rare tRNAs codons (AUA, AGG, AGA, CUA, CCC, GGA and CGG) from pRARE2 plasmid. iii) Enhance disulfide bond formation in cytoplasm.	Δ(<i>ara-leu</i>)7697 Δ <i>lacX74</i> Δ <i>phoA PvuII phoR araD139 ahpC galE galK rpsL</i> (DE3) F' [<i>lac⁺ lacI^q pro</i>] <i>gor522::Tn10 trxB</i> pRARE2 (Cm ^R Str ^R Tet ^R)	Novagen® Cat. No. 71351
LOBSTR-BL21 (DE3)	His ₆ -tagged Caf1R overexpression from pET28a ⁺ plasmid based constructs. Feature: i) BL21 (DE3) derivative. ii) Deficient in major <i>E. coli</i> contaminants, ArnA (74.3 kDa) and SlyD (21 kDa) (Robichon et al., 2011).	F ⁻ <i>ompT hsdSB</i> (r _B ⁻ , m _B ⁻) <i>gal dcm</i> (DE3) <i>araA- slyD</i> ⁻	(Andersen et al., 2013); Kerafast Cat. No EC1001
LOBSTR-BL21 (DE3)-RIL	His ₆ -tagged Caf1R overexpression from pET28a ⁺ plasmid based constructs. Feature: i) LOBSTR-BL21 (DE3) derivative. ii) Contains extra copies of genes coding for <i>argU</i> , <i>ileY</i> , and <i>leuW</i> tRNA.	F ⁻ <i>ompT hsdSB</i> (r _B ⁻ , m _B ⁻) <i>gal dcm</i> (DE3) <i>araA- slyD- araU ileY leuW</i> (Cm ^R)	(Andersen et al., 2013); Kerafast Cat. No EC1002
K12-ER2508	MBP-tagged Caf1R overexpression from pMALc2x based constructs. Feature: i) RR1 derivative with <i>lon</i> mutation. ii) Deficient in major ATP-dependent protease. iii) Improves stability of heterologous proteins in cytoplasm.	F ⁻ <i>ara-14 leuB6 fhuA2</i> Δ(<i>argF-lac</i>)U169 <i>lacY1 lon::miniTn10</i> (Tet ^R) <i>glnV44 galK2 rpsL20</i> (Str ^R) <i>xyl-5 mtl-5</i> Δ(<i>malB</i>) <i>zjc::Tn5</i> (Kan ^R) Δ(<i>mcrC-mrr</i>) _{HB101}	NEB; Cat no. E4127

SHuffle® T7	His ₆ -tagged Caf1R overexpression from pET28a+ plasmid based constructs. Feature: i) K12 derivative ii) Expresses DsbC chaperone in cytoplasm that enhances cytoplasmic disulfide bonds formation.	<i>F' lac, pro, lacI^Q / Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt::pNEB3-r1-cDsbC (Spec^R, lacI^Q) ΔtrxB rpsL150(Str^R) Δgor Δ(malF)3</i>	NEB Cat. No. C3026H
--------------------	--	--	------------------------

* Assigned stock number

2.1.2 Plasmids and their respective constructs

A descriptive list of all vectors used and plasmids constructed during this study is tabulated in Table 2.2.

Table 2.2| **Plasmid constructs used and designed during this study**

Plasmid	Description	Source
pFMA1	Contains only <i>caf1M</i> , <i>caf1A</i> and <i>caf1</i> genes of the <i>caf</i> locus under the control of IPTG-inducible Ptrc promoter. Amp ^R	(MacIntyre et al., 2001); Lab stock
pBADHisA (*R638)	Low copy number, pBR322-derived expression plasmid vector (4.1 kb) contains N-terminal His ₆ -tag and L-arabinose inducible P _{BAD} promoter. Amp ^R	Invitrogen; (Lab stock)
pBADhCaf1R _{E98G} (R1110)	pBADHisA based plasmid construct contains gene for mutated Caf1R _{E98G} , suncloned from pFS2 plasmid between SacI and BglII restriction sites.	Lopez-Tolman, A. (Lab stock)
pBADhCaf1R (R1117) (hCaf1R, 40.9 kDa)	Derivative of pBADhCaf1R _{E98G} with repaired E98G mutation (G98E, wild type).	Lopez-Tolman, A. (Lab stock)
pBADΔhCaf1R (R1121/R1119)	Derivative of pBADhCaf1R with deleted His6-tag and Enterokinase cleavage-site fragment. Sequence confirmed for two transformants, 4 (R1121) and 8 (R1119), both stocked and number 4 (R1121) was used.	This study
pACYCDuet-I	Medium copy number (10-12), P15A-derived cloning and expression plasmid vector (4.0 kb), contains two multiple cloning sites (MCS) preceded by IPTG inducible T7 promoter. Genes cloned into MCS1 can be sequenced by ACYCDuetUP1 and DuetDOWN1 primers whereas for MCS2 cloned gene DuetUP2 and T7 terminator primers can be used. Cm ^R	Novagen® (Lab stock)
pACYCF1 _{SpM} (R1101b)	pACYCDuet-I based plasmid construct contains spontaneous mutant <i>caf</i> locus (5.128 kb) with 188 and 175 bp flanking ends (relative to <i>caf1R</i> and <i>caf1</i> stop codon). Complete fragment was subcloned (in between NcoI/SacI of MCS1) from pFS2 plasmid (Drozdov et al., 1995; Galyov et al., 1990) under native regulatory system (<i>caf1R-caf1M</i> intergenic region and mutant Caf1R _{E98G} regulator).	Leonard, M. (Lab stock)
pACYC-MA1	Derivative of pACYCF1 _{SpM} with deleted Caf1R _{E98G} . Expression of <i>cafMA1</i> is controlled by <i>caf1R-caf1M</i> intergenic region.	This study

	Sequence confirmed transformant 5 was used.	
pACYC-R _{E98G}	Derivative of pACYCF1 _{SpM} with deleted <i>cafMA1</i> ; <i>caf1R</i> expression is controlled by <i>caf1R-caf1M</i> intergenic region. Sequence confirmed two transformants, 1 and 5 were stocked and 1 was used.	This study
pACYCF1 (R1122b)	Derivative of pACYCF1 _{SpM} with repaired Caf1R _{E98G} .	This study
pACYC-R	Derivative of pACYCF1 with deleted <i>cafMA1</i> ; <i>caf1R</i> expression is controlled by <i>caf1R-caf1M</i> intergenic region. Sequence confirmed two transformants, 1 and 5 were stocked and 5 was used.	This study
pACYC-R _{E98K}	Derivative of pACYC-R, with Caf1R _{E98K} substitution. Sequence confirmed two transformants, 2 and 4 were stocked and number 2 was used.	This study
ΔP _M pACYCF1	Derivative of pACYCF1 with TAAAAT to GCCGGC substitution in -10 element of visually identified P _M promoter. Sequence confirmed three transformants, 1, 2 and 3 were stocked and used all.	This study

Site-specific substitution mutations in *caf1R* of pACYCF1

Helix-2

pACYCF1-R _{I31A}	I31A. Sequence confirmed four transformants, 2-5 were stocked and number 3 was used.	This study
pACYCF1-R _{D32A}	D32A. Sequence confirmed two transformants, 1 and 2 were stocked and number 1 was used.	This study
pACYCF1-R _{C33A}	C33A. Sequence confirmed ten transformants, 3-12 were stocked and number 6 was used.	This study

Loop between Helix-2 and 3

pACYCF1-R _{F40A}	F40A. Sequence confirmed four transformants, 1-4 were stocked and number 4 was used.	This study
---------------------------	--	------------

Helix-3

pACYCF1-R _{R42A}	R42A. Sequence confirmed transformant number 4 was stocked and used.	This study
pACYCF1-R _{R43A}	R43A. Sequence confirmed four transformants, 1-4 were stocked and number 4 was used.	This study
pACYCF1-R _{Y44A}	Y44A. Sequence confirmed four transformants, 1-4 were stocked and number 3 was used.	This study

Helix-4

pACYCF1-R _{R62A}	R62A. Sequence confirmed three transformants, 1-3 were stocked and number 3 was used.	This study
pACYCF1-R _{R62S}	R62S. Sequence confirmed three transformants, 1, 2 and 4, were stocked and number 4 was used.	This study

Helix-6

pACYCF1-R _{Q93A}	Q93A. Sequence confirmed four transformants, 1-4 were stocked and number 1 was used.	This study
pACYCF1-R _{R97A}	R97A. Sequence confirmed three transformants, 1-3 were stocked and number 2 was used.	This study
pACYCF1-R _{E98K}	E98K. Obtained during the repair of Caf1R _{E98G} in pACYCF1 _{SpM} . Sequence confirmed transformants, 7, 11 and 14 were stocked and number 7 was used.	This study
pACYCF1-R _{E98A}	E98A. Sequence confirmed three transformants, 1, 2 and 4, were stocked and number 4 was used.	This study
pACYCF1-R _{E98T}	E98T. Sequence confirmed two transformants, 1 and 2 were stocked and number 1 was used.	This study

Loop between Helix-6 and 7

pACYCF1-R _{T106A}	T106A. Sequence confirmed transformant number 3 was stocked and used.	This study
----------------------------	---	------------

Helix-7

pACYCF1-R _{R108A}	R108A. Sequence confirmed three transformants, 1-3 were stocked and number 1 was used.	This study
pACYCF1-R _{Q109A}	Q109A. Sequence confirmed transformant number 3 was stocked and used.	This study

Constructs used for tagged Caf1R expression and purification

pMAL-c2x	Medium copy number (~20), pMB1-derived expression plasmid vector (6.646 kb), contains N-terminal maltose-binding protein (MBP)-tag and <i>ptac</i> promoter. Specific for over-expression of cytoplasmic protein, induction by IPTG; Amp ^R	NEB
pMALc2-Caf1R (MBPCaf1R, 79.3 kDa)	pMAL-c2x based plasmid construct, contains N-terminally MBP-tagged wild type native Caf1R (MBPCaf1R). Sequence confirmed three transformants, 23, 24, and 26 were stocked and number 23 was used.	This study
pMALc2-Caf1R _N (MBPCaf1R _N , 57.3 kDa)	pMAL-c2x based plasmid construct, contains N-terminally MBP-tagged DBD of wild type native Caf1R (MBPCaf1R _N). Sequence confirmed three transformants, L1-L3 were stocked and L1 was used.	This study
pET28a ⁺	High copy number (~40), pBR322-derived expression plasmid vector (5.369 kb), contains N-terminal His ₆ -tag and T7 promoter, induction by IPTG; Kan ^R	Novagen
pEThCaf1R (hCaf1R ^T , 40 kDa)	pET28a ⁺ based plasmid construct, contains N-terminally His ₆ -tagged wild type native Caf1R. Sequence confirmed two transformants, 6 and 8 were stocked and number 8 was used.	This study
pEThCaf1R _N (hCaf1R ^T _N , 16.4 kDa)	pET28a ⁺ based plasmid construct, contains N-terminally His ₆ -tagged DBD of wild type native Caf1R. Sequence confirmed eight transformants, 1-8 were stocked and 8 th was used.	This study

pEThCaf1R _{gs} (hCaf1R ^T _{gs} , 38.5 kDa)	pET28a ⁺ based plasmid construct, contains N-terminally His ₆ -tagged synthetic codon-optimised Caf1R; subcloned in between EcoRI and NheI sites.	GenScript, USA
pEThCaf1R _{gsE98G} (hCaf1R ^T _{gsE98G} , 38.5 kDa)	Derivative of pEThCaf1R _{gs} with Caf1R _{E98G} mutation.	GenScript, USA
pBADhCaf1R _{gs} (hCaf1R _{gs} , 37.5 kDa)	Derivative of pEThCaf1R _{gs} , contains N-terminally His ₆ -tagged synthetic codon-optimised wild type Caf1R in the pBADHisA plasmid; subcloned in between NheI and EcoRI sites.	This study
pMALc2-Caf1R _{gs} (MBPCaf1R _{gs} , 78.9 kDa)	Derivative of pEThCaf1R _{gs} , contains N-terminally MBP-tagged synthetic codon-optimised wild type Caf1R in pMALc2x plasmid.	This study
pRS415	Multicopy number, pNK678 derived transcription fusion plasmid vector (10.752 kb), contains promoter less <i>lacZ</i> . Amp ^R	(Simons et al., 1987)

pRS415 plasmid based promoter-*lacZ* fusion constructs to localise *caf1R* promoter(s) and Caf1R binding upstream of *caf1R*

pRScaf1M'-R'- <i>lacZ</i>	Contains reverse complement of complete <i>caf1R-caf1M</i> intergenic region (327 bp) plus 192 and 158 bp from the 5' of <i>caf1R</i> (ATG start) and <i>caf1M</i> , respectively. Complete fragment was amplified from pACYCF1 _{SPM} and subcloned in between EcoRI and BamHI sites (upstream of <i>lacZ</i>) of pRS415 plasmid.	Qahtani, A. and Pereira, I. (Lab stock)
---------------------------	---	---

Contains -xxx bp to +192 bp from Caf1R start (ATG), Infused upstream of promoter less *lacZ*

pRScaf1R' ₋₁₃₇₊₁₉₂ - <i>lacZ</i>	-137 bp, contains R1, R2 and R3 potential Caf1R binding repeats, and all three predicted promoters, P _R ¹ , P _R ² and P _R ^K for <i>caf1R</i> transcription. Sequence confirmed transformant number 7 was stocked and used.	This study
pRScaf1R' ₋₁₁₄₊₁₉₂ - <i>lacZ</i>	-114 bp, possesses R1 and R2 repeat motifs and all three predicted promoters, P _R ¹ , P _R ² and P _R ^K for <i>caf1R</i> . Sequence confirmed transformant number 7 was stocked and used.	This study
pRScaf1R' ₋₇₀₊₁₉₂ - <i>lacZ</i>	-70 bp, possesses only R1 repeat and two predicted promoters, P _R ¹ and P _R ² for <i>caf1R</i> . Sequence confirmed transformant number 8 was stoked and used.	This study
pRScaf1R' ₋₆₅₊₁₉₂ - <i>lacZ</i>	-65 bp, possesses only half of the R1 repeat and two predicted promoters, P _R ¹ and P _R ² for <i>caf1R</i> . Sequence confirmed transformant number 8 was stocked and used.	This study
pRScaf1R' ₋₄₆₊₁₉₂ - <i>lacZ</i>	-46 bp, possesses no repeat but contains partial sequence of both P _R ¹ and P _R ² promoters. Sequence confirmed transformant number 8 was stocked and used.	This study
pRS550	Multicopy number pNK678 derived transcription fusion plasmid (12.459 kb) contains promoter less <i>lacZ</i> gene. Identical to pRS415 plasmid except reverse orientation of EcoRI and BamHI cloning site. Amp ^R and Kan ^R	(Simons et al., 1987)

pRS550 plasmid based promoter-*lacZ* fusion constructs to localise *caf1M* promoter(s) and Caf1R binding upstream of *caf1M*

pRScaf1R'-M'- <i>lacZ</i>	Contains complete <i>caf1R-caf1M</i> intergenic region (327 bp) plus 192 and 158 bp from the 5' of <i>caf1R</i> (ATG start) and <i>caf1M</i> , respectively. Complete fragment was amplified from pACYCF1 _{SpM} and subcloned in between BamHI and EcoRI sites (upstream of <i>lacZ</i>) of pRS550 plasmid.	Qahtani, A. and Pereira, I. (Lab stock)
---------------------------	---	---

Contains -xxx bp to +158 bp from Caf1M start (ATG), Infused upstream of promoter less *lacZ*

pRScaf1M'-.270+158- <i>lacZ</i>	-270 bp, possesses R1, R2, R3, R3' and R4' potential Caf1R binding repeats, and all three predicted promoters, P _M , P _M ^B and P _M ^{K1-2} for <i>caf1M</i> transcription. Sequence confirmed transformant number 1 was stocked and used.	This study
pRScaf1M'-.197+158- <i>lacZ</i>	-197 bp, possesses R3' and R4' repeats, and all three predicted promoters, P _M , P _M ^B and P _M ^{K1-2} for <i>caf1M</i> . Sequence confirmed transformant const1col2_3 was stocked and used.	This study
pRScaf1M'-.184+158- <i>lacZ</i>	-184 bp, possesses only R4' repeat but all three predicted promoters, P _M , P _M ^B and P _M ^{K1-2} for <i>caf1M</i> . Sequence confirmed transformant number 3 was stocked and used.	This study
pRScaf1M'-.169+158- <i>lacZ</i>	-169 bp, possesses part of R4' repeat but all three predicted promoters, P _M , P _M ^B and P _M ^{K1-2} for <i>caf1M</i> . Sequence confirmed transformant number 2 was stocked and used.	This study
pRScaf1M'-.141+158- <i>lacZ</i>	-141 bp, possesses no repeat but has -35 element of P _M and complete P _M ^B and P _M ^{K1-2} predicted promoters for <i>caf1M</i> . Sequence confirmed transformant number 3 was stocked and used.	This study
pRScaf1M'-.102+158- <i>lacZ</i>	-102 bp, possesses no repeat motifs but has only P _M ^{K1-2} predicted promoter for <i>caf1M</i> . Sequence confirmed transformant number 1 was stocked and used.	This study

pRS550 plasmid based promoter-*lacZ* fusion constructs to localise *caf1* promoter(s) and to test Caf1R binding upstream of *caf1*

Contains -xxx bp to +88 bp (relative to *caf1* ATG start), Infused upstream of promoter less *lacZ*

pRScaf1A'-1'- <i>lacZ</i>	-259, possesses 179 bp of <i>caf1A</i> 3' end (relative to TGA stop codon), complete <i>caf1A-caf1</i> intergenic region (80 bp) and 88 bp of <i>caf1</i> (relative to ATG start). Sequence confirmed transformant Inf_3 was stocked and used.	This study
pRScaf1'-.181+88- <i>lacZ</i>	-181 bp, possesses R1 ^c , R2 ^a and R2 ^b repeats, and P ₁ ¹ , P ₁ ² , P ₁ ^B and -35 element of P ₁ ³ predicted promoters for <i>caf1</i> transcription. Sequence confirmed transformant number 1 was stocked and used.	This study
pRScaf1'-.158+88- <i>lacZ</i>	-158 bp, possesses R1 ^b , R1 ^c , R2 ^a and R2 ^b repeats, and P ₁ ¹ , P ₁ ² and P ₁ ^B predicted promoters for <i>caf1</i> . Sequence confirmed transformant number 1 was stocked and used.	This study
pRScaf1'-.61+88- <i>lacZ</i>	-61 bp, possesses R1 ^c , R2 ^a and R2 ^b repeat motifs, and P ₁ ¹ and P ₁ ^B predicted promoters for <i>caf1</i> . Sequence confirmed transformant number 2 was stocked and used.	This study

pRScaf1' ₋₃₈₊₈₈ -lacZ	-38 bp, possesses R1 ^c , R2 ^a and R2 ^b repeats, and P ₁ ^B and -10 element of P ₁ ¹ predicted promoters for <i>caf1</i> . Sequence confirmed transformant number 1 was stocked and used.	This study
Contains -259 bp to -xx bp (relative to <i>caf1</i> ATG start), Infused upstream of promoter less <i>lacZ</i>		
pRScaf1' ₋₂₅₉₋₆₁ -lacZ	-61 bp, possesses R1 ^a and R1 ^b repeats, and P ₁ ³ and P ₁ ² predicted promoters for <i>caf1</i> . Sequence confirmed transformant number 2 was stocked and used.	This study
pRScaf1' ₋₂₅₉₋₂₂ -lacZ	-22 bp, possesses R1 ^a and R1 ^b and part of R2 ^a repeats, and P ₁ ³ , P ₁ ² , P ₁ ¹ and -35 of P ₁ ^B predicted promoters for <i>caf1</i> . Sequence confirmed transformant number 1 was stocked and used.	This study
pRS550 plasmid based promoter-<i>lacZ</i> fusion constructs to test Caf1R binding (cis-orientation of <i>caf1R</i>) at <i>caf1R</i> or <i>caf1M</i> promoter(s)		
pRScaf1R _{E98G} -M'-lacZ	Contains mutant <i>caf1R</i> (E98G) and wild type <i>caf1R-caf1M</i> intergenic region plus 159 bp of <i>caf1M</i> (relative to ATG). Complete fragment was amplified from pACYCF1 _{SpM} and subcloned in between BamHI and EcoRI sites (upstream of <i>lacZ</i>) of pRS550 plasmid. Sequence confirmed transformant 9b2 was used and stocked.	Qahtani, A. and Pereira, I. (Lab stock)
pRScaf1R _{E98K} -M'-lacZ	Derivative of pRScaf1R _{E98G} -M'-lacZ containing <i>caf1R</i> with G98K mutation. Sequence confirmed 4_2 was stocked and used.	This study
pRScaf1R-M'-lacZ	Derivative of pRScaf1R _{E98K} -M'-lacZ containing wild type <i>caf1R</i> . Sequence confirmed transformants 11 and 16 were stocked and 11 (Q5-E98_11) was used.	This study

* Assigned stock number.

2.1.3 Chemicals and enzymes

Chemicals used in this study were of Molecular biology grade or higher and were sourced from Sigma-Aldrich, Melford Laboratories and Thermo Fisher Scientific. Restriction enzymes were sourced from NEB and Thermo Scientific. DNA polymerases were sourced from Takara Clontech, Bionline, NEB and Agilent Technologies. All enzymes were used according to manufacturer's specification with the appropriate buffer(s).

2.1.4 Centrifuges

For all centrifugation steps involving volumes ≤ 1.5 ml, a benchtop eppendorf centrifuge (Eppendorf, Germany) was used. For volumes 1.5-50 ml and centrifugation speeds ≤ 4000 rpm, a benchtop Megafuge1.0R centrifuge was used. For larger volumes (50-1000 ml) and centrifugation speed $\leq 18,000$ rpm, a Sorvall™ centrifuge was used with appropriate rotor(s). For ultra-centrifugation (50,000 rpm = 134877.8 $\times g$) a benchtop Beckman Optima TLX ultracentrifuge was used with TLA-100.3 rotor.

2.2 *In vitro* DNA techniques

2.2.1 Preparation of plasmid DNA

Plasmid DNA was routinely prepared (using Qiagen mini-prep kit) from 5 ml culture, grown at 37°C for 15-18 h with shaking (225 rpm). For temperature-dependent plasmid construct(s) such as pACYCF1 and pACYCF1-R mutants, culture(s) were grown at 26°C for at least 24 h with shaking (225 rpm) to avoid unwanted mutation(s) in the *caf* locus by metabolic burden due to fast growth rate at 37°C. DNA was routinely eluted into 35 μ l nH₂O followed by quantification and storage at -20°C.

2.2.2 Quantification of plasmid DNA

i) Nanodrop spectrophotometer

An estimation of plasmid DNA concentration was routinely carried out from 2 μ l sample, using a Nanodrop spectrophotometer-ND1000 (Thermo Scientific). It gives concentration in ng/ μ l with an indication of purity from the ratio of OD₂₆₀/OD₂₈₀.

ii) Agarose gel electrophoresis

Quality and concentration of DNA was routinely confirmed by agarose gel electrophoresis (AGE) using a horizontal agarose gel electrophoresis system (BioRad). Regularly DNA fragments were resolved on 0.75-0.8% (w/v) agarose gels (Melford Labs), prepared in TAE buffer (40 mM Tris-HCl, 1 mM EDTA, pH 8.0). Samples were applied with 1 \times DNA loading buffer blue (Bionline) and electrophoresed at 80 volts for 90 min at room temperature (RT). One Kilobase (1 kb) DNA HyperLadder™ I (Bionline, **Fig. 2.1**) was used throughout, as a size marker. Nucleic acid stain, GelRed (Cambridge Bioscience) was added to the gel (1:20000) to stain the DNA. DNA bands were visualised under ultraviolet light using G-Box (Syngene). Gel pictures were saved in .tif and .sgd formats. If required, DNA concentration was also calculated using the .sgd file of the gel picture by GeneTool, a DNA/protein band quantification software (Syngene). DNA quantity in band of interest was generally quantified by comparing with 20 ng of HyperLadder™ I (**Fig. 2.1**).

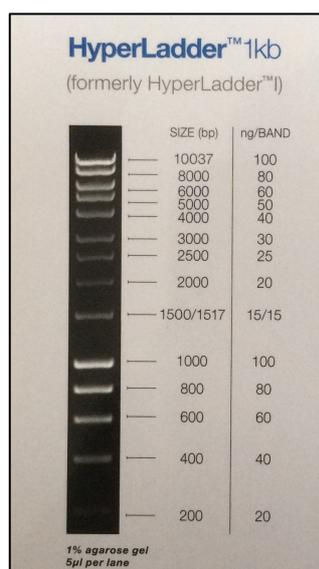


Figure 2.1| **DNA HyperLadder 1™ (Bioline).**

The molecular size (bp) and quantity (ng) of each separated band is shown from a 5 μl sample, on 1% agarose gel in 1× TAE.

2.2.3 Oligonucleotide primers

Oligonucleotide primers listed in Table 2.3 were purchased from Eurofins-Genomics (Germany) and routinely resuspended in nH₂O to a stock concentration of 200 pmol/μl. Designing strategies for deletion, Infusion, mutagenesis and sequencing primers were different.

For deletion primers, the length was kept ideally between 18-33 nucleotides (nt) with a melting temperature (T_m) range of 59-67°C and %GC content about 42-66%. Both T_m and %GC were calculated by primer properties scan tool (Eurofins-Genomics). T_m difference between forward and reverse primers was kept at $\leq 4^\circ\text{C}$ to obtain optimal amplification. One to three guanine (G) or cytosine (C) were included at the 3' end of each primer. A unique restriction enzyme site was inserted at the 5' end of each primer with some upstream nt for optimal action of the respective restriction enzyme.

Infusion primers were designed using Clontech primer design tool with the following features: The 5' end of each primer was taken from 15 bases of the linearised plasmid vector (linearised either by restriction digestion or inverse PCR). The 3' end of each primer was insert specific with a length of 18-25 bases, GC-content 40-60% and T_m 58-65°C. The T_m difference between forward and reverse primers was kept $\leq 4^\circ\text{C}$ and 1-3 G or C were included at the 3' end of each primer. Complementarity within each primer was avoided to prevent hairpin structures, and primer dimer formation between primer pairs.

Mutagenesis primers were designed according to QuikChange Site-Directed mutagenesis XLII kit (Agilent) and Q5® Site-Directed mutagenesis kit (NEB) with following considerations. For QuikChange Site-Directed mutagenesis, desired mutation(s) was added by incorporating the desired nucleotide change(s) in the middle of the forward primer. On either side of incorporated

mutation(s), 12-15 bases were kept gene-specific. Reverse primer was obtained by reverse complementing the forward primer. Primer length was kept ≤ 45 bases with $T_m \geq 78^\circ\text{C}$. The T_m was calculated by following formula:

$$T_m = 81.5 + 0.41(\%GC) - 675/N - \% \text{ mismatch}$$

N, total number of bases

For Q5[®] Site-Directed mutagenesis, desired mutation was added near the 5' end of forward primer and 3' was extended up to 20 bases from the site of mutation. Thus forward mutagenic primer was generated with a total length of about 25-30 bases. Reverse primer had no mutation and thus was gene-specific. It was designed in such a way that the 5' ends of both forward and reverse primers anneal back-to-back. NEB T_m calculator (<http://tmcalculator.neb.com/#!/>) was used to calculate T_m of each primer.

Most sequencing primers were used from on-site sequencing services. When there was need of additional sequencing primers these were designed according to standard features such as 18-25 bases in length, which are homologous to the gene of interest and having 55-58 $^\circ\text{C}$ T_m with 40-60% GC-content.

Table 2.3| Oligonucleotide primers used in PCR reactions

Construct	Primer	Sequence (5'-3')	GC (%)	T _m (°C)
Inverse PCR primers				
pACYC-MA1	Invcaf1Rfor	ACCCatgg ¹ CAAAATAATAGCATTCT AGATAGTG <u>NcoI</u>	36.4	64.5
	Invcaf1Rrev	TCCcAtgGCAGGAGTCGCATAAGG <u>NcoI</u>	58.3	66.1
pΔhBADcaf1R	dhCaf1Rfor	TATCCATGGCGCTCATTGGGTTATA TTCATGC <u>NcoI</u>	42.4	67.0
	dhCaf1Rrev	CGATGCCATGGTTAATTCCTCCTGTT AG	46.4	65.1
pACYC-R and	pACYCcaf1Rf	GCCTAACCCCTCGAGCTCG <u>SacI</u>	66.7	60.5
pACYC-R _{E98G}	pACYCcaf1Rr	GCTTgAgCTCCTTACGGAATG <u>SacI</u>	52.4	59.8
Infusion primers²				
pEThCaf1R	pETCaf1Rf	CAGCCATATGGCTAGCAAG GCTAGCT GGGGATCCGAGCTCATTG <u>NheI</u>	57.7	68.0
	pETCaf1Rr	GACGGAGCTCGAATTC TTTGAATTCC CATATGGTACCAGC <u>EcoRI</u>	50.0	60.3
pEThCaf1R _N	pETCaf1RNf*	CAGCCATATGGCTAGCT TCCGAGCTCA TTTGGGTTATATTC	41.7	59.3
	pETCaf1RNr*	GACGGAGCTCGAATTC TTAAAAGGAC CAAAAAGGGATCATC <u>TTA-Stop</u>	36.0	58.1
pMALc2-Caf1R	pMALcaf1Rf	AAGGATTTT CAGAATTCGTTGAATTCA TGCTAAAACAGATG <u>EcoRI</u>	33.3	55.9
	pMALcaf1Rr	CGACTCTAGAGGATCC TATGGATCCA GCTGCAGATC <u>BamHI</u>	50.0	57.3
pMALc2-Caf1R _N (L1)	pMALCaf1RNifF	AAGGATTTT CAGAATTC TCCGAGCTCA TTTGGGTTATATTC	41.7	59.3
	pMALCaf1RNinFR	CGACTCTAGAGGATCC TTAAAAGGAC CAAAAAGGGATCATC <u>TTA-Stop</u>	36.0	58.1
pMALc2-Caf1R _{gs}	pMALCaf1R_GSf	AAGGATTTT CAGAATTCATGCTGAAAC AAATGACCGTGAAC	41.7	59.3
	pMALCaf1R_GSr	CGACTCTAGAGGATCC TTATGATTTTC GGCAGACCCAGCGT	50.0	62.7

¹Nucleotides in small letters indicate the change(s) to make desired restriction sites.

² Nucleotides from vector ends are indicated in bold capital text and T_m is shown only from gene specific part, according to the Infusion method (Clontech).

pRScaf1R' _{-137+192-lacZ}	Delta7f	TGGGGATCGGAATT TTGCGCAAATAG CTAATTTTG	33.3	52.0
	Deltapcaf1Rr	CGTTTGTCCGGATC CAGCAGCTCTAC TAGCCC	61.1	58.2
pRScaf1R' _{-114+192-lacZ}	Delta6f	TGGGGATCGGAATT ATTGATGCGCGT GCACAAATAG	45.5	58.4
	Deltapcaf1Rr	CGTTTGTCCGGATC CAGCAGCTCTAC TAGCCC	61.1	58.2
pRScaf1R' _{-70+192-lacZ}	Delta5f	TGGGGATCGGAATT TGCGCAAATAGC TAAGTCCC	50.0	57.3
	Deltapcaf1Rr	CGTTTGTCCGGATC CAGCAGCTCTAC TAGCCC	61.1	58.2
pRScaf1R' _{-65+192-lacZ}	Delta16f	TGGGGATCGGAATT AAATAGCTAAGT CCCGGC	52.6	56.7
	Deltapcaf1Rr	CGTTTGTCCGGATC CAGCAGCTCTAC TAGCCC	61.1	58.2
pRScaf1R' _{-46+192-lacZ}	Delta17f	TGGGGATCGGAATT CCGGCCCACTAT CTAGAATG	55.0	59.3
	Deltapcaf1Rr	CGTTTGTCCGGATC CAGCAGCTCTAC TAGCCC	61.1	58.2
pRScaf1M' _{-270+158-lacZ}	pRSDeltaFor	CAGGAATTGGGGATC AGCTATTTGCG CATACCCAG	50.0	57.3
	pRS550(13-17)R	GTCCGGATCGGAATT CTAACGGGTAT ATGATCCTACTCTC	44.0	61.3
pRScaf1M' _{-197+158-lacZ}	pRS550(13-16)F; Delta1for	CAGGAATTGGGGATC TGCGCAACAAG CAAGTGGAGTG	50.0	62.7
	pRS550(13-16)R	GTCCGGATCGGAATT CTAACGGGTAT ATGATCCTACTCTC	44.0	61.3
pRScaf1M' _{-184+158-lacZ}	Delta 15F	CAGGAATTGGGGATC AGTGGAGTGCG C	66.7	40.0
	Delta 550R	GTCCGGATCGGAATT CTAACGGGTATA TGATCC	44.4	51.4
pRScaf1M' _{-169+158-lacZ}	Delta 14F	CAGGAATTGGGGATC AAGCTAAACTT TGTGTG	35.3	45.5
	Delta 550R	GTCCGGATCGGAATT CTAACGGGTATA TGATCC	44.4	51.4
pRScaf1M' _{-141+158-lacZ}	Delta4for	CAGGAATTGGGGATC AAAATTGTTCT CAGTGAGGCTG	40.9	56.5
	pRS550(13-16)R	GTCCGGATCGGAATT CTAACGGGTAT ATGATCCTACTCT	44.0	61.3
pRScaf1M' _{-102+158-lacZ}	pRS550(13-17)F or Delta2	CAGGAATTGGGGATC TCCCCTTCATT TGTTACCCACC	44.4	51.4
	pRS550(13-16)R	GTCCGGATCGGAATT CTAACGGGTAT ATGATCCTACTCTC	44.0	61.3

pRScaf1A'-1'-lacZ	pRS(11-12)infF	CAGGAATTGGGGATC CTGGCATTGTC GGAGATAATAGCG	50.0	62.7
	pRS(11-12)infR	GTCCGGATCGGAATT GCAGTGGTGCTT GCAGTTAAATC	47.8	60.6
pRScaf1'-181+88-lacZ	Delta11f	CAGGAATTGGGGATC AGAGATAAAAA TCAATC	23.5	40.7
	Deltapcaf1r	GTCCGGATCGGAATT GCAGTGGTGCT TGCAG	62.5	54.3
pRScaf1'-158+88-lacZ	Delta10f	CAGGAATTGGGGATC ATCTAATGTAG TTCTACC	36.8	50.2
	Deltapcaf1r	GTCCGGATCGGAATT GCAGTGGTGCT TGCAG	62.5	54.3
pRScaf1'-61+88-lacZ	Delta9f	CAGGAATTGGGGATC CAGGACACAAG CCCTCT	63.2	61.0
	Deltapcaf1r	GTCCGGATCGGAATT GCAGTGGTGCT TGCAG	62.5	54.3
pRScaf1'-38+88-lacZ	Delta8f	CAGGAATTGGGGATC ATTTGTTTCGTG GATTGG	44.4	51.4
	Deltapcaf1r	GTCCGGATCGGAATT GCAGTGGTGCT TGCAG	62.5	54.3
pRScaf1'-259-61-lacZ	Deltapcaf1f	CAGGAATTGGGGATC CTGGCATTGTC GGAG	60.0	50.6
	Delta12r	GTCCGGATCGGAATT GAAATAAACAT CCGTTTC	33.3	46.9
pRScaf1'-259-22-lacZ	Deltapcaf1f	CAGGAATTGGGGATC CTGGCATTGTC GGAG	60.0	50.6
	Delta13r	GTCCGGATCGGAATT CCAATCCACGA ACAAATTCG	45.0	55.2

Mutagenesis primers³

pACYC-R _{E98K}	pRS550(8-10)Lys98for	CCAAATATTTTCTTAAAT TTt TCTGGT GAATGTCTGTTGCGAATC	31.8	76.9
	pRS550(8-10)Lys98rev	GATTCGCAACAGACATT CACCAGaaA ATTTAAGAAAATATTTGG	31.8	76.9
ΔP _M pACYCF1	Pcaf1M_2f	CTTTGTGTGCATTTTTTAAAG GCCGCCT GTTCTCAGTGAGGCTGTGC	48.9	73.2
	Pcaf1M_2r	GCACAGCCTCACTGAGAACAG GGCGGC TTTTAAAATGCACACAAAG	48.9	73.2

³ Desired mutation(s) in the sequence are indicated by bold text with desired change(s) in small letter(s) and T_m, according to QuikChange II XL site-directed mutagenesis (Agilent).

pRScaf1R-M'-lacZ	G98Efor	CTTAAAT Tt cTCTGGTGAATGTCTGTT GCG	44.4	63.9
	G98Erev	AAAATATTTGGTTATACCCACGGCA G	40.7	61.9
pACYCF1	Caf1R _{G98E} for	GCAACAGACATTCCACCAGAg AA TTTA AGAAAATATTTGGTTATAC	31.1	67.6
	Caf1R _{G98E} rev	GTATAACCAAATATTT Tt cTTAAATTC TCTGGTGAATGTCTGTTGC	31.1	67.6
pACYCF1-R _{I31A}	Caf1R-I31Af	CGAAATTCATTAAC gct TGACTGTTTG GTTTTG	37.5	69.5
	Caf1R-I31Ar	CAAACCAAACAGTC Agc GTTAATGA ATTTGC	37.5	69.5
pACYCF1-R _{D32A}	Caf1R-D32Af	GTCGAAATTCATTAACATT Gc CTGTT TGGTTTTGTATTC	33.3	75.2
	Caf1R-D32Ar	GAATACAAAACCAAACA Gg CAATGTT AATGAATTTTCGAC	33.3	75.2
pACYCF1-R _{C33A}	Caf1R-C33Af*	GAAATTCATTAACATTGAC gct TTTGG TTTTGTATTCAGGATTCA	33.3	75.7
	Caf1R-C33Ar*	CTGAATCCTGAATACAAAACCAA Agc GTCAATGTTAATGAATTTTC	33.3	75.7
pACYCF1-R _{F40A}	Caf1R-F40Afor	GGTTTTGTATTCAGGAg c CAGCAGAA GGTATTTGCAAATTTCC	41.9	78.3
	Caf1R-F40Arev	GGAAATTTGCAAATACCTTCTGCT Gg ct CCTGAATACAAAACC	41.9	78.3
pACYCF1-R _{R42A}	Caf1R-R42Afor	GGTTTTGTATTCAGGATTCAGC gc AA GGTATTTGCAAATTTCC	39.5	77.3
	Caf1R-R42Arev	GGAAATTTGCAAATACCT Tgc GCTGA ATCCTGAATACAAAACC	39.5	77.3
pACYCF1-R _{R43A}	Caf1R-R43Afor	GGTTTTGTATTCAGGATTCAGCAGAg c GTATTTGCAAATTTCC	39.5	77.3
	Caf1R-R43Arev	GGAAATTTGCAAATA Cgc TCTGCTGA ATCCTGAATACAAAACC	39.5	77.3
pACYCF1-R _{Y44A}	Caf1R-Y44Afor	GGTTTTGTATTCAGGATTCAGCAGAA GG gc TTTGCAAATTTCC	41.9	78.3
	Caf1R-Y44Arev	GGAAATTTGCAA Agc CCTTCTGCTGA ATCCTGAATACAAAACC	41.9	78.3
pACYCF1-R _{R62A}	R62Afor	GCCTATTGGAACATATATT cg AGTTA GAAGGGCTAGTAGAGCGC	44.4	80.2
	R62Arev	GCAGCTCTACTAGCCCTTCTAAC Tcg AATATATGTTCCAATAGGC	44.4	80.2
pACYCF1-R _{R62S}	R62Sfor	GCCTATTGGAACATATATT AGc GTTA GAAGGGCTAGTAGAGCTGC	44.4	82.4
	R62Srev	GCAGCTCTACTAGCCCTTCTAAC gCT AATATATGTTCCAATAGGC	44.4	82.4

pACYCF1-R _{Q93A}	Caf1R-Q93Af	TGATTTCGCAAgcGACATTCACCAGAG	50.0	68.3
	Caf1R-Q93Ar	CTCTGGTGAATGT Cgc TTGCGAATCA	50.0	68.3
pACYCF1-R _{R97A}	Caf1R-R97Af	GCAACAGACATTCACC gc AGAATTTA AGAAAATATTTGG	35.9	73.7
	Caf1R-R97Ar	CCAAATATTTTCTTAAATTCT gc GGT GAATGTCTGTTGC	35.9	73.7
pACYCF1-R _{E98A}	E98Afor	CAGACATTCACCAGA Gc ATTTAAGAA AATATTTGGTTATACCCC	36.4	78.8
	E98Arev	GGGGTATAACCAAATATTTTCTTAAA Tg CTCTGGTGAATGTCTG	36.4	78.8
pACYCF1-R _{E98T}	E98Tfor	CAGACATTCACCAGA ac ATTTAAGAA AATATTTGGTTATACCCC	34.1	75.6
	E98Trev	GGGGTATAACCAAATATTTTCTTAAA Tgt CTCTGGTGAATGTCTG	34.1	75.6
pACYCF1-R _{T106A}	Caf1R-T106Af	GAAAATATTTGGTTAT g CCCCACGGC AGTATAGG	44.1	76.7
	Caf1R-T106Ar	CCTATACTGCCGTGG GGc ATAACCAA ATATTTTC	44.1	76.7
pACYCF1-R _{R108A}	Caf1R-R108Af	GGTTATACCCC gc GcAGTATAGGAT GATCC	51.6	74.4
	Caf1R-R108Ar	GGATCATCCTATACTG Cgc tGGGGTA TAACC	51.6	74.4
pACYCF1-R _{Q109A}	Caf1R-Q109Af	GGTTATACCCCACGG gc GTATAGGAT GATCC	54.8	75.7
	Caf1R-Q109Ar	GGATCATCCTATAC gc CCGTGGGGTA TAACC	54.8	75.7

Sequencing Primers

To confirm

Deletion of His ₆ -tag plus enterokinase cleavage-site from pBADhCaf1R.	pBADF	ATGCCATAGCATTTTTATCC	35.0	58.0
	pBADR	GATTTAATCTGTATCAGG	33.0	45.0
Repair and generated mutation in <i>caf1R</i> in pACYCF1 _{SpM} or pACYCF1 or pRScaf1R-M'-lacZ or pRScaf1R _{E98K} -M'-lacZ.	F1_1	ATGTTGGGTCGAACATAAATCG	40.9	56.5
<i>caf1R</i> and <i>caf1R-caf1M</i> intergenic region from pACYCF1 _{SpM} or pACYCF1 and pACYC-R or pACYC-R _{E98G} or pACYC-R _{E98K} .	ACYCDuetUp1	GGATCTCGACGCTCTCCCT	63.0	55.0

<i>caf1R</i> and <i>caf1R-caf1M</i> intergenic region from pACYC-R or pACYC-R _{E98G} or pACYC-R _{E98K} .	DuetDown1	GATTATGCGGCCGTGTACAA	50.0	52.0
<i>caf1R-caf1M</i> intergenic region.	F1_6	CCTTCTGCTGAATCCTGAATAC	45.0	53.0
<i>caf1A-caf1</i> intergenic region.	F1for11	GTGGATCCCTGGCATTGTCTGGAGATA ATAGCG	53.0	62.7
	F1rev12	CGGAATTTCGCAGTGGTGTGCTTGCAGTT AAATC	54.0	60.6
<i>caf1R</i> , <i>caf1R_N</i> in pET28a ⁺ plasmid.	T7F	TAATACGACTCACTATAGGG	40.0	48.0
	T7R	GCTAGTTATTGCTCAGCGG	53.0	51.0
<i>caf1R</i> , <i>caf1R_N</i> and <i>caf1R_{gs}</i> in pMALc2x plasmid.	<i>malE</i> for	GGTCGTCAGACTGTTCGATGAAGCC	58.0	61.0
<i>lacZ</i> -fusion constructs and <i>caf1R</i> and its derivatives in pMALc2x plasmid	M13F	TGTAACACGACGGCCAGT	50.0	48.0

2.2.4 PCR amplification

An Eppendorf mastercycler machine was used for all PCR amplification reactions. The PCR amplification products were routinely analysed by AGE and where required, were either purified using Qiagen PCR purification kit or gel-excised and purified using GeneJET Gel Extraction kit (Thermo Scientific), according to manufacturer's protocol.

Four types of PCR amplification were carried out during this study, 1) Simple PCR, to amplify a gene or a segment of a gene to be used for subsequent cloning, 2) Colony PCR, to screen for insertion of desired gene or gene fragment in the plasmid vector of choice, by direct amplification from single bacterial colonies transformed with respective ligation or Infusion mix, 3) Inverse PCR, to delete a particular region from the desired plasmid and 4) Mutagenic PCR, to introduce the desired mutation(s) into the target plasmid(s). PCR reaction(s) and conditions for each were as follows:

i) Subcloning or Infusion PCR

Gene or gene-fragment to be used for Infusion or ligation were amplified by this PCR type. PCR reaction mixture (25 µl) contained 50 ng plasmid DNA template, 0.5 pmol of each forward and reverse primer and 12.5 µl of 2× CloneAmpHiFi PCR Premix (Clontech) containing proof reading CloneAmpHiFi DNA polymerase, optimised dNTPs mix and reaction buffer. The reaction mixtures were always prepared on ice. The PCR-cycles used were routinely: 98°C/1 sec, followed by 35 cycles of 98°C for 10 sec, 55°C for 15 sec and 72°C for 5 sec/kb.

ii) Colony PCR

In order to identify positive recombinant clones within a large number of bacterial colonies, a short fragment (≈ 1 kb) was routinely amplified from 10-35 single individual colonies using appropriate primer pairs. The individual colonies were picked with a sterile pipette tip and mixed into 10 μ l of sterile nH_2O in an eppendorf tube. Five microliters of this colony re-suspension was mixed into 15 μ l of PCR reaction mixture, containing 1.25 units of DreamTaq DNA polymerase (Thermo Scientific), 2 mM dNTPs (5 μ l, 0.2 mM of each), appropriate forward and reverse primer (10 pmol of each), 10 \times DreamTaq buffer (5 μ l) in a final reaction volume of 50 μ l. PCR cycles were as follows: 95 $^{\circ}C$ /1 min (initial denaturation), followed by 30 cycles of 95 $^{\circ}C$ for 30 sec (denaturation), $T_m - 5^{\circ}C$ for 30 sec (annealing) and 72 $^{\circ}C$ for 1 min/kb (extension). The final extension was carried out at 72 $^{\circ}C$ for 5 min with a single cycle. Following colony PCR, the remaining 5 μ l colony re-suspension was added to 100 μ l sterile LB and incubated at 37 $^{\circ}C$ until positive clone(s) were identified by AGE. Positive colonies, in LB were then inoculated into 10 ml selective LB and incubated overnight at 37 $^{\circ}C$ with shaking at 225 rpm for plasmid DNA isolation and screening by restriction digestion and/or sequencing.

iii) Inverse PCR for fragment deletion

Inverse PCR was carried out using BIO-X-ACTTM short DNA polymerase (Bioline) and Q5 HiFi DNA polymerase (NEB). The PCR reaction mixtures and cycle conditions for each were as follows:

a) BIO-X-ACTTM short DNA polymerase

BIO-X-ACTTM enzyme (Bioline) was used to delete the His₆-tag and enterokinase cleavage site from the pBADhCaf1R plasmid in order to design pBAD Δ hCaf1R plasmid construct. PCR reaction mixture was prepared by adding 62 ng of pBADhCaf1R template, 10 \times Opti buffer (5 μ l), 50 mM MgCl₂ (2 μ l), 100 mM dNTPs (2.5 μ l), Hi space additive (2.5 μ l), dhCaf1Rfor and dhCaf1Rrev primers (20 pmol of each, 2 μ l) and one unit (1 μ l) of Bio-X-ACTTM DNA polymerase in a final reaction volume of 50 μ l. The PCR cycles were as follows: 95 $^{\circ}C$ /3 min (initial denaturation), followed by 30 cycles of 94 $^{\circ}C$ for 30 sec (denaturation), 57.5 $^{\circ}C$ for 30 sec (annealing) and 72 $^{\circ}C$ for 5 min (extension). The final extension was carried out at 72 $^{\circ}C$ for 10 min with a single cycle.

b) Q5 HiFi DNA polymerase

Q5 HiFi enzyme (NEB) was used to delete both the *caf1R* and *caf1MA1* genes from the pACYCF1_{SpM} and pACYCF1 plasmid templates in order to construct pACYC-MA1, pACYC-R and pACYC-R_{E98G}.

i) pACYC-MA1 construction

PCR reaction mixture and cycles to prepare pACYC-MA1 were as follows: pACYCF1_{SpM} DNA template (50 ng, 1.65 μ l) was mixed with 5 \times Q5 reaction buffer (5 μ l), Inv*caf1R*for and Inv*caf1R*rev primers (12.5 pmol of each, 1.25 μ l from 10 pmol/ μ l stock), 5 \times Q5 High GC enhancer and one unit of Q5 HiFi DNA polymerase in a final reaction volume 25 μ l. The PCR cycles were as follows: 98 $^{\circ}C$ /30 sec (initial denaturation) followed by 25 cycles of 98 $^{\circ}C$ for 5 sec

(denaturation), 55.1 and 60.8°C for 15 sec (annealing) and 72°C for 6 min (extension). The final extension was carried out at 72°C for 2 min with a single cycle.

ii) pACYC-R and pACYC-R_{E98G} construction

PCR reaction mixture and cycles to design pACYC-R_{E98G} and pACYC-R from pACYCF1_{S_{PM}} and pACYCF1 templates were as follows: pACYCF1_{S_{PM}} or pACYCF1 plasmid-DNA template (100 ng, 1.5 µl) was mixed with 5× Q5 reaction buffer (5 µl), pACYC*caf1Rf* and pACYC*caf1Rr* primers (10 pmol of each, 1.0 µl from 10 pmol/µl stock), 5× Q5 High GC enhancer (5.0 µl) and one unit of Q5 HiFi DNA polymerase (NEB) in a final reaction volume 25 µl. The PCR cycles were as follows: 98°C/30 sec (initial denaturation), followed by 25 cycles of 98°C for 5 sec (denaturation), 59.9°C for 15 sec (annealing) and 72°C for 4 min (extension). The final extension was carried out at 72°C for 2 min with a single cycle.

iv) Mutagenic PCR

Three kinds of mutagenic PCRs were used to add the desired substitution mutation(s) in the plasmid(s) of interest. First, by using QuikChange site-directed mutagenesis (QC-SDM) II XL kit (Agilent) second, by using KOD Hot-Start DNA Polymerase® (Novagen) and third by using Q5® site-directed mutagenesis (Q5®-SDM) kit (NEB). PCR reaction mixture and cycles for each were as follows:

a) QC-SDM II XL

QC-SDM XLII kit was used to add site-specific mutation(s) in *caf1R* of the pACYCF1 and also used to construct ΔP_M-pACYCF1 and pACYC-R_{E98K} mutants. PCR reaction for QC-SDM II XL was set up as follows: 5 µl of 10× reaction buffer, 50 ng of corresponding plasmid-DNA template, 125 ng of each forward and reverse mutagenic primers, 1 µl of dNTPs mix, 3 µl of Quik solution and one unit (1 µl) of Pfu-Ultra HF enzyme in a final reaction volume of 50 µl. The PCR cycles were as follows: 95°C/1 min (initial denaturation), followed by 18 cycles of 95°C for 50 sec (denaturation), 60°C for 15 sec (annealing) and 68°C for 1 min/kb (extension). The final extension was carried out at 68°C for 7 min with a single cycle. Mutagenic PCR products were routinely digested (25 µl) with 0.5 µl DpnI (from kit, unless otherwise stated) at 37°C for 25 min or unless otherwise specified to remove methylated template from newly synthesised mutant strands. Following DpnI treatment, 2 µl of treated PCR product was transformed into 50 µl of competent DH5α or Top10 cells.

b) KOD Hot-Start DNA Polymerase®

The plasmid construct pR*Scaf1R_{E98K}-M'-lacZ* was designed using KOD Hot-Start DNA polymerase (Novagen). PCR reaction and cycles were as follows: 70 ng of pR*Scaf1R_{E98K}-M'-lacZ* plasmid-DNA template, 5 µl of 10× reaction buffer, 10 pmol of each pRS550(8-10)Lys98for and pRS550(8-10)Lys98rev primers, 3 µl of 50 mM MgCl₂, 1 µl of 10 mM dNTPs mix and one unit (1 µl) of KOD Hot-Start DNA polymerase in final reaction volume of 50 µl. The PCR cycles were as follows: 94°C/2 min (initial denaturation), followed by 18 cycles of 98°C for 10 sec (denaturation), 57°C for 30 sec (annealing) and 68°C for 6.0 min (extension).

c) Q5®-SDM

The plasmid construct pRScaf1R-M'-lacZ was designed using Q5-SDM kit (NEB). PCR reaction and cycles were as follows: 50 ng of pRScaf1R_{E98K}-M'-lacZ plasmid-DNA template, 12.5 µl of 2× Q5 Hot-Start HF master mix, 12.5 pmol of each G98Efor and G98Erev primers in a final reaction volume of 25 µl. The PCR cycles were as follows: 98°C/30 seconds (initial denaturation), followed by 25 cycles of 98°C for 10 sec (denaturation), 65°C for 20 sec (annealing) and 72°C for 6.23 min (extension). The final extension was carried out at 72°C for 2 min with a single cycle. Following PCR, the PCR product was treated with an enzyme mixture containing Kinase, Ligase and Dpnl (KLD) at RT for 5 min and then 5 µl of KLD treated PCR product transformed into 50 µl of competent Top10 cells.

2.2.5 Restriction digestion

Restriction endonuclease enzymes used to digest either PCR product(s) or plasmid(s) were purchased from either NEB or Thermo Scientific. For restriction digestion, 100-700 ng of PCR product or plasmid-DNA was combined with 10 units of desired restriction enzyme in the presence of 1× enzyme-specific buffer. The final reaction volume was adjusted to 25 µl by adding autoclaved nH₂O. Reaction mixtures were then incubated at 37°C for 15-60 min (dependent upon enzyme manufacturer). In case of simultaneous digestion by two restriction enzymes, 5 units of each enzyme were used and an appropriate buffer was selected to ensure compatibility of both enzymes with maximum activity.

2.2.6 Clean-up of restriction digestion reactions

This procedure was carried out using QIAquick PCR Purification Kit (Qiagen, 0.1-15 kb product size), according to the manufacturer's instructions. The purified digested products were eluted using sterile nH₂O and then stored at -20°C.

2.2.7 Purification of DNA from agarose gel slice

DNA to be used for cloning or Infusion was routinely purified from an agarose gel, using GenJET gel extraction kit (Thermo Scientific) according to the manufacturer's protocol. The purified DNA was typically eluted in 35 µl of sterile nH₂O.

2.2.8 Molecular cloning techniques

Two kinds of cloning procedures were used during this study. One based on T4 DNA-ligase, known as ligation, and the other was Infusion, which is independent of DNA-ligase.

i) Ligation

T4 DNA ligase (Thermo Scientific, UK) was used throughout for ligation of Inverse PCR product(s). The digested and purified PCR products (50-100 ng) were combined with 5 units of T4 DNA ligase in the presence of 1× T4 DNA ligase buffer and the reaction volume was adjusted to 20 µl with autoclaved nH₂O. Reaction mixtures were then incubated at room temperature for 30 min in order to allow ligation.

ii) Infusion

Ligation sometimes produces recombinant clones in the wrong orientation when a single restriction enzyme is used to digest the PCR product and the destination plasmid vector. To avoid this, either digestion of both PCR product and destination plasmid vector can be carried out using two restriction enzymes or another procedure such as Infusion cloning can be used to generate seamless clones with confirmed correct orientation. Hence, Infusion was used to prepare all promoter-*lacZ* fusion constructs in pRS415 and pRS550 plasmids. In addition, Infusion was also used to fuse *caf1R* or *caf1R_N* into both the pET28a⁺ and pMALc2x plasmid. For this purpose the Infusion[®] HD Cloning kit (Clontech, Takara) was utilised throughout to set up the Infusion reaction(s). Briefly, the destination plasmid vectors were digested by two restriction enzymes such as EcoRI and BamHI in case of pRS415, pRS550 and pMALc2x plasmids, and EcoRI and NheI in case of pET28a⁺ plasmid. The digested plasmids were purified using the Qiagen PCR purification kit and eluted into nH₂O, whereas the PCR products were gel-excised, purified by GeneJET gel extraction kit (Thermo Scientific) and eluted into nH₂O. To set up an Infusion reaction, 50 ng of purified PCR products were combined with 200 ng of linearised destination plasmid vector in the presence of 5× Infusion HD enzyme pre-mix (2 µl). The final reaction volume was adjusted to 10 µl with autoclaved nH₂O and the Infusion reaction mixtures were incubated at 50°C for 15 min. Routinely, 2.5 µl of Infusion reaction mixture(s) was transformed in 50 µl Stellar[™] competent cells (Clontech). A complete overview of this procedure for one example of plasmid construct, pMALc2-Caf1Rgs is depicted in **Fig. 2.2**.

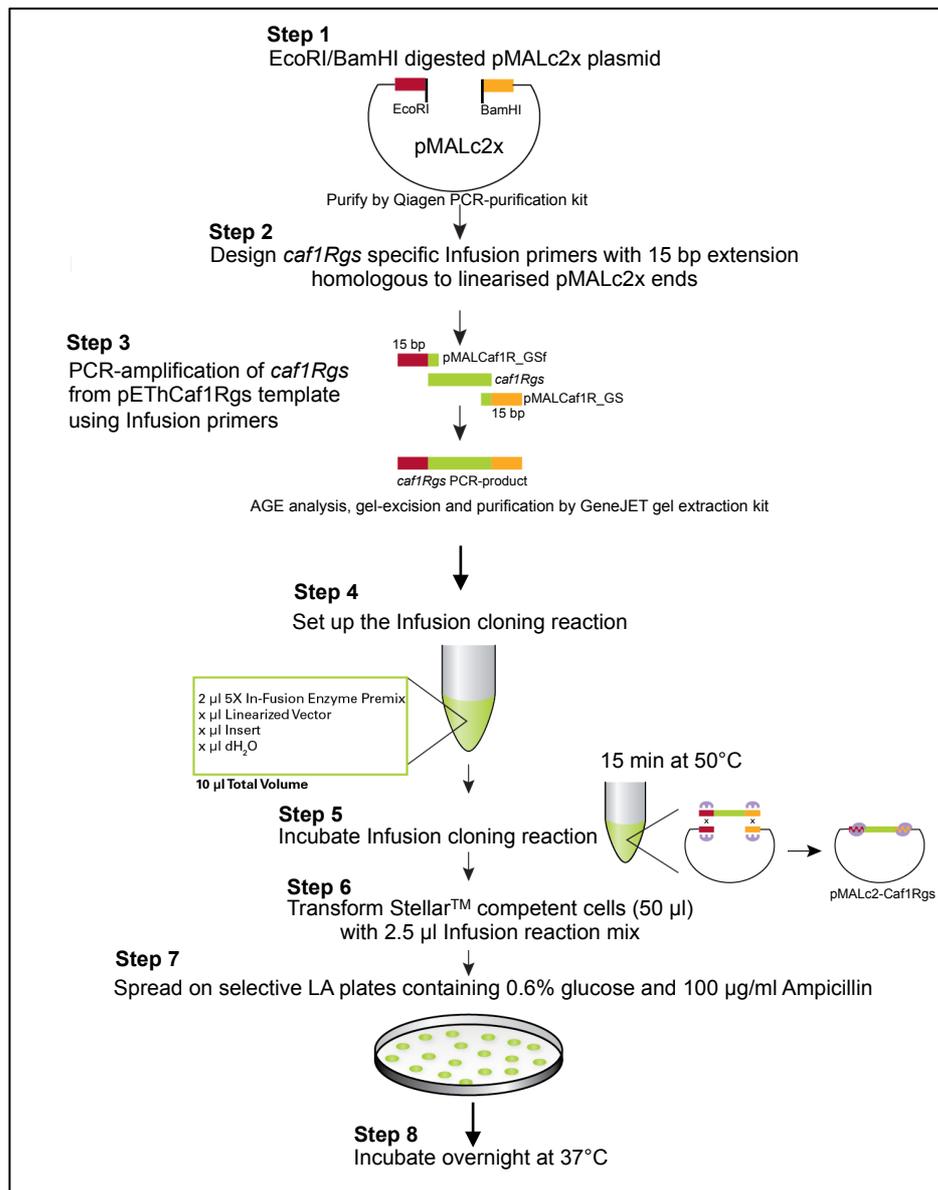


Figure 2.2| An example of plasmid construct designing by Infusion cloning method. Picture modified from Infusion® HD cloning kit manual (Clontech) for pMALc2-Caf1R designing.

2.2.9 Transformation

i) Preparing chemically competent cells

The *E. coli* cells were routinely made competent using the classical CaCl₂ method (Hanahan, 1983; Hanahan et al., 1991). The desired *E. coli* strains were grown overnight (~15 h) in 10 ml LB (no glucose) at 37°C with shaking at 225 rpm. Next day, 50 ml LB (no glucose) was inoculated (1/100) with overnight grown culture and incubated at 37°C with shaking (225 rpm) to 0.3-0.5 OD₆₀₀. Cells were recovered by centrifugation at 4000 rpm for 20 min at 4°C. The pelleted cells were resuspended in 15 ml of ice-cold RF1 buffer (160.9 mM KCl, 50.5 mM MnCl₂, 30.5 mM Potassium acetate, 10.2 mM CaCl₂, 60% glycerol; pH 5.8) and incubated on ice for 15 min. Following incubation on ice, the cell suspension was centrifuged at 4000 rpm for 10 min at 4°C and the RF1 treated cells were then resuspended into 5 ml of ice-cold RF2 buffer (20.1 mM KCl, 74.8 mM CaCl₂, 500 mM MOPS, 60% glycerol; pH 6.8) and incubated on ice for 30 min. After 30

min incubation on ice, the RF2 treated cell suspension was aliquoted (50 μ l) into ice-cold eppendorf tubes then snap frozen in liquid nitrogen. All aliquots were stored at -80°C.

ii) Transformation of chemically competent *E. coli* cells

An aliquot of competent cells was thawed on ice and up to 3 μ l of plasmid DNA (10-50 ng/ μ l) was added to 50 μ l cells. The tube was then gently flicked in order to mix the DNA with the competent cells, incubated on ice for 30 min, and then heat shocked at 42°C for 45 sec followed by incubation on ice for a further 2 min. Pre-warmed (at 37°C) 500 μ l LB (with 0.6% glucose) was then added and incubated at 37°C for 1 h with shaking at 225 rpm. Cells were then spreaded onto selective LA plates containing appropriate antibiotic(s). Plates were usually incubated at 37°C overnight except pACYCF1 and pACYCF1-R plasmids which, where were always incubated at 26°C for at least 48 h. Infusion mixtures were transformed into specialized competent cells, Stellar™ (Clontech), otherwise homemade competent DH5 α and/or Top10 cells were used.

ii) Co-transformation or complementation

Co-transformation was done for a combination of pACYC+pBAD or pACYC+pRS based plasmids. Equimolar concentrations of each plasmid (20-25 ng) were mixed and transformed into 50 μ l of competent cells of choice. The co-transformation mixture was spreaded on LA plates containing two corresponding selective antibiotics and 0.6% glucose. Plates were often incubated at 37°C overnight, unless otherwise stated.

2.2.10 DNA sequencing

Following transformation, several individual colonies of each kind of transformant were picked and either screened by Colony PCR and/or the plasmid-DNA was directly isolated for subsequent screening by restriction digestion and/or confirmation by sequencing. Routinely, 50-100 ng/ μ l of plasmid-DNA was sent either to Source Bioscience, Oxford or Eurofin-Genomic (Germany), along with the required concentration of desired sequencing primers (if not available onsite). Whenever plasmid DNA was not in the required range, then onsite template amplification service from Source Bioscience, Oxford was used. Sequencing results were provided as .ab1 and .seq files from Source Bioscience, Oxford, while from Eurofin-Genomic (Germany) sequencing files were obtained in .scf, .seq and .pdf format. The sequencing results from .seq and .scf files were analysed by DNADynamo and FinchTV, sequence analysis software.

2.3 Transcriptional fusions and β -galactosidase assay

2.3.1 Construction of promoter-*lacZ* fusion plasmids

Two transcriptional fusion vectors pRS415 and pRS550 (Simons et al., 1987) (Appendix 1) were used to identify promoter activity within the *caf* locus. In these two vectors, a promoterless *lacZ* gene, with its own RBS and start codon, is located downstream from the EcoRI and BamHI cloning sites. Insertion in pRS550 was used to test promoter activity upstream of *caf1M* and *caf1* whereas insertion in pRS415 was used to test promoter activity of the complementary strand and to identify *caf1R* promoter. Initial constructs with the *caf1R-caf1M* intergenic region cloned in pRS415 (pRS*caf1M'*-*R'*-*lacZ*) and pRS550 (pRS*caf1R'*-*M'*-*lacZ*), respectively, were already available and were used to test promoter activity upstream of *caf1R* and *caf1M*. To test promoter

activity upstream of *caf1*, pRScaf1A'-1'-lacZ was constructed by Infusion using the pACYCF1 template and PCR amplification/Infusion primers, pRS(11-12)infF and pRS(11-12)infR to amplify a 347 bp fragment encompassing the *caf1A-caf1* intergenic region (80 bp), 88 bp of *caf1* and 179 bp of *caf1A*. The pRS550 plasmid was linearised with EcoRI/BamHI and column purified. Insert and vector were then Infused and transformed to generate pRScaf1A'-1'-lacZ. Three promoter-lacZ fusion constructs, pRScaf1R-M'-lacZ, pRScaf1R_{E98G}-M'-lacZ and pRScaf1R_{E98K}-M'-lacZ containing the same DNA fragment as in pRScaf1R'-M'-lacZ plus the complete wild type *caf1R*, *caf1R* with E98G and E98K substitutions, respectively, were used to test the impact of wild type and mutant *caf1R* on *caf1M* transcription, in a *cis* orientation. The wild type pRScaf1R-M'-lacZ was difficult to make so initially the pRScaf1R_{E98K}-M'-lacZ mutant was constructed by using previously designed pRScaf1R_{E98G}-M'-lacZ mutant template and then pRScaf1R_{E98K}-M'-lacZ was used as template to create wild type pRScaf1R-M'-lacZ (section 2.2.4 (iv) (c)). A pACYCF1_{SPM} plasmid fragment (1519 bp) encompassing mutant *caf1R* (E98G), *caf1R-caf1M* intergenic region and the first 158 bp of *caf1M* was previously subcloned in between EcoRI/BamHI sites of the pRS550 plasmid and thus constructed pRScaf1R_{E98G}-M'-lacZ. The remaining two constructs, pRScaf1R_{E98K}-M'-lacZ and pRScaf1R-M'-lacZ were constructed by site-specific mutagenesis of pRScaf1R_{E98G}-M'-lacZ and pRScaf1R_{E98K}-M'-lacZ templates, respectively (section 2.2.4 (iv) b and c). Constructs used to test progressively smaller fragments for promoter activity were constructed by PCR amplification and Infusion, using the corresponding PCR amplification/Infusion primers and linearized pRS415 or pRS550 as required. Sequences of insert for all constructs were confirmed by sequencing with the M13F primer for both pRS415 and pRS550 based constructs. The sequence of *caf1R* in pRScaf1R-*caf1M*'-lacZ and pRScaf1R_{E98K}-M'-lacZ was additionally confirmed with F1_1 sequencing primer.

2.3.2 Monitoring β -galactosidase activity

In order to quantify the strength of predicted promoters (identified within the intergenic region of the *caf* locus) and to monitor the impact of Caf1R on their activity, the β -galactosidase activity of each promoter-lacZ fusion construct was measured \pm Caf1R (expressed from pACYC-R) from three individual colonies as follows:

i) Growth curve and sample collection

Single colonies of freshly transformed *E. coli* Top10 cells, carrying promoter-lacZ fusion construct \pm pACYC-R were inoculated, in triplicate, into 10 ml LB containing 100 μ g/ml Amp (+10 μ g/ml Cm for cells also carrying pACYC Duet-I vector or pACYC-R/pACYC-R_{E98G}/pACYC-R_{E98K}). Cells were grown for 24 h at 26°C with shaking at 225 rpm. On the following day, OD₆₀₀ was measured and one sample of cells equivalent to 0.5 OD unit was harvested by centrifugation (13,000 rpm - 10 min), as 0 time sample. A second sample equivalent to 0.5 OD unit was added into 50 ml pre-warmed (at 37°C) selective LB (100 μ g/ml Amp \pm 10 μ g/ml Cm) in 250 ml sterile flasks and incubated at 37°C with shaking at 225 rpm for 24 h. OD₆₀₀ was monitored and 0.5 OD unit of culture harvested every 2 h until cultures reached stationary phase (10 h). A final sample was harvested at 24 h. During sampling, all samples were kept on ice to halt growth while OD₆₀₀ was measured. All recovered cell pellets were immediately frozen at -20°C until required for the measurement of β -galactosidase activity.

ii) Cell lysis and β -galactosidase assay

Frozen pelleted cells were resuspended in 100 μ l of Bugbuster master mix (Novagen[®]) solution and incubated at 37°C for 20 min with gentle shaking. The cell lysate was then centrifuged (13,000 rpm - 5 min) and the supernatant transferred to another sterile 1.5 ml eppendorf tube. Lysed samples were stored on ice, processed the same day to measure β -galactosidase activity, and were never refrozen. Four microliter lysate samples were loaded into a 96-well plate (Grainer BIO-ONE, flat bottom) and 196 μ l of reactant solution⁴ was added using a multichannel pipettor immediately before inserting the plate into a SpectraMax Plus Microplate reader (Molecular Devices corps) and to begin recording. SoftMaxT program was used to measure the reaction rate (V_{max}) across each well of the plate, for an hour at 30°C with readings at OD₄₂₀ at 30 sec intervals. The V_{max} values and plot of each plate was exported as .pdf file and used to calculate the β -galactosidase activity as Abs₄₂₀ min⁻¹ OD⁻¹ of cells, as follows:

$$\beta\text{-galactosidase activity} = V_{max} = 1000 \times \Delta\text{Abs}_{420} \text{ min}^{-1}$$

$$\Delta\text{Abs}_{420} \text{ min}^{-1} = V_{max}/1000$$

$$\begin{aligned} \Delta\text{Abs}_{420} \text{ min}^{-1} \text{ OD}^{-1} &= (V_{max} \times 25 \times 2)/1000 \\ &= V_{max}/20 \text{ (for undiluted sample)} \\ &= \text{Abs}_{420} \text{ min}^{-1} \text{ OD}^{-1} \text{ (Units)} \end{aligned}$$

Factor 25 is from 100/4: 4 μ l of the 100 μ l Bugbuster lysed cells used to test activity

Factor 2 is from 1/0.5: To calculate activity per OD and take account of 0.5 OD cells used in the assay.

2.4 *In vitro* protein techniques

2.4.1 F1 analysis in whole cell

To analyse F1 expression in whole cells, a single bacterial colony or glycerol stock culture of an individual colony was inoculated in 10 ml selective LB (containing 0.6% glucose and appropriate antibiotic/s) and incubated at 26°C for 22 h with shaking at 225 rpm. Next day, 1 ml of this overnight culture was pelleted by centrifugation (13,000 rpm-10 min) and washed with fresh LB (no glucose) by re-suspension and re-centrifugation (13,000 rpm-10 min). This pellet was then resuspended in 10 ml selective LB (containing only appropriate antibiotics) and grown at 37°C for 6 h, unless otherwise stated. After 4 and 6 h thermo-induction (26 to 37°C), 1 ml sample was recovered by centrifugation (13,000 rpm- 20 min) and the pelleted cells mixed in 200 μ l PBS/OD₆₀₀, unless otherwise specified. On some occasions, pelleted cells were mixed in 100 μ l PBS/OD₆₀₀ as stated in the corresponding result sections. F1, from the co-transformed constructs, was extracted after induction with L-ara (0.02%) at 37°C for 4-6 h. During F1 sample preparation extreme care was taken to remove the culture supernatant (LB) from the pelleted

⁴ 1.0 mM MgSO₄·7H₂O, 10 mM KCl, 60 mM Na₂HPO₄·2H₂O, 40 mM NaH₂PO₄·2H₂O, 14.9 mM ONPG, 24.9 mM DTT, 8.85% PBS; pH 7.5.

cells, ensuring no loss of fluffy F1, on the surrounded cell surface. The recovered cells were stored at -20°C for further use.

2.4.2 F1 extraction from cell surface

To extract F1 from the cell surface, PBS mixed cell suspensions (section 2.4.1) were heat-shocked at 57°C for 1 h with intermittent mixing every 10-15 min to detach F1 from the associated Caf1M chaperone (Chapman et al., 1999; MacIntyre et al., 2001). Following heat shocking, cells were removed by centrifugation at 13,000 rpm for 10 min and the F1 containing supernatant was recovered (Zavialov et al., 2003). Samples were stored at -20°C for further analysis.

2.5 Expression and purification of tagged Caf1R

Caf1R was overexpressed and purified from its native and synthetic codon optimised sequence in three expression plasmids, pBADHisA, pET28a⁺ and pMALc2x. A brief overview of expression and purification of tagged Caf1R was as follows:

2.5.1 Expression, sampling and lysis of recombinant *caf1R* constructs in *E. coli*

Ten ml (small-scale) or 40-500 ml (large-scale) LB, containing the appropriate selective antibiotic and 0.6% glucose, was inoculated with either a single fresh transformant or from a glycerol stocked culture of the construct of interest. Cultures were grown at 37°C on a rotary shaker (225 rpm) for 16-17 h. Next day, cells from 1-5 ml culture were pelleted-down (13,000 rpm/5 min), washed by mixing in fresh LB (no glucose) and re-centrifuged (13,000 rpm/5 min). The washed cells were then resuspended in the same original volume of LB (no glucose) and subcultured (1/100) in selective LB (containing the appropriate antibiotic ± 1.0% glucose)⁵. Small-scale (10 ml) preparations were carried out in 20 ml universals while large-scale (40-500 ml) preparations were in 250 ml to 2-liter Erlenmeyer flasks. Subcultured cells were grown at 37°C on a rotary shaker (225 rpm) until about 0.5 OD₆₀₀. At this point cells were induced with either a range of IPTG concentrations (pET28a⁺ and pMALc2x based constructs) or L-ara (pBADHisA based constructs) as specified. During induction, cultures were grown at 37°C, or at an alternate temperature as mentioned, and samples taken at the time points indicated. For small-scale analysis of expression directly in whole cells, cells (1 OD unit) were recovered at 13,000 rpm/5 min at RT and directly mixed routinely into 100 µl of 4× SDS-PAGE sample buffers. For preparative large-scale procedures, cells were recovered at 4,000 rpm/20 min at 4°C (large-scale) and resuspended directly in the lysis buffer of choice (100 µl/OD, unless otherwise stated). When lysis was followed by affinity purification, cells were generally lysed in the appropriate column buffer. To small screens, Bug buster was often used. Use of other lysis buffers are indicated in the respective Results section. For pBADHisA and pET28a⁺ based construct, EDTA-free and for pMALc2x based constructs EDTA-plus proteases inhibitor was added during lysis. Cells were lysed by sonication or French press, directly or following storage at -20°C, as described below. Cells mixed in Bug buster master mix (Novagen) do not require sonication or French press.

⁵ After confirming IPTG-mediated toxicity pEThCaf1Rgs or pEThCaf1Rgs_{E98G} plasmid containing cells were always grown in LB containing 1.0% glucose throughout irrespective of induction.

2.5.2 Cell lysis

Cells were routinely lysed either by Bug buster master mix (Novagen) or sonication or French press followed by mixing into appropriate lysis buffer. Bug buster master mix was used to lyse pelleted cells harvested from 500 μ l to 5 ml culture. Bug buster master mix contains protein extraction reagents, benzonase nuclease and rLysozyme™ in one convenient formulation allowing gentle disruption of the *E. coli* cell wall to liberate **active** proteins without the use of any mechanical methods of cells lysis i.e. French press or sonication. To lyse cells using Bug buster master mix, recovered cells were regularly mixed in 100 μ l/OD₆₀₀ of Bug buster master mix, unless otherwise stated, and incubated at RT with gentle shaking for about 15-20 min. The unlysed cells were then removed by centrifugation, as described in the subsequent section. To lyse cells by sonication, Sonic Vibra™ sonicator (Sonics) was used to lyse cell-suspensions from 500 μ l to 10 ml. Samples were always kept on ice and cells were routinely lysed with appropriate-sized probe for about 3-5 min, with 10 sec pulse ON/OFF mode at amplification-1, 50% energy setting. To lyse 40-50 ml of cell suspension, prepared in the affinity column binding buffer supplemented with EDTA-free proteases inhibitor (Pierce), a French press (Thermo Scientific) with pre-chilled (at 4°C) 45 ml cell was regularly used to at 12,000 psi, 4 \times .

2.5.3 Clarification of cell lysates: preparation of soluble and insoluble fractions

Unless otherwise specified, following cell-lysis, lysates were centrifuged at 15-20,000 rpm/15-20 min in an Eppendorf or MSE bench top centrifuge to remove any remaining unlysed cells, cells debris (membranes) and inclusion bodies (pellet, insoluble fraction) from the soluble protein fraction (supernatant). Where indicated, lysates were clarified by ultracentrifugation (50,000 rpm⁶/60 min at 4°C) using a Beckman Optima TLX ultracentrifuge with TLA-100.3 rotor. This should remove all aggregates, leaving soluble protein in the supernatant fraction. Lysates to be used in subsequent protein purification were always ultracentrifuged to obtain the soluble fraction. After separation of supernatant, pelleted cells were resuspended in the volume of buffer used for lysis. Pellet and supernatant fractions were analysed by SDS-PAGE using 12.5% acrylamide for all MBP-tagged constructs and 14% acrylamide for His-tagged constructs.

2.5.4 Purification

Affinity-chromatography based methods were used to purify His₆-tagged and MBP-tagged Caf1R using an AKTA purifier10 (GE Healthcare) at 15-18°C. For purification of tagged Caf1R, soluble fractions were always prepared by ultracentrifugation (50,000 rpm/60 min at 4°C) to ensure removal of all insoluble aggregates.

i) Purification of hCaf1R/hCaf1R^Tgs by Ni²⁺ affinity chromatography

His₆-tagged Caf1R was prepared from 250 or 100 ml culture of Top10/pBADhCaf1R (induced with 0.02% L-ara at 30°C-6 h) or BL21(DE3)/pEThCaf1Rgs (induced with 0.35 mM IPTG at 37°C for 2 h 30 min). Cells were resuspended in 10-40 ml of HisTrapFF column binding buffer supplemented with 1 \times EDTA-free protease inhibitor (Pierce) and lysed by French press

⁶ 50,000 rpm = 134,877 $\times g$ or rcf

(hCaf1R) or sonication (hCaf1R^{Tgs}) followed by ultracentrifugation. The clarified lysate supernatants (10-40 ml) were fractionated using a pre-packed 1 ml HisTrapFF column and His trap buffer kit (GE Healthcare). The column was pre-equilibrated with 10-column volume (CV) binding buffer (20 mM NaPO₄ buffer, 500 mM NaCl, 5 mM DTT, 20 mM imidazole; pH 7.4). Samples were applied at a flow rate of 0.5 ml/min. The column was then washed with 10-CV binding buffer at a speed of 1 ml/min. Flow-through (FT) and wash (W) fractions were collected to analyse unbound protein. Bound hCaf1R or hCaf1R^{Tgs} was eluted with 20-CV of a linear gradient of 0-250 mM imidazole in binding buffer, at a flow rate of 1 ml/min. Column was washed with 5 ml binding buffer containing 250 mM imidazole (100% B). One ml fractions were collected and stored on ice or at 4°C until required.

ii) Purification of MBPCaf1R_{gs} by Maltose affinity chromatography

Likewise a pre-packed MBPTrapHP column (1 ml size, GE Healthcare) fitted to an AKTA purifier was regularly used to purify MBPCaf1R_{gs} from the clarified lysate supernatant. Lysate-supernatant was prepared from the IPTG (0.3 mM) induced culture of *E. coli* K-12 2508/pMALc2-Caf1R_{gs} (500 ml), recovered after 5 h at 25°C post-induction. Recovered pelleted cells were mixed in 40 ml of MBPTrapHP binding buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA and 10 mM β-Mercaptoethanol; pH 7.45) supplemented with 1× protease inhibitor (Pierce) followed by sonication and ultracentrifugation (50,000 rpm/60 min at 4°C). Before applying the clarified lysate supernatant, the column was equilibrated with 10-CV of MBPTrap column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA and 10 mM β-Mercaptoethanol; pH 7.45). The sample was then applied at a rate of 0.5 ml/min. The flow-through (FT) was collected and the column washed with 10-CV of column buffer at 1 ml/min. Unbound and washed samples were collected for analysis. The column-matrix bound MBPCaf1R_{gs} was eluted with 12-CV of the column buffer containing maltose over a linear gradient of 0-100% 10 mM maltose, collecting 12 fractions (1 ml each). Eluted fractions were kept on ice and analysed by SDS-PAGE.

2.5.5 Solubilisation of inclusion bodies by urea

A culture (50 ml) of BL21(DE3)/pEThCaf1R_{gs} was induced with 0.1 mM IPTG at 37°C for 5 h. Recovered cells (90 OD units) were resuspended in 2 ml 20 mM Na₂PO₄, 500 mM NaCl; pH 7.4 (buffer 1) and lysed by sonication (20 min 10 sec ON/OFF pulse). Cell lysates were cleared at 4000 rpm/10 min/4°C to remove any unlysed cells and then centrifuged at low speed (3000 rpm/20 min at 4°C) to pellet inclusion bodies. This pellet was resuspended in 2 ml solubilisation buffer 2 (20 mM NaPO₄, 500 mM NaCl, 0.5% Triton-X; pH 7.4), aliquoted (400 μl ×5) and centrifuged (2000×g/20 min at 4°C), to wash the pelleted protein (Hancock, 2001). These pellets were mixed in 400 μl of buffer 1, (control, no urea) or buffer 1 containing 4 M, 6M or 8 M urea (500 mM NaCl, 20 mM NaPO₄ buffer; pH 8.0) and sonicated for 1 min 40 sec with 10 sec ON/OFF pulse on ice. Supernatants (urea solubilised protein) and pellets (urea insoluble protein) were separated by centrifugation (14,000×g/30 min at RT) and kept on ice. Resulting final pellets were mixed in 400 μl of 4× SDS-PAGE sample buffer and analysed by SDS-PAGE along with final supernatants.

2.6 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

All protein samples were routinely analysed by discontinuous Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) (Laemmli, 1970). A vertical gel electrophoresis system, mini-PROTEAN II (Bio-Rad) was used throughout with freshly prepared gels (see gel making procedure in section 2.6.1) of varying acrylamide concentration (%) in the resolving gel. The most common concentrations of acrylamide used were 14 or 16% for F1 subunit (Caf1) analysis and 12.5 or 14% for Caf1R as specified in respective result sections. Before analysis, samples were mixed with an appropriate volume of 4× sample buffer (10% (w/v) glycerol, 62.5 mM Tris-HCl pH 6.8, 8% (w/v) SDS, 0.1 mg/ml bromophenol blue) to give a final concentration of ×1. For analysis of Caf1R, routinely DTT was added to the sample buffer to a final concentration of 50 mM. Samples were heat-denatured (97.5°C for 10-15 min) before applying to gel except for F1 polymer analysis, which was performed using unheated samples. Gels used were of 0.75 mm thick and routinely run at 200 volts for 45 min at RT, in 1x SDS-PAGE running buffer (0.25 M Tris-HCl base, 0.2 M glycine, 0.1 (w/v) SDS).

Following SDS-PAGE, the resolved proteins were either stained with Coomassie Brilliant Blue R-250 (CB) stain (45% (v/v) methanol, 10% (v/v) glacial acetic acid and 0.25% (w/v) CB stain) and/or processed for western immunoblotting (section 2.7). For CB staining, gels were submerged in CB stain in a plastic box and kept on a gentle shaking platform for about an hour at RT followed by de-staining of excess CB stain by transferring the gels into another plastic box containing de-staining solution (40% (v/v) methanol, 10% (v/v) glacial acetic acid). De-staining was at RT with gentle shaking for about 2 h with 2-3 times replacement of de-staining solution. Finally, gels were equilibrated with nH₂O and imaged in G-Box (Syngene, UK). Gel pictures were always saved in .tif and .sgd file formats. Sometimes, gels were dried between cellophane sheets (Bio-Rad) using GelAir dryer system (Bio-Rad) for storage.

2.6.1 Preparation of discontinuous SDS-PAGE gel

As the name suggests, discontinuous SDS-PAGE gels contain two different concentrations of acrylamide and pH on a single gel to resolve proteins according to their molecular weight and charge. Hence, two gels, resolving (with varying concentration of acrylamide, pH 8.8) and stacking (contains 5% (w/v) acrylamide, pH 6.8) were prepared one after the other and a single SDS-PAGE gel was made by combining both as follows:

i) Resolving and stacking gel preparation

Initially, the resolving and stacking gel buffers⁷ were prepared which were then used to make the corresponding gel. An example of a 14% resolving gel (0.75 mm thickness) is as follows: in a clean 15 ml Falcon® tube, the following components were added in the following order: 4× resolving gel buffer (950 µl), 30% (w/v) acrylamide (acrylamide: bisacrylamide, ratio 37.5:1, Severn Biotech Ltd) (1.68 ml), nH₂O (1.0 ml), 10% (w/v) APS (56 µl) and TEMED (2.5 µl). Similarly the stacking gels were prepared by adding the following components in another clean

⁷ Resolving gel buffer (4×): 1.5 M Tris-HCl (pH 8.8), 0.4%(w/v) SDS
Stacking gel buffer (4×): 0.5 M Tris-HCl (pH 6.8), 0.4%(w/v) SDS

15 ml size Falcon® tube: 4× stacking gel buffer (333 µl), 30% acrylamide (225 µl), nH₂O (750 µl), 10% (w/v) APS (21 µl) and TEMED (2.5 µl).

ii) Preparation of SDS-PAGE gel

The resolving gel mixture was mixed and poured in between the assembled glass plates in the Mini-PROTEIN II gel casting apparatus (Bio-Rad). Isopropanol was layered on top of the gel in order to exclude air bubbles and the gel was left to set for about 20 min at RT. After the resolving gel had set, the isopropanol was decanted and rinsed with nH₂O. Now the stacking gel was applied on top of the pre-set resolving gel and a plastic comb (Bio-Rad) was inserted to form wells in the stacking gel. The gel was left to set again for about 20 min at RT. Routinely a coloring agent, bromophenol blue (1% (w/v)) was added in the stacking gel to give a good contrast.

2.7 Western Immunoblotting

Western Immunoblotting or WB (Towbin et al., 1979) was routinely used to detect the low expression level or to confirm the identity of tagged Caf1R (from pBADHisA, pET28a+ and pMALc2x based Caf1R/Caf1R_N encoding constructs) and F1 subunit (Caf1) expression from pACYCF1 (WT) or pACYCF1-R_{E98K} mutant. All WBs were carried out using a Mini-PROTEIN II Mini Trans-blot module (Bio-Rad). Following SDS-PAGE, the gels were equilibrated with ×1 electroblotting transfer buffer (0.025 M Tris-HCl, 0.2 M glycine) along with two fiber sponges, four pieces of adsorbent Whatman's paper and an electro blotting membrane per gel. Two types of electroblotting membranes, Nitrocellulose (Amersham Biosciences) (NCM) and PVDF (Thermo Scientific) both 0.45 µm pores size were used. PVDF membrane was activated with 95% methanol prior to equilibration with transfer buffer, unlike NCM that does not need such activation.

The gel-membrane sandwich was assembled on the black side of the blotting cartridge in the following order: equilibrated sponge, two Whatman's filter paper, SDS-PAGE gel, blotting membrane, two Whatman's paper and finally the second sponge. The assembled cartridge was placed into the gel tank (Bio-Rad) filled with transfer buffer, along with an ice container and electrophoresed at 36 volts for 90 min at RT. Following electrophoretic transfer, the sandwich was disassembled and the membrane removed and incubated face-side up in blocking buffer⁸ for either an hour or overnight at 4°C in order to prevent non-specific antibody binding. Following blocking, the membrane was rinsed with TBST and incubated with protein-specific primary antibody solution, prepared in the TBST for an hour at RT with gentle shaking. After probing with primary antibody, the membrane was washed with TBST for 15 min, with every 5 min changing of TBST on a shaking platform. Following washing, the membrane was re-probed with HRP-conjugated secondary antibody for an hour, similar to primary antibody and likewise washed after that. For hCaf1R probing, a single antibody, monoclonal antiHis-HRP antibody (Roche) was used at either 1 in 10,000 or 1 in 20,000 dilution as specified in the corresponding result sections. For Caf1, two antibodies were used and each are stated in the corresponding

⁸ 5% (w/v) dried milk powder (Marvel) in TBST (0.05 M Tris-HCl (pH 7.5), 0.15 M NaCl, 0.1% (v/v) Tween-20

result sections. The probed antibody/ies was/were visualised by a chemiluminescence substrate, ChemiFast (Syngene) according to the manufacturer's protocol. Blotted membranes were put in between a plastic sheet and imaged using G-box (Syngene, UK). Blot images were also saved as .tif and .sgd file formats.

2.8 Protein quantification

After determining the level of protein purity *via* SDS-PAGE, the protein was quantified from the solution as well as from the corresponding band on SDS-PAGE as follows:

2.8.1 Quantification from solution

Nanodrop ND-1000 spectrophotometer (Fisher Thermo Scientific, UK) was used to estimate the protein concentration in the purified tagged Caf1R samples (average of 3 × 2 µl). It gave concentration values in mg/ml, which were further converted into pmol/µl for subsequent use in DNA-Protein binding assay, EMSA.

2.8.2 Quantification from a band on SDS-PAGE

This procedure was used for quantification of Caf1, either extracted from the cell surface or directly analysed from the whole cell of *E. coli* DH5α expressing either pACYCF1 or pACYCF1-R mutants. ImageJ software (Hartig, 2013) was used to quantitate the CB stained Caf1 bands of the G-box imaged gels saved as .tif files. ImageJ calculates the peak area of the band to be quantified, in grey scale. This was readily identified by location on the gel and can be easily distinguished from background bands. The average area under the peak was calculated from 3 different transformants and then converted into relative percentage (%) of WT Caf1 extracted from *E. coli* DH5α/pACYCF1, following growth at 37°C for both 4 h and 6 h. For quantification, an equal OD unit of the cells was applied to gels in order to get valid comparison between samples.

2.9 Protein characterisation procedure

2.9.1 Concentration

Following purification of wild type His₆-tagged hCaf1R or hCaf1R^T and MBP-tagged MBPCaf1R, the eluted fractions (1 ml) that showed tagged-Caf1R in relatively high amount with less background (checked by SDS-PAGE and confirmed by western blot) were selected for subsequent concentration. In this regard, Fn-18 from hCaf1R (30 µg/ml; **Fig. 4.4**), Fn-10 from hCaf1R^Tgs (40 µg/ml; **Fig. 4.21**) and Fn-3 from MBPCaf1R_{gs} (380 µg/ml; **Fig. 4.25**) were concentrated using 10 kDa MWCO Viva spin concentrators (Vivaproducts). Concentrators were pre-equilibrated with nH₂O for 5 min at RT and then the purified fraction of choice was applied (500 µl × 2) on top of the concentrator membrane followed by centrifugation (14,000 rpm/30 min at RT). Tenfold-concentrated (1 ml to 100 µl) samples were remixed and carefully removed from the concentrator ensuring not to touch the membrane. Protease inhibitor (Thermo Scientific Pierce, UK) was added to each concentrated protein sample (EDTA-free for His-tagged Caf1R and EDTA plus for MBP-tagged Caf1R) and all were stored 4°C.

2.9.2 Buffer exchange using viva spin concentrator

Excess salt or imidazole from the concentrated hCaf1R was removed using a Viva spin concentrator (1 ml size). The concentrated hCaf1R (100 μ l) was mixed in 500 μ l of exchange buffer (50 mM NaPO₄ buffer, 150 mM NaCl, 5 mM DTT; pH 7.0) and applied to the concentrators, which had been pre-treated with nH₂O. The concentrator was centrifuged at 14,000 rpm/30 min at RT and 200 μ l buffer exchanged hCaf1R recovered. After exchanging buffer, 50 μ l was stored at 4°C (EDTA-free proteases inhibitor added) and the remaining 150 μ l (200 μ g/ml) was directly used for further gel filtration chromatography.

2.9.3 Buffer exchange using Mini D-tube dialyser

Purified and concentrated hCaf1R^{Tgs} (250 μ g/ml) and MBPCaf1R_{gs} (750 μ g/ml) along with purchased hCaf1R^P (500 μ g/ml) were dialysed against 20 mM Tris, 300 mM NaCl, pH 8.0; adapted from (Lowden et al., 2010). Mini D-tube dialysers (Novagen®), with MWCO 6-8 kDa were used for this purpose. Prior to dialysis, three D-tubes were pre-treated with nH₂O (100 μ l) for 20 min at RT. Protein samples (100 μ l) were then loaded into equilibrated Mini D-tube dialysers and the dialysers, inserted into a Styrofoam rack, were placed in a beaker containing 50 ml dialysis buffer (20 mM Tris, 300 mM NaCl, pH 8.0). Samples were dialysed for at least 12 h at 4°C with 2 changes of buffer. Following dialysis, the protein samples were removed (\approx 90 μ l) from the D-tube dialyser, and the protein concentration of each determined using a Nanodrop spectrophotometer (hCaf1R^{Tgs}, 245 μ g/ml; MBPCaf1R_{gs}, 740 μ g/ml; hCaf1R^P, 500 μ g/ml). Samples were stored at 4°C after adding 1 \times proteases inhibitor (Thermo Scientific Pierce, UK) (EDTA-free for hCaf1R^T and EDTA plus for MBPCaf1R_{gs}).

2.9.4 Size exclusion chromatography

A pre-packed Superose™12 10/300GL column (GE Healthcare), and AKTA purifier (GE Healthcare) was used to estimate oligomerisation state of hCaf1R. The column was equilibrated with 2-column volumes (CV) of elution buffer (50 mM NaPO₄, 150 mM NaCl, 5 mM DTT; pH 7.0) at a rate of 800 μ l/min. The elution profile of four protein markers, Blue dextran (2000 kDa), Albumin (67 kDa) Ovalbumin (45 kDa) and Ribonuclease A (13.7 kDa) from GE Healthcare, (150 μ l, 1 mg/ml each) were monitored individually. All four markers were then mixed at an equimolar concentration and 150 μ l was applied to the column to calibrate. Concentrated and buffer exchanged hCaf1R (150 μ l) was applied to the column and 1 ml fractions were collected at an elution rate of 0.8 ml/min. Position of hCaf1R was monitored by immunoblotting.

2.10 Electrophoretic mobility shift assay (EMSA)

EMSA has been used extensively for studying DNA-protein interactions (Carolina and Celso, 2012). Principally, this technique is based on the fact that DNA-protein complexes migrate slower than non-bound DNA in a native polyacrylamide gel, resulting in a shift in migration of the labeled DNA band (Carolina and Celso, 2012). A LightShift® Chemiluminescent EMSA kit (Thermo Scientific, Pierce) was used to test the ability of binding of His₆- and MBP tagged Caf1R to bind specific and non-specific *caf* DNA fragments.

2.10.1 Preparation of specific and non-specific *caf* DNA probes for use in EMSA

Biotin-TEG (at 3') labelled, HPLC purified oligonucleotides (47 bp) of both strands for specific (F2-B/F2c-B) and non-specific (F1-B/F1c-B) DNA fragments were purchased from MWG, Eurofin (Germany) (See Fig. 5.14 for details). F2 oligonucleotides were also purchased unlabeled (F2/F2c). Following purchase, each strand was resuspended in nanopure nH₂O to give a stock of 200 pmol/μl and stored at -20°C. For subsequent use in EMSA, the complementary strands of the desired probe (Biotin labelled F1-B/F1c-B or F2-B/F2c-B or unlabelled F2/F2c) were annealed by mixing 20 μl of working stock (100 pmol/μl) of each strand in a final volume of 50 μl annealing buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA). Samples were incubated at T_m+10°C for 10 min in an Eppendorf Mastercycler PCR machine and then cooled down at RT for 1 h. A dilution series of each annealed probe (40 pmol/μl, ds DNA) was prepared in 1× annealing buffer down to a concentration of 5 fmol/μl. Stock and diluted probes were stored at -20°C, for up to one year.

2.10.2 Preparation of cell lysate for EMSA

Overnight culture of *E. coli* BL21 (DE3) carrying pEThCaf1R_{gs}, pEThCaf1R_{gsE98G} or pET28a⁺ (negative control) or cultures of *E. coli* K12-ER2508 carrying pMALc2-MBPCaf1R_{gs}, pMALc2-MBPCaf1R_N or pMALc2x plasmid (negative control), were subcultured (1/100) in 10 ml LB plus kanamycin (30 μg/ml) or ampicillin (100 μg/ml), as appropriate and grown at 37°C with shaking at 225 rpm. At 0.5 OD₆₀₀ IPTG (0.30 mM final concentration) was added to induce expression of hCaf1R^T_{gs} or hCaf1R^T_{gsE98G} and MBPCaf1R_{gs} or MBPCaf1R_N. Following a further 37°C/5.0 h incubation (pET28a⁺ plasmid based constructs) or 2.5 h incubation (pMALc2x plasmid based constructs), 1.0 OD unit of induced cells were harvested by centrifugation (13,000 rpm – 5 min), resuspended in 500 μl EMSA lysis buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, 10% Glycerol and 1× EDTA-free protease inhibitor (Pierce) and lysed by sonication (on ice) for 3 min with 25 sec pulse ON/OFF mode. Following lysis, the cell lysates were clarified by centrifugation (21,000 rpm/15 min/4°C), and the soluble supernatant fraction used directly to set up EMSA binding reactions.

2.10.3 Purified, concentrated and dialysed protein samples used in EMSA

From section 2.9.1 and 2.9.3, purified, concentrated and dialysed hCaf1R^T_{gs} (40 kDa), MBPCaf1R_{gs} (78.9 kDa) and hCaf1R^P (38.5 kDa) protein samples, stored at 4°C with 1× proteases inhibitor (EDTA-free for hCaf1R^T_{gs}/hCaf1R^P and EDTA plus for MBPCaf1R_{gs}) were used in corresponding final concentration, 6.31, 9.37 and 12.87 nmol/ml.

2.10.4 Setting up EMSA binding reaction

Binding reactions were set up in 1.5 ml eppendorf tubes at RT. In a standard 20 μl reaction, the reaction components were added in the following order: autoclaved nanopure nH₂O (xx μl), 10× binding buffer⁹ (2 μl), 1 μg Poly dI.dC (1 μl), freshly prepared cell lysate (4 or 6 μl) or varying amount of protein sample (0.10-55 pmol/μl), 3' Biotin-labeled F1-B or F2-B *caf* DNA probe (5 to 12.5 fmol in 2 μl). For competition reactions 100-fold molar excess unlabeled specific (F2) *caf*

⁹ 100 mM Tris, 500 mM KCl, 10 mM DTT; pH 7.5

DNA probe (2 μ l) was included, prior to addition of purified protein or lysate sample and biotin-labeled specific (F2-B) *caf* DNA probe. After adding each component, the reaction mix was gently mixed by pipetting once or twice and then incubated at RT for 20 min. Prior to running the gel, 5 μ l of 5 \times loading buffer (Thermo Scientific Pierce, UK) was added to each reaction. Following incubation all reactions were resolved on polyacrylamide DNA-retardation gels followed by electro transfer on positively charged Nylon membrane, and detection with Streptavidin-HRP.

To discriminate binding (shift) of tagged Caf1R from lysate(s) or purified/dialysed sample(s), a negative control of binding reaction was always included in the assay. For example, the negative control for cell lysate(s) expressing tagged Caf1R was freshly prepared cell lysate(s) from the corresponding cells expressing the respective vector (pET28a⁺ or pMALc2x) whereas the negative control for binding reaction(s), set up by using purified/dialysed tagged Caf1R was reaction in the absence of Caf1R.

2.10.5 Polyacrylamide Gel Electrophoresis and Streptavidin-HRP based detection

Bolt™ Mini DNA-retardation gels (6% acrylamide with 12 \times 25 μ l wells) (Life technologies) were used to monitor gel shift. Gels were pre-run at 100 volts for 1 h at 4°C in 0.5 \times TBE buffer (44.57 mM Tris-base, 44.47 mM Boric acid, 0.025 mM EDTA; pH 8.0) using a Bolt™ Mini gel electrophoresis system (Life technologies). The entire binding reaction was directly applied to the pre-run gel and electrophoresis continued at 4°C without changing buffer, until the blue tracking dye was 0.5 cm from the bottom of the gel (approximately 1 h 30 min). DNA was then electro-transferred to positively charged Nylon Hybond XL membrane (GE Healthcare) at 30 volts for 1 h at 4°C in fresh 0.5 \times TBE. The membrane was then cross-linked using UV_{302nm} for 5 min with DNA side down on a T-20 UVP Dual-intensity trans-illuminator. After UV cross-linking, the membrane was blocked in EMSA blocking buffer (20 ml) for 15 min at RT with shaking, then transferred into a clean square petri dish. The membrane was submerged in 20 ml EMSA blocking buffer, supplemented with streptavidin-HRP conjugate (1:300) and incubated at RT for 15 min with gentle shaking. The membrane was then rinsed with 1 \times EMSA wash solution, washed four times (5 min each) with 1 \times EMSA wash solution (20 ml), with gentle shaking. The membrane was then carefully removed, the edge blotted on a paper towel to remove excess buffer and then transferred to another clean plastic box containing 30 ml EMSA substrate equilibration buffer. This was incubated for 5 min with gentle shaking. In the mean time a fresh EMSA substrate-working solution (1 ml/membrane) was prepared by mixing an equal amount (500 μ l each) of Luminol enhancer solution and stable peroxidase solution in 1.5 ml eppendorf tubes, keeping the substrate in the dark. The membrane was removed, the edge again blotted to remove excess buffer and then placed onto a clean plastic sheet face-up (with the intense dye-front facing up). The freshly prepared EMSA substrate-working solution was then pipetted over the entire surface and the membrane incubated for 5 min at RT in the dark, without shaking. Excess solution was removed by blotting the edge of membrane on a paper towel. The moist membrane was then placed between 2 clean transparent plastic sheets (Rexel, reinforced polished plastic pockets – no 12265), avoiding bubbles and imaged immediately using G-box (Syngene) with automatic exposure time, under chemiluminescent substrate-specific immunoblotting image setting. The images were saved in .tif file format.

2.11 Bioinformatic analysis

2.11.1 DNA sequence(s) analysis, primers and constructs design

Sequence analysis software, DNAdynamo (<http://www.bluetractorsoftware.co.uk/DNAdynamo/>) was routinely used to design primers, plasmid-constructs or mutants and analysis of sequencing results. Raw sequencing data (peaks chromatogram) were regularly analysed by DNA sequencing data viewer tool, FinchTV (<http://www.geospiza.com/FinchTV/>).

2.11.2 Promoter prediction

A bioinformatic and visual analysis of *caf1R-caf1M* (327 bb) and *caf1A-caf1* (80 bp) intergenic region was carried out to predict promoter(s) for *caf1R*, *caf1M* and *caf1*. The complete intergenic region of choice with a few hundred bases of the genes on either side joining was analysed using the bacterial promoter prediction program, BPROM (Solovyev, 2011) (<http://www.softberry.com/berry.phtml/BPROM>) with default parameters in order to identify -10 and -35 elements. BPROM identifies σ^{70} type promoters (one of the major class of *E. coli* promoters) based on a linear discriminant function (Korzheva et al.) algorithm. This algorithm is based on nt composition of the predicted -10 and -35 elements and the distance between them. In addition to identifying -10 and -35 elements, BPROM also identifies transcriptional factor(s) binding site(s), surrounding the predicted -10 and -35 elements. It has been suggested that BPROM identifies -10 and -35 elements with about 80% accuracy and specificity, when tested on equal number of sequence sets containing promoter and non-promoter sequences (Solovyev, 2011). For visual prediction of promoters, upstream of *caf1R*, *caf1M* and *caf1*, the upstream region of choice was initially screened for -10 element consensus (TATAAT) and then looked for sequence motif with highest similarity with -35 element consensus (TTGACA), at 16-22 nt upstream. The following criteria were used to identify optimum likely promoters with an optimum distance (spacer) between -10 and -35 elements, 17 nt (from the end of -10 element to the start of -35 element) using as a guide the consensus frequency percentage (%) for each nucleotide within both elements as follows: T_{80%}A_{95%}T_{45%}A_{50%}A_{60%}T_{96%} (-10) and T_{82%}T_{84%}G_{78%}A_{65%}C_{54%}A_{48%} (-35) (Lisser and Margalit, 1993).

2.11.3 Identifying potential Caf1R binding sites

The *caf1R-caf1M* and *caf1A-caf1A* intergenic region along with few hundred bases of the adjoining genes was screened to identify nucleotide-repeat sequences, using the repeat finder program, REPFIND (<http://zlab.bu.edu/REPFIND/>) with default parameters. Identified repeat sequences (3-7 nt long) were further analysed visually on both strands to increase the length of each repeat including mismatches. The direction of each repeat sequence was assigned with respect to closeness to the preceding gene.

2.11.4 Prediction of potential Shine-Dalgarno sequence for *caf1R*, *caf1M* and *caf1*

Identification of potential Shine-Dalgarno (SD) sequences or ribosomal binding sites (Quade et al.), 5-10 bp upstream of the assigned *caf1R*, *caf1M* and *caf1* translation initiation codon (ATG (Met) for *caf1M* and *caf1*; ATG and TTG (Met) for *caf1R*) was visually analysed for the sequence

motif with highest degree of similarity to the *Y. pestis* predicted SD consensus motif, AGGAGGT (Karlyshev et al., 1992; Parkhill et al., 2001; Starmer et al., 2006).

2.11.5 Prediction of potential transcriptional start site for *caf1R*, *caf1M* and *caf1*

After localisation of an active promoter for the corresponding gene, 10th base downstream from start of the identified -10 element of each promoter was assigned as +1.

2.11.6 Prediction of RNA thermometers (RNATs)

Nucleotides from assigned +1 of P_R², P_R^K and P_M promoters to a string of 1-10 amino acids of *caf1R* or *caf1M* was screened with Mfold2.3 (Zuker, 2003) at different temperatures with default parameters.

2.11.7 Physiochemical properties prediction of Caf1R

Entrez gene database, Prosite database and ProtPrm were used to confirm gene and protein entry of Caf1R along with its physiochemical properties, respectively.

2.11.8 Caf1R homologs search and their multiple sequence alignment

The NCBI accession number of full-length Caf1R (Q65AJ3, relative to TTG start codon) was analysed by NCBI protein-BLAST (BLASTp) server (<https://blast.ncbi.nlm.nih.gov>) against non-redundant (nr) database with an E-value threshold, 1e-3 and excluding the *Y. pestis*. The 500 sequences were retrieved. Identical sequences were removed manually and the remaining sequences were analysed from their genomic records and genome organisation in order to identify regulators associated to CU system. Following this analysis, selected sequences were organised according to similarity percentage (most similar on top) using SIAS web server (<http://imed.med.ucm.es/Tools/sias.html>) and Jalview 2.9 (Waterhouse et al., 2009) was used to view multiple sequence alignment.

2.11.9 Caf1R secondary structure prediction

The amino acid sequence of Caf1R (relative to TTG start) was analysed by PSIPRED (McGuffin et al., 2000), with default parameters, in order to predict the secondary structure of the full-length Caf1R and DNA binding predicted recognition helix-6 with E98G/A/T/K mutations.

2.11.10 Modelling of Caf1R and its DNA-binding domain (DBD) mutants

The amino acid sequences of Caf1R (relative to TTG start) and all DBD mutants were submitted to IntFOLD3.0 (McGuffin et al., 2015; Roche et al., 2011). In each case, model was retrieved from the lig2.pdb file (protein model associated with two DNA ligands). This in turn was based on the top scored model with a quality score of ≥ 0.788 over 1.0, probability of incorrectness or P-value, $\leq 2.328E-4$ and reliability score, 'CERTAIN'. Unless otherwise stated, almost all protein models were based on the Rob protein template with two DNA ligands, one from cocrystal structure of the *micF*-Rob complex (PDB-1d5y) (Kwon et al., 2000) and the other from cocrystal structure of the *mar*-MarA complex (PDB-1bl0) (Rhee et al., 1998). For further analysis, *micF*-DNA was removed from all models and thus kept modelled Caf1R associated only with the *mar* DNA in order to analyse Caf1R-DNA interactions.

To assess the impact of site specific mutation(s) on Caf1R-DNA interactions, modelled *mar* DNA-Caf1R complex and mutant model of interest (after removing both DNA ligands) were superposed by TM align web server (Zhang and Skolnick, 2005). Full-atom superposition for the entire protein chain was retrieved to analyse DNA-Caf1R interactions, using MacPyMOL1.3.

Chapter 3

Demonstration of Caf1R as a positive regulator of the *caf* locus

3.1 Introduction

Unlike the study of F1 assembly *via* classical chaperone/usher system (Dubnovitsky et al., 2010; Kersley et al., 2003; Knight, 2007; MacIntyre et al., 2001; Yu et al., 2009; Yu et al., 2012; Zavialov et al., 2001; Zavialov et al., 2003) the mechanisms controlling expression of F1 have received very little attention. Sequencing of the *caf* locus (Karlyshev et al., 1992) and bioinformatic analysis of AraC/XylS-like regulators identified Caf1R as a member of the AraC/XylS family of transcription regulators (Gallegos et al., 1997; Ibarra et al., 2008; Schuller et al., 2012). But there have been no studies on the way in which Caf1R regulates expression of the *caf* gene cluster or *caf* locus. Transcriptional control *via* Caf1R regulator is likely to be fundamental in controlling F1 expression. Previous studies have used recombinant plasmids, pFS2 and pFMA1 to study the mechanism of F1 assembly (Chapman et al., 1999; Di Yu et al., 2012; Dubnovitsky et al., 2010; Kersley et al., 2003; MacIntyre et al., 2001; Yu et al., 2009; Yu et al., 2012; Zavialov et al., 2003; Zavialov et al., 2005). The *ptrc99* based plasmid, pFMA1 contains *caf1M*, *caf1A* and *caf1* under tight control of *ptrc* promoter (induction by IPTG). It produces a high-level of periplasmic chaperone (Caf1M), which in turn has been useful to trap accumulated F1 subunit, Caf1 in the periplasm in assembly defective mutants (Dubnovitsky et al., 2010; MacIntyre et al., 2001; Yu et al., 2009). However, high-level of production of the outer membrane usher, Caf1A from this plasmid is toxic to cells, requiring short induction times (Chapman et al., 1999; MacIntyre et al., 2001). The complete *caf* locus was originally subcloned from *Y. pestis* pFra plasmid into a high-copy number plasmid, pFS2 (Drozdov et al., 1995; Galyov et al., 1990). Spontaneous loss of F1 expression was observed in strains carrying this plasmid following storage at -80°C (Galyov et al., 1991; Karlyshev et al., 1992). In order to minimise spontaneous loss in F1 expression and maintain high levels of production of F1 without toxicity, the complete *caf* locus from pFS2 was subcloned into a medium copy number plasmid, pACYCDuet-I (Novagen) to generate the construct pACYCF1 (**Fig. 3.1**). A surprisingly low level of F1 expression was observed from this new construct. Sequencing of the entire *caf* locus in both the pACYCF1 construct and the original pFS2 plasmid identified a single nucleotide change (A→G) in *caf1R*, causing an amino acid substitution Glu98→Gly98 (E98G) (**Fig. 3.1**) within the N-terminus of Caf1R (Lopez-Tolman, A.; unpublished data). It was concluded that this single Caf1R_{E98G} mutation must have been responsible for the low level of F1 in both plasmids. Hence this pACYCF1 construct was assigned the name pACYCF1_{S_{PM}} to reflect that it has a spontaneous mutation.

The aim of this first chapter was to unveil the impact of this spontaneous mutation and to begin to understand the properties of Caf1R and its role in regulation of F1 expression. Different approaches were employed including generation of a stable recombinant *caf* construct to monitor Caf1R regulated F1 expression, bioinformatic analysis of Caf1R to identify closest homologs and in particular those linked to chaperone/usher systems, modelling the structure of Caf1R-DNA complex to address the impact of the E98G substitution and mutagenesis to identify the contribution of key residues including E98 on transcription of the *caf* locus and subsequent assembly of the F1.

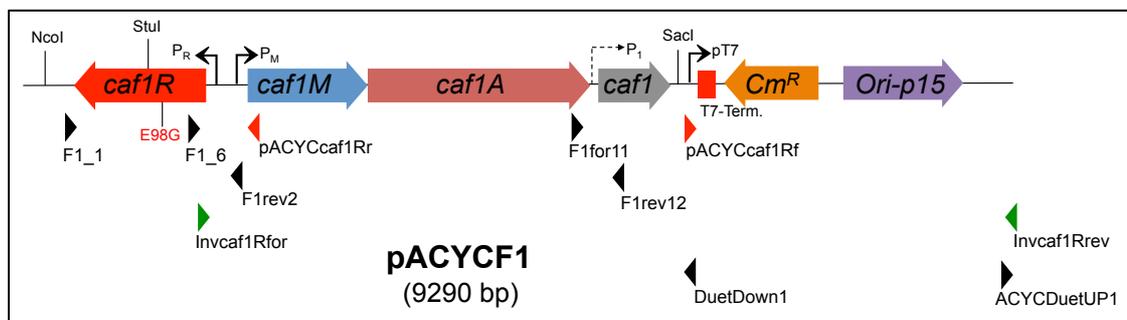


Figure 3.1| **Plasmid map of pACYCF1.**

The *caf* gene cluster 5128 bp plus 115 and 82 bp from the end of *caf1R* and *caf1*, respectively, PCR-amplified from pFS2 plasmid and subcloned into NcoI and SacI sites of pACYCDuet-I vector (Novagen). Encoded products include the positive regulator (Caf1R), chaperone (Caf1M), outer membrane usher (Caf1A) and structural subunit of F1 capsule (Caf1). Additional features include, chloramphenicol resistance marker (Cm^R), origin of replication (Ori-p15), T7 promoter (pT7) and T7 terminator (T7-term). Red and green arrowheads indicate primers used to create pACYC-R and pACYC-MA1. Black arrowheads indicate sequencing primers. Locations of the E98G spontaneous mutation and StuI site within *caf1R* are indicated. P_R and P_M , indicate proposed promoters for *caf1R* and *caf1M*, respectively. Putative promoter for *caf1* (P_I) is indicated by dotted arrow. Expression from the *caf* locus is controlled by Caf1R regulator and is under thermoregulation, ON at 37°C and OFF at 26°C (see Chapter-6).

Results and Discussion

3.2 Generation of a stable construct expressing recombinant *caf* locus

3.2.1 Caf1R_{E98G} mutation is complemented by co-expression with wild type Caf1R

To confirm the negative impact of the Caf1R_{E98G} mutation on F1 expression, a complementation approach was taken. The hypothesis was that the mutated regulator, Caf1R_{E98G} encoded within pACYCF1_{SpM} could be complemented by wild type Caf1R_{G98E}, and that this would restore a high level of F1. Therefore, recombinant *E. coli* Top10 cells carrying pACYCF1_{SpM} plasmid (encoding Caf1R_{E98G}) were transformed with pBADhCaf1R, which encodes His₆-tagged Caf1R (hCaf1R) under control of P_{BAD} promoter. F1 expression was monitored following induction of expression of hCaf1R with 0.02-2% arabinose (L-ara) at 37°C. Following induction, F1 polymer was heat extracted at 57°C from the cell surface of harvested cells (4.5 OD units), (section 2.4.2) and analysed by SDS-PAGE for the denatured F1 subunit, Caf1. SDS-PAGE analysis confirmed a very low level of Caf1 from cells carrying pACYCF1_{SpM} and pBADHisA vector (**Fig. 3.2a, lane 2**). A dramatically enhanced level of F1 from co-transformants carrying both pACYCF1_{SpM} and pBADhCaf1R (**Fig. 3.2a, lane 3-4**) was observed; confirming that the mutated Caf1R_{E98G} encoded in pACYCF1_{SpM} had been successfully complemented with pBADhCaf1R encoded wild type hCaf1R. No band at this position from the negative control, pBADHisA+pACYCDuet-I vectors (**lane 5**) and a small amount from the Caf1 positive control, pFMA1 (MacIntyre et al., 2001) (**lane 1**) confirmed identity of the Caf1 band and the high level from pACYCF1. Induction of hCaf1R with 0.02% L-ara seemed sufficient to produce an optimum level of F1 from these co-transformants.

It was possible that the enhanced level of F1 from pACYCF1_{SpM}+pBADhCaf1R co-transformants was a consequence of higher levels of Caf1R in the presence of 2 copies of the gene and that L-ara induced expression had little to do with the E98G mutation of Caf1R. To clarify this, F1 expression was monitored from pACYCF1_{SpM}, co-transformed with either pBADhCaf1R_{E98G} or

with pBADhCaf1R. The level of Caf1 produced in the presence of complementing wild type Caf1R was much higher (**Fig. 3.2b, lane 3**) than that where complementation was with mutated Caf1R_{E98G} (**Fig. 3.2b, lane 2**). Thus regain of F1 in pACYCF1_{SpM}+pBADhCaf1R co-transformants could not simply be attributed to an increase in copy number of the *caf1R* gene. This highlighted the loss-in-function phenotype of the E98G mutation of Caf1R and indicated a possible critical significance of residue E98 to Caf1R structure or function.

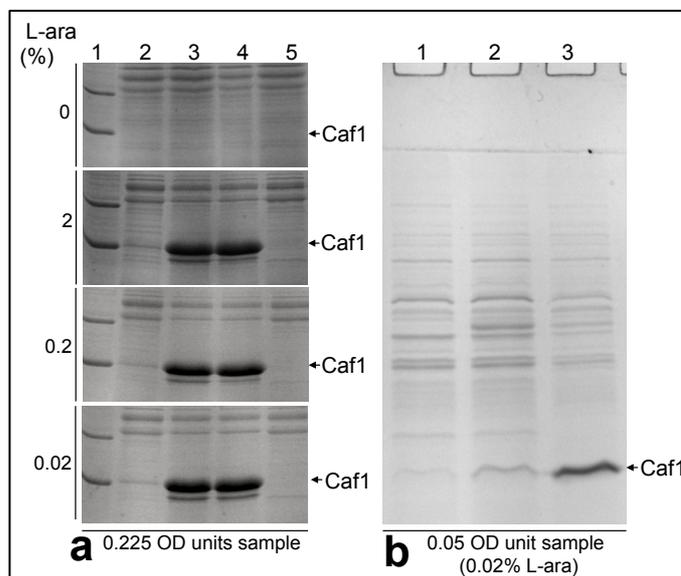


Figure 3.2| **Caf1R_{E98G} mutation is complemented by co-expression with wild type Caf1R.**

a) SDS-PAGE (14% acrylamide) analysis of heat denatured (98°C-15 min) F1 subunit Caf1 from cells carrying pACYCF1_{SpM} (Caf1R_{E98G}) complemented with pBADhCaf1R (WT Caf1R). F1 was extracted in 100 µl PBS from the cell surface of 4.5 OD units culture of *E. coli* Top10 cells, expressing either pACYCF1_{SpM}+pBADHisA (lane 2) or pACYCF1_{SpM}+pBADhCaf1R (lanes 3 and 4). Cells were harvested after 4 h induction (in LB containing 0.6% glucose) at 37°C with L-ara (2.0-0.02%), as indicated. Lane 1, positive control for Caf1, *E. coli* BL21 (DE3)/pFMA1, induced with IPTG (0.5 mM) at 37°C for 1.5 h and processed for F1 extraction in similar way. Lane 5, negative control, Top10/pBADHisA+pACYCDuet-1 vectors. Sample equivalent to 0.225 OD units cells was applied in each lane. **b)** Cells carrying pACYCF1_{SpM} complemented with either pBADhCaf1R (WT Caf1R) (lane 3) or pBADhCaf1R_{E98G} (mutant Caf1R) (lane 2) or pBADHisA vector (lane 1). F1 prepared from 1.0 OD unit cells harvested after 4 h induction with 0.02% L-ara at 37°C. Sample equivalent to 0.05 OD units cells was applied in each lane.

3.2.2 Caf1R regulator is an activator of F1 expression

To confirm the requirement of Caf1R for expression of F1, the entire *caf1R* gene was deleted from pACYCF1_{SpM} and replaced with an NcoI site by Inverse PCR (section 2.2.4(iii)(b)(i)). The inverse PCR primers, *Invcaf1Rfor* and *Invcaf1Rrev* were used with Q5 HiFi DNA polymerase (NEB) at two annealing temperatures (55.1 and 60.8 °C) and gave products of the expected size (≈ 8.366 kb) (**Fig. 3.3a**). Both products were gel-excised, column purified, digested with NcoI (**Fig. 3.3b**) and then religated with T4 DNA ligase (Thermo Scientific) prior to transformation of *E. coli* DH5α. Seven individual transformants, 3 from the 55.1°C PCR reaction and 4 from the 60.8°C PCR reaction were screened by NcoI and StuI restriction digestion. NcoI confirmed the correct linear size (8.366 kb) of pACYC-MA1 in all 7 transformants. Absence of StuI digestion confirmed deletion of *caf1R*, as StuI is a unique restriction enzyme site within *caf1R* (**Fig. 3.3c**). Retention of the intact and correct DNA sequence of the *caf1R-caf1M* intergenic, regulatory region was confirmed for transformant 5, which was then tested for expression of F1 in whole

cells. SDS-PAGE analysis showed that no CB stainable Caf1 was detected from pACYC-MA1 following 4 h growth at 37°C (**Fig. 3.4, lane 2**), although a very low level of F1 expression was confirmed from pACYC-MA1 by immunoblotting (data not shown). In contrast, *caf1R* complemented pACYC-MA1 (pBADhCaf1R+pACYC-MA1) produced a very high level of Caf1 (**Fig. 3.4, lane 4**), confirming gain-in-function upon supply of Caf1R regulator in *trans*. Thus these results confirm the role of Caf1R as an activator of expression from the *caf* locus. The role of Caf1R as regulator is studied in depth in Chapter-5.

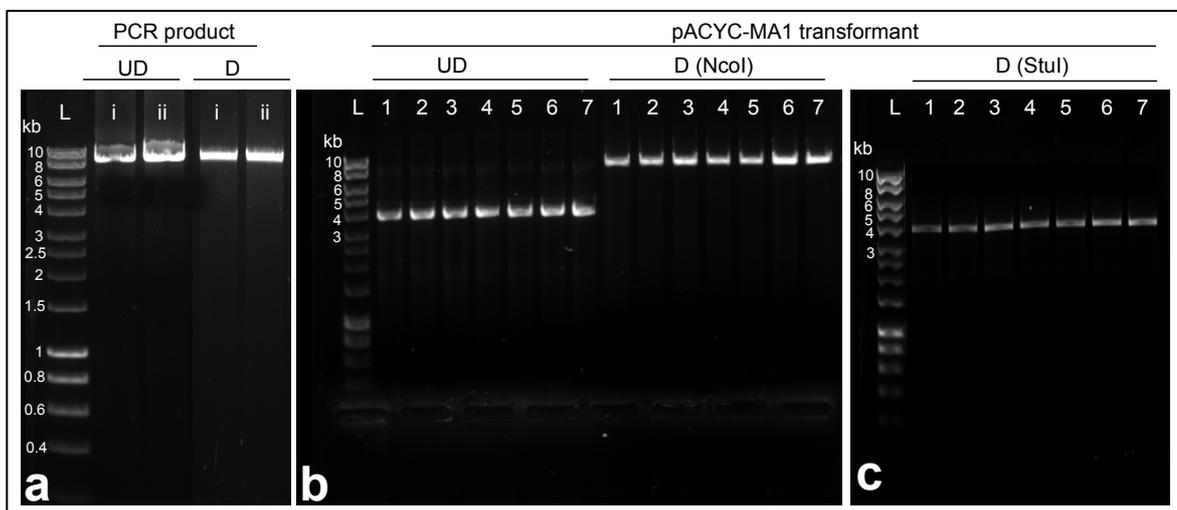


Figure 3.3| **Construction of pACYC-MA1.**

a) AGE (0.75% agarose) analysis of inverse PCR products (100 ng each) from 55.1°C (i) and 60.8°C (ii) annealing temperature from pACYCF1_{SpM} template. UD, undigested and D, NcoI digested. **b-c)** Restriction digestion analysis of 7 pACYC-MA1 transformants with NcoI, StuI or undigested (UD). Transformants 1-3 and 4-7 are from PCR reactions (i) and (ii), respectively. L, DNA Hyper ladder-I (20 ng).

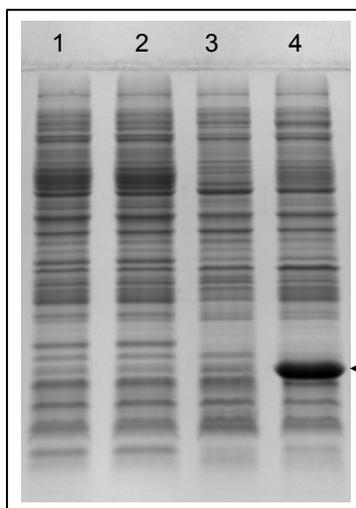


Figure 3.4| **F1 expression is abolished in pACYC-MA1.**

SDS-PAGE (16% acrylamide) of F1 extracted in 100 µl PBS from 5 OD units culture of Top10 cells containing pACYC Duet-I empty plasmid (lane 1); pACYC-MA1 (lane 2); pACYC-MA1+pBADHisA (lane 3) and pACYC-MA1+pBADhCaf1R (lane 4) plasmids. Samples in lane 1 and 2 are from noninduced cultures while samples in lane 3 and 4 are from 0.02% L-ara induced cultures. Cells from both noninduced and induced cultures were harvested at 4 h post-induction at 37°C. The equivalent of 0.25 OD unit cells was applied to each lane. Arrowhead indicates denatured F1 subunit, Caf1.

3.2.3 Impact of N-terminal His₆-tag on Caf1R function

The His₆-tag and enterokinase cleavage site were included at the N-terminus of Caf1R to aid identification, purification and subsequent characterisation of Caf1R. The protease cleavage site aids in subsequent cleavage of protein from the tag. On the other hand, addition of a His-tag extension can interfere with the overall protein folding process if it is attached to a critical region involved in folding (Halliwell et al., 2001; Ledent et al., 1997; Wu and Filutowicz, 1999), for example, residues contributing to the hydrophobic core (Murphy et al., 2015). This can destabilise the protein. hCaf1R expressed from pBADhCaf1R was clearly functional (**Fig 3.4**). However, as the construct had been designed for purification and use in *in vitro* activity studies, activity of hCaf1R was compared to that of Caf1R (without the His₆-tag and enterokinase site).

The DNA sequence encoding His₆-tag and enterokinase cleavage site was deleted by Inverse PCR (section 2.2.4(iii)(a)) using pBADhCaf1R template, deletion primers, dhCaf1Rfor and dhCaf1Rrev and Bio-X-ACT short DNA polymerase (Bioline). The PCR product (**Fig. 3.5a**) was digested with NcoI, purified by spin column purification, religated with T4 DNA ligase and transformed into Top10 competent cells. Six transformants were screened by restriction digestion using NcoI and SacI enzymes (**Fig. 3.5b**). NcoI was used to determine the linear size of pBADΔhCaf1R (5.5 kb) and SacI to confirm the deletion of His₆-tag and the enterokinase cleavage site. A restriction site for SacI is present within the codons for the enterokinase recognition sequence, thus the deletion construct should no longer contain SacI site and should remain supercoiled (2.8 kb) in the presence of this enzyme. As expected, digestion with NcoI produced a single band at about 5.5 kb, except for transformant 1, for which two bands were observed, at 5.5 kb and 2.8 kb, suggesting partial digestion. SacI as expected was unable to digest plasmid in all transformants tested (**Fig. 3.5b**). The complete *caf1R* sequence and precise deletion of the coding sequence for the His₆-tag plus enterokinase site was confirmed for transformants 2 and 4 by DNA sequencing using pBADF and pBADR primers. Both transformants were stocked and number 4 was used to test expression.

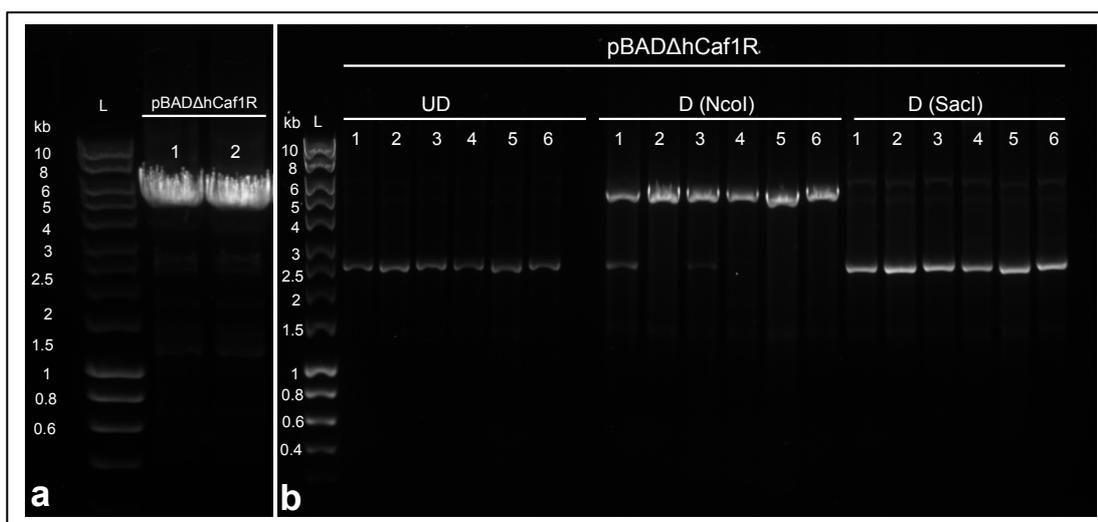


Figure 3.5| **Construction of pBADΔhCaf1R.**

a) PCR product from inverse deletion PCR; undigested (1) and NcoI digested, purified (2, 250 ng), which was used for religation and transformation. **b)** Restriction digestion analysis (0.75% agarose) of plasmids from transformants 1-6; undigested (UD), NcoI or SacI digested (D), as indicated. DNA bands at about 5.5 and 2.8 kb indicate linear and supercoiled plasmid pBADΔhCaf1R. L, Hyper DNA ladder-I (20 ng).

The impact of presence/absence of the His₆-tag plus enterokinase site on Caf1R function was monitored by comparing F1 expression in cells carrying the defective pACYCF1_{SpM} complemented with either pBADhCaf1R (His₆-tagged Caf1R) or pBADΔhCaf1R (no His₆-tag). F1 was extracted from the cell surface of a noninduced and a 0.02% L-ara induced culture of each. SDS-PAGE analysis of F1 polymer (migrates in the stacking gel) and Caf1 subunit (15.5 kDa) (**Fig. 3.6**) showed that upon induction the level of F1 from cells carrying pACYCF1_{SpM}+pBADhCaf1R (H) was actually slightly better than F1 in cells carrying pACYCF1_{SpM}+pBADΔhCaf1R (ΔH), suggesting that addition of the His₆-tag extension does not interfere with Caf1R-mediated activation of F1 expression. F1 was also produced in the noninduced samples (no L-ara) consistent with a high basal level of expression in the absence of glucose induced catabolite repression. In these noninduced samples, H co-transformant produced a much higher level of Caf1 than ΔH co-transformant (**Fig. 3.6**). It is possible that presence of the N-terminal His₆-tag extension actually stabilises Caf1R leading to slightly better levels of F1. Finally, it was concluded that hCaf1R encoded on plasmid pBADhCaf1R was fully functional and a suitable choice for subsequent studies involving complementation as well as for purification of hCaf1R.

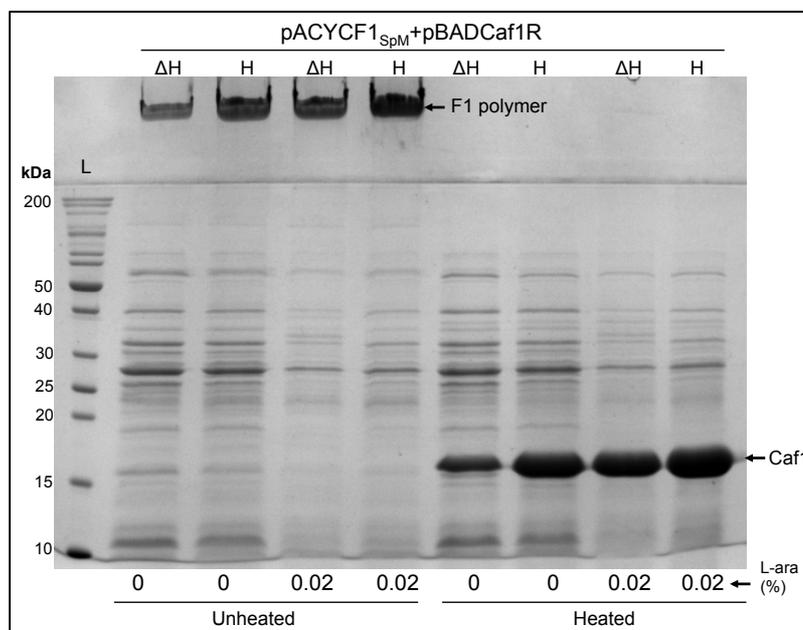


Figure 3.6| N-terminal His₆-tag plus enterokinase cleavage site does not affect Caf1R function.

SDS-PAGE (14% acrylamide) analysis of F1 polymer extracted in 100 μ l PBS from 5 OD units culture of *E. coli* Top10 cells containing pACYCF1_{SpM} complemented with either pBADhCaf1R (H; N-terminal His₆-tag) or pBADΔhCaf1R (ΔH - N-terminal His₆-tag plus enterokinase site deleted). For F1 polymer analysis (remains in stacking gel), an equal amount (10 μ l) surface extracted F1 was mixed in SDS-PAGE sample buffer and applied directly to gel without heating. Denatured subunit (Caf1) was resolved following heat denaturation (97.5°C-15 min) of samples in SDS-PAGE sample buffer. Induction with L-ara in LB (no glucose), as indicated. Cells from both noninduced and induced cultures were harvested at 4 h postinduction at 37°C. The equivalent of 0.25 OD unit cells was applied to each lane.

3.2.4 Repair of E98G mutation in pACYCF1_{SpM}

Results in earlier sections clearly demonstrated the requirement of wild type Caf1R for F1 expression and loss-of-function with the E98G spontaneous mutation. Therefore, it was decided to repair Caf1R_{E98G} in the pACYCF1_{SpM} template in order to generate a working model to study the regulation of F1 expression as well as for studies on the assembly of F1. Site-directed mutagenesis was applied to repair this mutation using the Quikchange Lightning kit (Agilent) (section 2.2.4 (iv) (a)) with the mutagenic primers, Caf1RG98Efor and Caf1RG98Erev. These primers introduce a single nucleotide change in Gly98 codon (TCC to TTC) to revert this residue to wild type Glu98 codon. Following mutagenesis methylated or hemimethylated pACYCF1_{SpM} template was removed by digestion with Fast-DpnI at 37°C for 5 min according to the manufacturer's protocol (Agilent) (**Fig. 3.7**) and 3 µl of DpnI treated PCR product transformed into competent *E. coli* Top10 (50 µl). Screening of transformant plasmids by sequencing with F1_1 primer revealed a number of unwanted mutations. Hence DpnI treatment was repeated using standard DpnI (Bioline) for 4 h as had been used previously in the lab. The *E. coli* strain used (Top10) could also have contributed to the problem, as this strain has lower efficiency of repairing nicked DNA and stabilising bigger plasmids as in the case of pACYCF1_{SpM} (9.33 kb). Therefore, the PCR product digested with standard DpnI (Bioline) was transformed into *E. coli* DH5α. From this set of transformants, eight transformants (numbered 15-22) were picked, plasmids prepared and the DNA sequenced with F1_1 sequencing primer. Sequencing results revealed that the TCC (G98) → TTC (E98) repair was present in all 8 transformants. The complete *caf1R*; *caf1R-caf1M* and *caf1A-caf1* intergenic regions were confirmed by sequencing with F1_1, F1_6, ACYCDuetUP1, F1for11 and F1rev12 primers (**Fig. 3.1**) from transformants 16, 17, 19 and 20. Strains and plasmid DNA of all were stocked and transformant 17 was assigned as pACYCF1 and used in all subsequent studies.

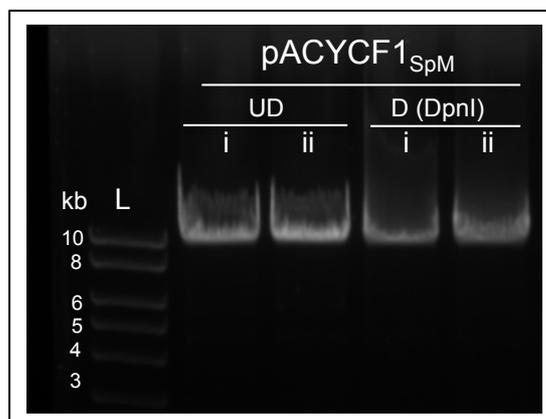


Figure 3.7| **Site-directed mutagenesis of pACYCF1_{SpM}.**

AGE (0.75% agarose) analysis of site-directed (Caf1R_{E98G}) mutagenesis PCR products of pACYCF1_{SpM}. Undigested (UD) and DpnI digested (D) products from two individual PCR reactions. UD (i and ii), ≈100 ng each from 50 and 100 ng of pACYCF1_{SpM} template. D, Fast (i) and normal DpnI digested (ii) product (≈100 ng each) from 50 and 100 ng of pACYCF1_{SpM} template. L, Hyper DNA ladder-I (20 ng).

F1 expression from the newly created pACYCF1 plasmid, encoding the complete wild type *caf* locus was compared to expression from the original plasmid pACYCF1_{SpM}. As F1 expression is known to be downregulated at lower temperatures (Cao et al., 2012; Simpson et al., 1990; Straley and Perry, 1995; Titball et al., 1997), all cultures containing pACYCF1 were grown at 26°C until induction was required. For induction, the growth temperature was switched to 37°C. Production of F1 in high levels, from pFS2, was known to result in a dense capsule of F1 polymer around the cells, making the cells less dense. Recovering these cells requires an extended centrifugation time (13,000 rpm for 20 min) and results in pellets with a fluffy white appearance (Moslehi, E., and MacIntyre, S., unpublished data). Therefore, following temperature upshift from 26 to 37°C, 5 OD units of *E. coli* DH5α carrying either pACYCF1 or pACYCF1_{SpM} were recovered at 4 h and 6 h and the pelleted cells were visually inspected. Cells containing pACYCF1 exhibited the typical fluffy white pellet (**Fig. 3.8a**), indicating a high level of production of F1 polymer. Gain-in-function of the repaired *caf1R* in pACYCF1 was confirmed by SDS-PAGE (**Fig. 3.8b**) analysis of whole cells and of surface extracted F1, as described in section 2.4.

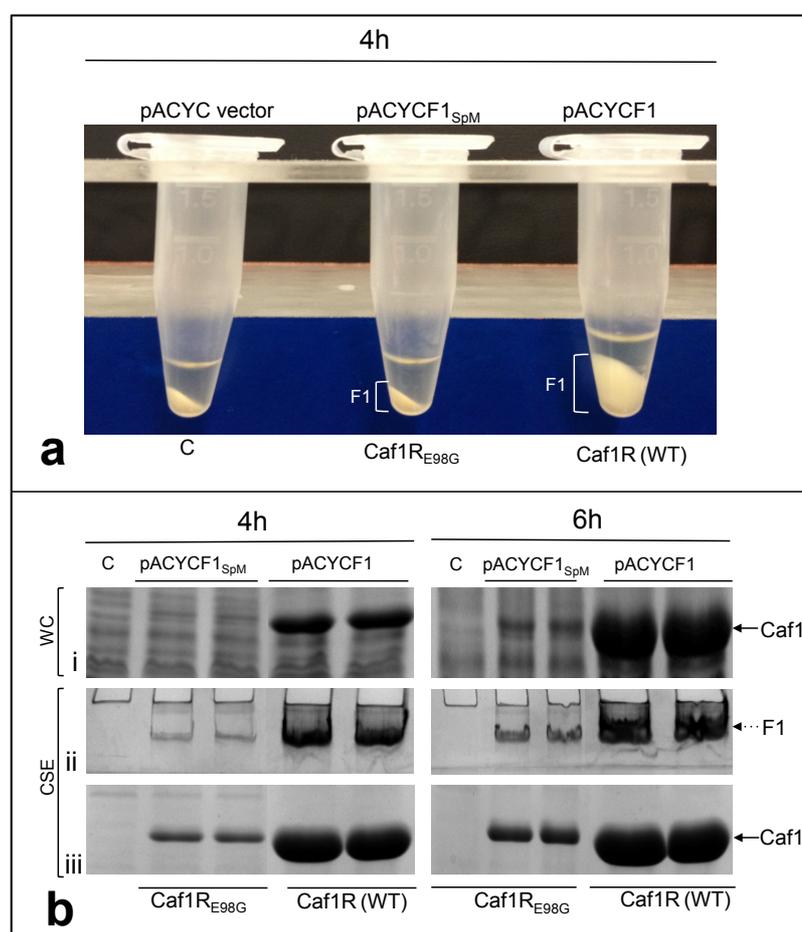


Figure 3.8| **Comparative analysis of F1 from repaired pACYCF1 to spontaneous mutant pACYCF1_{SpM}.** **a)** Visual analysis of pelleted (13,000 rpm-20 min) whole cells, recovered after 4 h thermoinduction at 37°C. **b)** SDS-PAGE (16% acrylamide) analysis from whole cell (WC) (i) and from cell surface extract (CSE) (ii and iii) after 4 and 6 h thermoinduction at 37°C. C, Negative control, *E. coli* DH5α expressing pACYC Duet-1 vector; pACYCF1_{SpM}, *caf1R*_{E98G}; pACYCF1, *caf1R* (WT). 5 OD units cells, mixed in 100 μl PBS were processed for both **(a)** and **(b)**. The equivalent of 0.25 OD unit cells was applied to each lane.

The level of Caf1 subunit was quantified (section 2.8.2) by imaging coomassie blue (CB) stained acrylamide (16%) gels of a lower level of whole cells (1 OD unit, suspended in 200 μ l PBS). The level of F1 from three individual transformants of *E. coli* DH5 α /pACYCF1_{S_{PM}} (Caf1R_{E98G}, mutant) and of *E. coli* DH5 α /pACYCF1 (Caf1R, WT) was monitored after 4 and 6 h thermoinduction (**Fig. 3.9**). The level of F1 increased by 45% in cells with Caf1R and by 58% in cells with Caf1R_{E98G} between 4 h and 6 h. At both time points cells expressing Caf1R_{E98G} showed a marked reduction in F1. After 4 h induction, the level of Caf1/OD cells in samples expressing Caf1R_{E98G} was only 13.05% \pm 0.62 of the level in cells expressing wild type Caf1R (100% \pm 4.38). While after 6 h induction, the level for the Caf1R_{E98G} sample was 14.16% \pm 0.32 that of wild type Caf1R samples (100% \pm 3.09). These values highlight the extent of loss-in-function of the mutated regulator Caf1R_{E98G} with respect to production of F1. The high level of F1 production from wild type pACYCF1 was consistent over all of these tests, indicating a stable working model, to study F1 expression and assembly under the native thermoinducible promoter.

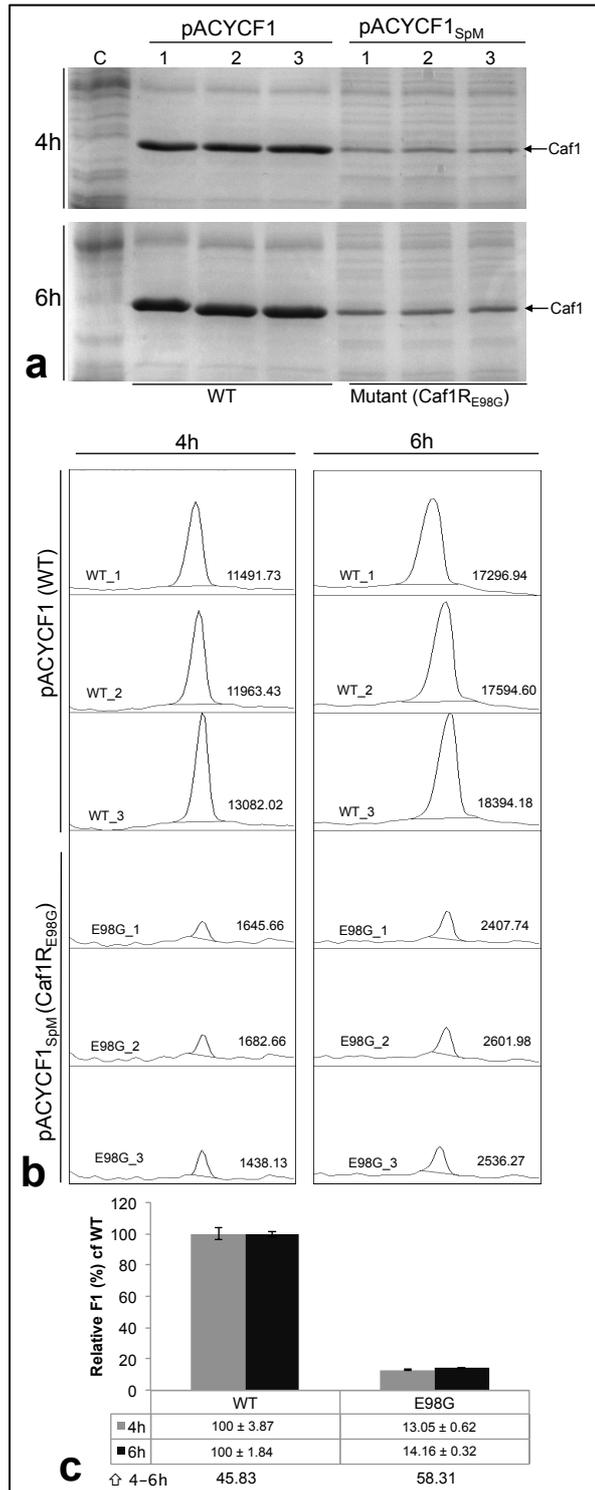


Figure 3.9| **Quantification of F1 in pACYCF1 (WT) and pACYCF1_{SpM} (Caf1R_{E98G} mutant).**

a) SDS-PAGE (16% acrylamide) analysis of F1 levels from three individual colonies of *E. coli* DH5 α carrying pACYCF1 (WT) or pACYCF1_{SpM} plasmid. Expression was monitored from whole cells after 4 and 6 h of thermoinduction at 37°C. *Note:* for quantification 1 OD unit of harvested cells were mixed and heated in 200 μ l PBS + sample buffer and thus an equivalent of 0.04 OD unit sample applied to each lane. The average OD₆₀₀ of cultures of the three replicates at 4 and 6 h was 1.27 and 1.52 (control; *E. coli* DH5 α /pACYC Duet-1 vector); 0.85 and 1.47 (WT); 1.15 and 1.34 (Caf1R_{E98G} mutant), respectively. **b)** Caf1 band quantification by imageJ as described in section 2.8.2. Value shown is the total peak area for the Caf1 band. Line at bottom of the peak, excluded background and unrelated proteins. **c)** The F1 peak area from the triplicate cultures of WT at 4 and 6 h were assigned as 100% and the relative recovery from the Caf1R_{E98G} mutant calculated for both time points. Standard error on each bar is \pm SEM.

3.2.5 Effect of Caf1R_{E98G} on transcriptional activity of *caf1R* and *caf1M* promoter(s)

Results in the earlier sections demonstrated that the Caf1R_{E98G} mutation is a loss-in-function mutation, resulting in a dramatically decreased level of F1 production. As Caf1R is a regulator of the AraC/XylS family many of which activate transcription by binding close to their cognate operator/promoter region (Gallegos et al., 1997), it was hypothesised that Caf1R_{E98G} was most likely ineffective in activating transcription of the *caf* gene cluster (*caf* locus). This could be due to weak binding at the *caf1M* and/or the *caf1R* promoter(s) leading to ineffective or reduced activation of transcription. To test this hypothesis the region encompassing the *caf1R-caf1M* intergenic region subcloned into the two transcriptional fusion plasmids, pRS415 or pRS550 (Simons et al., 1987) were used to monitor promoter activity (Section 2.3). pRS*caf1M'*-*R'*-*lacZ* tests for *caf1R* promoter(s) and pRS*caf1R'*-*M'*-*lacZ* tests for the *caf1M* promoter as depicted in **Fig. 3.10a**. These plasmids carry a promoterless *lacZ* gene, hence the impact of mutated Caf1R was tested by complementation with pACYC-R encoded Caf1R_{E98G} and monitoring β-galactosidase activity (promoter activity indicator).

To create pACYC-R_{E98G} and the wild type control pACYC-R the DNA fragment containing the complete sequence for *caf1MA1* was deleted by inverse PCR from pACYCF1_{SPM} and pACYCF1 as template. This left only *caf1R_{E98G}* (or *caf1R*) and the *caf1R-caf1M* intergenic region. PCR was performed using *caf1MA1* deletion primers, pACYC*caf1Rf* and pACYC*caf1Rr* and Q5 HiFi DNA polymerase (NEB). The PCR products (5.9 kb) (**Fig. 3.10b**) were digested with SacI-HF restriction enzyme (NEB), gel purified and ligated with T4 DNA ligase (Thermo Scientific) followed by transformation into *E. coli* DH5α. Plasmid DNA was isolated from five individual transformants of each and analysed by AGE (data not shown). Based on AGE, a plasmid from a single transformant of each, pACYC-R_{E98G} (transformant 1) and pACYC-R (transformant 5) was digested with BamHI (unique site) to confirm the linear size (~5.9 kb) (**Fig. 3.10c**). In these pACYC-R constructs, *caf1R_{E98G}/caf1R* is under native control from the *caf1R-caf1M* intergenic region. Therefore, the complete sequence of *caf1R* and the *caf1R-caf1M* intergenic region was confirmed by sequencing using ACYCDuetUp1 and DuetDown1 sequencing primers. Both transformants were stocked and used for complementation studies.

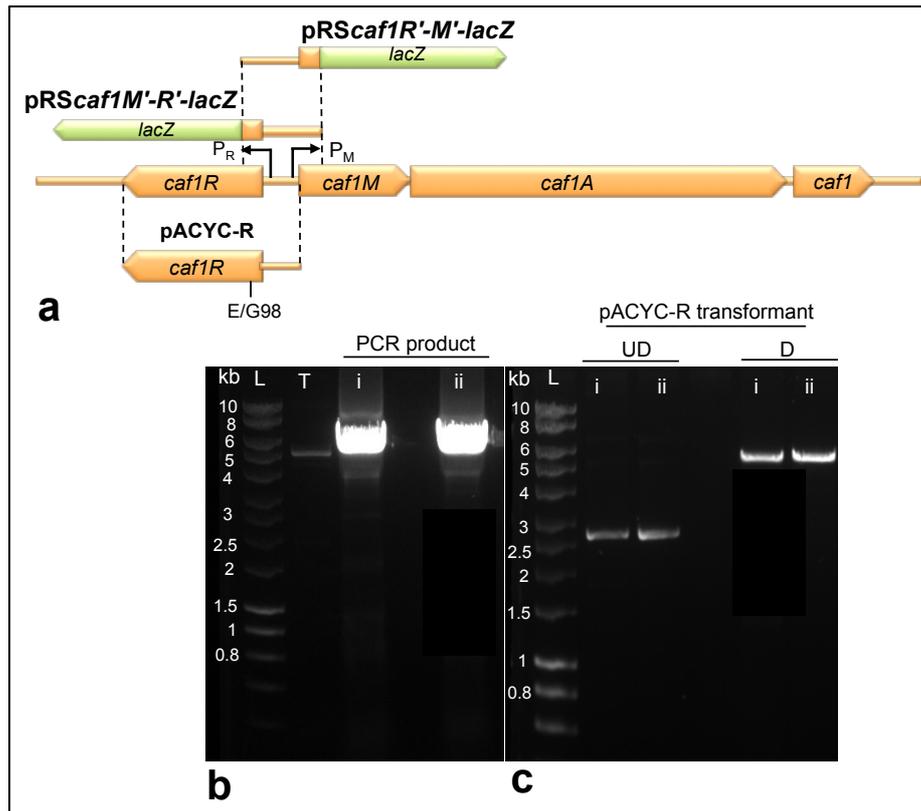


Figure 3.10| Strategy to compare impact of wild type and mutant (E98G) Caf1R on transcription.

a) Diagram depicting *caf* locus, promoter-*lacZ* fusion constructs, $pRScaf1M'-R'-lacZ$ (contains *caf1R* promoter/s) and $pRScaf1R'-M'-lacZ$ (contains *caf1M* promoter/s), and the complementing plasmids, $pACYC-R$ and $pACYC-R_{E98G}$, which encodes wild type and mutant (E98G) Caf1R, respectively. The *caf* locus regions indicated in between dotted lines were PCR amplified along with backbone plasmid $pACYC$, in case of $pACYC-R$ and $pACYC-R_{E98G}$ or subcloned in between *EcoRI/BamHI* sites at 5' UTR of promoter less *lacZ* gene in the transcriptional fusion plasmids, $pRS415$ and $pRS550$ (Simons et al., 1987) to construct $pRScaf1M'-R'-lacZ$ (contains – strand of *caf* locus) and $pRScaf1R'-M'-lacZ$ (contains + strand of *caf* locus), respectively. **b)** Construction of $pACYC-R$ and $pACYC-R_{E98G}$. AGE (0.8% agarose) of *SacI*-HF digested PCR products (1 μ g each) of $pACYC-R_{E98G}$ (i) and $pACYC-R$ (ii) along with one template (T), $pACYCF1$ (30 ng). **c)** Restriction digestion (D) analysis of (i) and (ii) with single cutter or unique *BamHI* enzyme along with undigested (50 ng) corresponding plasmid.

The impact on *caf1R* transcription was tested using co-transformants of *E. coli* Top10 carrying $pACYC-R_{E98G}+pRScaf1M'-R'-lacZ$ (Caf1R_{E98G} mutant), $pACYC-R+pRScaf1M'-R'-lacZ$ (WT Caf1R) or $pACYC+pRScaf1M'-R'-lacZ$ (no Caf1R; control). Promoter activity from the Caf1R_{E98G} co-transformants was only 1.2- (2 h), 1.9- (4 h) and 1.5- (6 h) fold increased over the control. In contrast, activity from the wild type co-transformants was enhanced 8.2- (2 h), 10.9- (4 h) and 6.5-(6 h) fold (**Fig. 3.11a**). This corresponded to an average of 5.6-fold reduction in Caf1R mediated activation at the *caf1R* promoter(s) in the presence of Caf1R_{E98G}. The impact on promoter activity upstream of *caf1M* was similarly studied using *E. coli* Top10 co-transformants carrying the $pRScaf1R'-M'-lacZ$ promoter fusion combined with $pACYC-R_{E98G}$, $pACYC-R$ or $pACYC$. Activity in the presence of Caf1R_{E98G} was 3.0- (2 h), 5.2- (4 h) and 9.8- (6 h) fold increased compared to the control; whereas, activity in the presence of wild type was enhanced 12.7- (2 h), 24.5- (4 h) and 40.0-fold (6 h) (**Fig. 3.11a**). Again this highlighted a significant reduction (4.3 fold average) in Caf1R-mediated promoter activation with the E98G mutation. Thus these results

clearly show a negative effect of mutated Caf1R_{E98G} on the activation of transcription at both *caf1R* and *caf1M* promoter(s). The impact of this mutation on the ability of Caf1R to bind at *caf1R* and *caf1M* promoter(s) is addressed in Chapter-5. In conclusion, the reduced level of F1 from pACYCF1_{SpM} can be explained by poor transcriptional activation by Caf1R_{E98G}.

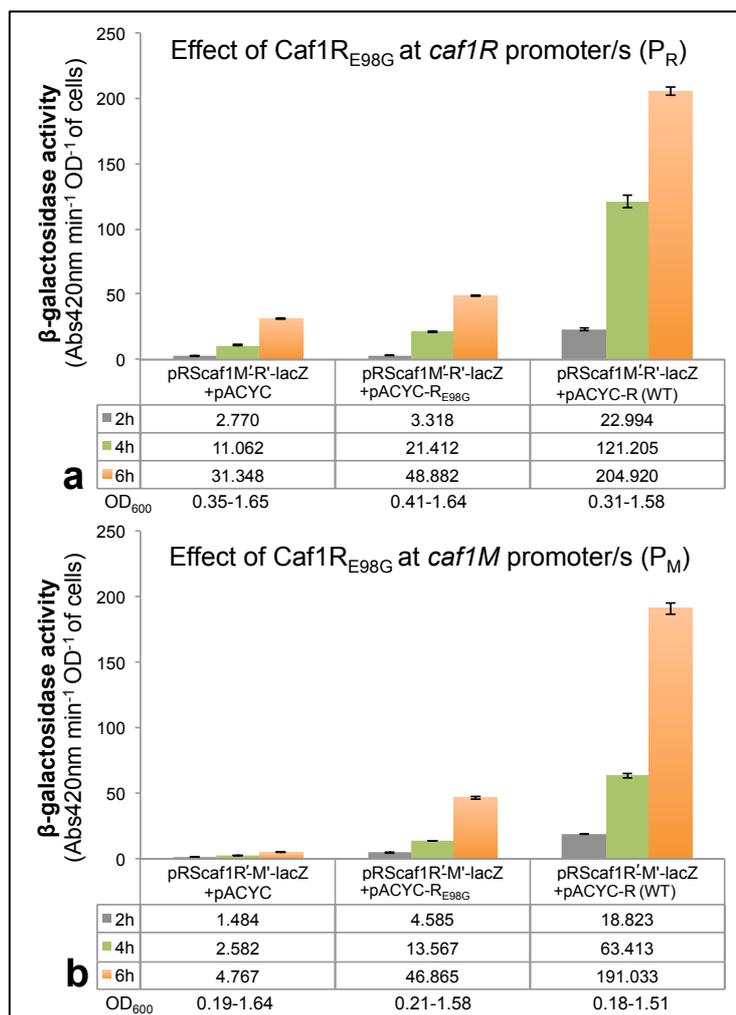


Figure 3.11| Caf1R_{E98G} lower the transcription activation at both *caf1R* and *caf1M* promoter(s).

a) Effect of mutant Caf1R_{E98G} of wild type on the activity of *caf1R* promoter/s (P_R). **b)** Effect of mutant Caf1R_{E98G} of wild type on the activity of *caf1M* promoter/s (P_M). Promoter or β -galactosidase activity from *E. coli* Top10 carrying the indicated combination of plasmids was monitored from cells recovered after 2, 4 and 6 h growth at 37°C along with respective control. Standard error on the activity bars shows \pm SEM of activity from three individual co-transformant of each. The range of OD₆₀₀ (2-6 h) from *E. coli* Top10 carrying respective combination of plasmids is indicated underneath activity values.

3.3 Bioinformatic analysis of Caf1R

3.3.1 Characterisation of Caf1R

Annotations of *Y. pestis* *caf* locus identify two possible start sites for Caf1R. One encodes a product of 301 amino acids (MWt 36.04 kDa), starting with ATG (Met) codon, having NCBI accession number, CAA43969.1; structural organization of *Y. pestis* virulence-associated plasmids and Caf1R (Hu et al., 1998; Karlyshev et al., 1992). The second predicted start site was annotated six codons upstream, starting at the rare start codon TTG (Leu or ^fMet) and encodes a product of 307 amino acids (MWt 36.84 kDa). This start site was used with NCBI accession

number CAB55263.1; genome sequence of *Y. pestis*, strain CO92 (Parkhill et al., 2001) and AAS58717.1, genome sequence of *Y. pestis*, biovar Microtus strain-91001 (Song et al., 2004). As there had been no experimental data confirming the functional start site of Caf1R and the larger open reading frame (ORF) was predicted in the published genome sequences, bioinformatic analysis was performed using this larger ORF (Swiss-Prot accession number, Q65AJ3). Initial analysis was carried out using the ExpASY-ProtParam tool (<http://web.expasy.org/protparam>) to determine relevant physicochemical properties. The theoretical pI of Caf1R, 9.17, indicated a highly basic protein. Furthermore, Ile and Tyr are the most frequent amino acids with 10.1 and 8.1% compositions, respectively. Total numbers of negatively charged amino acids (Asp+Glu) are 32 whereas the total numbers of positively charged amino acids (Arg+Lys) are 43; consistent with it being a positively charged protein at pH 7.0. The instability index was computed to be 50.34, which classifies Caf1R as an unstable protein. The PROSITE database (<http://expasy.org/prosite/>) entry, PS01124 of AraC/XylS family regulators confirms Caf1R inclusion within AraC/XylS family regulators as any proteins with a score of 12.52-30.74 are classified as a member of AraC/XylS regulator family and for Caf1R this score was 28.810. Analysis with Pfam tool (<http://pfam.xfam.org/protein/Q65AJ3>) shows that Caf1R is a two-domain protein, containing N-terminal DNA-binding domain (DBD) encompassing residues 1-112 and a C-terminal GyrI-like domain, encompassing residues, 131-270. The N-terminal position of the DBD is uncommon (Gallegos et al., 1997; Ibarra et al., 2008; Schuller et al., 2012) but Caf1R is predicted to have the two characteristic helix-turn-helix (HTH) motifs, encompassing residues, 31-53 (HTH1) and 80-103 (HTH2). The presence of a GyrI-like domain suggests the possibility that Caf1R may bind small molecule(s) like other regulators of AraC/XylS family such as the sugar metabolite binding of AraC (Schleif, 2010), RhaS and RhaR (Kolin et al., 2008), XylR (Ni et al., 2013), XylS (Gallegos et al., 1996) and fatty acid binding of ToxT (Lowden et al., 2010). Seven Cys residues (C33, C129, C137, C148, C164, C269 and C277) were identified in the entire chain of Caf1R, this is quite unusual as the analysed regulators belonging to AraC/XylS family contain much less Cys residues (**see Fig. 3.15**). For example, the DBD of Rob, MarA and SoxS don't contain any Cys residue (**see Fig. 3.17**). The N-terminal DBD of Caf1R contains a single Cys residue (C33) whereas the remaining 6 were located in the putative C-terminal GyrI-like domain.

3.3.2 Identification of Caf1R homologs

AraC/XylS family contains regulators controlling the expression of genes or gene cluster(s), involved in a diverse range of cellular processes such as carbon metabolism, stress responses and virulence in both Gram-negative and Gram-positive bacteria (Gallegos et al., 1997; Ibarra et al., 2008). In Gram-negative bacteria, many different surface molecules and secretion systems are critical to virulence, reviewed in (Costa et al., 2015) including fimbriae or pili (Busch and Waksman, 2012; Thanassi et al., 2012). A review of the literature (Sections 1.4.1-4; **Fig. 1.6**) had identified four different AraC/XylS regulators controlling CU pathways, AfrR (Cantey et al., 1999), AggR (Nataro et al., 1994), LdaA (Scaletsky et al., 2005) and Rns (Munson and Scott, 2000). Pairwise Blast of Caf1R with each of these four regulators gave E-values of 3e-61 (AfrR), 4e-06 (AggR), 1e-68 (LdaA) and 3e-06 (Rns). LdaA and AfrR shared 121/276 (44%) and 116/274 (42%) identical amino acids with Caf1R whereas AggR and Rns shared only 40/121 (41%) and 22/56 (39%) identical amino acids to Caf1R. This indicates that LdaA and AfrR are

likely to be relatively close homologs of Caf1R and that AggR and Rns are more distantly related. Hence, only LdaA and AfrR were considered in further sequence analysis with Caf1R.

In order to investigate the closest relatives of Caf1R, initially *Y. pestis* entries to NCBI were searched for homologs of full length Caf1R (Q65AJ3, the longer 307 aa predicted protein), using the BLASTp server (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) with an E-value threshold of 1e-3. All Caf1R regulator entries were 100% identical for full length Caf1R. After these the closest entry was annotated as Rob (Accession no WP_002209231.1, E-value 4e-25, 83/287 identity (29%) and 45% similarity). Hence, there was no protein closely related to Caf1R encoded within the *Y. pestis* genome or associated plasmids. The non-redundant (nr) protein database was then searched for all proteins related to Caf1R using the same parameters, but excluding *Y. pestis* entries. The BLASTp output for Caf1R showed that the closest homolog (WP_042110517.1) was from a whole genome sequence (WGS) project, contig-2016 of uropathogenic *E. coli* (UPEC), strain-126 (Salipante et al., 2015). This protein bears 61% identity over 87% of the Caf1R sequence with an E-value of 4e-105. The second homolog (WP_001697817.1) was also from a WGS project, contig-134 and assigned as a putative F1 regulator from avian pathogenic *E. coli* (APEC), strain-008 (Rojas et al., 2013). This exhibited 53% identity over 87% of the sequence and a slightly lower E-value (1e-79).

Analysis of the genomic context of the closest homolog (WP_042110517.1) revealed a divergently transcribed gene annotated as a carbohydrate transporter approximately 500 bp upstream of this Caf1R related regulator. Immediately upstream of this was a glycosyl transferase. There was nothing encoded immediately downstream of the regulator, before a small hypothetical protein followed by a gene encoding a putative plasmid stabilisation protein, and a CopG family transcriptional regulator (**Fig. 3.12a**). Thus, this closest homolog appeared to be linked to carbohydrate metabolism rather than a CU system. The second homolog (WP_001697817.1) was on a small contig of 900 bp and hence neighbouring genes could not be identified (**Fig. 3.12b**). A more detailed analysis of these homologs revealed that the N-terminal DBD of Caf1R showed a particularly close similarity to these proteins, 74% and 67% amino acid identity and 82% and 77% similarity with the first (WP_042110517.1) and the second (WP_001697817.1) homolog, respectively.

The third top homolog (WP_024182496.1) was also from a WGS project, contig-11_1 of diarrheagenic *E. coli*, strain-DEC11A (Hazen et al., 2012). This exhibited 44% identity over 89% sequence coverage to Caf1R with an E-value of 2e-68. The genomic context of this homolog showed an associated CU system with four fimbrial subunits, three unassigned and one annotated as K88-like minor fimbrial subunit, an outer membrane usher, annotated as FaeF and a periplasmic chaperone, annotated as ClpE (**Fig. 3.13a**). Aside from the associated CU system, this homolog is surrounded by a few hypothetical proteins, one transposase, one recombinase and one additional transcriptional regulator of unknown family and a repFib replication protein A (**Fig. 3.13a**). After excluding identical entries, the genomic context of the remaining entries within the top 500 homologs were analysed. Twelve homologs of Caf1R, including LdaA and AfrR, associated with ORFs belonging to a complete or partial CU system, were identified. LdaA (AAX78184.1) was identified as 7th top homolog and AfrR (WP_032489301.1) was present at

position 35. Sequence entries between the 3rd and 7th closest homolog and many others were hypothetical or unassigned regulators of AraC/XylS family on short contigs with a single ORF that could not be analysed.

The sequences of all 12 CU associated regulators were retrieved and arranged on a neighbor-joining phylogenetic tree (based on % identity) along with AggR and Rns, and some non CU-type regulators, which showed $\geq 30\%$ amino acids identity to Caf1R over $\geq 85\%$ sequence coverage and an E. value cut off, $\leq 6e-32$ (**Fig. 3.14**). A multiple sequence alignment was generated from the 12 CU-type regulator sequences along with sequences of the closest two (non-CU) homologs (**Fig. 3.15**). It was observed that the putative DBD of all regulators is highly conserved with an equal distribution of hydrophobic and charged residues, which would be consistent with a similar structure. Within the HTHs motifs, the sequence conservation was more pronounced; HTH1 has seven and HTH2 contains eight identical or highly similar amino acids. The highly conserved GxSSR in HTH1 and DSxQ and RE*FxK in HTH2 (x, variant residue; E*, Glu98) (**Fig. 3.15**), suggest that these residues may be essential for DNA-protein interactions. In addition, this includes Glu98 the mutated residue in the poorly functioning Caf1R regulator. This may indicate a similar mode of DNA binding.

The sequence identity of these regulators over the GyrI-like domain of Caf1R was lower. The significance of this domain to regulator function has not been addressed in any of these systems. It remains to be seen if this C-terminal domain is involved in sensor binding or dimerisation. The connecting region between the putative DBD and GyrI-like domain (113-130 aa) was highly variable, consistent with a linker segment although unique properties of each regulator should not be excluded. Conclusively, 12 CU-linked AraC-like regulators were identified all sharing a high degree of similarity with Caf1R although none of them seem to be linked to CU systems from FGL class (or Y3 family) (Nuccio and Baumler, 2007; Zav'yalov et al., 2010; Zavialov et al., 2007) to which the *caf* locus belongs. The high degree of similarity indicates a common underlying regulatory mechanism among these regulators.

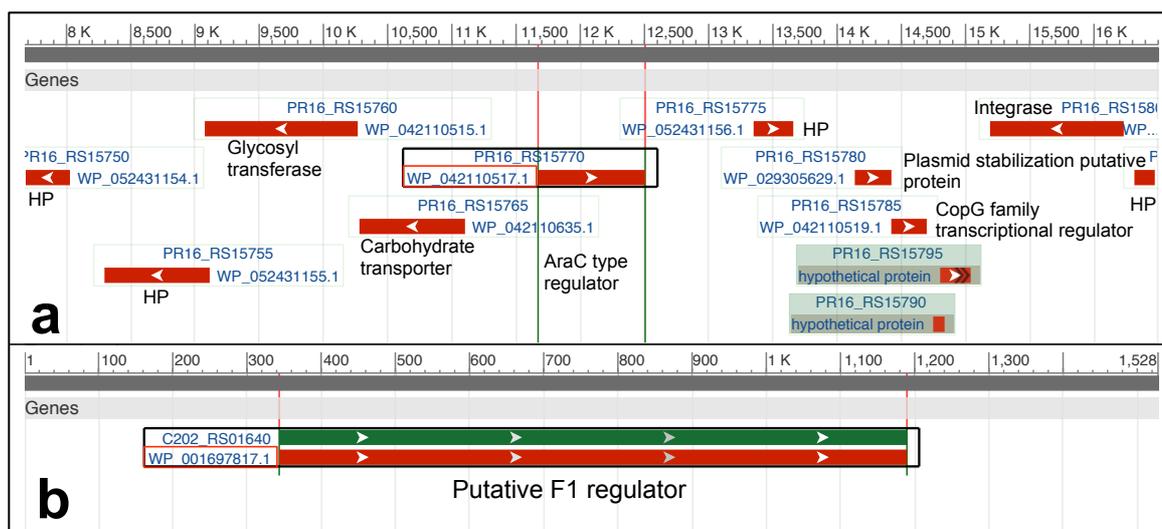


Figure 3.12| Closest homologue of Caf1R is not associated with CU system.

Genome context of closest 2 homologs of Caf1R. a) Accession no WP_042110517.1 (closest NCBI entry) and b) Accession no WP_001697817.1 (second closest NCBI entry). HP, hypothetical protein.

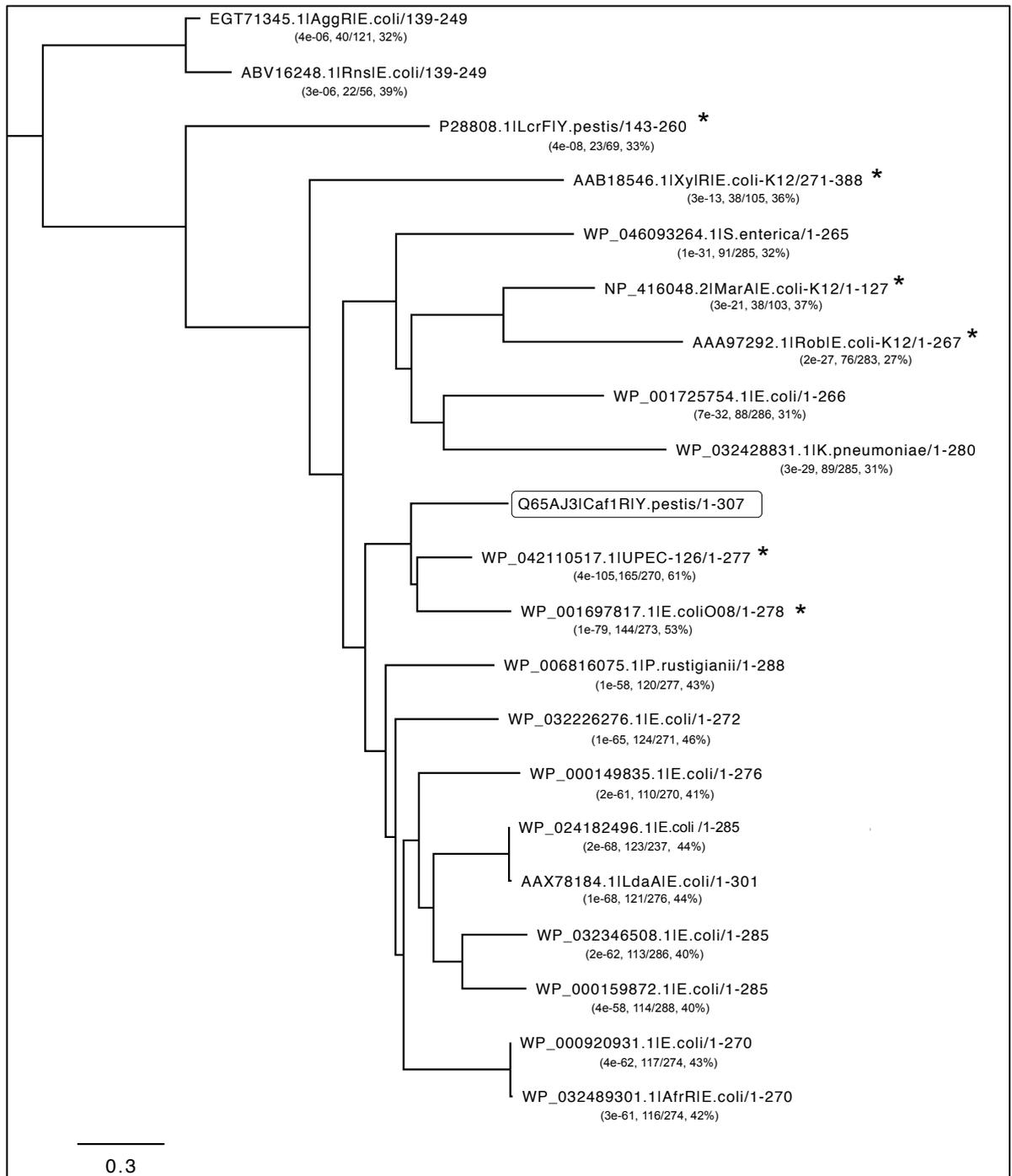


Figure 3.14| **Phylogenetic tree of Caf1R homologs from AraC/XylS family and CU-type regulators.**

NCBI accession number of each is indicated followed by regulator (if identified), name of organism and the sequence limit. Caf1R is boxed. The E-value, query coverage and amino acid identity % to complete Caf1R for each are shown in parentheses. Non CU-type AraC/XylS family regulators are indicated by *. All others are associated with a CU system. Phylogenetic distance (0.3) is indicated by bar. APEC: Avian pathogenic *E. coli* and UPEC: uropathogenic *E. coli*. NCBI tree viewer1.6 program was used to construct phylogenetic tree from the multiple sequence alignment (by NCBI-COBALT program) of the retrieved sequences. Neighbor joining method with default parameters was used to construct tree.

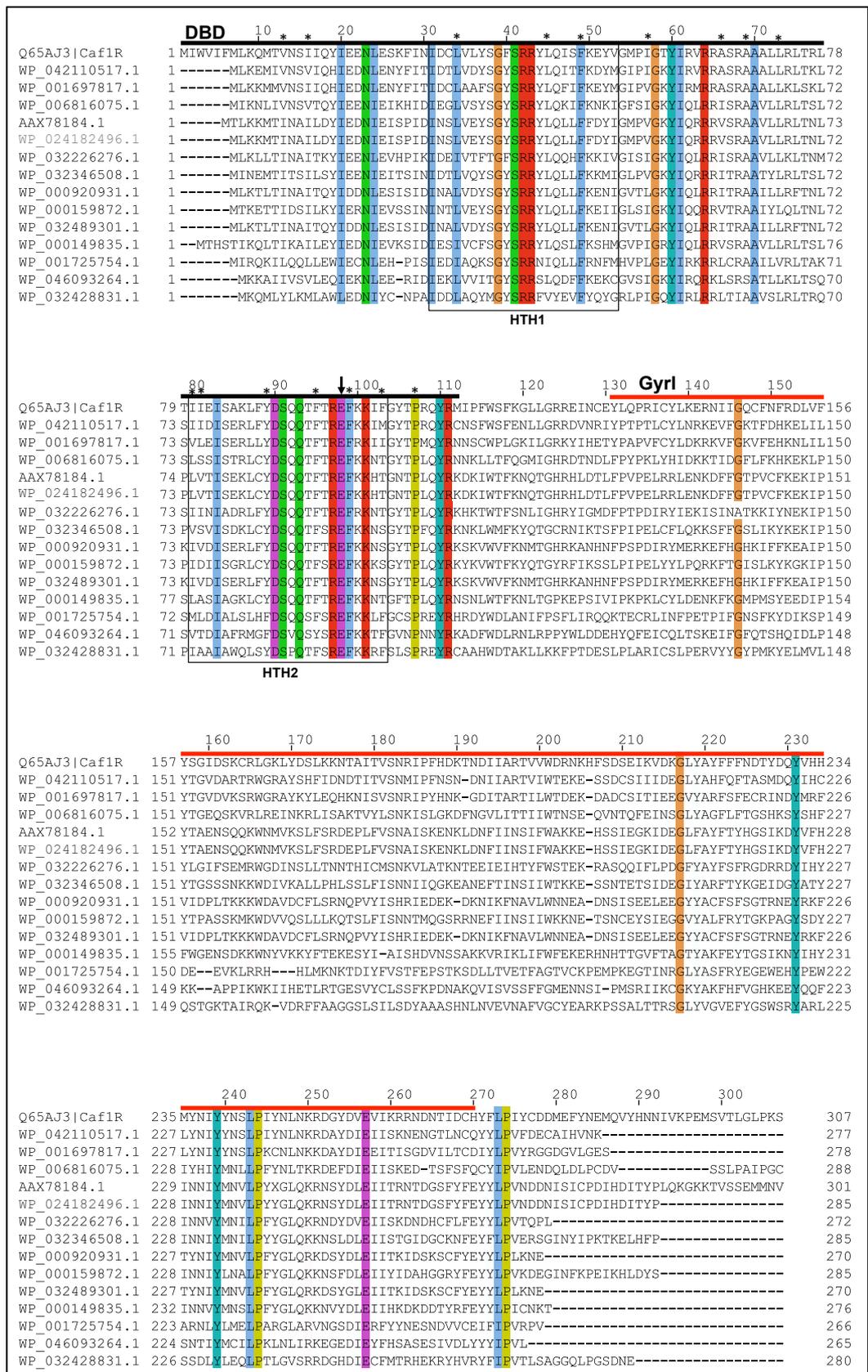


Figure 3.15| Multiple sequence alignment of selected Caf1R homologs.

Sequences with respective NCBI accession number are organised according to similarity to Caf1R with a range of 61% (top) to 31% (bottom) amino acid identity to complete Caf1R (1-307 aa) and 74-44% to N-terminal DBD (1-112 aa). Black line, location of predicated DBD; red line, Gyrl-like domain. Putative helix-turn-helix (HTH) motifs 1 and 2 are boxed. The conserved E98 residue is indicated with an arrow. Residues highlighted in different colors represent features of absolutely conserved (100% identical) and similar amino acid according to ClustalX color scheme (Livingstone and Barton, 1993) I, L, A, F: Hydrophobic; N, S, Q: Polar; G: Hydrophobic and small; R, K: Polar and positive charged; Y: Polar, hydrophobic and aromatic; P: Pro (small and nonpolar)

and D, E: Polar and negatively charged. (*), indicates residues assumed to contribute to the hydrophobic core of DBD, based on *mar*-MarA complex (PBD-1bl0). Similarity percentage % corresponds to complete Caf1R and N-terminal DBD alone is indicated in parentheses following description of each NCBI accession number. WP_042110517.1, uncharacterised regulator from uropathogenic *E. coli*, strain-126 (65/82); WP_001697817.1, putative F1 regulator from avian pathogenic *E. coli*, strain-O08 (58/78); WP_006816075.1, uncharacterised regulator from *P. rustigianii* (53/71); AAX78184.1, LdaA, regulator of locus for diffuse adherence of enteropathogenic *E. coli* (52/71); WP_024182496.1, uncharacterised regulator from Diarrheagenic *E. coli*, strain-DEC11A (52/71); WP_032226276.1, uncharacterised regulator from *E. coli* (51/71); WP_032346508.1, uncharacterised regulator from *E. coli* (48/69); WP_000920931.1, uncharacterised regulator from *E. coli* (50/72); WP_000159872.1, uncharacterised regulator from *E. coli* (49/65); WP_032489301.1, AfrR regulator from *E. coli* (49/72); WP_000149835.1, uncharacterised regulator from *E. coli* (48/69); WP_0001725754.1, uncharacterised regulator from *E. coli* (41/65); WP_046093264.1, uncharacterised regulator from *S. enterica* (40/61); WP_032428831.1, uncharacterised regulator from *K. pneumoniae* (38/55). Jalview2.9 (Waterhouse et al., 2009) was used to generate multiple sequence alignment and SIAS web server (<http://imed.med.ucm.es/Tools/sias.html>) used to calculate similarity (%) over complete and N-terminal DBD alone.

3.3.3 Identification of Caf1R homologs within PDB database

The Caf1R accession number (Q65AJ3) was analysed against the protein data bank (PDB) database to identify the closest AraC/XylS regulators that have been characterised structurally. Eleven proteins were found to have a 3D fold to which Caf1R might be similar. Description of each along with their accession number is shown in Table 3.1. Out of 11, eight of these proteins belong to the AraC/XylS regulator family and the remaining three are two-component response regulators. Rob protein from *E. coli*-K12 gave the highest score (229, E-value, 1.7e-21) followed by *E. coli* MarA (176, E-value, 2.3e-15). Although the amino acid identity of the entire Caf1R with Rob was only 27.5%, the two proteins (Caf1R and Rob) shared 36.6% identity within N-terminal domains, slightly lower than the identity with MarA, which was 37.3%. MarA contains only the DBD. This can explain why the identity with MarA was highest, while the highest score was with Rob, which like Caf1R has a C-terminal domain. The score from XylR, AraC and other regulators was much lower than of Rob or MarA, suggesting that structures of Rob and/or MarA are the best on which to model the Caf1R 3D structure.

Caf1R secondary structure elements (α -helix, β -strand and turn) were predicted (**Fig. 3.16**) by PSIPRED (McGuffin et al., 2000) and added above the alignment. Ten α -helices and 9 β -strands were predicted in the complete Caf1R. Out of 10 α -helices, 7 were present in the N-terminus (1-112 aa) and the remaining 3 were located in the C-terminus (113-307 aa). The β -strands (9) were predicted only in the C-terminus (**Fig. 3.16**). The result from this secondary structure prediction suggested that Caf1R has a typical AraC/XylS type regulator fold with 6 α -helices in the N-terminal DBD, which are organised into two HTH motifs (α 2-T- α 3) and (α 5-T- α 6) connected by a long linker (helix-4), and a pocket of antiparallel β -strands in the C-terminal domain, essential for either sensing or dimerisation. Despite these common themes among AraC/XylS-like regulators, the domain location within Caf1R (like Rob) is in the reverse orientation when compared to most AraC/XylS regulators (Gallegos et al., 1997). However, this N-terminal DBD is common to all CU-type regulators of AraC/XylS family, analysed above except AggR and Rns, where the DBD is C-terminal.

Table 3.1| **Predicted structural homologs of Caf1R identified in PDB database.**

Dark color shaded regulators represent AraC/XylS family regulators whereas histidine kinases/two-component response regulators are indicated by light color shade.

UniProt Entry	Protein name	Description	Score
P0ACI0	Rob	Right origin-binding protein from <i>E. coli</i> -K12	E-value: 1.7E-21 Score: 229 Identity: 27.5%
P0ACH5	MarA	Multiple antibiotic resistance protein MarA from <i>E. coli</i> -K12	E-value: 2.3E-15 Score: 176 Identity: 37.3%
Q89YQ8	Histidine Kinase	From <i>B. thetaiotaomicron</i>	E-value: 1.4E-9 Score: 143 Identity: 31.6%
A9KJQ6	AraC type regulator	From <i>C. phytofermentans</i>	E-value: 2.4E-9 Score: 138 Identity: 34.3%
P0ACI3	XylR	Xylose operon regulatory protein from <i>E. coli</i> -K12	E-value: 22E-9 Score: 132 Identity: 36.2%
Q8RGT8	YesN	Two-component response regulator from <i>F. nucleatum</i>	E-value: 150E-9 Score: 124 Identity: 32.6%
Q7NTG7	AraC type regulator	From <i>C. violaceum</i>	E-value: 190E-9 Score: 124 Identity: 30.4%
Q88H39	AraC type regulator	From <i>P. putida</i>	E-value: 0.0000026 Score: 115 Identity: 28.7%
P0A9E0	AraC	Arabinose operon regulatory protein from <i>E.coli</i> -K12	E-value: 0.0000030 Score: 114 Identity: 28.3%
A9KIW7	AraC type regulator	Two component regulator from <i>C. phytofermentans</i>	E-value: 0.0000034 Score: 115 Identity: 33.0%
Q89YR8	Histidine kinase	Two-component sensor histidine kinase/response from <i>B. thetaiotaomicron</i>	E-value: 0.000013 Score: 111 Identity: 29.7%

To fully understand the function of AraC/XylS type regulators, characterisation of DNA-protein interactions is essential and in this regard only Rob and MarA have been characterised with their cognate DNA (Kwon et al., 2000; Rhee et al., 1998). Therefore in further analysis, the remaining nine proteins were excluded. The amino acid sequences of MarA, the N-terminal DBD of Rob and the N-terminal DBD of Caf1R, were aligned along with one other well-characterised MarA homologue, SoxS (Griffith and Wolf, 2002) to establish a structure-based functional relationship for Caf1R (**Fig. 3.17**). Following multiple sequence alignment, Caf1R residues corresponding to residues of *mar*-MarA DNA-protein complex (PDB-1bl0) involved in DNA-protein interactions were identified (Rhee et al., 1998). Within HTH1 motif (α 2-T- α 3), five amino acids of Caf1R, G39, S41, L45, Q46 and F49 shared identity with MarA, Rob and SoxS whereas in HTH2 motif (α 5-T- α 6), eight amino acids, I80, S91, Q92, Q93, T94, R97, F100 and F103 shared identity to MarA, Rob and SoxS. The higher amino acid conservation of HTH2 has been taken as an indication of common DNA binding features within AraC/XylS regulators and the greater variation in HTH1 as a possible indication of specificity of DNA-protein interaction (Gallegos et al., 1997).

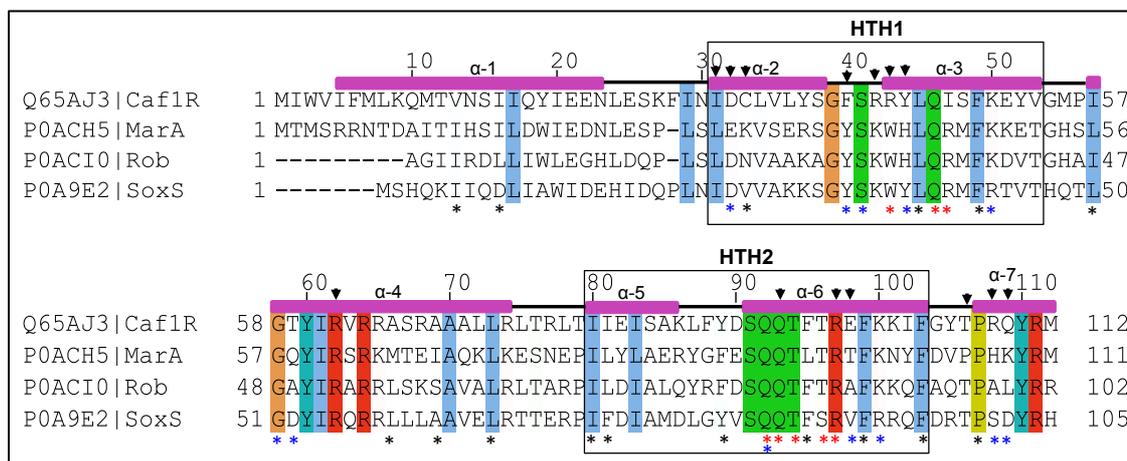


Figure 3.17| Multiple sequence alignment of Caf1R DBD with MarA, Rob and SoxS DBD.

Secondary structure elements, α -helices (1-7; pink rods) and turn (black lines) indicated above Caf1R were predicted by PSIPRED (McGuffin et al., 2000). Putative Helix-turn-helix (HTH) motifs (HTH1, 31-53 aa and HTH2, 80-103 aa) assigned according to *mar*-MarA complex (Rhee et al., 1998) are indicated by black boxes. Alignment numbering is relative to Caf1R sequence. Colored asterisks below alignment indicate the assigned role of residues in *mar*-MarA complex (PDB-1bl0). *: Hydrophobic core residues; *: residues interacting with phosphate group of DNA backbone and *: residues interacting with DNA bases (see Chapter 1, section 1.5.2 for further detail). Jalview2.9 program (Waterhouse et al., 2009) was used for multiple sequence alignment with ClustalX color scheme (Livingstone and Barton, 1993). Residues highlighted in different colors represent features of absolutely conserved (100% identical) and similar amino acids. I, L, A, F: Hydrophobic; S, Q, T: Polar; G: Hydrophobic and small; R: Polar and positive charged; P: Pro (small and nonpolar) and Y: Polar, hydrophobic and aromatic. Arrowheads, location of site-specific mutations within Caf1R-DBD (Section 3.4 and onward).

3.3.4 Modelling of Caf1R

In order to identify the location and influence of Glu98 in the Caf1R structure, the structure of Caf1R was modelled. A working model of Caf1R would also assist in identifying key residues within the DBD, essential for activation of transcription. The amino acid sequence of Caf1R (relative to TTG start codon) was submitted to a protein-modelling server, IntFOLD3.0 (McGuffin et al., 2015; Roche et al., 2011) that uses template based method from the crystal structures in PDB database, representing the correct folds. Following submission of Caf1R to IntFold3.0, 103 models were predicted and almost all of them were based on either Rob (PDB-1d5y) or MarA (PDB-1b10) or both protein templates. This is supported by Caf1R analysis against the PDB database (discussed earlier). The highest scored model was based on the Rob template with $\geq 70\%$ non-redundancy (**Fig. 3.18a**). The reliability of this model can be judged by the score of modeling-confidence, which was categorised as 'certain' with a P-value (probability of incorrectness) of 1.921E-4 and an overall global model quality score of 0.8133/1.0. The high reliability of the model is also indicated by the excellent β -factor represented by color intensity from blue (high accuracy) to red (low accuracy) (**Fig. 3.18a**). The N-terminal domain (DBD) is apparently a well-predicted model; this was also reflected in the low level of intrinsic disorder and residue error (in Å). The prediction for the C-terminal domain was also good, aside from β -strands 4 and 5 (residues 173-206) which were slightly less reliable and the extreme C-terminus (residues 278-307), which was unstructured. This can be seen by the much higher peaks and histogram bars for the C-terminus, compared to the N-terminus in the DISCOClust disorder prediction plot and the per residue error plot (**Fig. 3.18d-e**).

One advantage of the IntFOLD3.0 server is that it also predicts ligand(s) associated with the submitted protein and potential binding site(s) within both protein and the associated ligand. Not surprisingly, the 2 DNA ligands from the solved cocrystal structures of *mar*-MarA (Rhee et al., 1998) and *micF*-Rob (Kwon et al., 2000) DNA-protein complexes were modelled as ligands of Caf1R. However, the *micF* DNA was bound to the modelled Caf1R DBD in only one major DNA groove *via* Caf1R helix-3 (**Fig. 3.18b**) and therefore *micF*-Caf1R model was not used further. In contrast, modeling with *mar* DNA from *mar*-MarA complex, generated the expected interaction of regulator with two major grooves within the DNA. In the *mar* DNA-Caf1R model, helix-3 of Caf1R interacts with one major groove and helix-6 with the other (**Fig. 3.18c**). This requires bending of the DNA, as shown for MarA binding to *mar* promoter (Rhee et al., 1998). Three Caf1R residues were predicted to interact with the bound *mar* DNA ligand; R43 in helix-3 interacting with a G-base, and Q93 and R97 (both in helix-6) interacting with a T-base (**Fig. 3.18b-c**). Residues in the positions corresponding to Caf1R- R43 and R97 are known to be critical in DNA binding of MarA (Rhee et al., 1998), Rob (Kwon et al., 2000) and SoxS (Griffith and Wolf, 2002). Therefore, R43, Q93 and R97 might reasonably be considered as key DNA binding residues of Caf1R. Because the model with *mar* DNA is consistent with current evidence of interaction of regulators with 2 binding sites, BS1 and BS2 adjacent to HTH1 and HTH2, respectively, the *mar* DNA-Caf1R model was used as a working model for analysis of Caf1R-DNA interactions and designing a mutagenesis strategy.

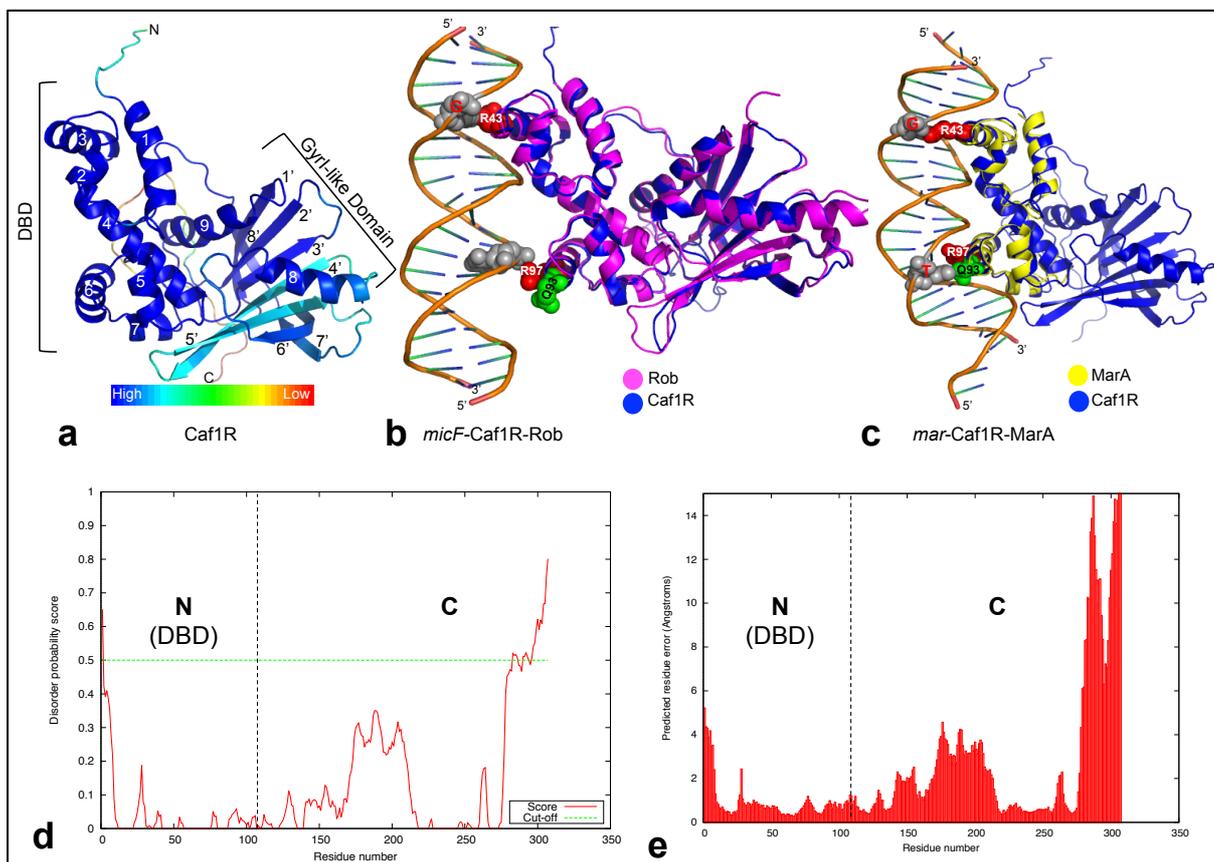


Figure 3.18| Caf1R modelling.

a) Ribbon diagram of IntFOLD3.0 predicted model of full-length Caf1R, indicating quality of prediction by β -factor color intensity. Predicted N-terminal DNA-binding domain (DBD) (helices, 1-7) and C-terminal GyrI-like domain (helices (8-9) and β -strands (1'-8')) are annotated. **b-c)** Caf1R-DNA interaction with *micF* (**b**) and *mar* (**c**) DNA indicating modelled interactions of R43 (in helix-3) with a G-base in both cases and Q93 and R97 (in helix-6) with a T-base only with *mar* DNA. Caf1R model with *micF* DNA indicates helix-6 interaction with DNA backbone adjacent to a T-base. The retrieved *micF* and *mar* DNA-Caf1R models (from lig2.pdb file) were superposed by Tm align web server (Zhang and Skolnick, 2005) on the corresponding protein template (Rob and MarA) in order to show complementarity between modelled Caf1R and template. **d)** Disorder prediction and **e)** per-residue error in Å of the selected top scored Caf1R model based on Rob template. The default cut-off value (0.5) of good quality prediction is indicated by green dotted line in (**d**). A vertical black dotted line in (**d**) and (**e**) indicates division between N and C-terminal domains of Caf1R. MacPyMOL1.3 was utilised to visualise and generate the ribbon diagrams of Caf1R and Caf1R-DNA models.

3.4 Mutagenesis supports role of Glu98 in bridge interaction

3.4.1 Impact of E/G98 residues on Caf1R structure

The repair of the spontaneous E98G mutation within Caf1R DBD identified E98 as a potentially key residue, essential for high levels of F1 production, but the question remained why was this residue so important? How and where does it interact, in a DNA-protein or protein-protein interaction? This Glu98 is highly conserved in all of the analysed CU-type regulators (**Fig. 3.15**) whereas in comparison to Rob, MarA and SoxS, this position is highly variable (**Fig. 3.17**). Therefore the modelled *mar* DNA-Caf1R DBD complex was analysed with respect to this residue. Glu98 appeared to be involved in a bridge-type interaction (**Fig. 3.19a**) with R62 (in helix-4) on one side, which in turn interacts with the DNA backbone. On the other side E98 is located on helix-6 neighbouring R97, which is highly conserved in all of these AraC/XylS regulators (**Figs. 3.15 and 3.17**). In the *mar* DNA-Caf1R model R97 retains the interaction with a T-base in the DNA major groove, adjacent to helix-6. This interaction appears to be stabilized by interaction of

Q93 (also in helix-6) with the same T-base in the major DNA groove. Thus, this bridge-type interaction (DNA backbone::R62-E98-R97-Q93::T-base), could play a critical role in stabilising the overall Caf1R-DNA interactions and also in correctly orientating helix-6 and the predicted DNA specific binding residues R97 and Q93.

In order to address the impact of the G98 mutation on the bridge interaction, the amino acid sequence of Caf1R_{E98G} was submitted to IntFOLD3.0 and the top scored model retrieved for analysis. Caf1R_{E98G} was modelled on two templates, *micF*-Rob complex (PDB-1d5y) (Kwon et al., 2000) and *mar*-MarA-RNAP ternary complex (PDB-1xs9) (Dangi et al., 2004). With E98G substitution, Intfold3.0 predicted three residues, R43, Y44 and R97, which might be essential for DNA-Caf1R interactions. Selection of the additional residue, Y44 as a binding residue could be related to the fact that the Caf1R_{E98G} model is based on two protein templates unlike the single protein template (Rob) on which the wild type Caf1R model was based. This extra residue, Y44 of helix-3 was predicted with a possible interaction with a G-base of DNA. Superposition of the modelled Caf1R_{E98G} structure on the wild type model of Caf1R indicated no change in position of helices even between DNA recognition helix-3 and helix-6 in either case. Although slight changes in the side chains of the respective amino acids were observed in both cases, for example R43 and Y44 (**Fig. 3.19b-c**). The DNA-Caf1R interactions were compared by superposing Caf1R_{E98G} model on the *mar* DNA-Caf1R modelled complex. As would be expected, the bridge-type interaction was disrupted upon substitution of E98 with G98 (**Fig. 3.19b**). The small size of the Gly residue and absence of side chain abolishes the interaction with R62 in the *mar* DNA-Caf1R_{E98G} modelled complex. The interaction of R97 and Q93 with predicted T-base was unaffected in the model, however, the small size and conformational flexibility of Gly means that it generally has a low propensity for α -helix formation. Therefore, using PSIPRED web server (McGuffin et al., 2000) the secondary structure of Caf1R helix-6 was predicted with E98 (wild type) and E98G substitution (**Fig. 3.19b-c**). The E98G substitution gave a substantially lower confidence of α -helix formation and predicted a slightly shorter helix-6 (**Fig. 3.19b-c**). Hence, an impact of the Gly98 substitution on helix-6 length and stability might also directly affect correct spatial orientation of binding residues within helix-6 as well as through loss of the R62 backbone interaction.

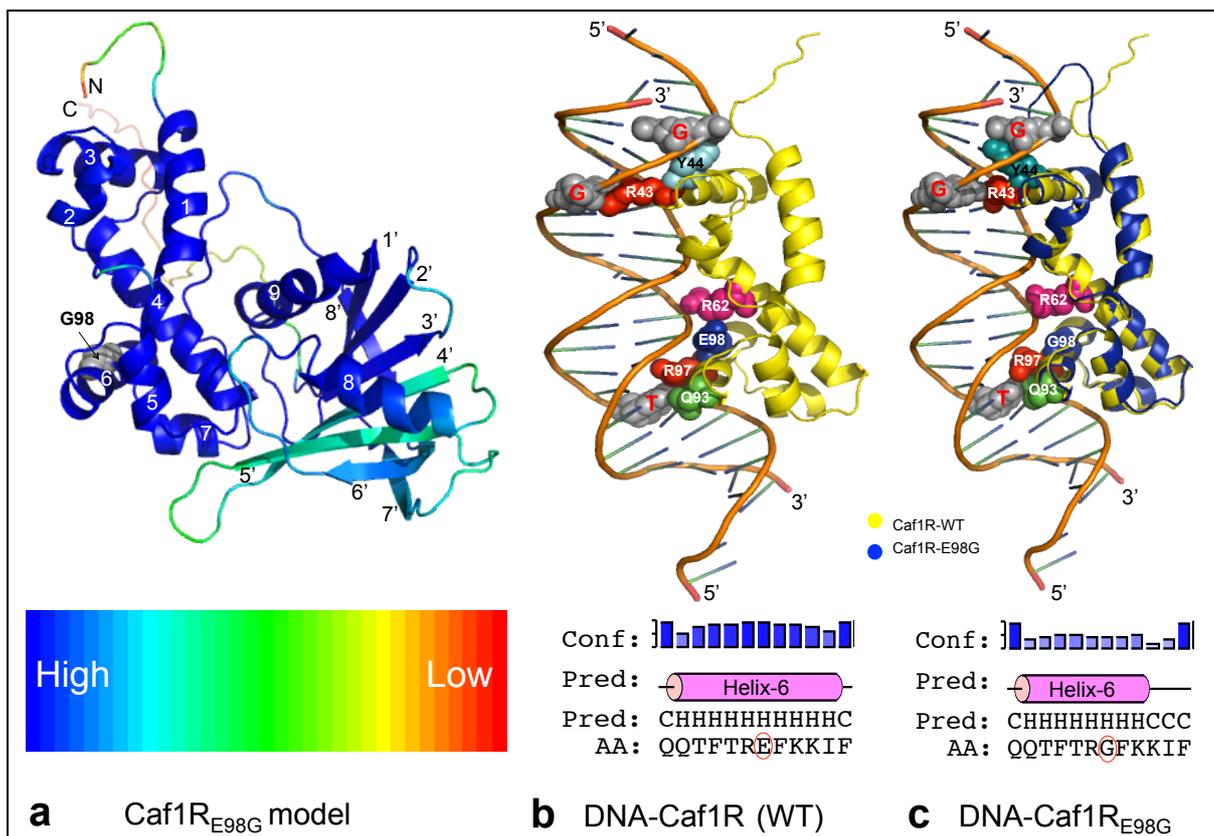


Figure 3.19 | **Impact of E/G98 residues on Caf1R structure.**

a) Ribbon diagrams of the predicted Caf1R model with G98 substitution (grey-balls). Quality of prediction is indicated by β -factor color intensity (underneath). Model quality score was 0.806 with P-value of 2.054E-4 and ‘certain’ reliability score. Modelled *mar* DNA-Caf1R (WT) (**b**) and *mar* DNA-Caf1R_{E98G} (**c**) indicating interactions of Intfold3.0 predicted (R43, Y44, Q93 and R97) and visually identified (R62 and E/G98) residues of Caf1R to DNA-bases (G, G and T). To generate the *mar* DNA-Caf1R_{E98G} complex, Tm align web server (Zhang and Skolnick, 2005) was used to superpose modelled structure of Caf1R_{E98G} on WT modelled complex of *mar* DNA-Caf1R. For clarity, only the N-terminal DBD is shown in both cases. Using PSIPRED (McGuffin et al., 2000) secondary structure was predicted for helix-6 with E/G98 substitution as indicated. MacPyMOL1.3 was utilised to visualize and generate the pictures.

3.4.2 E98 bridges critical DNA binding residues, R62, Q93 and R97

As reviewed in section 1.5.2 and **Fig. 1.11**, the corresponding residues of Caf1R- Q93 and R97 in MarA (Q92 and R96) make extensive van der Waals interactions and/or hydrogen bonds with the *mar* DNA bases in the cocrystal structure of *mar*-MarA complex. More so, alanine substitution of MarA-R96 showed a major impact on all six MarA-regulated promoters tested, although substitution of MarA-Q92 had less of an effect, only two of six MarA-regulated promoters were affected (Gillette et al., 2000). From the modelled *mar* DNA-Caf1R complex, Caf1R- R62, Q93 and R97 appeared to be critical residues, linked together by a bridge-type interaction through E98. Therefore, it was predicted that mutation of Q93 and R97 would have a major impact on Caf1R activity and that substitutions of the bridge related residues R62 and E98 would destabilise the Caf1R-DNA interaction and regulatory impact regulator activity.

Initially, Caf1R-R62 was substituted *in silico* to the smaller and non-polar or polar residues, Ala and Ser whereas Q93 and R97 were substituted only to Ala and Caf1R was again modelled with each mutation. Each of these models was superposed on the *mar* DNA-Caf1R complex model to analyse the impact on the bridge-type interaction and interaction of the key helix-6 residues with DNA. Each substitution was found to be responsible for disruption of the bridge-type interaction and would be predicted to destabilise the Caf1R-DNA interaction (**Fig 3.20**).

The impact of each of the R62, Q93 and R97 substitutions on activation of F1 production was assessed following construction of site-specific mutations in pACYCF1. F1 expression was quantified either from cell surface extracts (pACYCF1-R_{R62A} and pACYCF1-R_{R62S}) or directly from whole cells (pACYCF1-R_{Q93A} and pACYCF1-R_{R97A}) from three transformants of each following thermoinduction at 37°C (**Fig 3.20**). Following 4 h induction, cells carrying the mutated plasmids, pACYCF1-R_{R62A} or pACYCF1-R_{R62S} produced only 76.35% ± 3.86 and 58.85% ± 0.20 the amount of F1 produced with wild type Caf1R (**Fig. 3.20a (i-ii)**). At 6 h post-induction, the level was comparable to that of wild type, 96.55% ± 1.27 from pACYCF1-R_{R62A} and 92.97% ± 2.94 from pACYCF1-R_{R62S} (**Fig. 3.20a (i-ii)**). This reduction in F1 expression from both mutants at 4 h is consistent with the modelled loss of interaction of residue A/S62 with E98. The model also indicated loss of interaction of A/S62 with the phosphate backbone. Both effects could contribute to decreased activation of transcription by Caf1R (**Fig. 3.20a (iii-iv)**). With the mutation R_{Q93A}, the F1 level was very low at both time points tested, 6.99% ± 0.36 (4 h) and 33.90% ± 1.77 (6 h) of wild type levels (**Fig. 3.20b (i-ii)**). Similarly, the level of F1 from pACYCF1-R_{R97A} mutant was again drastically reduced at 4 h (4.79% ± 1.83) although there was more at 6 h, 86.53% ± 4.78 of wild type levels (**Fig. 3.20b (i-ii)**). The dramatically enhanced level of F1 at 6 h from pACYCF1-R_{R97A} mutant appears unusual in light of the critical role of this conserved Arg in MarA (R96) (Gillette et al., 2000; Rhee et al., 1998). The reduced level of F1 from each of these mutants (including R97A at 4 h) supports the involvement of these key residues in optimising Caf1R-DNA interactions.

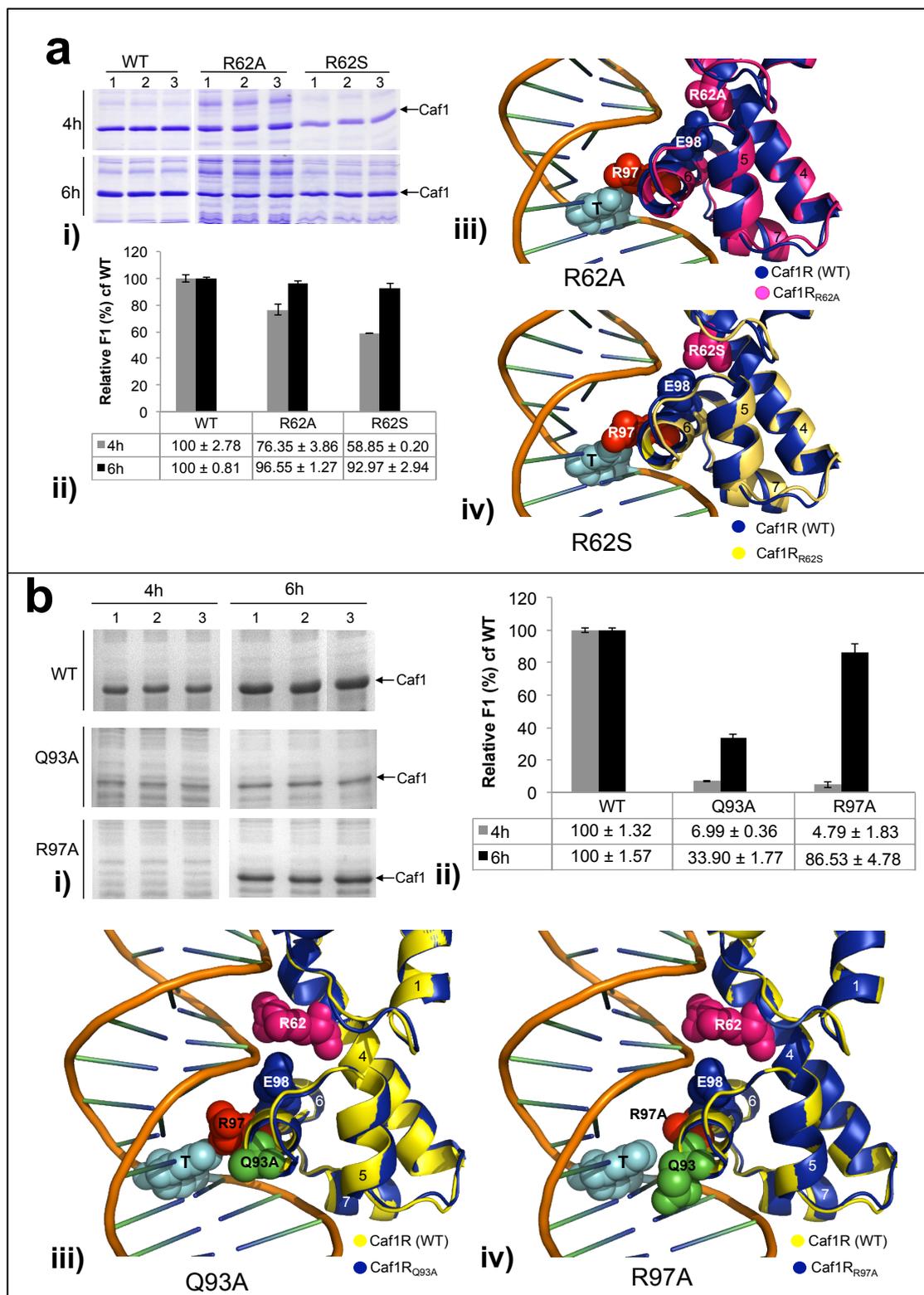


Figure 3.20| Caf1R- R62, Q93 and R97 are essential for bridge-type interaction and F1 production.

a) Effect of R62A and R62S substitution on F1 recovery and E98-mediated bridge-type interaction. **(i-ii)** SDS-PAGE (16% acrylamide) analysis **(i)** and F1 subunit, Caf1 band (←) quantification **(ii)** following F1 recovery from the ‘cell surface’ of recombinant DH5 α cells carrying pACYCF1-R_{62A/S} mutant plasmids (R62A or R62S) or wild type (WT), pACYCF1. F1 expression was monitored after 4 and 6 h of thermoinduction at 37°C and an equivalent of 0.04 OD unit sample of each was applied to gel following heat denaturation (97.5°C-15 min). Average of cell densities (OD₆₀₀) of triplicate cultures from either WT or mutants plasmid carrying cells was similar at both selected time points, 0.96-1.05 (4 h) and 1.09-1.19 (6 h). Error bars are based on \pm SEM of F1 recovery from three transformants of each. **(iii-iv)** Effect of R62A and R62S substitutions on bridge-type interaction. Superposed model of Caf1R_{62A} **(iii)** and Caf1R_{62S} **(iv)** on modelled *mar* DNA-Caf1R complex indicating a gap (disruption of bridge) between R62A or R62S (in helix-4) and the E98 (in helix-6) as shown by colored spheres and labels. **b)** Effect of Q93A and R97A substitution on F1 recovery and E98-mediated bridge-

type interaction. **(i-ii)** SDS-PAGE analysis **(i)** and Caf1 band (\leftarrow) quantification **(ii)** following F1 recovery in 'whole cell' of recombinant DH5 α cells carrying either pACYCF1-R_{Q93A} (Q93A) or pACYCF1-R_{R97A} (R97A) mutant plasmid or wild type (WT), pACYCF1 were carried out as in **(a)**. Average of cell densities (OD₆₀₀) of triplicate cultures from either WT or mutants plasmid carrying cells at 4 and 6 h were 0.84 and 1.04 (WT), 0.96 and 1.03 (Q93A), and 1.23 and 1.32 (R97A) **(iii-iv)**. Effect of Q93A and R97A substitutions on E98-mediated bridge-type interaction. Superposed model of Caf1R_{Q93A} **(iii)** and Caf1R_{R97A} **(iv)** on modelled *mar* DNA-Caf1R complex indicating a gap (disruption of bridge) between R62 and the E98 in both cases as shown by colored spheres and labels. *Note*: Each mutant model was predicted on Rob protein template (PDB-1d5y) with model quality score $\geq 0.788/1.0$, P-value $\leq 2.328E-4$ and 'certain' reliability score.

3.4.3 Effect of additional E98 substitutions on bridge interaction and F1 production

The evidence from modelled Caf1R and mutagenesis was consistent with a key role of the helix 6 residues (Q93 and R97) in binding in the major DNA groove. Binding of R62 was also important. Therefore to further explore the role of E98 as a bridge between these key residues, additional less drastic substitutions were made at residue 98. E98 was substituted with the analogous residues present in Rob (A88) and MarA (T97) to create the plasmids pACYCF1-R_{E98A} and pACYCF1-R_{E98T}. F1 expression was monitored from cell surface extracts of three individual transformants as above. The level of F1 from cells carrying either of the mutated plasmids was substantially reduced in comparison to cells carrying WT pACYCF1 plasmid. After 4 h thermoinduction at 37°C, the F1 level was 58.85% \pm 2.15 of wild type with pACYCF1-R_{E98A} and 67.38% \pm 1.57 with pACYCF1-R_{E98T}. After 6 h thermoinduction, the F1 level of both increased in comparison to wild type, 95.24% \pm 3.24 (with pACYCF1-R_{E98A}) and 73.12% \pm 2.44 (with pACYCF1-R_{E98T}) **(Fig. 3.21a (i) and b)**. The reduced level of F1 from both E98A and E98T substitutions reflects loss of the E98-mediated bridge interaction with R62 in the modelled Caf1R mutants **(Fig. 3.21c (i-ii))**. Additionally, secondary structure prediction of helix-6 with E98A and E98T substitutions showed a good score of helix-6 propensity **(Fig. 3.21c (i-ii))**, suggesting that loss of the E98-mediated bridge interaction is responsible for destabilising Caf1R-DNA interactions to ensure optimum activation of transcription and F1 production.

These results were further supported by an 'unwanted' mutation in *caf1R*, encoding Caf1R_{E98K}, which led to virtual complete loss of F1 production. There was no CB stainable Caf1 in whole cells **(Fig. 3.21a (ii))**. This was further confirmed by immunoblotting **(Fig. 3.21a (iii))**. Following co-transformation of pACYCF1-R_{E98K} with, pBADhCaf1R (WT Caf1R), F1 production was restored to wild type levels **(Fig. 3.21a (ii))**, confirming that loss of F1 production could be attributed to the E98K mutation within Caf1R. In the IntFOLD3.0 generated model of Caf1R_{E98K} **(Fig. 3.21c (iii))** a single base-amino acid (A-R43) interaction was predicted. Interactions between helix-6 and DNA were no longer predicted. Following analysis of superposed model, the proposed bridge-type interaction was completely lost in Caf1R_{E98K} **(Fig. 3.21d (ii))**. This could be due to repulsive interactions between the positively charged R62 and K98 residues. Additionally, slight displacement of both Q93 and R97 were observed in the Caf1R_{E98K} model of wild type **(Fig. 3.21d (ii))**, suggesting an additional impact on base-specific interactions with helix-6 residues. To test the impact of the E98K mutation on transcription, *E. coli* Top10 was co-transformed with the promoter-*lacZ* fusion plasmids, pRScaf1M'-R'-*lacZ* (contains *caf1R* promoter(s); P_R) or pRScaf1R'-M'-*lacZ* (contains *caf1M* promoter(s); P_M) together with pACYC-R_{E98K} and promoter activity or β -galactosidase level was monitored **(Fig. 3.22)**. There was very low activity at both P_M and P_R promoter(s) upon transactivation by pACYC-R_{E98K} compared to

transactivation by wild type pACYC-R. This is consistent with a much lower affinity of the mutated Caf1R_{E98K} for the *caf* DNA as indicated in the model. The effect was much more severe at the P_M promoter(s) (32.62-fold reduction at 10 h) than at the P_R (5.52-fold reduction) promoter(s) (**Fig. 3.22**, solid lines). There was no dramatic change on the growth of cells in either case (**Fig. 3.22**, dotted lines). These results confirm that the dramatic reduction in F1 production with the E98K mutation is a consequence of very poor activation of transcription at both *caf1R* and *caf1M* promoter(s) of the *caf* locus.

These studies support the hypothesis of a role for E98 as a linker residue between 2 key Caf1R-DNA interacting sites generating the DNA backbone::R62-E98-R97-Q93::T-base bridge interaction. This bridge interaction would be predicted to ensure optimum spatial orientation of the key DNA binding residues and stability of the Caf1R-DNA complex.

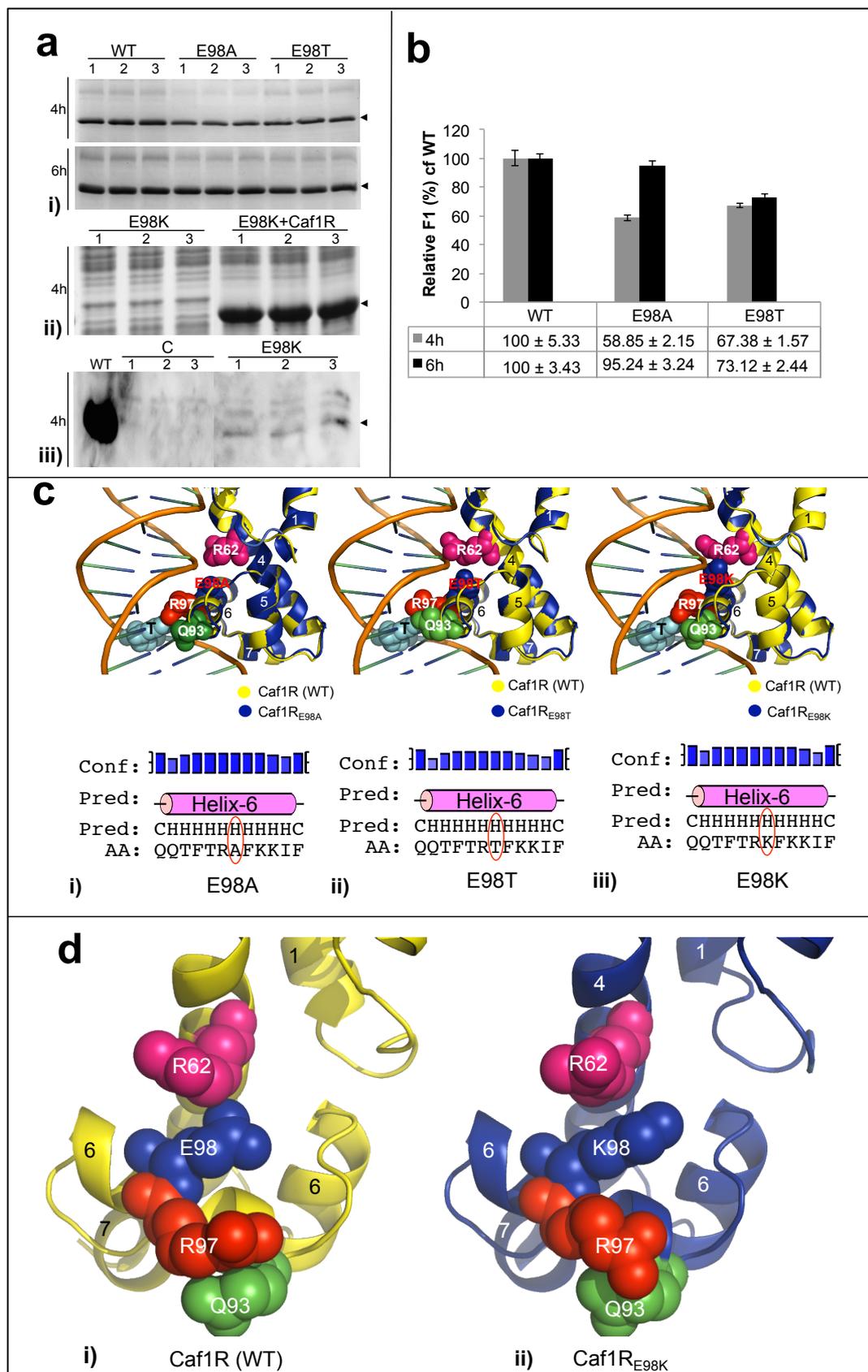


Figure 3.21| Caf1R- 98th residue is essential for bridge-type interaction and F1 production.

a-b) Effect of Caf1R- E98A, E98T and E98K substitutions on F1 recovery. SDS-PAGE (16% acrylamide) analysis of F1 subunit, Caf1 (indicated by arrowheads) from pACYCF1-R_{E98A} (E98A) and pACYCF1-R_{E98T} (E98T) (i) and from pACYCF1-R_{E98K} (E98K) (ii). F1 recovery in (i) was monitored and quantitated (b) from 'cell surface' of recombinant DH5 α cells carrying either mutant plasmids (E98A or E98T) or wild type pACYCF1 (WT) following 4 and 6 h of thermoinduction at 37°C. An equivalent of 0.04 OD unit sample of each was applied to gel following heat denaturation (97.5°C-15 min). Average of cell densities (OD₆₀₀) of triplicate cultures from either WT or mutants plasmid carrying cells at corresponding 4 and 6 h were 0.84 and 1.04 (WT);

0.81 and 0.96 (E98A); 0.89 and 0.95 (E98T). Indicated error bars are in **(b)** are based on \pm SEM of F1% from three biological replicates of each. **ii)** F1 expression from three individual co-transformants of Top10/pACYCF1-R_{E98K}+pBADHis (E98K) and wild type Caf1R complemented Top10/pACYCF1-R_{E98K}+pBADhCaf1R (E98K+Caf1R). F1 expression was monitored from 'whole cell' recovered from 5 OD units induced culture (0.02% L-ara arabinose, 4 h at 37°C induction), mixed in 100 μ l PBS and an equivalent of 0.37 OD unit sample of each was applied to gel following heat denaturation (97.5°C-15 min). **ii)** Western blot (on NCM) analysis of F1 expression from pACYCF1-R_{E98K} mutant along with negative (pACYC empty vector; C) and positive (pACYCF1; WT) control. F1 expression was monitored (37°C-4 h) from the 'whole cell' samples (5 OD units) of recombinant DH5 α carrying pACYCF1-R_{E98K} mutant plasmid or positive or negative control plasmid. Level of F1 expression (Caf1 subunit, from heat denatured 0.37 OD unit sample of each was detected by blotting with anti-Caf1sc (1:5000) primary antibody and anti-Rabbit IgG-HRP (1:40,000) secondary antibody. **c)** Effect of E98A, E98T and E98K substitutions on bridge-type interaction and helix-6 propensity. Superposed model of Caf1R_{E98A} **(i)**, Caf1R_{E98T} **(ii)** and Caf1R_{E98K} **(iii)** on modelled *mar* DNA-Caf1R complex indicating a gap (disruption of bridge) between R62 and the E98A and E98T substitution whereas no such effect in **(iii)** as indicated by colored spheres and labels. No dramatic effect was observed on helix-6 propensity upon E98A/T/K substitutions as indicated by identical length of helix and equal prediction score in each case. **d)** Bridge-type interaction analysis from another angle of modelled complex of wild type (WT) and mutant (E98K) Caf1R, indicate clear disruption of E98-mediated bridge-type interaction upon E98K substitution. For picture clarity DNA was hide in both cases.

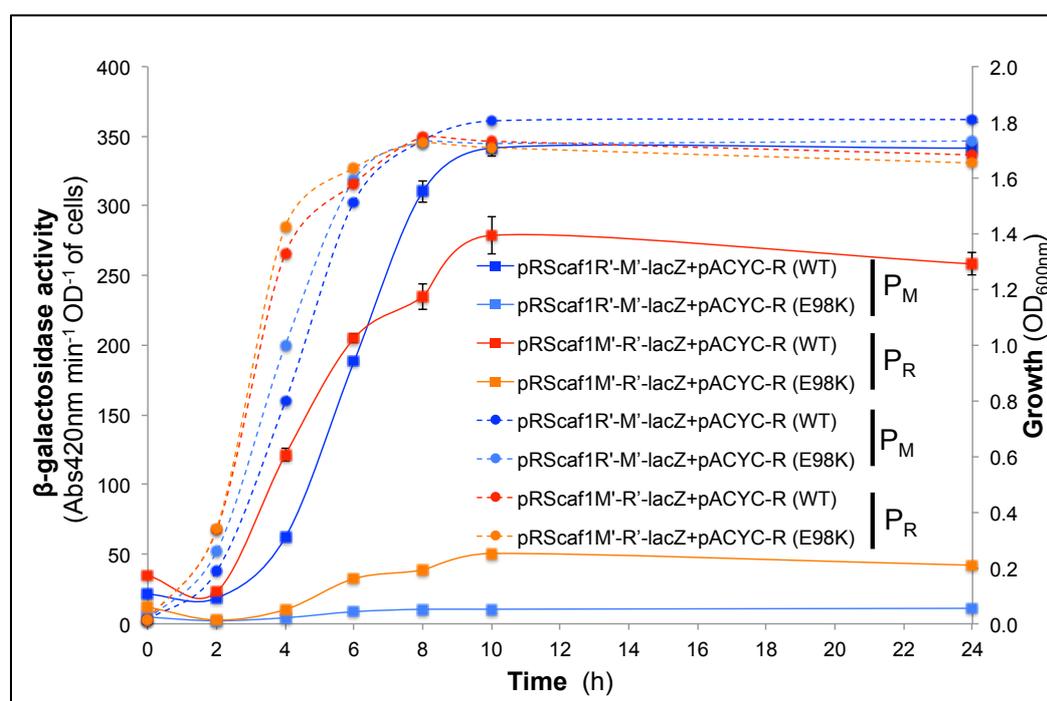


Figure 3.22| **Caf1R_{E98K} causes reduced transcription activation at both P_M and P_R promoter(s).**

β -galactosidase activity from recombinant Top10 cells carrying respective *caf1M* (P_M) and *caf1R* promoters (P_R) containing promoter-*lacZ* fusion constructs, pRScaf1R'-M'-*lacZ* (P_M) and pRScaf1M'-R'-*lacZ* (P_R) was tested in presence of wild type pACYC-R (WT) and mutant pACYC-R_{E98K} (E98K) as indicated. Standard error bars, on activity lines are based on \pm SEM of activity from three biological replicates of each. Growth (OD₆₀₀) of corresponding sample is indicated by dotted lines of identical color.

3.5 Analysis of other key binding residues within N-terminal DBD of Caf1R

3.5.1 Localisation of other key residues modelled as interacting with DNA

The requirement for Caf1R- R62 (helix 4) and E98, Q93 and R97 (HTH2) for stable DNA-protein interaction in the major DNA groove was addressed in earlier sections. In order to identify other key residues, essential for DNA-protein interactions, the modelled *mar* DNA-Caf1R complex was analysed for all residues, which appear to interact with the *mar* DNA ligand (**Fig. 3.23**). In HTH1 (H2-T-H3) I31, D32, C33, F40, R42, R43 and Y44 were identified as interacting with the associated DNA. I31, D32 and C33 are located in helix-2, F40 is within the connecting loop of helix-2 and 3 and residues, R42, R43 and Y44 in helix-3 (**Fig. 3.23**). Only R43 and Y44 interact with nucleotides in DNA binding site, BS1. The remaining five residues, I31, D32, C33, F40, R42 of HTH1 interact with the DNA backbone (summarised in Table 3.2). This is in contrast to the *mar*-MarA crystal structure where 3 residues W42, Q45 and R46 (corresponding to Caf1R- R43, Q46 and I47) make extensive base specific interactions in BS1. In addition MarA-H43 (corresponds to Caf1R-Y44) interacted with the DNA backbone (Rhee et al., 1998). This one shared base-interacting position (R43 in Caf1R and W42 in MarA) was proposed to contribute to sequence specific interaction within the AraC/XylS regulators (Rhee et al., 1998). Each of the DNA backbone interactions corresponds to similar locations of interaction in the *mar*-MarA cocrystal structure. Close to HTH2, there were four additional DNA binding residues, K100, T106, R108 and Q109 (**Fig. 3.23**). K100 is within helix-6 while R108 to Q109, lie within the predicted helix-7 and T106 is located in the connecting loop of helix-6 and 7 (**Fig. 3.23**). All of these residues are surface exposed. K100, R108 and Q109 would appear to have the same function as the corresponding position in MarA, where they interact with the phosphate group of the DNA backbone in *mar*-MarA complex. Caf1R- T106 corresponds to MarA (P105) which faces away from the DNA in the *mar*-MarA complex (Rhee et al., 1998).

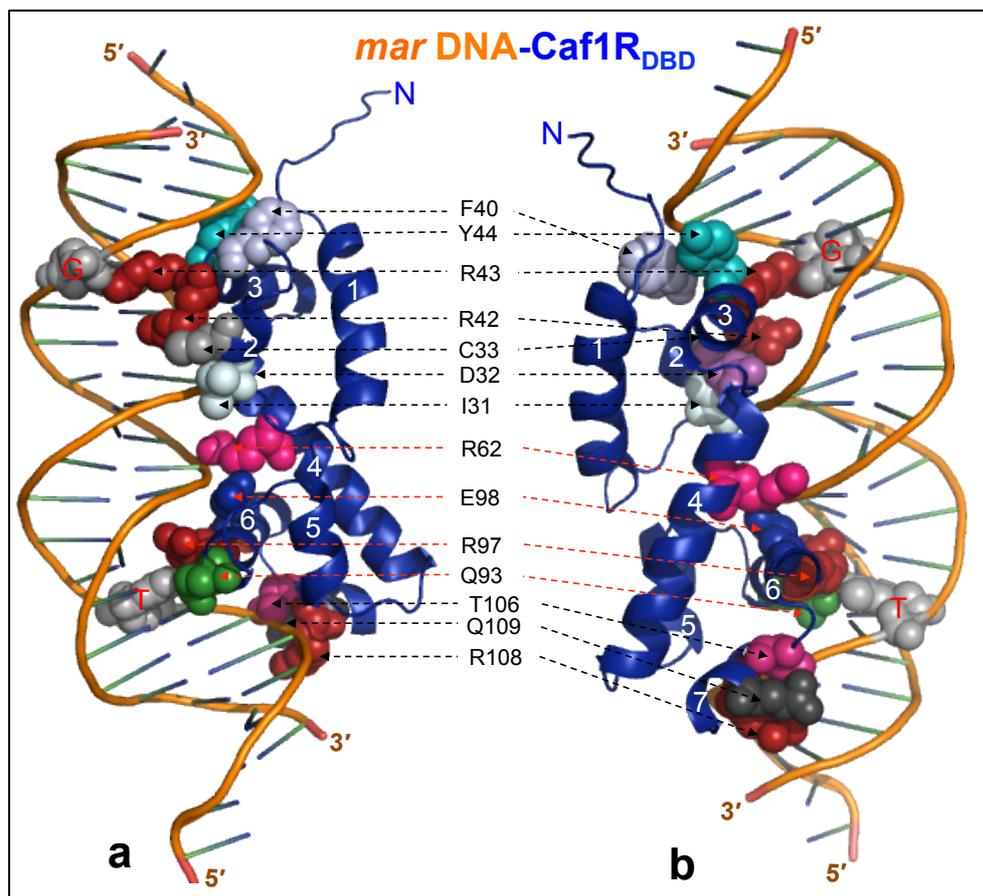


Figure 3.23| Caf1R-DBD mutagenesis model in complex with DNA.

IntFOLD3.0 modelled wild type Caf1R_N (blue colored ribbons with indicated helices, 1-7) at 90° (a) and 180° (b) is shown in complex with *mar* DNA from MarA-DNA complex (PDB-1bl0). Caf1R residues from HTH1 (H2-T-H3) and HTH2 (H5-T-H6) along with termini (Helix-7 and loop between helix 2 and 3), which seems to interact with associated DNA are shown in different colored balls and labeled. Residues presented by red dotted lines are screened in earlier sections. DNA bases interacting with R43 (G) and R97 (T) are indicated by grey colored balls and labeled. N & C and 5' & 3' show the termini of Caf1R_N and DNA, respectively. MacPyMOL1.3 was used to generate the model picture.

3.5.2 Effect of predicted DNA binding residues of Caf1R-DBD on F1 expression

Following identification of putative DNA-interacting residues within *mar* DNA-Caf1R complex, each identified residue was mutated to the small and non-reactive alanine (A) residue in the pACYCF1 template. Following construction of this library of pACYCF1-R mutants by SDM (section 2.2.4 (iv) (a)), the level of F1 expression was monitored by SDS-PAGE either directly in whole cells or from the cell surface extract and the intensity of the F1 subunit (Caf1) bands quantitated (Fig 3.24a-b) as described in sections 2.4.1 and 2.4.2. The level of F1 at both 4 and 6 h from the selected mutants of DBD is presented as before, as percent recovery relative to wild type Caf1R (Fig. 3.24). The data is summarised in Table 3.2.

It was observed that upon alanine substitution of the 2 nucleotide contacting residues of helix-3, R43 and Y44, the level of F1 was moderately reduced at both time points compared to WT; 79.48% ± 4.02 and 66.35% ± 0.66 (4 h and 6 h, respectively, pACYCF1-R_{R43A}) and 72.06% ± 1.39 and 76.44% ± 2.34 (4 h and 6 h, respectively, pACYCF1-R_{Y44A}) (Fig 3.24a and c). This provides support to the model that both R43 and Y44 of Caf1R are equally essential residues to interact with DNA bases. Of the predicted DNA backbone contacting residues of HTH1, I31, D32, C33, F40

and R42 four mutants, pACYCF1-R_{I31A}, pACYCF1-R_{D32A}, pACYCF1-R_{C33A} and pACYCF1-R_{F40A} significantly affected F1 production (**Fig. 3.24a and c**). A wild type phenotype was observed from pACYCF1-R_{R42A} at both time points, suggesting no impact of R42A mutation on Caf1R mediated transcription activation. Among the DNA-backbone interacting residues, F40A substitution had the biggest effect; producing only 4.18% ± 0.33 the level of WT F1 at 4 h and 5.53% ± 0.06 at 6 h. F40 is located in the loop region between helix-2 and 3. Aside from the interaction with DNA it is surface exposed. Hence, the impact it has cannot be readily explained from the modelled complex. The reduction of F1 production also from pACYCF1-R_{I31A}, pACYCF1-R_{D32A} and pACYCF1-R_{C33A} mutants at 4 h suggests an importance of I31, D32 and C33 in maintaining protein-DNA backbone interactions to enable efficient activation of transcription.

Substitution of additional DNA-backbone interacting residues close to HTH2, for examples T106 (present in connected loop of helix-6 and 7) and Q109 (in helix-7) appeared to have little impact, suggesting a negligible impact on the Caf1R-DNA interaction. In contrast, substitution of R108 (in helix-7) led to about 50% reduction in F1 recovery at both 4 and 6 h. This may reflect a critical role in anchoring the HTH-2 motif in place. MarA-K99 corresponds to Caf1R K100 (in helix-6). MarA (K99A) generated a substantial impact on transcription activation (Gillette et al., 2000). But attempts to construct pACYCF1-R_{K100A} were not successful therefore the impact of this DNA backbone interacting residues could not be assessed.

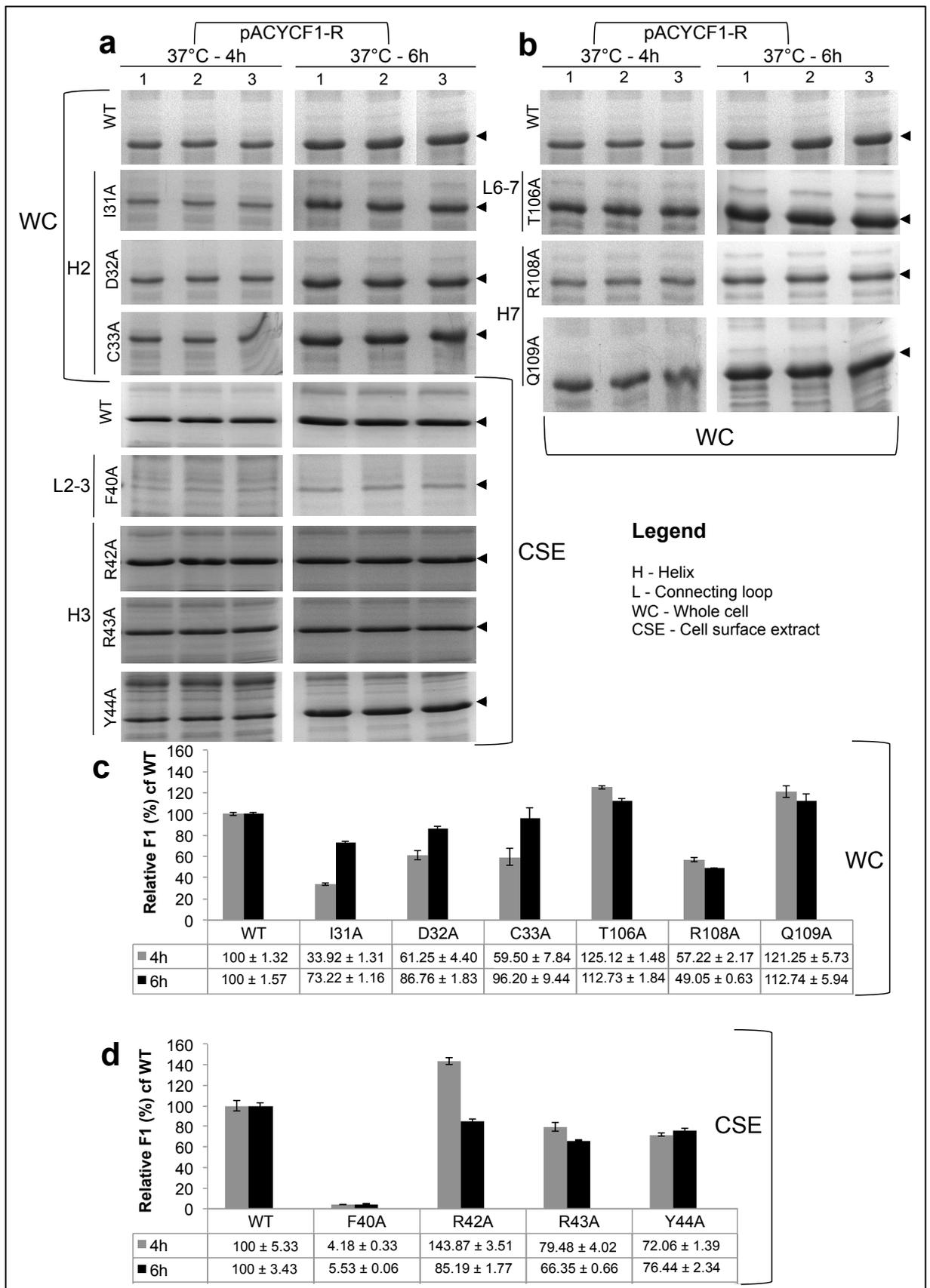


Figure 3.24 | **Quantification of F1 recovery from DH5 α /pACYCF1-R mutants.**

a-b) SDS-PAGE (16% acrylamide) analysis of F1 expression from the selected Caf1R mutants. Samples for F1 expression analysis were prepared either from 'whole cell' culture (1 ml) or cell surface extract of 5 OD units cells. Samples were harvested after 4 and 6 h of thermoinduction at 37°C, from an OD₆₀₀ range 0.85-1.23 (4 h) and 0.95-1.46 (6 h). The recovered pelleted whole cells (from 1 ml culture) were mixed in 200 μ l/OD PBS to give 5 OD units. Heat-denatured (97.5°C-15 min) 0.04 OD unit sample of each was analysed by SDS-PAGE. Arrowheads indicate F1 subunit, Caf1 band and expression level from respective mutant or wild type. Label of

each mutant is indicated on left. WT DH5 α /pACYCF1. **c)** Caf1 band quantification by imageJ software (Hartig, 2013). Relative recovery (%) with respect to 100% from corresponding WT was calculated from the average of three individual transformants (1-3) of each at 4 and 6 h of post-induction. Standard error on recovered F1 (%), added according to \pm SEM of F1 recovery from three independent transformants of each.

Table 3.2| Impact of Caf1R-DBD site-specific mutations on the expression of F1.

The relative % of recovered F1 from 0.04 OD sample, either from cell surface extract (CSE) or in whole cell (WC) from each constructed mutant was calculated with respect to corresponding WT (100%) at 4 and 6 h of thermoinduction at 37°C. MacPyMOL1.3 predicted possible interaction(s) of each mutagenesis-targeted residue with associated DNA within modelled *mar* DNA-Caf1R_N complex. Interatomic distance (Å; from center to center) between closest atom of DNA and the mutagenesis-targeted residue is calculated by PyMOL1.3.

pACYCF1-R mutant	F1 recovery (%)		Location of mutation in Caf1R _N	Interaction(s) in DNA-Caf1R _N model
	4 h	6 h		
WT	100 \pm 1.32 (WC)	100 \pm 1.57 (WC)	Bridge-type; DNA backbone :: R62-E98-R97-Q93 :: DT	
WT	100 \pm 5.33 (CSE)	100 \pm 3.43 (CSE)		
I31A	33.92 \pm 1.31 (WC)	73.22 \pm 1.16 (WC)	Helix-2	DNA backbone; (3.1Å)
D32A	61.25 \pm 4.40 (WC)	86.76 \pm 1.83 (WC)	Helix-2	DNA backbone; (3.5Å)
C33A	59.50 \pm 7.84 (WC)	96.20 \pm 9.44 (WC)	Helix-2	DNA backbone; (5.4Å)
F40A	4.18 \pm 0.33 (CSE)	5.53 \pm 0.06 (CSE)	Connecting loop, 2-3	DNA backbone; (3.4Å)
R42A	143.87 \pm 3.51 (CSE)	85.19 \pm 1.77 (CSE)	Helix-3	DNA backbone; (4.9Å)
R43A	79.48 \pm 4.02 (CSE)	66.35 \pm 0.66 (CSE)	Helix-3	(DG :: R43) (Overlap)
Y44A	72.06 \pm 1.39 (CSE)	76.44 \pm 2.34 (CSE)	Helix-3	(DG :: Y44) Overlap
R62A	76.33 \pm 3.66 (CSE)	96.55 \pm 1.27 (CSE)	Helix-4	Bridge-type
R62S	58.85 \pm 0.20 (CSE)	92.97 \pm 2.94 (CSE)	Helix-4	Bridge-type
Q93A	6.99 \pm 0.36 (WC)	33.90 \pm 1.77 (WC)	Helix-6	Bridge-type
R97A	4.79 \pm 1.83 (WC)	86.53 \pm 4.78 (WC)	Helix-6	Bridge-type
E98G	13.05 \pm 0.62 (CSE)	14.16 \pm 0.32 (CSE)	Helix-6	Bridge-type
E98A	58.85 \pm 2.15 (CSE)	95.24 \pm 3.24 (CSE)	Helix-6	Bridge-type
E98T	67.38 \pm 1.57 (CSE)	73.12 \pm 2.44 (CSE)	Helix-6	Bridge-type
E98K	00 (WC)	00 (WC)	Helix-6	Bridge-type
T106A	125.12 \pm 1.48 (WC)	112.73 \pm 1.84 (WC)	Connecting loop, 6-7	DNA backbone; (Overlap)
R108A	57.22 \pm 2.17 (WC)	49.05 \pm 0.63 (WC)	Helix-7	DNA backbone; (Overlap)
Q109A	121.25 \pm 5.73 (WC)	112.74 \pm 5.94 (WC)	Helix-7	DNA backbone; (3.6Å)

3.6 Conclusion

Original cloning and identification of the *caf* locus (Galyov et al., 1990; Karlyshev et al., 1992) had identified the *caf1R* gene located upstream of *caf1M* in the opposite orientation. The gene product was known to be a regulator belonging to the AraC/XylS family of transcriptional regulators (Gallegos et al., 1997) and is required for the biogenesis of F1. Regulators belonging to this family have a conserved DNA binding domain (DBD) that is usually present at the C-terminus of the protein (Gallegos et al., 1997; Ibarra et al., 2008; Schuller et al., 2012). Unlike most AraC/XylS family, the putative DBD of Caf1R is located at the N-terminus (Karlyshev et al., 1992; Schuller et al., 2012). Prior to this study, a single point mutation of Caf1R- E98G leading to virtual loss of F1 production was identified in a plasmid, pFS2 encoding the complete recombinant *caf* locus (unpublished). Spontaneous loss in F1 expression in cells carrying recombinant native *caf* locus had been previously noticed (Galyov et al., 1991; Karlyshev et al., 1992), but characterisation of mutations leading to loss of F1 had not been investigated.

In this study, in order to minimise loss of F1 the native *caf* locus subcloned into a medium-copy number plasmid, pACYCF1, was used. Combined with temperature control of the *caf* genes, expression off at 26°C and on at 37°C (Cao et al., 2012; Simpson et al., 1990; Straley and Perry, 1995; Titball et al., 1997), this permitted study of stable, controllable expression and assembly of F1. The E98G mutation had initially been subcloned in pACYCF1_{spM}. Repair of this mutation, together with complementation studies with pBADhCaf1R encoded Caf1R, confirmed that the E98G mutation was solely responsible for the dramatic loss in F1 production from pACYCF1-R_{E98G}. Furthermore, the E98G mutation had a dramatic impact on transcription activation at both *caf1R* (P_R) and *caf1M* (P_M) proposed promoters.

Secondary structure prediction and modelling of Caf1R (on Rob template (Kwon et al., 2000)) showed this mutation is located within a HTH DNA binding motif, more specifically within DNA recognition helix-6 of HTH2. Analysis of the modelled Caf1R with *mar* DNA, from *mar*-MarA complex (Rhee et al., 1998) identified a potential role for E98 as a bridging residue, interacting with R62 (in helix-4) on one side and with Q93 and R97 (in helix-6) on the other side. In the modelled DNA-Caf1R complex, R62 interacts with the DNA-backbone. At the other end of the bridge, Q93 and R97 were predicted by IntFold3.0 to be base interacting residues and in the model both interact with *mar* DNA base (T) in the major groove adjacent to helix-6. On substitution of E98 with a Gly residue, the bridge interaction is modelled as disrupted due to the very small size of Gly. Superposition of wild type Caf1R model on Caf1R_{E98G} mutant model showed no structural difference, although the secondary structure prediction of helix-6 with E/G98 residue indicated a substantial decrease in helix propensity and shortening of helix-6 by G98. Thus, destabilisation of helix-6 in Caf1R_{E98G}, combined with loss of the bridge interaction, is likely to have a synergistic negative effect on Caf1R mediated activation of transcription.

The role of E98 as a bridging residue was further supported by E98 substitution with residues, which showed a helix-6 propensity score comparable to that of wild type. For example E98T (corresponding to MarA T97), E98A (corresponding to Rob A88) and E98K (replaces negative with positive charge) substitutions. Modelling of all 3 mutated Caf1R protein-DNA complexes

showed loss of a bridge interaction with R62 and there was a considerable reduction, of 33-42% in F1 production from the 2 mutated plasmids, pACYCF1-R_{E98A} and pACYCF1-R_{E98T}. The fact that E98K virtually abolished F1 expression can be explained by location of this positively charged side chain between 2 positively charged Arg residues, R62 and R97. This resulted in displacement of the key base interacting residues R97 and Q93, as well as loss of the E98-R62 interaction. This again highlights the strict requirement for optimal orientation of the base interacting residues of helix-6 within BS2.

The role of R62 in stabilising the DNA-Caf1R complex *via* a helix-4-DNA backbone interaction was also validated by R62A and R62S substitutions. The significant loss in recovery of F1 by 24-42% reflected the modelled loss of R62-DNA backbone interaction. This was combined with loss of E98 interaction. Similarly, the critical role of the invariant base-binding residues, Q93 and R97, was confirmed by the drastic reduction of 66-95% in F1 production following independent Ala substitution of each residue. Thus, these results provide strong evidence that Q93 and R97 are key residues in specific base interaction at BS2, that R62-backbone DNA interaction is important in stabilising the DNA-regulator interaction and that, in Caf1R, E98 additionally stabilises the complex *via* a bridge interaction between helix-6 and R62 of helix-4. The residue corresponding to E98 in the MarA/ Rob/Sox family is highly variable. The importance of this residue, at least in Caf1R, is most likely related to optimum spatial orientation of HTH2 residues within the major DNA groove to ensure efficient activation of transcription and F1 expression.

Modelling and mutagenesis also identified other key Caf1R residues interacting with DNA. In HTH1 (H2-T-H3), both R43 and Y44 were modelled as making specific DNA-base interactions in BS1 and alanine substitution of these 2 residues resulted in a moderate reduction of 21-34% in F1 production from pACYCF1. Within HTH1, I31, D32, C33, F40 and R42 were all identified as residues interacting with the DNA backbone. Alanine substitution of each of these residues, with the exception of R42, affected F1 production. F40A was notable in the pronounced effect that this substitution had, reducing F1 production by 95%. Similarly, with residues close to HTH2 (H5-T-H6), K100, T106, R108 and Q109 were all modelled as being involved in DNA backbone interactions close to BS2. Alanine substitution of these residues appeared to have less of an impact. T106A and Q109A had negligible effect on F1 production. However, R108A did have a substantial effect reducing F1 production by about 50%. Attempts to construct pACYCF1-R_{K100A} were not successful hence, the impact of this DNA backbone interacting residue could not be assessed.

A homology search revealed that Caf1R homologs are distributed throughout the *Enterobacteriaceae* family of Gram-negative bacteria with many being identifiable within whole genome sequence data of different strains of *E. coli*. The closest homologs were an AraC/XylS family regulator of 1) uropathogenic *E. coli*, strain-126 (Salipante et al., 2015) and 2) avian pathogenic *E. coli*, strain-008 (Rojas et al., 2013). These shared 61% and 53% amino acid identity over the entire length of Caf1R and >71% amino acid identity with Caf1R-DBD alone, reflecting high conservation of the DBD. Unexpectedly, the closest homologue (WP_042110517.1) appeared to be linked to metabolism, specifically carbohydrate transport and not associated with any CU-system. This regulator has not been characterised. However,

based on the ability of other AraC/XylS-like regulators to bind small substrates *via* the 'sensing' domain (Gallegos et al., 1997) the possibility exists that Caf1R may bind to a small molecule. No information could be obtained regarding the genome context of the second top homolog (WP_001697817.1). However phylogenetic analysis indicated a close relationship to (WP_042110517.1), hence it would be interesting to understand more about this regulator as well.

The third top homolog (WP_024182496.1) was identified as a CU-linked regulator from diarrheagenic *E. coli*, strain-DEC11A (Hazen et al., 2012), with 44% amino acid identity and 52% similarity over 89% coverage to Caf1R. An additional 11 further CU-linked regulators were identified with $\geq 30\%$ amino acid identity and $\geq 38\%$ similarity over $\geq 85\%$ coverage to Caf1R. Among these only two, LdaA (44% identity, 52% similarity) and AfrR (42% identity, 49% similarity) of *E. coli* have been studied so far (Cantey et al., 1999; Scaletsky et al., 2005). LdaA controls expression from the locus of diffusive adherence (*lda*) fimbriae of enteropathogenic *E. coli* and AfrR regulates the expression of AF/R1 fimbriae of attaching/effacing *E. coli*. Interestingly, the predicted DBD of all of these CU-linked regulators is located at N-terminus and is highly conserved. Conservation of residues was particularly pronounced in the predicted HTH1 and HTH2 motifs, and the HTH2 sequence R97 E98 F99 was invariant. This suggests a very similar mode of DNA-protein interaction particularly at BS2 and also a shared requirement for E98. Based on this, one would expect a very similar DNA binding sequence for each of these regulators.

In conclusion, through modelling, bioinformatic analysis and mutagenesis, this Chapter has identified key residues involved in Caf1R-DNA interaction. The mutation Caf1R- E98G was shown to drastically reduce Caf1R dependent activation of transcription at both P_R and P_M promoters. This was attributed to loss of correct orientation of HTH2 residues due to reduced helix-6 length or stability combined with loss of a bridge-type interaction with R62 of helix-4.

Chapter 4

Recombinant expression and purification of tagged Caf1R

4.1 Introduction

Caf1R, a regulator of the *caf* locus, belongs to the AraC/XylS family of transcriptional regulators (Gallegos et al., 1997; Karlyshev et al., 1992; Schuller et al., 2012). Many proteins of this family are expressed in low-level and when overexpressed often make inclusion bodies and aggregates, consequently making it difficult to purify the soluble active protein (Gallegos et al., 1997; Schleif, 2010). The problem of aggregates can be prevented to a certain extent before, during and after purification, as has been described in a comprehensive study by Bondos and Bicknell (Bondos and Bicknell, 2003). Researchers have taken several different approaches to overexpress and purify the active form of regulators belonging to the AraC/XylS family. For example, MarA and Rob, the only proteins of this family whose high-resolution crystal structure has been solved with their cognate promoter DNA (Kwon et al., 2000; Rhee et al., 1998) were overexpressed as N- and C-terminal His₆-tagged proteins in the pET15b and pET28b expression plasmids (Novagen), respectively (Jair et al., 1995; Kwon et al., 2000). However, MarA was purified from insoluble inclusion bodies by urea solubilisation and subsequently refolding to obtain the active form, whereas Rob was purified as a soluble protein avoiding any treatment with denaturant like urea. Both proteins were purified by Ni²⁺ affinity chromatography. Final preparations of MarA were about 0.6 mg/ml whereas those of Rob were about 19 mg/ml. Likewise, XylR (another homologue of Caf1R putative-structure; 36% identity), was recently overexpressed and purified as a C-terminal His₆-tagged recombinant protein from the pET28a vector (Novagen). It was purified in the soluble form by growing BL21 (DE3) cells containing *xyIRpET28a* at 15°C/16 h with 0.5 mM IPTG induction, followed by Ni²⁺ affinity chromatography (Ni et al., 2013). SoxS, a protein closely related to MarA (41% identity, 67% similarity) (Duval and Lister, 2013) was overexpressed as a soluble maltose binding protein (MBP) fusion and purified by amylose affinity chromatography (Fawcett and Wolf, 1994). These proteins share 37% (MarA), 27% (Rob) and 36% (XylR) identity with Caf1R. Therefore similar approaches were used to overexpress and purify functional recombinant Caf1R.

Hence, this chapter describes the different approaches taken to express and purify soluble tagged Caf1R to be used initially in *in vitro* DNA binding studies (Chapter-5) and optimise overexpression, irrespective of solubility for use in future anti-Caf1R antibody production and structural analysis. N-terminally His₆-tagged Caf1R functionality in F1 production had already been confirmed using the pBADhCaf1R construct (Chapter-3). Therefore, initially expression, purification and characterisation of His₆-Caf1R (hCaf1R) was carried out from this construct. Attempts were then made to maximise the level of soluble protein by cloning native and synthetic *caf1R* into pET28a⁺ and pMALc2x expression plasmid vectors. Expression was monitored in a range of *E. coli* strains with different conditions of induction and cell lysis.

Results and Discussion

4.2 Expression and purification of tagged Caf1R from native gene

4.2.1 Expression and purification of hCaf1R from pBADhCaf1R

i) Induction and solubility of hCaf1R

As functional His₆-tagged Caf1R had already been shown by complementation of pACYCF1_{S_{PM}} with the pBADhCaf1R plasmid construct (Section 3.2.1), this plasmid was initially tested for the recovery of soluble hCaf1R. The final protein sequence and strategy that had been used to prepare this construct is depicted in **Fig. 4.1**. The pBADhCaf1R encodes hCaf1R, a product with 6 N-terminal Histidine (His) residues followed by the Xpress epitope and with a calculated molecular weight of 40.98 kDa.

For testing induction, a small-scale (10 ml) expression study (section 2.5.1) was carried out using *E. coli* Top10/pBADhCaf1R (pBADhR) and its negative control, Top10/pBADHisA (empty vector, C). Expression was monitored following induction with a range of L-ara concentrations (0.2-0.0002%) at 37°C for 4 h. At each time point, one OD unit of bacterial cells was recovered and lysed by sonication (section 2.5.2) in 100 µl PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄; pH 7.4). Supernatant (S) and pellet (P) fractions were then recovered by centrifugation at 15,000 rpm/20 min as described in section 2.5.3. Both fractions were analysed by SDS-PAGE (**Fig. 4.2a**). There was no clearly distinguishable band at the expected size of hCaf1R (40.98 kDa) in either S or P fraction following coomassie blue (CB) staining (**Fig. 4.2a**). A Western blot from the S fraction of 0.02% L-ara induced sample of Top10/pBADhCaf1R showed a clear distinguishable band at about 35-38 kDa (slightly smaller than the expected size of 40.98 kDa), confirming expression of hCaf1R (**Fig. 4.2b**). There was no such band at this location from the negative control (C), which validates this expression.

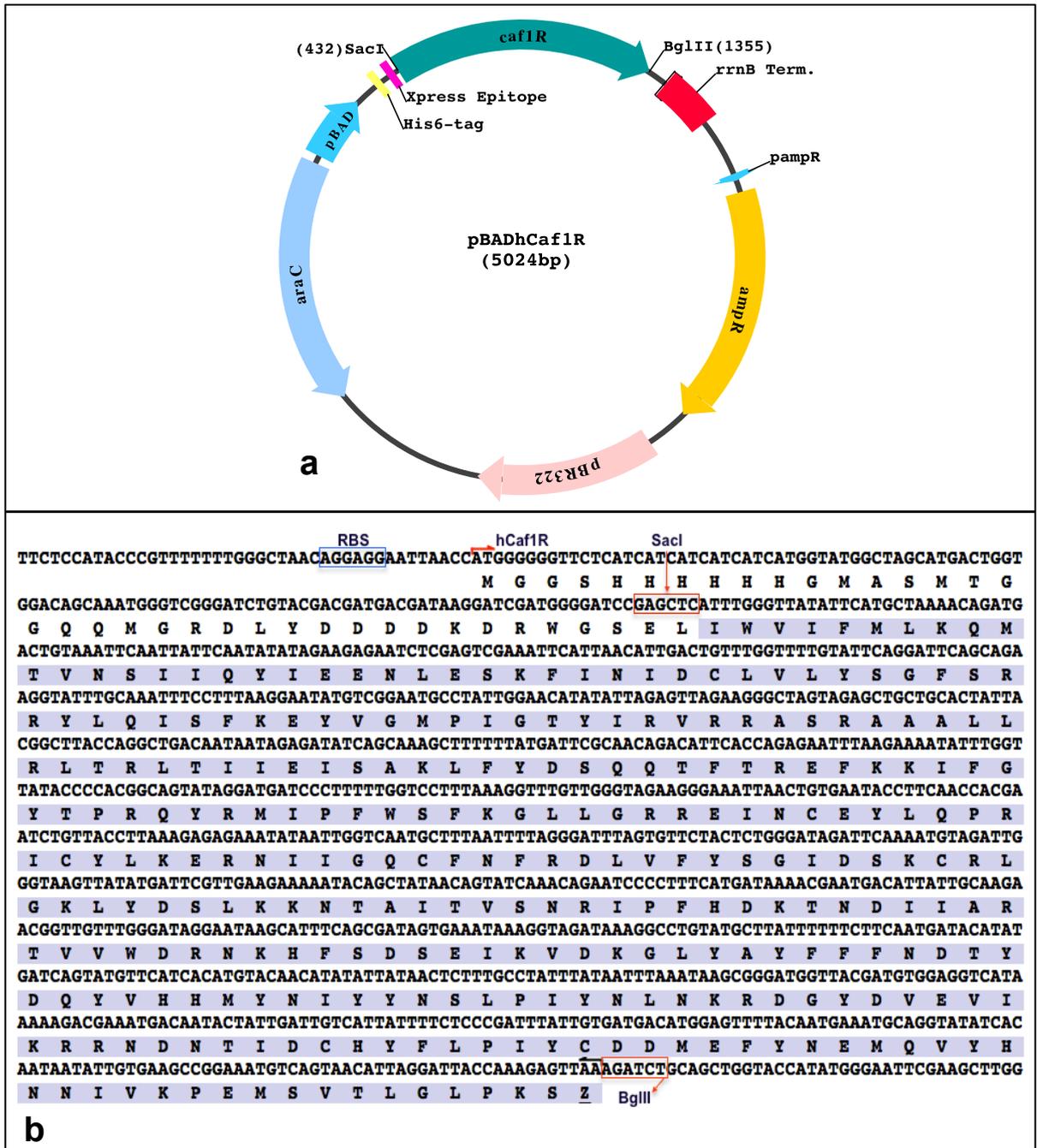


Figure 4.1| The pBADhCaf1R construct.

a) Circular plasmid map of pBADhCaf1R (5024 bp). **b)** Sequence of pBADhCaf1R, gene encoding Caf1R had been amplified from pFS2 plasmid (Drozdov et al., 1995) using primers with SacI and BglII sites and cloned between SacI/BglII sites (red boxes) of pBADHisA plasmid (Invitrogen) to create pBADhCaf1R (Lopez-Tolman, A. and MacIntyre, S., unpublished data). Translated sequence indicates the tagged Caf1R (hCaf1R) (345 aa; MWt 40.98 kDa). Coding sequence derived from *caf1R* is highlighted in blue and starts with a Leu start codon. Ribosomal binding site is enclosed in blue box.

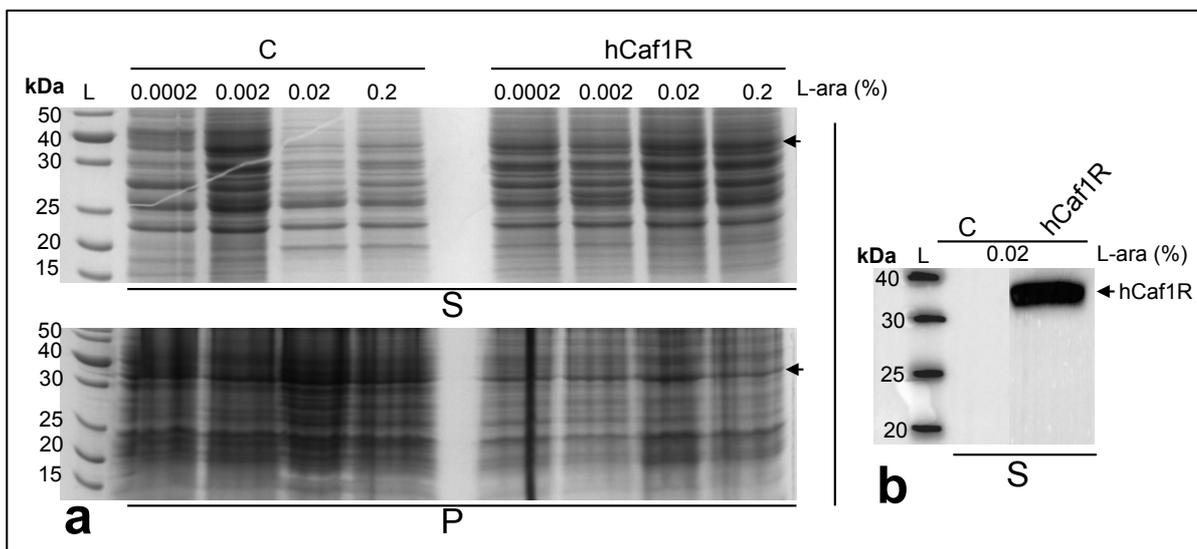


Figure 4.2| **Expression of native hCaf1R from *E. coli* Top10/pBADhCaf1R.**

a) SDS-PAGE (14% acrylamide) from the supernatant (S) and pellet (P) fractions of cells carrying pBADhCaf1R (pBADhR) or pBADHisA empty plasmid (C) induced (at 37°C/4 h) with a range of L-ara as indicated. Arrows, indicate the expected location of hCaf1R. **b)** Western blot (on NCM) from the supernatant (S) fractions of 0.02% L-ara induced sample of C (pBADHisA) and pBADhCaf1R. Arrow indicates hCaf1R, identified with antiHis-HRP monoclonal antibody (1:5000). L, protein size marker.

In order to determine solubility a large-scale (100 ml) preparation (section 2.5.1) was carried out to assess solubility of hCaf1R by ultracentrifugation and at the same time to test two different lysis buffers, L1 and L2. L1 was the binding buffer from the HisTapFF column (GE Healthcare) (20 mM NaPO₄ buffer, 500 mM NaCl; pH 7.4) supplemented with 3 mM DTT, 10% glycerol and 1× EDTA-free protease inhibitor (Pierce). L2 is a lysis buffer (20 mM TrisHCl, 200 mM KCl; pH 7.4) as identified by Bondos and Bicknell as a suitable buffer for maintaining a range of prokaryotic and eukaryotic proteins, including LacI regulator, in a soluble state (Bondos and Bicknell, 2003). Cells of *E. coli* Top10/pBADhCaf1R (2 × 50 ml) were induced with 0.02% L-ara at 37°C for 4 h (OD₆₀₀, 1.94), pelleted, mixed in 5 ml of either L1 or L2 and lysed by sonication. Unlysed cells were removed by centrifugation (5000 rpm/5 min/RT) and the resulting supernatant fractions were subjected to ultracentrifugation at 50,000 rpm for 60 min at 4°C to separate soluble hCaf1R (supernatant fractions, S1 and S2 from Lysis buffer L1 and L2, respectively) from insoluble hCaf1R in the corresponding pellet fractions, (P1 and P2). Despite the higher volume of sample processed for this analysis, there was still no clearly distinguishable hCaf1R band identifiable following CB staining of whole cells or lysed cell fractions (**Fig. 4.3a**). Western blotting confirmed presence of hCaf1R in all induced samples. A major host-protein band in the pellet samples and whole cells migrates at about 38 kDa and may mask low levels of hCaf1R (**Fig. 4.3b**). The results from Western blot also indicated that the supernatant fraction S2, recovered in L2 buffer contained relatively more soluble hCaf1R than L1 recovered S1 fraction (**Fig. 4.3b**). This could be due to protein aggregation caused by salt components of L1. The level of hCaf1R in the P1 and P2 fractions (insoluble form) was higher than that in the S1 and S2 fractions (soluble form), suggesting either initial aggregation into the insoluble form or degradation of soluble hCaf1R.

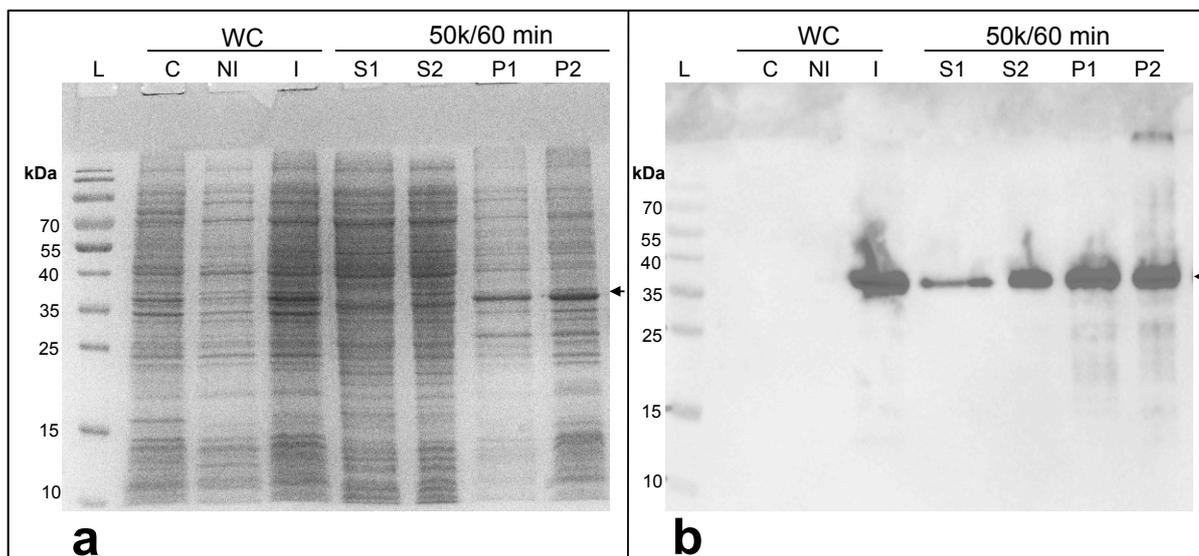


Figure 4.3| **Solubility analysis of hCaf1R by ultracentrifugation and 2 different cell lysis buffers.** CB stained SDS-PAGE (14% acrylamide) (a) and Western blot on NCM (b) of samples from *E. coli* Top10/pBADhCaf1R, induced with 0.02% L-ara for 4 h at 37°C and lysed by sonication in L1 (20 mM NaPO₄ buffer, 500 mM NaCl; pH 7.4 supplemented with 3 mM DTT, 10% glycerol and 1× EDTA-free protease inhibitor) or L2 (20 mM TrisHCl, 200 mM KCl; pH 7.4) lysis buffer. S1 and S2, Supernatant (soluble) fractions recovered following ultracentrifugation at 50k/60 min of lysed cells in L1 and L2 lysis buffers, respectively. P1 and P2, corresponding pellet fractions. WC, whole cells samples; C, sample from negative control (Top10/pBADHisA), NI and I, non-induced and L-ara induced whole cell samples from Top10/pBADhCaf1R. WC samples were directly mixed into 4× SDS-PAGE sample buffer (100 µl/OD). Identity of hCaf1R (in b) was confirmed by blotting with antiHis-HRP monoclonal antibody (1:5000, Roche). L, Protein size marker.

ii) Ni²⁺ affinity based isolation of soluble hCaf1R from *E. coli* Top10/pBADhCaf1R

It is well known that lowering the incubation temperature during induction enhances the possibility of soluble protein recovery. The lower temperatures prolong the time of protein folding and thus decrease the chances of aggregation which is favoured at high temperature due to fast misfolding and temperature dependent hydrophobic interactions (Baldwin, 1986; Makhatadze and Privalov, 1995; Noteborn, 1988; Rosano and Ceccarelli, 2014; Schellman, 1997; Vasina and Baneyx, 1997; Vera et al., 2007). Therefore with the aim of enhancing solubility of recombinant hCaf1R, a large batch (250 ml) of *E. coli* Top10/pBADhCaf1R culture was induced with 0.02% L-ara at 30°C for 6 h. Recovered cells were resuspended in 20 ml of HisTrapFF column binding buffer (20 mM NaPO₄ buffer, 500 mM NaCl, 5 mM DTT, 20 mM imidazole; pH 7.4), supplemented with EDTA-free proteases inhibitor, 1× (Thermo Scientific, Pierce) and lysed by French press. The lysate was clarified by ultracentrifugation and the supernatant fractionated by Ni²⁺ affinity chromatography as described (section 2.5.4(i)). The elution profile following Ni²⁺ affinity purification indicates two A₂₈₀ peaks, encompassing fractions, 2-4 (small peak) and 17-22 (large peak), respectively (**Fig. 4.4a**). Fractions, 2-4 were recovered in about 6% of buffer B, corresponding to about 15 mM imidazole whereas fractions, 17-22 appeared in about 47% of B, corresponding to 117.5 mM imidazole. Fractions 19-22 contained a large number of host background protein (not shown). A Western blot (**Fig. 4.4b**) confirmed identity of hCaf1R at about 38 kDa primarily in fractions 18 and 19 with a tail to fraction 22 and very low level in fractions 2-4.

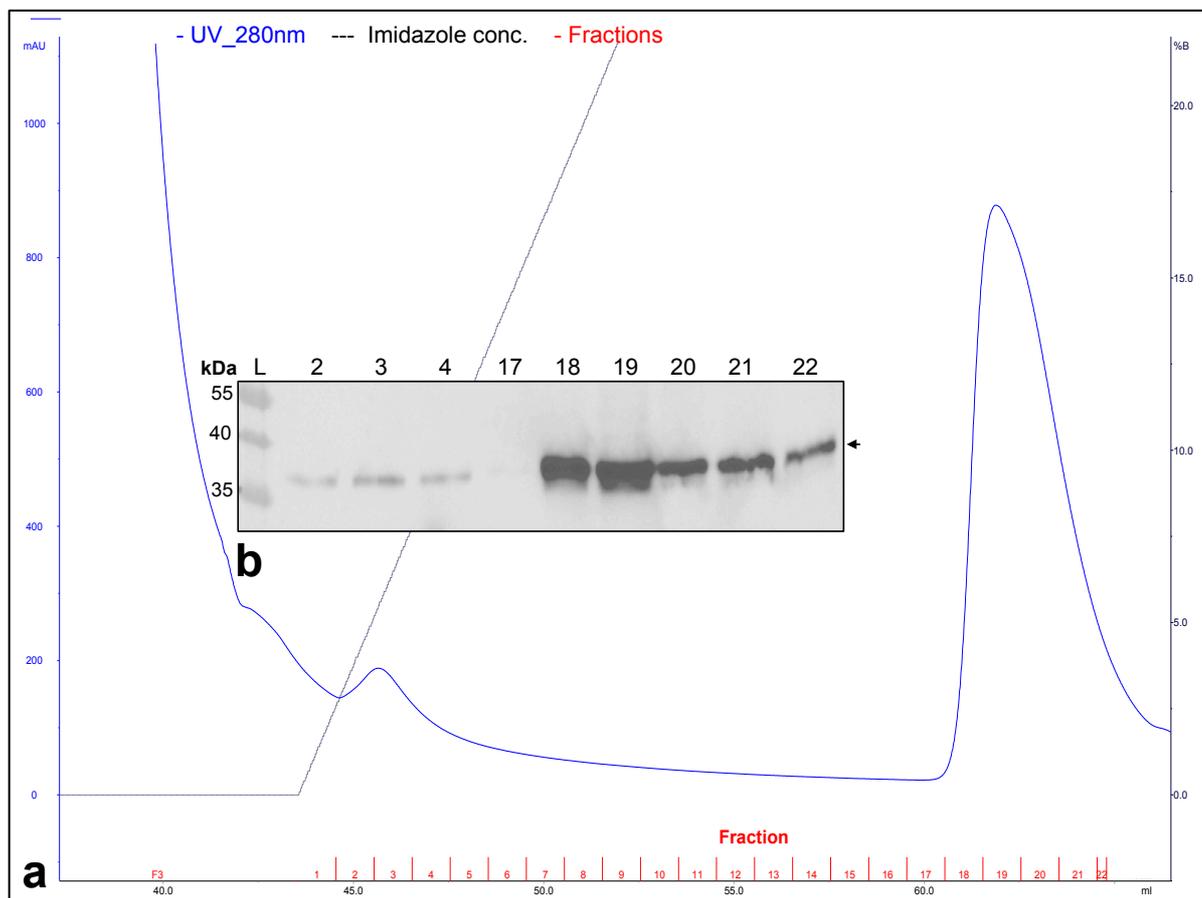


Figure 4.4| Ni^{2+} affinity chromatography of soluble hCaf1R.

Cells from 250 ml (438 OD units) induced culture of *E. coli* Top10/pBADhCaf1R were lysed and subjected to ultracentrifugation, as in text. Lysate supernatant (40 ml) was applied to a 1 ml HisTrapFF column. **a**) Elution profile (zoomed version) from HisTrapFF column with a linear gradient of 0-250 mM imidazole (100% buffer B, 20 mM NaPO_4 buffer, 500 mM NaCl, 5 mM DTT, 250 mM imidazole; pH 7.4) over 20 \times 1 ml fractions. **b**) Western blot analysis of peak-eluted fractions (2-4 and 17-22), blotted with monoclonal, antiHis-HRP conjugate (1:10,000). Arrowhead indicates the location and bands of hCaf1R.

iii) Estimation of oligomeric state of recombinant hCaf1R by size exclusion chromatography

Size exclusion or gel filtration chromatography is a technique used to separate proteins according to their size and oligomeric state. In order to estimate the size and form(s) of purified hCaf1R, a sample was analysed by gel filtration on a Superose™12 10/300GL column (GE Healthcare) as described in section 2.9.4. Initially, the elution profile of four molecular mass standards (Blue dextran, 2000 kDa; albumin, 67 kDa; Ovalbumin, 45 kDa and RibonucleaseA, 13.7 kDa) was monitored individually (data not shown) and then together (**Fig. 4.5a**). To analyse hCaf1R, fraction 18 (1 ml) from the Ni^{2+} affinity column (**Fig. 4.4**) was concentrated 10-fold and buffer exchanged against 50 mM NaPO_4 , 150 mM NaCl, 5 mM DTT; pH 7.0 using a Viva spin concentrator (10 kDa MWCO) (sections 2.9.1 and 2.9.2). Approximately 150 μ l of the concentrated sample was applied to the Superose™12 column and the 22 \times 1 ml fractions were monitored by Western blot (**Fig. 4.5b**). Fractions, 6-10, correspond to the column void volume (V_0) and hence likely contain aggregated forms of hCaf1R with high molecular weight, \geq 2000 kDa. The bands from fractions 11 and 12 could correspond to hCaf1R dimer (MWt 81.96 kDa). Fractions, 14 and 15 are consistent with a small amount of monomer (MWt 40.98 kDa). Thus

this analysis indicates only a small proportion of hCaf1R was recovered in the soluble monomeric state. While there may also be some dimeric hCaf1R, most of hCaf1R was in some higher size complex or aggregate.

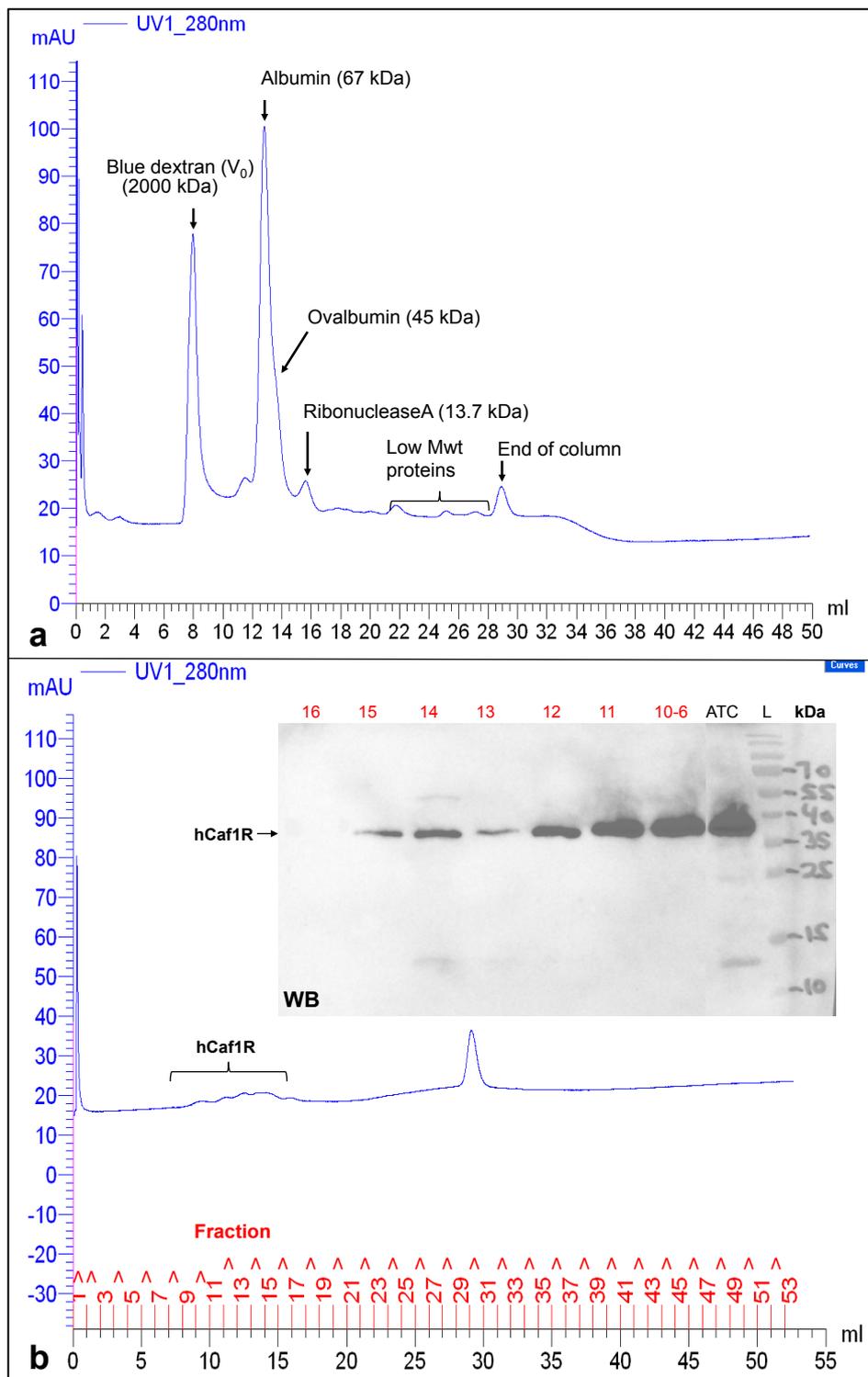


Figure 4.5| **Molecular size determination of hCaf1R by gel filtration chromatography.**

a) Elution profile of molecular mass standards from Superose™12 10/300GL column. Standards (150 μ l, 1 mg/ml each), molecular weight and elution volumes are as indicated. Flow rate 0.8 ml/min, fraction size, 1 ml. V_0 : void volume, identified with blue dextran. **b)** Elution profile and confirmation of fractionated hCaf1R by Western blot (WB; on NCM) analysis, using antiHis-HRP monoclonal antibody (1:10,000). ATC, applied sample (150 μ l) of Ni²⁺ affinity purified Fn-18 (Fig 4.4) that had been (concentrated and buffer exchanged as described in text). Elution conditions as in (a). Numbers correspond to 1 ml fraction number. 6-10, pooled fractions 6 to 10.

4.2.2 Expression of hCaf1R^T and hCaf1R^T_N from pET28a⁺ based vector

i) Construction of pEThCaf1R

As recovery of hCaf1R from *E. coli* Top10/pBADhCaf1R was low, to enhance the level of hCaf1R expression, the well-recognised expression plasmid, pET28a⁺ (Novagen) was tested for expression of *caf1R*. The strategy used to construct pEThCaf1R is depicted in **Fig. 4.6a**. The pEThCaf1R plasmid encodes hCaf1R^T, a product with N-terminal His₆-tag followed by a thrombin cleavage site and a calculated molecular weight of 40 kDa. Briefly, the *caf1R* gene (921 bp) plus either side pET28a⁺ vector Infusion joints (19- and 18 bp) was PCR-amplified from the pBADhCaf1R construct using Infusion primers, pET*caf1R*f and pET*caf1R*r as described in (section 2.2.4(i)). The gel purified PCR-product was Infused into EcoRI/NheI linearised pET28a⁺ plasmid (**Fig. 4.6b**, section 2.2.8(ii)). The resulting Infusion product was then transformed into competent StellarTM cells. Twenty-six transformants were screened by colony PCR (Section 2.2.4(ii)) using the same set of Infusion primers. Colony PCR-products from 23 transformants gave positive reaction with a band slightly smaller (\approx 800-900 bp) than the expected PCR-product (958 bp) (**Fig. 4.6c**). Plasmid DNA was isolated from 6 clearly positive transformants and analysed by EcoRI digestion (**Fig. 4.6d**) to confirm the quality of isolated DNA and the linear size of pEThCaf1R construct. All digested pEThCaf1R transformants showed a single band just below 8.0 kb while undigested transformants ran at about 3.5 kb, indicating the linear pEThCaf1R is running above the correct size (6.295 kb). Two transformants, 6 and 8 were sent for DNA sequencing with T7F and T7R primers. This confirmed the correct Infusion of *caf1R* into pET28a⁺ vector with intact fusion joints at both ends and the correct *caf1R* sequence (**Fig. 4.6e**). Both transformants were stocked and transformant 8 was used for overexpression and purification of hCaf1R^T.

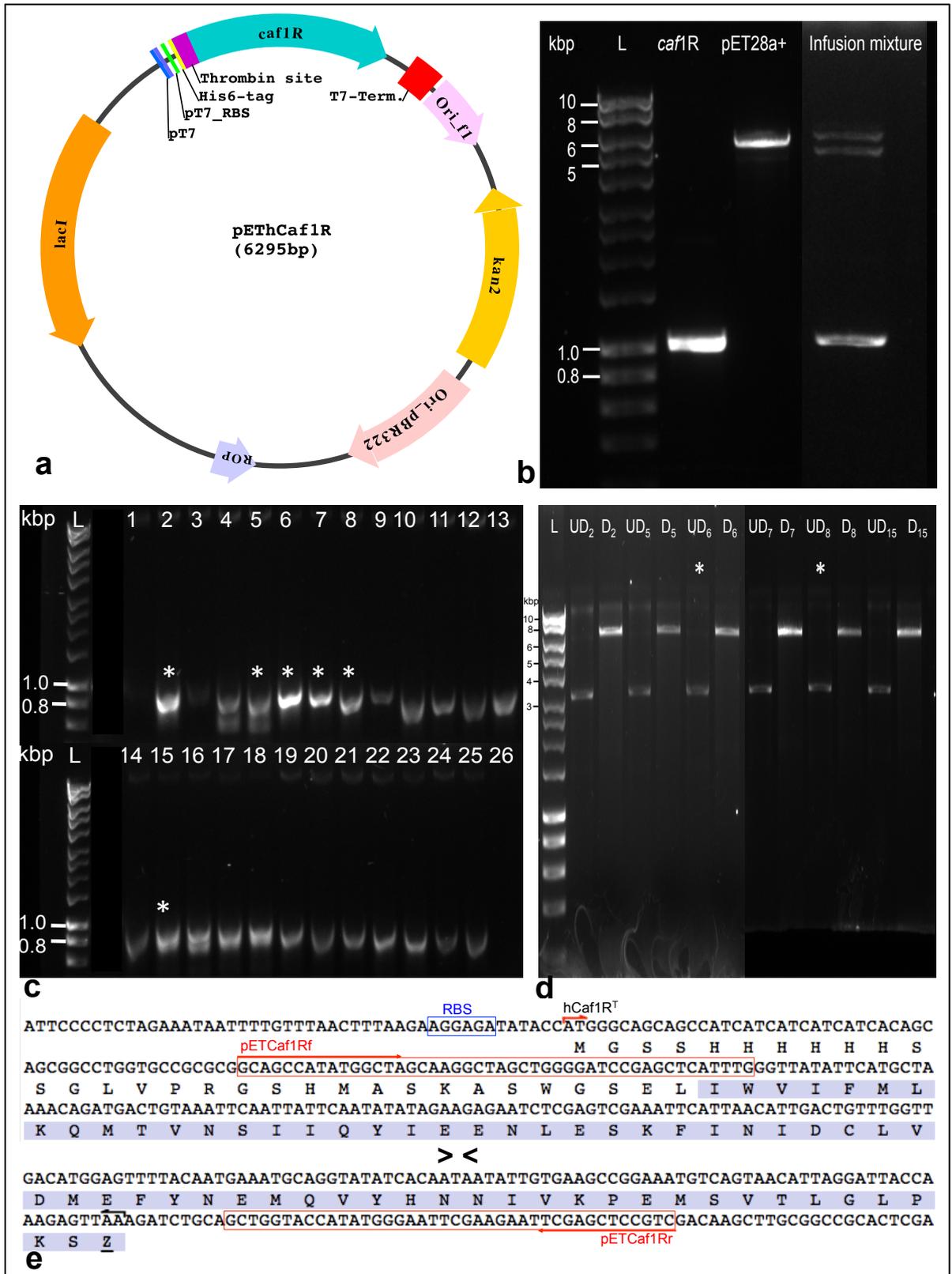


Figure 4.6| Construction of pEThCaf1R.

a) Circular plasmid map of pEThCaf1R (6295 bp). **b**) AGE analysis of *caf1R*+pET28a⁺ Infusion components and reaction mix. *caf1R*, gel purified PCR-product of *caf1R* with Infusion joints (100 ng; 958 bp), pET28a⁺, NheI/EcoRI digested and gel purified pET28a⁺ vector (50 ng). Infusion mixture, *caf1R* plus pET28a⁺ vector (2.5 μl, ~ 50 ng). **c**) AGE analysis of colony PCR-product, amplified with primers, pET*caf1R*f and pET*caf1R*r from transformants 1-26. *, Plasmids analysed in **(d)**. **d**) EcoRI digestion. UD, undigested and D, digested plasmid from transformants 2, 5, 6, 7, 8 and 15. Transformants 6 and 8, selected for DNA sequencing. L, DNA Hyper ladder-I (20 ng). **e**) DNA sequence of transformant 8, showing fusion joints and encoded hCaf1R^T fusion (338 aa; 40.0 kDa). Deduced amino acids from *caf1R* are highlighted in blue. ><, Represents sequence not shown. Infusion primers are enclosed in red boxes.

ii) Expression of hCaf1R^T at 28 and 37°C with 0.1 to 1.0 mM IPTG

To assess expression and recovery of hCaf1R^T, small-scale (10 ml) expression (section 2.5.1) was carried out with *E. coli* BL21(DE3) carrying pEThCaf1R (pEThR) or pET28a⁺ (control, C). Cells were induced with 0.1 mM to 1 mM IPTG at 37 or 28°C. Induced cells were lysed by either sonication or Bug-busting as described in **Fig. 4.7** and Section 2.5.2. Following lysis, the soluble (S) and insoluble (P) fractions were isolated by centrifugation at 16, 000×g/20 min and analysed by SDS-PAGE and Western blot. At 37°C/4 h with 1.0 mM IPTG induction, no distinguishable band of hCaf1R^T was observed by CB staining at the expected location of about 40 kDa in either soluble or pellet fraction (sonicated samples) (**Fig. 4.7a; CB**). WB of the soluble fractions confirmed presence of hCaf1R^T at about 38 kDa from both non-induced (NI) and induced (I) samples of BL21(DE3)/pEThR (**Fig. 4.7a; WB**). A prominent band at this location from NI samples suggests a high basal level of expression (**Fig. 4.7a; WB**). A faint band at about 24 kDa (←) from both (NI) and (I) samples, and a strong band at about 16 kDa (←←) only from (I) samples suggest truncation of hCaf1R^T most likely at the His₆-tagged N-terminal domain (hCaf1R^T_N, MWt 16.448 kDa). Thus, induction with 1.0 mM IPTG at 37°C/4 h showed a marked deleterious effect on expression of full-length hCaf1R^T as most of the recovered protein is truncated. Other possible reasons for this truncation could be the temperature and incubation time. Hence, the induction temperature was reduced to 28°C/16 h and induction tested with both 1.0 mM and 0.1 mM IPTG. Under both conditions, most of the blottable hCaf1R^T protein was observed in the insoluble pellet fractions (**Fig. 4.7b-c; WB**). Moreover, upon induction with either 1 mM or 0.1 mM IPTG, a major proportion of hCaf1R^T was degraded (↔) or formed aggregates (→) (**Fig. 4.7b-c; WB**). The presence of full-length hCaf1R^T from (NI) samples suggests leaky expression. Further analysis at shorter induction times (28°C/6 h and 37°C/3 h) with 0.35 and 0.5 mM IPTG resulted in complete hCaf1R^T recovery in the pellet fractions, following bug buster lysis, irrespective of temperature/time and IPTG concentrations (**Fig. 4.7d; WB**). Throughout, no blottable band was detected from the negative control, BL21(DE3)/pET28a⁺ (C) validating identification of hCaf1R^T.

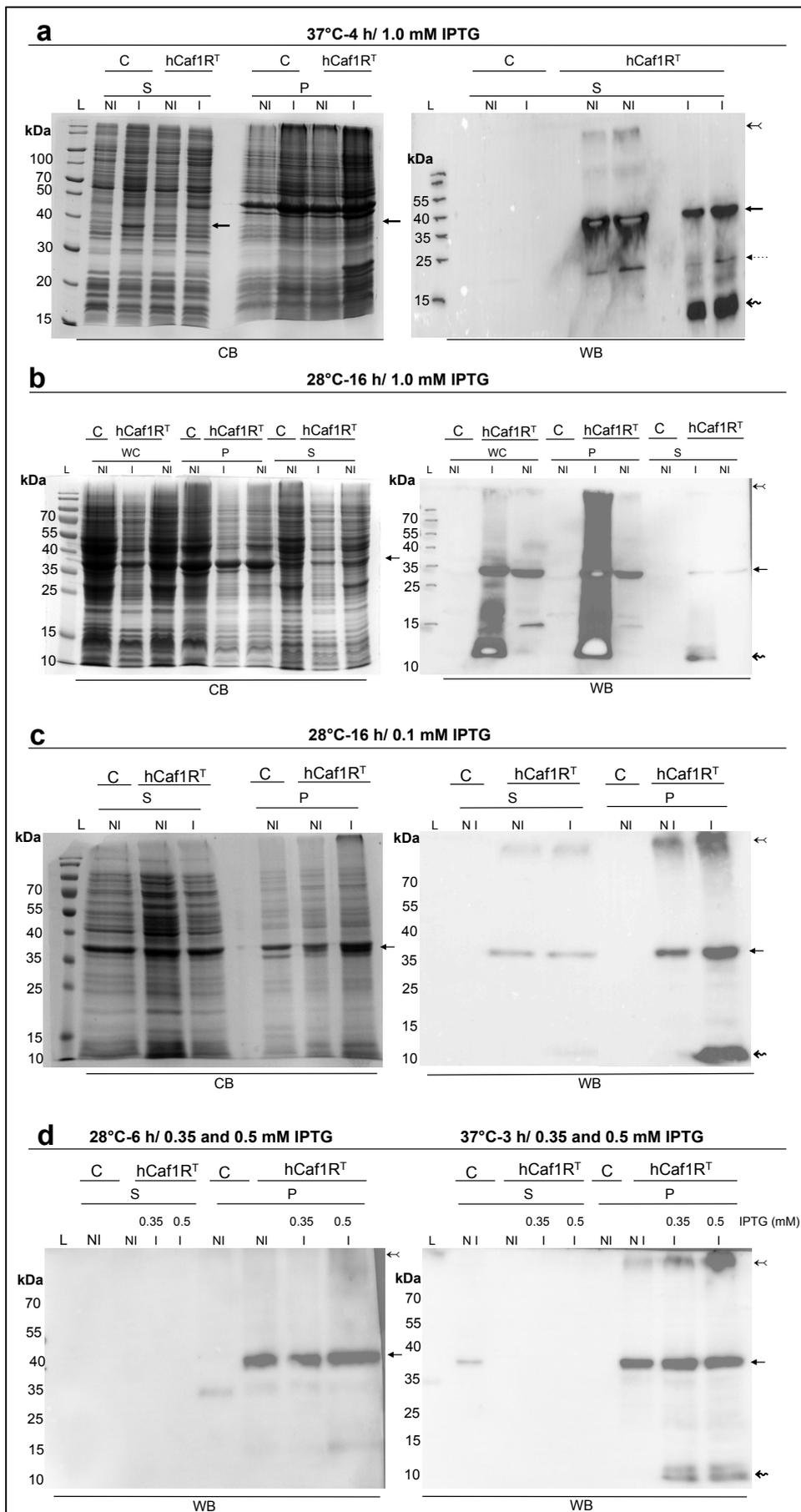


Figure 4.7 | Expression of hCaf1R^T from *E. coli* BL21(DE3)/pETHCaf1R at two different temperatures and IPTG concentrations.

SDS-PAGE (14% acrylamide, CB) and Western blot (WB, on NCM) analysis of hCaf1R^T expression from BL21(DE3)/pETHCaf1R. Cells were induced with 0.1-1.0 mM IPTG as indicated at 37°C/3-4 h or 28°C/6-16 h.

At the indicated time, cells (1 OD unit) were recovered and mixed in 100 μ l of PBS supplemented with 1 \times proteases inhibitor (EDTA-free) and lysed by sonication (**a-c**), or in 100 μ l bugbuster (**d**). S, soluble supernatant and P, insoluble pellet fractions isolated following centrifugation of cells lysate at 15k rpm/20 min. Identification of hCaf1R^T on WB was confirmed by blotting with antiHis-HRP monoclonal antibody (1:10,000) and comparison with negative control (C, BL21(DE3)/pET28a⁺). IPTG induced and non-induced samples are indicated by (I) and (NI). \leftarrow , \leftarrow & \leftarrow and \leftarrow arrows show location (expected location in CB gels) of full-length, truncated version and aggregates forms of hCaf1R^T, respectively. L, protein size marker indicating molecular mass in kDa.

iii) Construction of pEThCaf1R_N

With BL21(DE3)/pEThCaf1R, degradation and aggregation of hCaf1R^T were the main problems. It was considered that the C-terminal domain of Caf1R, which putatively acts as a sensing or oligomerisation domain, might be responsible for degradation and aggregation of hCaf1R^T. Therefore it was decided to express only the N-terminal DBD of hCaf1R^T, with the aim of enhancing recovery of soluble hCaf1R^T. The construct, pEThCaf1R_N contains the sequence encoding the complete N-terminal DBD and His₆-tag in pET28a⁺ vector (**Fig. 4.8**). Briefly, the *caf1R_N* fragment (377 bp) was PCR amplified from pBADhCaf1R template using Infusion primers, pET*caf1RNf** and pET*caf1RNr** as described in section 2.2.4 (i). The purified PCR product was Infused into pre-linearised (NheI/EcoRI digestion) pET28a⁺ vector and transformed into competent *E. coli* StellarTM cells. Plasmid DNA was prepared from eight individual transformants and screened by sequencing using sequencing primers, T7F and T7R. Successful Infusion was confirmed in all transformants. All were stocked and plasmid DNA from transformant 8 was used to test hCaf1R_N expression in *E. coli* BL21(DE3) cells.

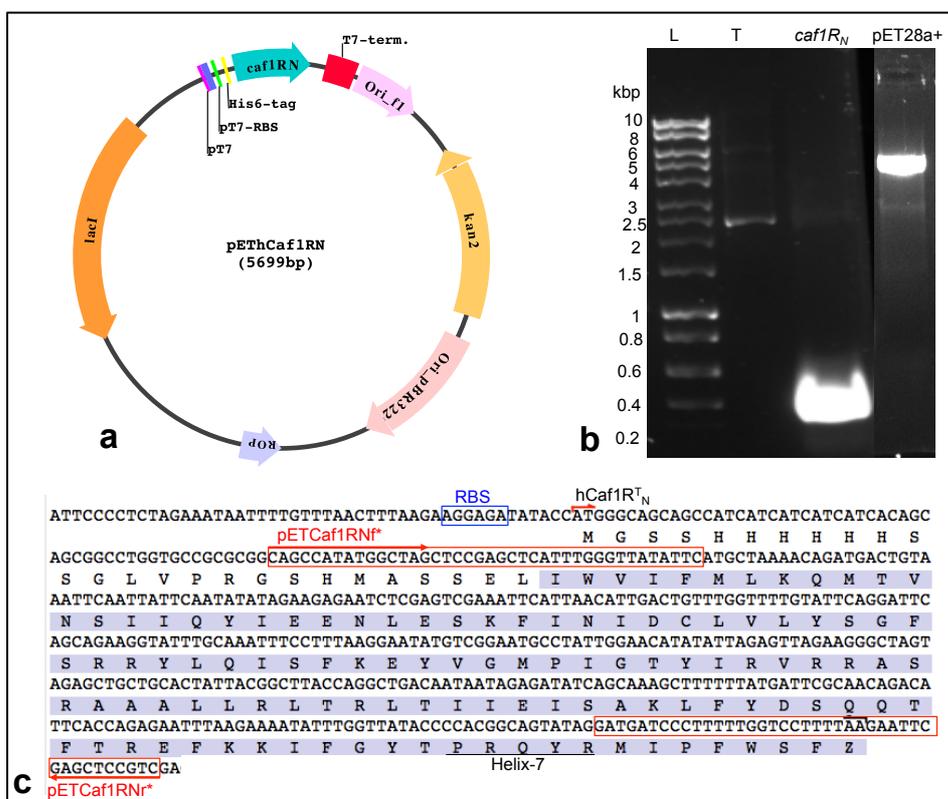


Figure 4.8| Construction of pEThCaf1R_N.

a) Circular plasmid map of pEThCaf1R_N (5699 bp). **b**) AGE analysis (0.8% agarose). T, pBADhCaf1R template (50 ng); *caf1R_N*, purified PCR-product of *caf1R_N* (354 bp, 750 ng); pET28a⁺, NheI/EcoRI digested linearised plasmid (5369 bp, 300 ng) and L, DNA Hyper ladder-I (20ng). **c**) Confirmed sequence of hCaf1R_N^T (144 aa);

MWt 16.86 kDa). Translated sequence in blue is derived from *caf1R*. Infusion primers and RBS are enclosed in red and blue boxes, respectively. The last α -helix (Helix-7) of Caf1R DBD is underlined.

iv) Expression of hCaf1R^{T_N} at 27 and 37°C with 1.0 mM IPTG induction

A small-scale (10 ml) expression trial was carried out with *E. coli* BL21(DE3) cells containing pEThCaf1R_N (pETR_N) or pET28a⁺ (C) plasmid. Expression was tested at 27 and 37°C with 1.0 mM IPTG induction. Induced cells were harvested at 2, 4, 6, 8 and 22 h post-induction. One set of recovered cells (1 OD) was mixed in 100 μ l of HisTrapFF binding buffer and sonicated. The other set was lysed by Bug-buster master mix. Following cell lysis, the S and P fractions from the sonicated samples were recovered by ultracentrifugation and from bug-busted samples by routine centrifugation at 15,000 rpm/20 min/RT. WB identified a strong band of hCaf1R^{T_N} in pellet (P) fractions from both samples (**Fig. 4.9**), corresponding to a possible band in CB stained gel. However, there was negligible hCaf1R_N in supernatant (S) from either set of samples indicating aggregation of the N-terminus alone. Aggregation was evident in samples treated with bug buster and clarified at low speed of centrifugation. Extensive aggregation of hCaf1R_N was visible at the top of the gel. In conclusion, the production of soluble protein was no better for hCaf1R^{T_N} compared to hCaf1R^T, irrespective of the temperature and IPTG concentration tested.

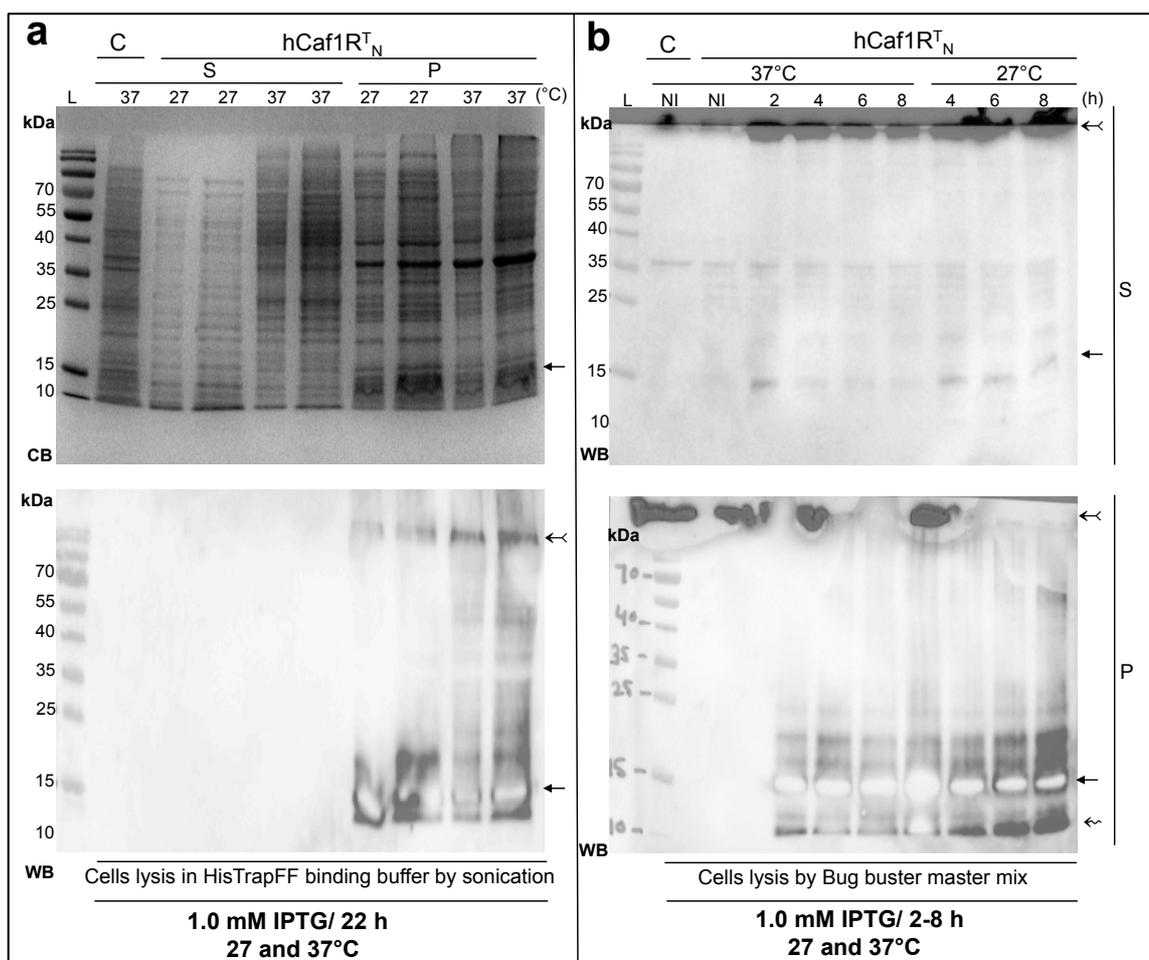


Figure 4.9| The hCaf1R^{T_N} is insoluble following expression at both 27 and 37°C.

E. coli BL21(DE3)/pEThCaf1R_N was induced with 1 mM IPTG at either 27 or 37 for 0-8 h. Samples were processed by sonication and ultracentrifugation (**a**) or by bugbuster and bench top centrifugation (**b**) See text for details. Coomassie Blue (CB) stained SDS-PAGE (14% acrylamide) and Western blot (on PVDF, WB) shown in (**a**) analysis of soluble (S) and insoluble (P) fractions; (**b**) WB only. NI, non-induced/0 h samples; C, negative control (BL21(DE3)/pET28a⁺) and pETR_N, BL21(DE3)/pEThCaf1R_N. →, ↘ and ↗ arrow indicates full-length,

aggregates and degraded product of hCaf1R^T_N, respectively. Identification of hCaf1R^T_N on WB was confirmed by blotting with monoclonal antiHis-HRP conjugated antibody (1:5000).

4.2.3 Expression of MBPCaf1R and MBPCaf1R_N

Only low-levels of hCaf1R/hCaf1R^T or hCaf1R^T_N were recovered using the pBADHisA and pET28a⁺ vectors. Moreover, the recombinant His₆-tagged Caf1R was mainly in the insoluble form with aggregates and degraded product, especially using pET28a⁺. These problems could be related to misfolding of Caf1R or Caf1R_N. Therefore, it was decided to overexpress Caf1R and Caf1R_N with the soluble protein-tag, maltose-binding protein (MBP). The MBP-fusion expression system has been widely used in the overexpression and purification of AraC/XylS family regulators such as SoxS (Fawcett and Wolf, 1994), Rns (Munson and Scott, 1999, 2000) and PchR (Lin et al., 2013; Michel et al., 2005). It was hoped that the MBPCaf1R and MBPCaf1R_N fusion may enhance solubility and stability of Caf1R and Caf1R_N. A cytoplasmic-expression version of pMAL vector, pMALc2x (NEB) was used to In-fuse *caf1R* and *caf1R_N* fragment downstream of the MBP coding gene, *malE* thus constructing pMALc2-Caf1R and pMALc2-Caf1R_N (**Fig. 4.10**). Caf1R homology with AraC/XylS family regulators (Chapter-3) would be consistent with the ATG (Met) codon being the start codon therefore MBPCaf1R was constructed assuming an MLKQ start. MBPCaf1R_N fusion retained the additional 6 residues, starting IWVIFMLKQ.

i) Construction of pMALc2-Caf1R and pMALc2-Caf1R_N

The 906 and 376 bp fragments of *caf1R* (coding from MLKQ...) and *caf1R_N* (coding from IWVIFMLKQ...) were PCR-amplified from pBADhCaf1R plasmid template using Infusion primer sets, pMAL*caf1R*f and pMAL*caf1R*r, and pMALCaf1RNifF and pMALCaf1RNifR, respectively as described in section 2.2.4 (i). PCR products were gel purified and Infused into pre-linearised (BamHI/EcoRI digested) pMALc2x plasmid by Infusion cloning method (Section 2.2.8(ii)). Plasmid DNA from three transformants of each (23, 24 and 26 in case of pMALc2-MBPCaf1R and 1-3 in case of pMALc2-MBPCaf1R_N) was sequenced with sequencing primers, *malE* for and M13F, confirming Infusion of both *caf1R* and *caf1R_N* with intact fusion joints. All of these transformants were stocked and DNA from transformant 23 (pMALc2-Caf1R) and transformant 1 (pMALc2-Caf1R_N) was used to monitor expression of MBPCaf1R and MBPCaf1R_N, respectively.

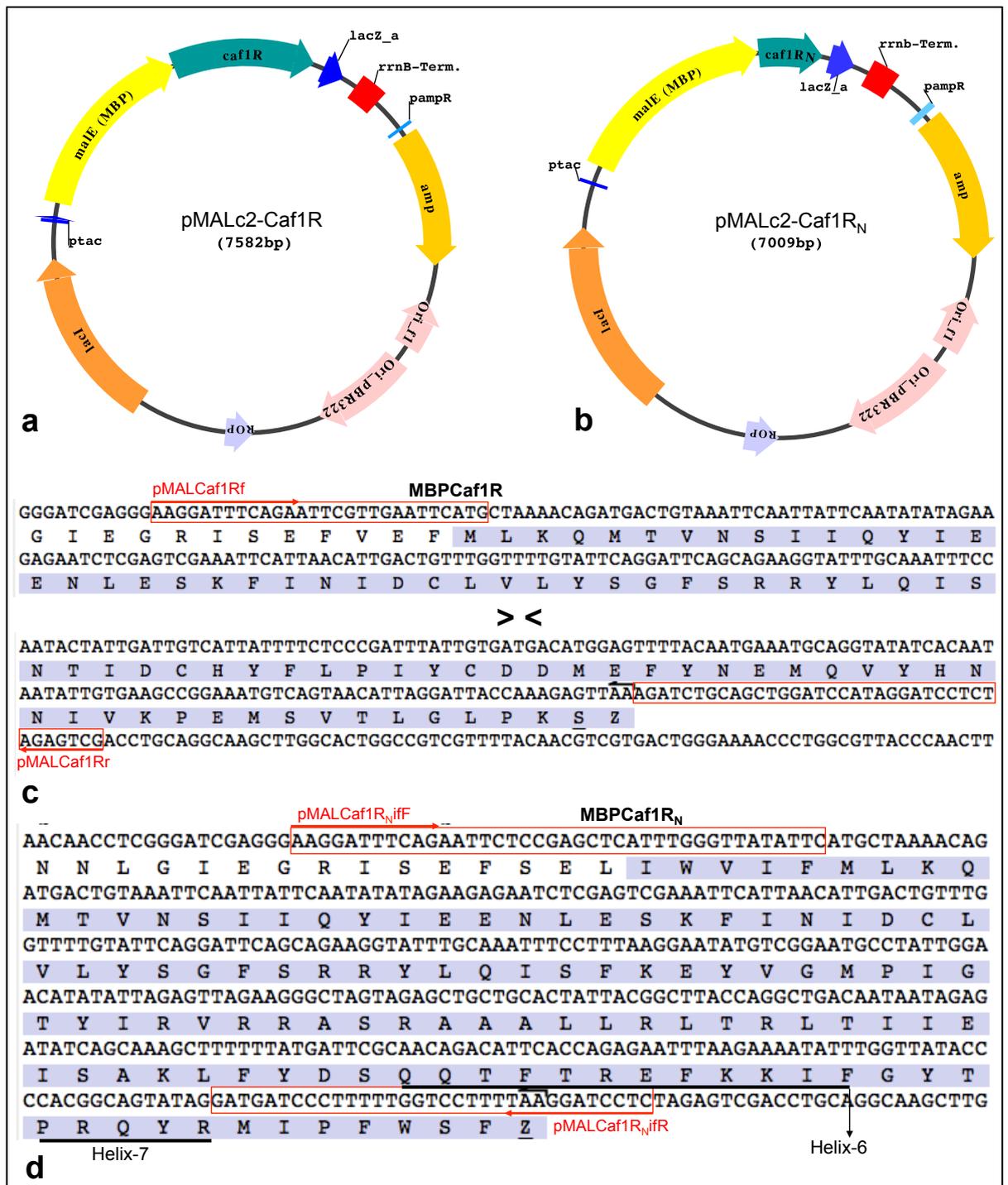


Figure 4.10| **Constructs, pMALc2-Caf1R and pMALc2-Caf1R_N.**
a-b) Circular plasmid map of pMALc2-Caf1R (7582 bp) and pMALc2-Caf1R_N (7009 bp). **c-d)** Confirmed sequences of, pMALc2-Caf1R (696 aa; MWt 79.36 kDa) and pMALc2-Caf1R_N (512 aa; MWt 57.34 kDa), respectively. Translated sequence derived from *caf1R* and *caf1R_N* (highlighted in blue); ><, Represents sequence not shown; Infusion primers on either end (red box). Residues of helix-6 and helix-7 of Caf1R DBD are underlined and annotated in (d).

ii) Expression of MBPCaf1R in *E. coli* BL21 at 37 and 25°C

A small-scale expression (10 ml) was carried out from BL21(DE3) cells carrying pMALc2-Caf1R or pMALc2x (empty plasmid). Induction was initially at 37°C and later at 25°C to test the impact of the two temperatures on recovery of soluble MBPCaf1R. Cells were induced with 0.3 mM IPTG at 37°C/2 h or 25°C/3-9 h. Cells (1 OD unit) were lysed by sonication in 100 µl of MPBTrapHP column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA and 10 mM β-ME; pH 7.45) with protease inhibitor, followed by ultracentrifugation (134,877×g/60 min/4°C). Production of large amounts of MBP (migrating at MWt 48 kDa) expressed from pMALc2x was evident on SDS-PAGE analysis of induced whole cell samples (**Fig. 4.11a**). Increased intensity of a band just above 70 kDa in cells carrying pMALc2-Caf1R would correspond with presence of the MBPCaf1R fusion (MWt 79.30 kDa). The SDS-PAGE profile of the lysed cell supernatant and pellet fractions indicated that this MBPCaf1R fusion was primarily in the pellet fraction with poor recovery of MBPCaf1R in the soluble form. In addition, the appearance of MBP in the MBPCaf1R pellet fraction revealed degradation of the fusion protein at 37°C.

In an attempt to enhance solubility and stability of the fusion protein, constructs were induced at the lower temperature of 25°C. As with 37°C samples, there was a high level of MBPCaf1R in pellet fractions at all time points tested. Interestingly, there was no evidence, on CB staining or immunoblots of degradation of the fusion (**Fig. 4.11b (iii-iv)**). Immunoblotting of the lysate supernatants revealed increasing levels of soluble MBPCaf1R with induction time. However, all samples showed evidence of degradation, suggesting that soluble MBPCaf1R fusion is more susceptible to proteolysis. Hence, these results indicate that expression of Caf1R N-terminal DBD fusion with MBP (MBPCaf1R_N) could be a better choice to avoid degradation.

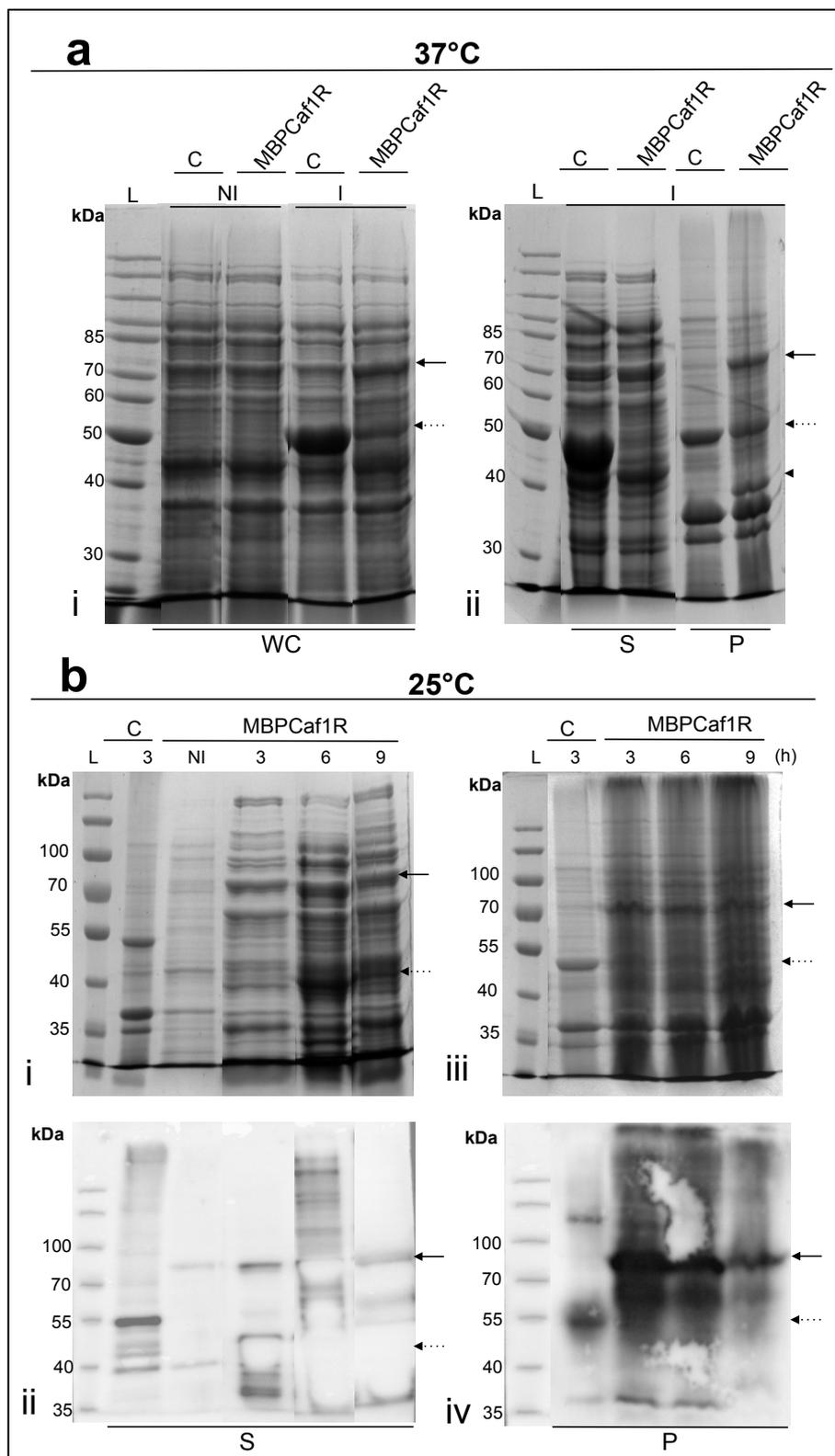


Figure 4.11| MBPCaf1R expression at 37°C and 25°C from *E. coli* BL21(DE3)/pMALc2-Caf1R.

a) Expression at 37°C, SDS-PAGE (12.5% acrylamide) analysis from whole cells (WC), soluble supernatant (S) and insoluble pellet (P) fractions from 37°C/2 h samples. NI and I, non-induced and IPTG induced (0.3 mM) samples. **b)** Expression at 25°C, SDS-PAGE and WB (on NCM) analysis from (S) and (P) fractions, recovered after 3, 6 and 9 h induction with 0.3 mM IPTG as indicated. Band for full-length MBPCaf1R at about 79.3 kDa and MBP alone (C, BL21(DE3)/pMALc2x.) at about 48 kDa is indicated by ← and ←- arrow, respectively. Degraded product in the pellet fraction from 37°C sample is indicated by arrowhead. Identity of MBPCaf1R fusion on WB was confirmed by blotting with primary monoclonal antibody, antiMBP (1:10,000) followed by blotting with secondary polyclonal antibody, antiRabbit-HRP (1:10,000).

iii) **Expression of MBPCaf1R_N in *E. coli* BL21 at 37 and 28°C**

One OD unit cells of *E. coli* BL21 (DE3)/pMALc2-Caf1R_N were recovered before and after IPTG (0.3 mM) induction at 37°C/2 and 3 h and at 28°C/4, 6 and 8 h. Recovered cells were mixed in 100 µl of MBPTrapHP buffer supplemented with 1× proteases inhibitor (EDTA-plus). Soluble (S) and insoluble (P) fractions were then isolated by sonication followed by ultracentrifugation (134,877×g/60 min/4°C). Following SDS-PAGE and CB staining (**Fig. 4.12, CB**) a very faint distinguishing band was observed in all samples just above 55 kDa, consistent with the size of MBPCaf1R_N fusion (57.3 kDa). This was confirmed by Western blot with anti-MBP, with reaction of a small band at approximately 57 kDa, running just above 55 kDa MBP band in the control sample (**Fig. 4.12, WB**). The relative level of MBPCaf1R_N fusion was slightly higher in the pellet fractions at both tested temperatures, suggesting no significant effect of lower temperature on the recovery of more soluble MBPCaf1R_N in the supernatant fractions. Although unlike MBPCaf1R fusion, in above section there was no degradation of MBPCaf1R_N, suggesting MBPCaf1R_N fusion is more stable than full-length MBPCaf1R.

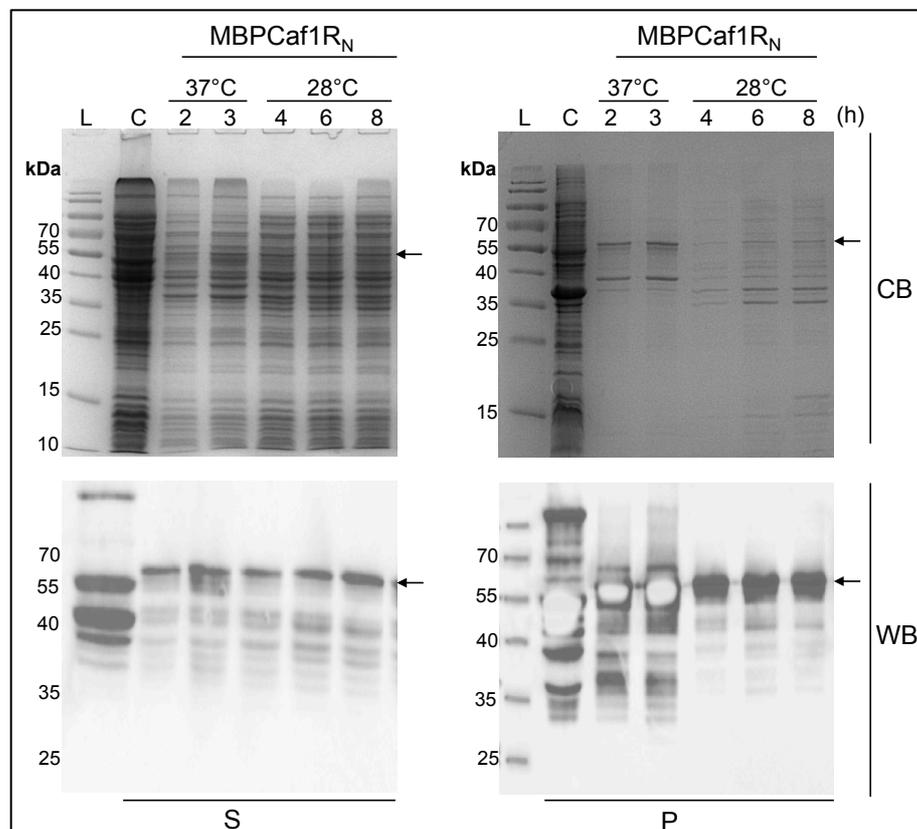


Figure 4.12| **MBPCaf1R_N expression from *E. coli* BL21(DE3)/pMALc2-Caf1R_N.**

SDS-PAGE (12.5% acrylamide; CB) and Western blot (on PVDF; WB) analysis of MBPCaf1R_N expression at 37°C and 28°C with 0.3 mM IPTG induction at various time points as indicated. Soluble (S) and insoluble (P) fractions were prepared by sonicating 1 OD unit cells in 100 µl of MBPTrapHP column buffer with proteases inhibitor followed by ultracentrifugation (50,000 rpm/60 min/4°C). Identity of MBPCaf1R_N band in each case was confirmed by blotting with antiMBP-HRP monoclonal antibody (1:15,000; NEB) and comparison with MBP alone (C) from BL21(DE3)/pMALc2x empty plasmid. ← indicates full-length MBP-R_N. NI, non-induced sample and L, pre-stained protein marker.

iv) Expression of MBPCaf1R and MBPCaf1R_N in *E. coli* K12-ER2508

In order to enhance the level of soluble MBPCaf1R and MBPCaf1R_N and to relieve degradation problems, an alternate *E. coli* strain, K12-ER2508, recommended for MBP-fusion protein expression (NEB) was tested. Briefly, expression of soluble fusion protein from this strain was tested at 37°C with 0.3 mM IPTG ±1% glucose¹⁰ induction for 7 h. At hourly intervals one OD unit cells was resuspended in 250 µl of MBPTrapFF column buffer supplemented with 1× cocktail cOmplete protease inhibitor (Roche), sonicated and centrifuged (20,000 rpm/30 min/4°C) in order to isolate soluble supernatant (S) and insoluble pellet (P) fractions. Both S and P fractions were analysed by SDS-PAGE (**Fig. 4.13**). With this strain of *E. coli*, the level of MBPCaf1R and MBPCaf1R_N was much higher in the soluble (S) fractions than in insoluble fractions (P). In addition, unlike MBPCaf1R expression in BL21(DE3), no degradation was observed from this strain and the basal level of expression (in non-induced samples; NI) was found negligible compared to BL21(DE3) both in the presence and absence of glucose. Inclusion of glucose produced better growth and slightly more fusion-protein/OD, but also more host background protein. It was noticed that the longer the induction time the more fusion protein ended up in the insoluble P fractions. For large-scale production and purification of soluble MBPCaf1R and MBPCaf1R_N, the shortest convenient induction time (i.e. 3-5 h) is probably best whether with or without glucose. In conclusion, promising results for recovery of soluble MBPCaf1R and MBPCaf1R_N were obtained by using the recommended host strain.

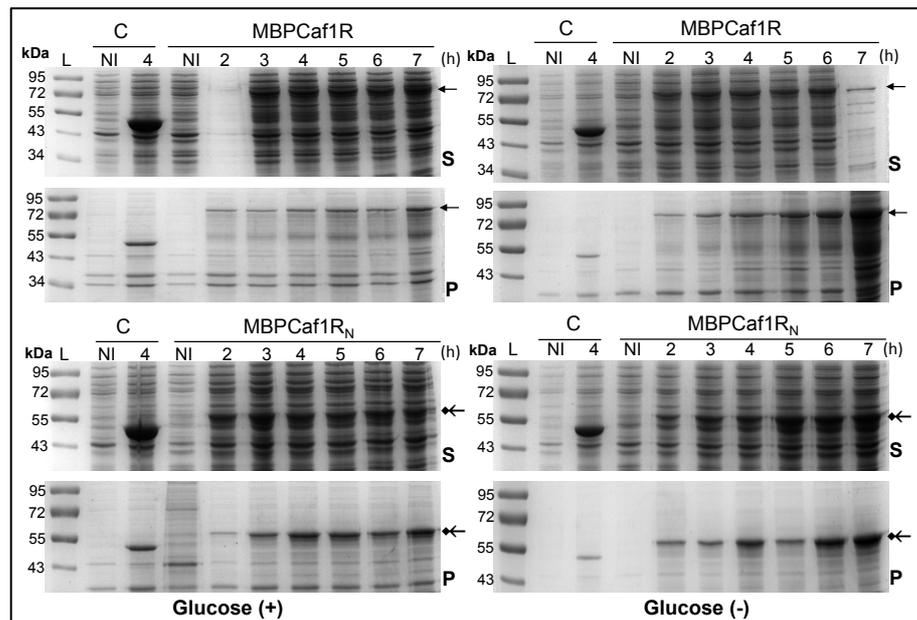


Figure 4.13| MBPCaf1R and MBPCaf1R_N expression from *E. coli* K12-ER2508.

CB stained SDS-PAGE (12.5% acrylamide) from 21k×g/30 min soluble (S) and insoluble (P) fractions of MBPCaf1R and MBPCaf1R_N in ± glucose (1%). NI, non-induced and 2-7 h, samples harvested after 2-7 h of IPTG (0.3 mM) induction. ← and ← arrows indicate the corresponding location of MBPCaf1R and MBPCaf1R_N fusion. Expression of MBP (C) from cells carrying empty pMALc2x plasmid can be seen at about 45 kDa.

¹⁰ Plus glucose would reduce basal level of expression as well as eliminate IPTG-mediated toxicity if any.

4.3 Expression and purification of tagged Caf1R from the synthetic gene

Due to difficulties faced during expression of native tagged Caf1R it was thought that one of the problems could be related to the presence of rare codons in the *caf1R* open reading frame (ORF). If a protein ORF contains rare codons, not often used by *E. coli*, then overexpression of that protein in *E. coli* may be severely diminished to a point of being undetectable (Rosano and Ceccarelli, 2009, 2014). In particular, the codons for Arg (R) (AGG, AGA, and CGA), Leu (L) (CTA), Ile (I) (ATA) and Pro (P) (CCC) can be a problem particularly if two of these codons are present together or in a tandem repeat (Rosano and Ceccarelli, 2009, 2014). Analysis of *caf1R* ORF (relative to **MLKQ...** start) by rare codons finder program, RaCC (<http://nihserver.mbi.ucla.edu/RACC/>) identified 32 rare codons (**Fig. 4.14**). Twenty out of thirty four code for (R), and among these eight make a tandem double repeat. Similarly, eleven rare codons for Ile, two for Leu and one for Pro throughout *caf1R* ORF suggest this could contribute to lower level of expression of native tagged Caf1R thus far. These predictions were further confirmed by another rare codons analysis tool from GenScript, USA (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis). The codon usage frequency for the expression of native *caf1R* in *E. coli* host strains was found to be 0.6, nearly half of the recommended 1.0 (GenScript, USA). Plasmid constructs, pEThCaf1Rgs, containing the codon optimised *caf1R* gene (*caf1Rgs*) (**Fig 4.14**) and also the variant E98G mutant, pEThCaf1Rgs^{E98G} were purchased from GenScript, USA. In total 162 codons of *caf1R* were optimised to obtain the recommended codon usage frequency and optimum GC content for overexpression in *E. coli* host strains. In this construct *caf1Rgs* (starting with **MLK...** codons) was subcloned downstream of the N-terminal His₆-tag between NheI and EcoRI sites of the pET28a⁺ vector. This codes for hCaf1R^{Tgs} (325 aa) with a calculated molecular weight of 38.5 kDa (**Fig. 4.15**). In the following sections only the WT version of *caf1Rgs* was used to optimise over-expression of tagged Caf1Rgs from pET28a⁺ (hCaf1R^{Tgs}), pBADHisA (hCaf1Rgs) and pMALc2x (MBPCaf1Rgs) plasmids. For subsequent purification, only the pEThCaf1Rgs (hCaf1R^{Tgs}) and pMALc2-MBPCaf1Rgs (MBPCaf1Rgs) were used. The mutant version (pEThCaf1Rgs^{E98G}) was used in the lysate assay for binding activity as described in Chapter-5.

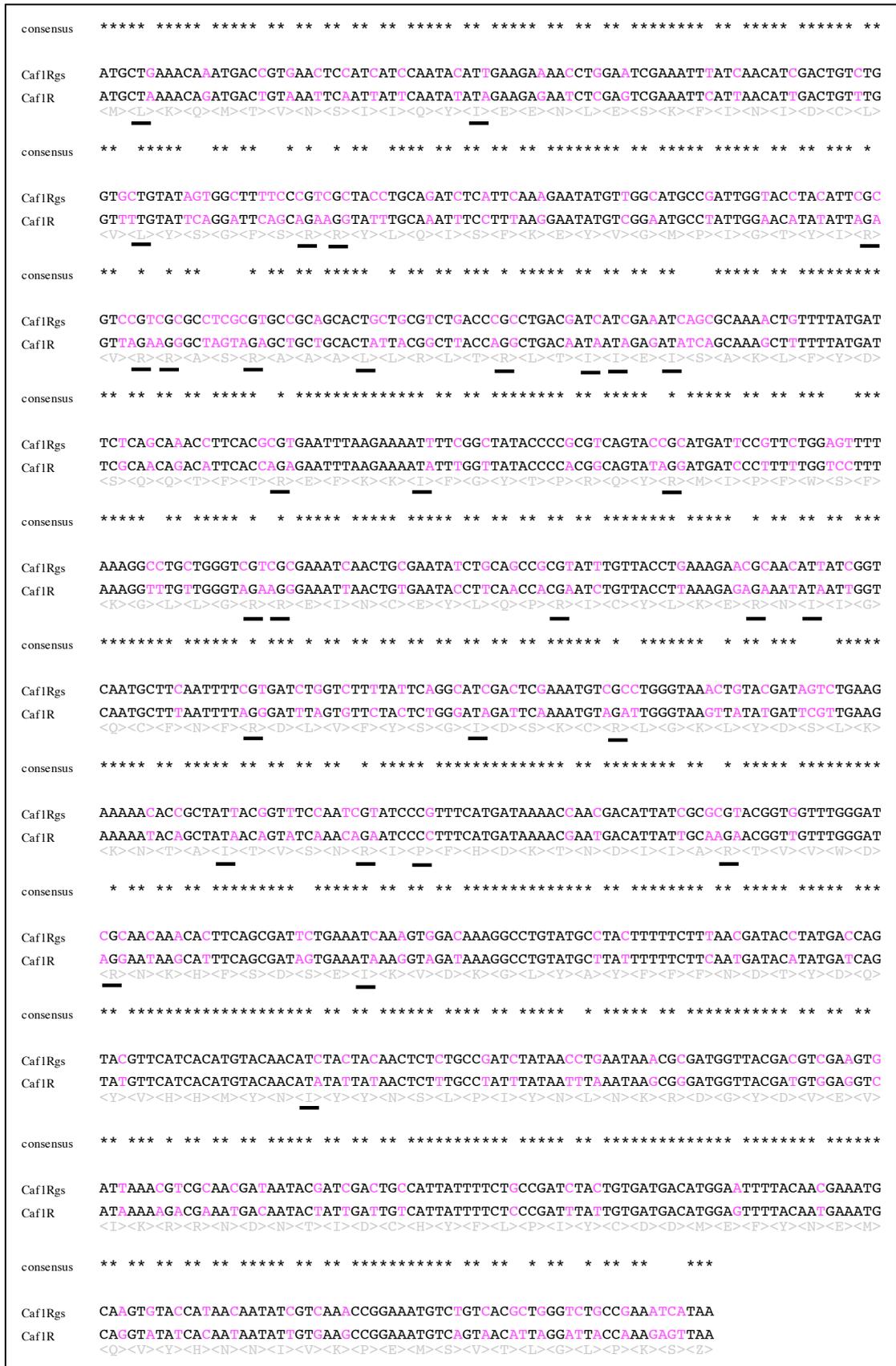


Figure 4.14| Pairwise sequence alignment of native *caf1R* with synthetic gene (*caf1Rgs*). Rare codons identified by RaCC calculator are indicated by black underlines (32 amino acids). Changes made by GenScript, USA throughout *caf1R* ORF are indicated in pink (232 nucleotides encoding 162 amino acids). Consensus nucleotides are indicated by *. Deduced amino acid sequence is indicated underneath of alignment. DNADynamo was used to generate sequence alignment.

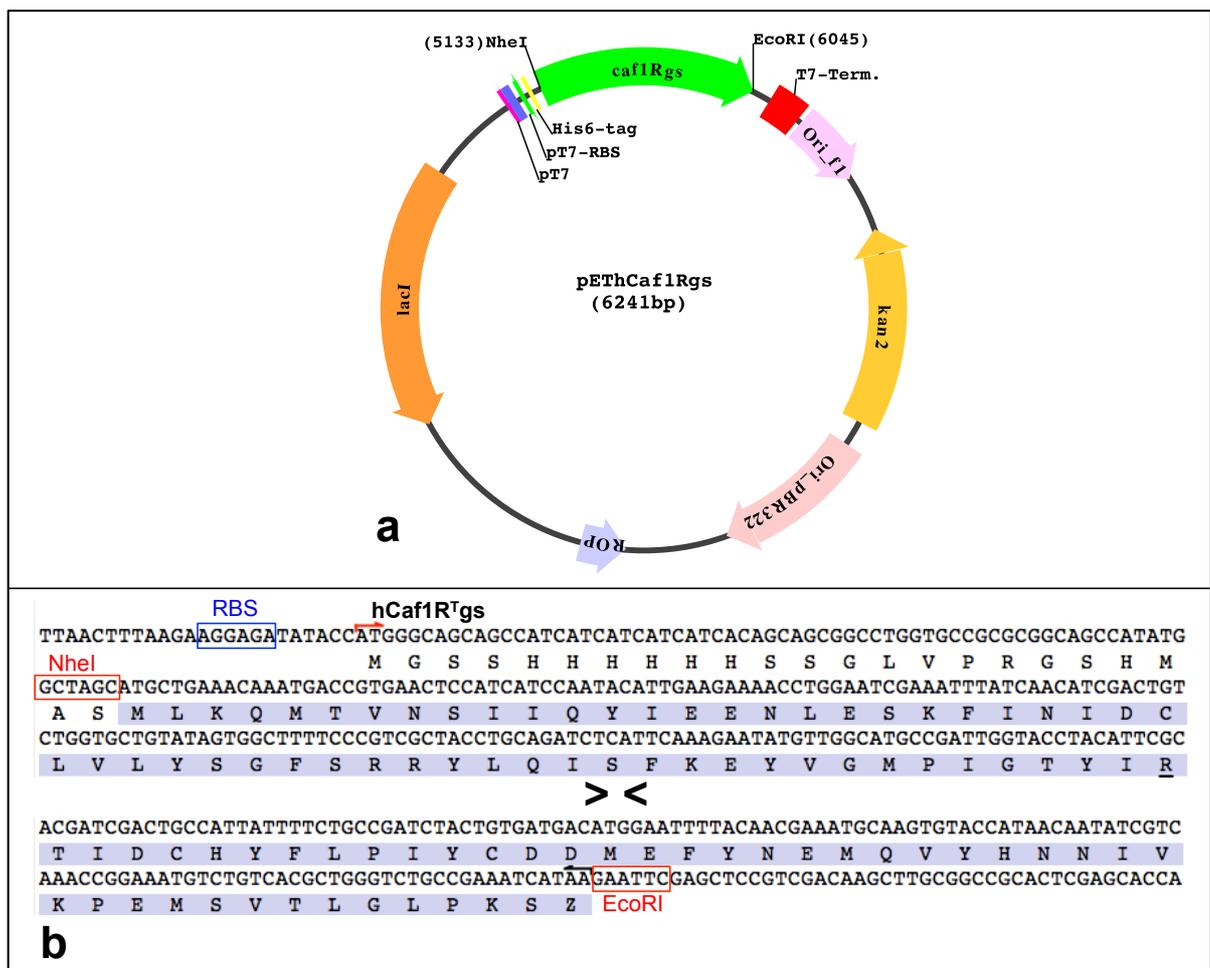


Figure 4.15| The pETHCaf1Rgs construct.

a) Circular plasmid map of synthetic Caf1R cloned in pET28a⁺ to create pETHCaf1Rgs (6241 bp). Plasmid purchased from GenScript, USA. **b).** Sequence of *caf1Rgs* in between NheI/EcoRI sites (red boxes) of pET28a⁺. The *caf1Rgs* derived amino acid sequence is highlighted in blue with upstream fusion to His₆-tag (hCaf1R^{Tgs}, 325 aa, MWt 38.5 kDa). Ribosomal binding site, RBS is indicated in blue box.

4.3.1 Expression and purification of hCaf1Rgs

i) Expression of hCaf1R^{Tgs} from pETHCaf1Rgs with different temperatures, IPTG concentrations and *E. coli* host strains

Initially hCaf1R^{Tgs} expression was monitored at 37°C using four different *E. coli* strains, BL21(DE3), Rosetta gami™ 2(DE3), LOBSTR-BL21(DE3) and LOBSTR-BL21(DE3)-RIL with 0.35 mM IPTG induction. Cells were recovered after 1.5, 3, 4.5, 6 and 18 h induction. Following SDS-PAGE and WB analysis, no hCaf1R^{Tgs} band was observed at the expected location, about 38.5 kDa. This was attributed to IPTG-mediated toxicity (data not shown). From this point onward, cells with plasmid containing the synthetic gene *caf1Rgs* were always grown in the presence of 1% (w/v) glucose irrespective of induction. Expression of hCaf1R^{Tgs} was re-tested in BL21(DE3)/pETHCaf1Rgs induced with glucose plus 0.35 mM IPTG at 37°C for 2-18 h. A clearly induced protein, corresponding to the expected location of hCaf1R^{Tgs} (38.5 kDa) was present in all induced samples of whole cells carrying pETHCaf1Rgs. A very thick band of hCaf1R^{Tgs} was present in the lysed cell pellet (P) fractions (**Fig. 4.16, P**), reflecting a very high level of expression but again recovery primarily in the insoluble form. There was no distinguishable

difference from the non-induced and induced supernatant samples, although any soluble protein could be masked by a host protein migrating in the same location (38.5 kDa).

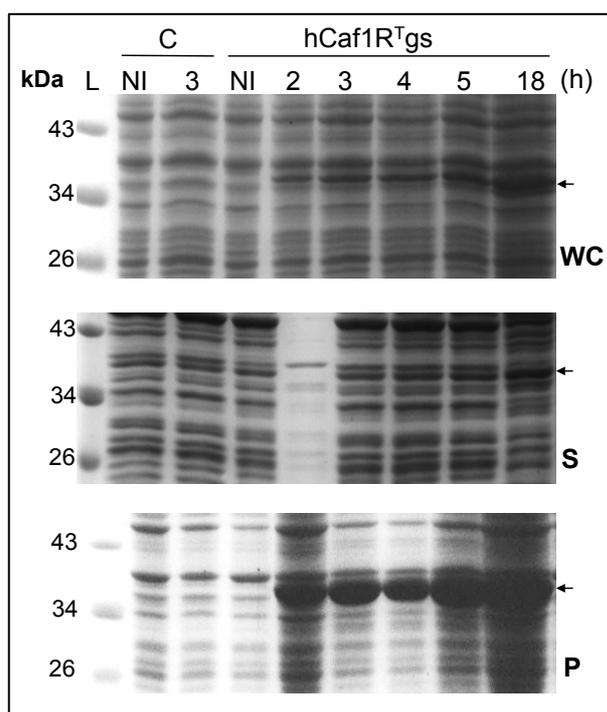


Figure 4.16| **Expression of hCaf1R^Tgs from *E. coli* BL21(DE3)/pEThCaf1Rgs.**

SDS-PAGE (14% acrylamide) analysis of hCaf1R^Tgs expression following induction with 0.35 mM IPTG + 1% glucose at 37°C (from 2-18 h). One OD unit of induced cells was mixed in 100 µl PBS for analysis of whole cells (WC) or in 100 µl Bugbuster master mix followed by centrifugation (20,000 rpm/15 min/4°C) to give soluble supernatant (S) and insoluble pellet (P) fractions. NI, non-induced samples; L, protein size marker. Arrows indicate hCaf1R^Tgs (about 38.5 kDa). All gels stained with Coomassie Blue.

Because of the very high levels of insoluble hCaf1R^Tgs recovered following induction at 37°C, expression of soluble hCaf1R^Tgs, from *E. coli* BL21(DE3)/pEThCaf1Rgs, was monitored at lower temperatures of 15 - 25°C with either 0.1 and 0.5 mM IPTG, and 1% glucose. Following induction at 21 or 25°C, a major prominent band of hCaf1R^Tgs was detected on SDS-PAGE analysis of the insoluble pellet fractions of all samples (**Fig 4.17a**). As before no hCaf1R^Tgs specific band could be identified in CB stained supernatant fractions and only a small band of hCaf1R^Tgs was identified by immunoblotting. Thus at 21 or 25°C, virtually all of the expressed hCaf1R^Tgs was recovered as insoluble protein. However, cells induced with 0.1 mM and 0.5 mM IPTG at 21 °C for 6 h did appear to have slightly more soluble hCaf1R^Tgs than either later time points or 25°C samples. When the induction temperature was decreased further to 18 and 15°C, as expected protein production slowed down (**Fig 4.17b, CB-P**), but while hCaf1R^Tgs was the major protein in the insoluble pellet fraction, it was still not identifiable in CB stained supernatant fractions. As with 21°C induction, immunoblots identified slightly more soluble protein at the shortest sampling time (4 h), but aggregates also accumulated at the top of the gel. Hence, while decreasing the rate of production of His₆-tagged Caf1R from pEThCaf1Rgs substantially improved relative recovery of soluble hCaf1R, it did not greatly enhance overall recovery of soluble protein.

Further two strains of *E. coli*, BL21 Star™ (DE3)pLysS and SHuffle® T7 were tested for the recovery of both soluble and insoluble level of hCaf1R^Tgs. Expression was tested from BL21 Star™(DE3)pLysS at 37°C with 1.0 mM IPTG induction and from SHuffle® T7 at 37, 30 and 16°C with 0.1 mM IPTG induction. BL21 Star™ (DE3) pLysS showed very tight control of hCaf1R^Tgs expression with no expression in non-induced samples (**Fig. 4.18a, WB**). There was even less expression from this strain, with product only visible in both soluble and insoluble pellet fractions by immunoblotting. Similarly, with the SHuffle® T7 strain there was no distinguishable CB stainable product at the expected location of hCaf1R^Tgs, in any of the supernatant fractions tested (**Fig. 4.18b, CB**), despite a large amount of product in pellet fractions (**Fig. 4.18b, CB and WB**). During expression studies extensive aggregation of Caf1R in 21,000×g/15 min supernatant fractions was often noticed. This appeared to be particularly prominent in bugbuster lysed cells. Prior to purification, ultracentrifugation would remove much of this aggregate.

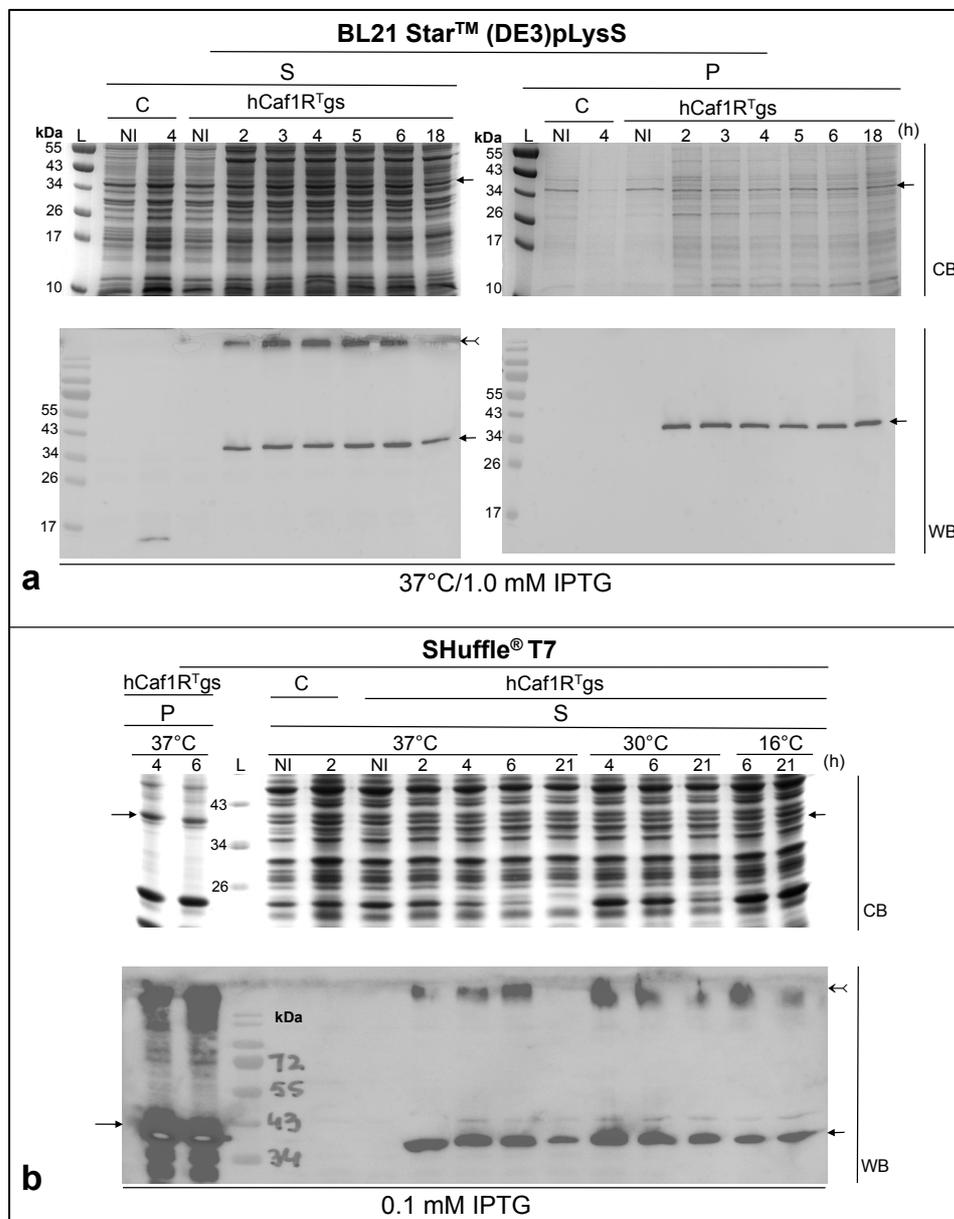


Figure 4.18 | Expression of hCaf1R^Tgs from *E. coli* BL21 Star™(DE3) pLysS and SHuffle® T7 carrying pEThCaf1Rgs plasmid.

SDS-PAGE (14% acrylamide; CB) and Western blot (WB, on NCM) of hCaf1R^Tgs expression from BL21 Star (DE3) pLysS at 37°C with 1.0 mM IPTG induction **(a)** and from SHuffle[®] T7 at 37, 30 and 16°C with 0.1 mM IPTG induction **(b)** at different time (h) as indicated. Soluble (S) and pellet (P) fractions were prepared by mixing 1 OD unit of recovered cells in 100 µl of Bugbuster master mix followed by centrifugation (20,000 rpm/15 min/4°C). Identity of hCaf1R^Tgs on WB was validated by blotting with antiHis-HRP monoclonal antibody (1:10,000), and comparison with negative control, C (corresponding cells strain containing pET28a⁺ empty vector). ←, ⇐ and ⇨ arrows indicate corresponding location and band of full-length hCaf1R^Tgs, aggregates and degradation products. L, protein size marker.

ii) Expression of hCaf1Rgs from *E. coli* Top10/pBADhCaf1Rgs at 37°C

Expression of hCaf1R from the tightly regulated, low-copy number plasmid pBADhcaf1R had resulted in soluble hCaf1R, but only low levels **(Fig. 4.3)**. Therefore expression from the pBADHis vector was again tested with the *caf1Rgs* gene. The gene *caf1Rgs* was excised from pEThCaf1Rgs (by NheI/EcoRI digestion) and ligated between the NheI and EcoRI sites of pBADHisA plasmid to construct pBADhCaf1Rgs. This construct encodes N-terminal His₆-tagged *caf1Rgs* (hCaf1Rgs) with a calculated molecular weight 37.55 kDa, as depicted in **Fig. 4.19**. Small scale (10 ml) expression of hCaf1Rgs from this construct, was tested in *E. coli* Top10 with a range (0-0.2%) of L-ara induction at 37°C until 8 h as described in section 2.5.1 and shown in **Fig. 4.20**. Following SDS-PAGE, no hCaf1Rgs band could be identified by CB staining of either supernatant or pellet fractions of the lysed cells; an apparently induced protein band just above 34 kDa in supernatant fractions was also present in the induced negative control (C; Top10/pBADHisA) **(Fig. 4.20, CB)**, suggesting expression of hCaf1Rgs (37.55 kDa) is masked by a host protein. Western blot confirmed recovery of reasonable levels of hCaf1Rgs in supernatant fractions following induction at all L-ara concentrations tested. The 0.02% L-ara induction at 4 h produced a reasonably good amount of full-length hCaf1Rgs with little or no aggregate, as did induction with 0.002% L-ara **(Fig 4.20, WB)**. Hence, production of soluble hCaf1R is slightly improved from this construct when compared to production from pBADhCaf1R with the native *caf1R* sequence. However, the level of soluble product is still low.

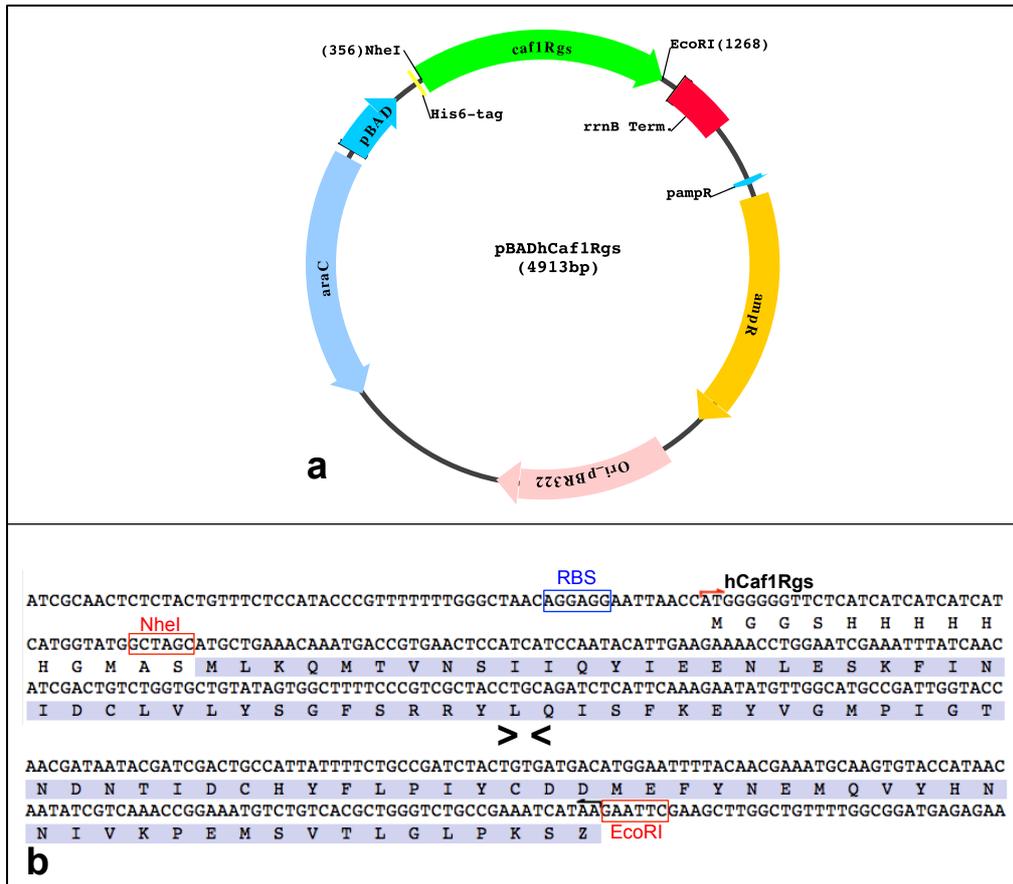


Figure 4.19| The pBADhCaf1Rgs construct.

a) Circular plasmid map of pBADhCaf1Rgs (4913 bp). **b)** Confirmed sequence of pBADhCaf1Rgs encoding hCaf1Rgs. Cloning sites (NheI and EcoRI) and RBS are indicated in red and blue boxes, respectively. N-terminal and C-terminal amino acid sequence of hCaf1Rgs (316 aa, MWt 37.55 kDa) is shown. ><, Indicates sequence not shown. Amino acids highlighted in blue indicate codons derived from *caf1Rgs*, those in black plasmid derived codons.

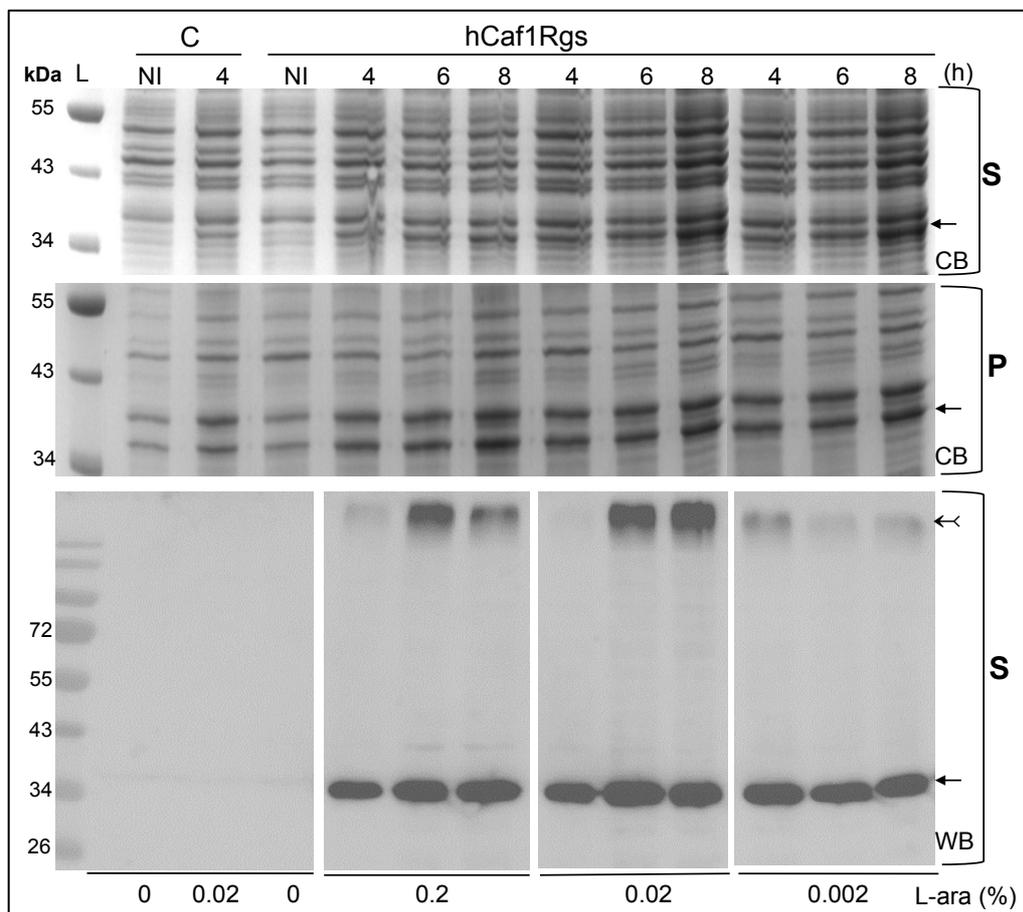


Figure 4.20| **Expression of hCaf1Rgs from *E. coli* Top10/pBADhCaf1Rgs.** SDS-PAGE (14% acrylamide, CB) and Western blot (WB, on NCM) analysis of hCaf1Rgs expression following induction of 10 ml culture with 0-0.2 % of L-ara at 37°C for 4-8 h, as shown. S, supernatant and P, pellet fractions prepared following lysis of 1 OD unit cells in 100 μ l of HisTrapFF column buffer (without imidazole), supplemented with 1 \times proteases inhibitor (EDTA-free, Pierce) and centrifugation (20,000 rpm/15 min/4°C). Identity of hCaf1Rgs on WB was confirmed by blotting with antiHis-HRP monoclonal antibody (1:10,000). ← and ←← arrows indicate the full-length and aggregated hCaf1Rgs. C, negative control (Top10/pBADHisA) and NI, not induced samples. L, protein size marker.

iii) IMAC purification of soluble hCaf1R^Tgs from *E. coli* BL21(DE3)/pEThCaf1Rgs

It was decided that highest levels of recovery of soluble hCaf1R would most likely be obtained from the pEThCaf1Rgs construct. Following, consideration of previous conditions tested and time constraints soluble hCaf1R^Tgs was prepared by sonication and ultracentrifugation of cells of *E. coli* BL21(DE3)/pEThCaf1R^Tgs culture (100 ml) that had been induced with 0.35 mM IPTG + 1% glucose for 2.5 h at 37°C. Purification was carried out by using HisTrapFF column (1 ml) with a linear gradient of 0-250 mM imidazole as described in **Fig. 4.21** and section 2.5.4(i). Elution profile of HisTrapFF column (**Fig. 4.21a**) indicates two very small peaks at Fns 3-5 and 14, but neither correlated with hCaf1R, which was eluted mainly between Fn 7-12, at about 31% of buffer B or 77.5 mM imidazole. Fns 3-6 contained high molecular weight (55-75 kDa) Histidine rich host proteins from host, which also contaminated Fn 7 and 8. No visible background bands were present in Fn 9, 10 and 11 with best recovery and maximum purity in fraction 10 (**Fig. 4.21b**). A Nano drop measurement from this fraction (10th) showed 40 μ g/ml concentration of hCaf1R^Tgs, suggesting it would be sufficient for subsequent DNA-protein binding assays upon concentration.

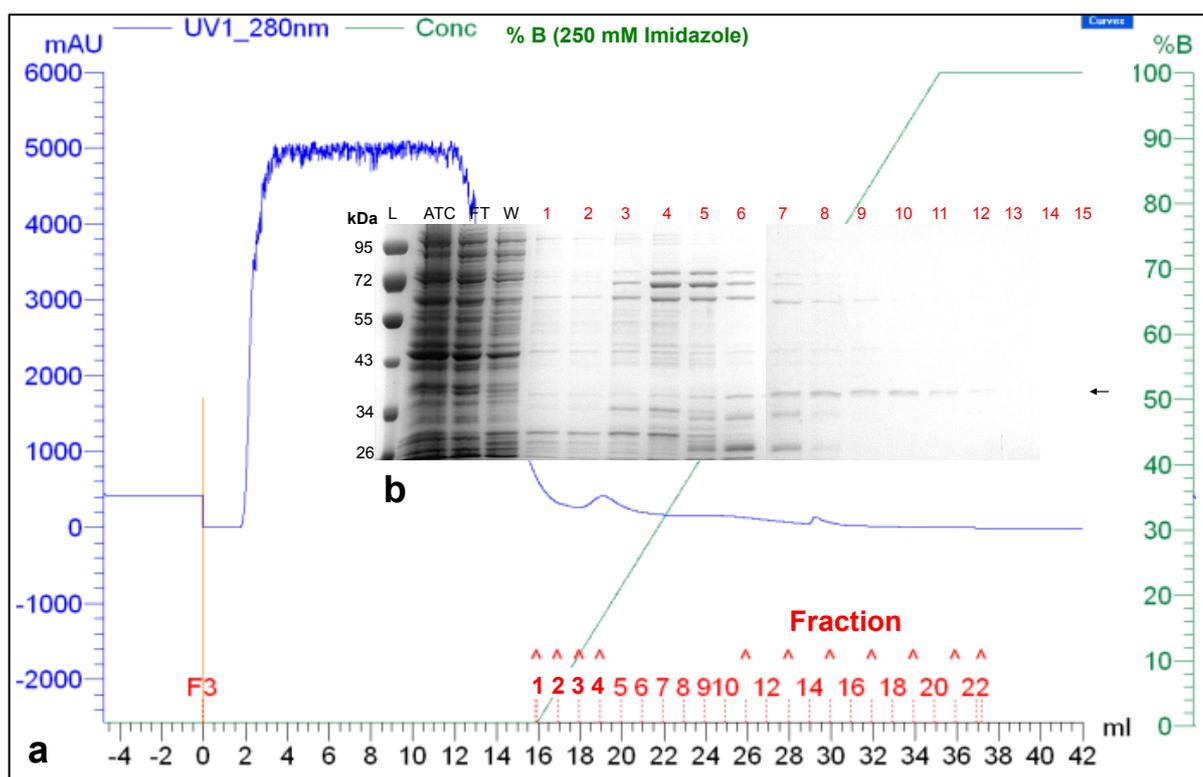


Figure 4.21| **IMAC purification of soluble hCaf1R^Tgs from *E. coli* BL21(DE3)/pEThCaf1Rgs.**

Induced cells of *E. coli* BL21(DE3)/pEThCaf1Rgs (100 ml culture) were lysed by sonication in 10 ml HisTrapFF column binding buffer (20 mM NaPO₄ buffer, 500 mM NaCl, 5 mM DTT, 20 mM imidazole; pH 7.4), supplemented with 1× proteases inhibitor (EDTA-free). The supernatant fraction following ultracentrifugation (50,000 rpm/60 min/4°C) was applied to a 1 ml HisTrapFF column. **a**) Elution profile with a linear gradient, 0-250 mM imidazole over 20 × 1 ml fractions, followed by wash with 250 mM imidazole (100% buffer B). **b**) SDS-PAGE (14% acrylamide) analysis of eluted fractions (1-15). ATC, applied sample to column, FT, flow through and W, wash fraction. ← arrow indicates hCaf1R^Tgs. L, protein size marker.

iv) Urea solubilisation of hCaf1R^Tgs from the inclusion bodies

As most of the recombinant hCaf1R^Tgs was recovered as insoluble inclusion bodies, solubilisation of hCaf1R^Tgs with urea was tested. This might be useful for anti-Caf1R antibody production or refolding of hCaf1R if recovery of soluble protein was too difficult. A batch of hCaf1R^Tgs inclusion bodies was prepared from a 50 ml culture of *E. coli* BL21(DE3)/pEThCaf1Rgs that had been induced with 0.1 mM IPTG at 37°C/5 h. Pelleted inclusion bodies were treated with a 0-8 M urea as described in section 2.5.5 and solubilisation assessed by centrifugation. The highest recovery of soluble hCaf1R^Tgs was with 8 M urea, with approximately 50-60% solubilised (**Fig. 4.22**). Urea solubilisation was not taken any further, but it was concluded that recovery of soluble hCaf1R^Tgs could be optimised either by adjusting the ratio of urea to protein or by using another method of solubilisation such as Guanidine hydrochloride, if needs to be, and optimising refolding conditions.

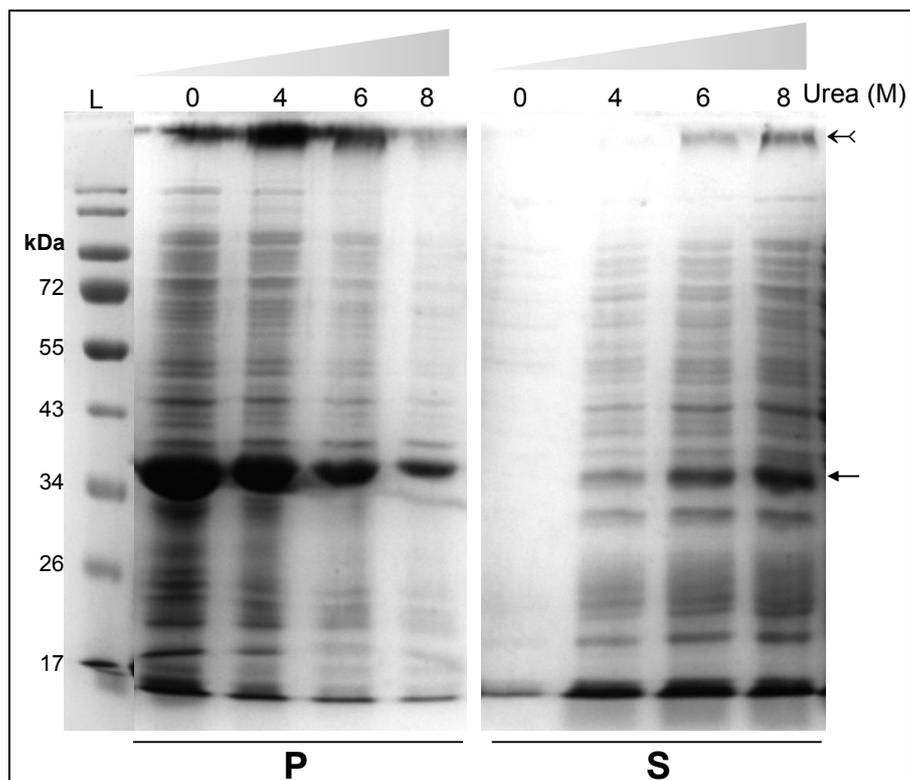


Figure 4.22| Solubilisation of hCaf1R^Tgs from inclusion bodies. of *E. coli* BL21(DE3)/pETHCaf1Rgs by urea treatment.

SDS-PAGE (14% acrylamide) analysis of hCaf1R^Tgs non-solubilised (P) and urea solubilised (S) Caf1R with a range of urea (M) as indicated. ← and ←← arrows indicate full-length hCaf1R^Tgs and aggregates, respectively. Inclusion bodies containing hCaf1R^Tgs were recovered from induced, lysed cells of *E. coli* BL21(DE3)/pETHCaf1Rgs and tested for solubilisation in urea as described in section 2.5.5.

4.3.2 Expression and purification of MBPCaf1R_{gs}

i) Construction of pMALc2-Caf1R_{gs} and expression of MBPCaf1R_{gs}

To optimise the solubility and stability of tagged Caf1R, *caf1Rgs* was amplified from pETHCaf1Rgs template, using Infusion primers, pMALCaf1R_GSf and pMALCaf1R_GSr and subcloned into EcoRI/BamHI digested pMALc2x plasmid using Infusion cloning method. The complete *caf1Rgs* sequence with intact fusion on either side was confirmed from three transformants using *malE* for and M13F sequencing primers. All were stocked and number 1 was assigned as pMALc2-Caf1R_{gs} (Fig. 4.23) and used.

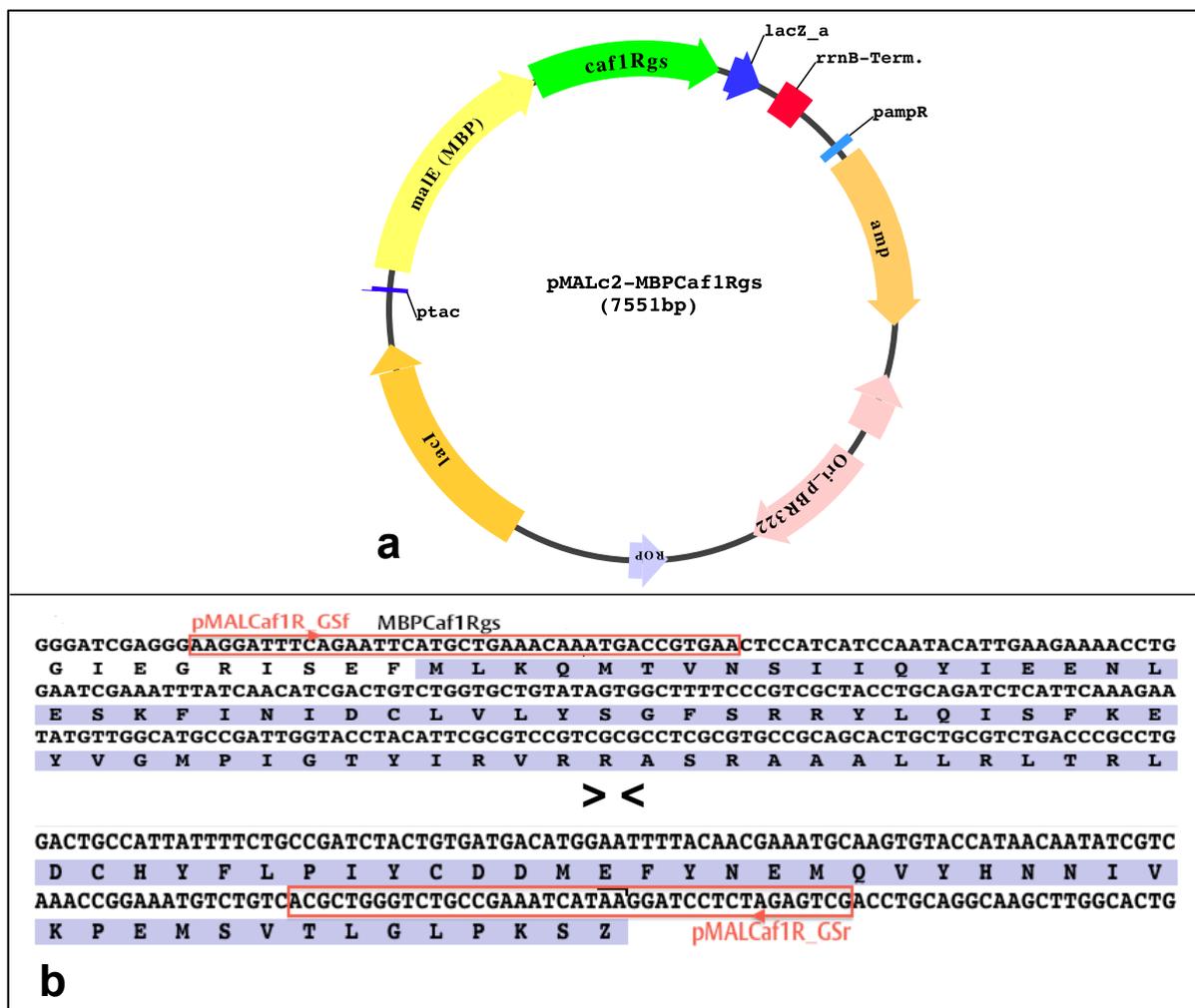


Figure 4.23| The pMALc2-Caf1R_{gs} construct.

a) Circular plasmid map of pMALc2-Caf1R_{gs} (7551 bp). **b)** Confirmed sequence of pMALc2-Caf1R_{gs}. Encoded amino acid sequence derived from *caf1Rgs* is highlighted in blue, fused to N-terminal MBP (only a few linker residues (black) are shown). Infusion primers on either side are shown in red boxes. Complete fusion of MBPCaf1R_{gs} codes 693 aa with a calculated molecular weight, 78.98 kDa.

A small-scale (10 ml) trial of recovery of soluble MBPCaf1R_{gs} over 7 h induction with 0.3 mM IPTG ± 1% (w/v) glucose at 37°C was performed from *E. coli* K12-E2508/pMALc2-MBPCaf1R_{gs}. Induced cells were lysed by sonication in MBPTrapHP column buffer and the supernatant (soluble fraction) recovered by centrifugation (20,000 rpm/15 min/4°C). Following SDS-PAGE analysis (**Fig. 4.24, CB**), a clearly distinguishable band correlating with MBPCaf1R_{gs} fusion (78.98 kDa) was observed between 72 and 95 kDa in all induced samples – supernatant and pellet induced in presence and absence of glucose. Identity of this was confirmed by immunoblotting of supernatant samples. This also highlighted the higher levels of MBPCaf1R_{gs} when induced in the presence of glucose. However, glucose grown cells also contained a much higher level of background host proteins (**Fig. 4.24, WB**). It was also noticed that as culture incubation time increased, the relative ratio of MBPCaf1R_{gs} in pellet fractions also increased. Hence, for maximum recovery of soluble MBPCaf1R_{gs}, with less background, 3-5 h induction time in the absence of glucose would be most suitable.

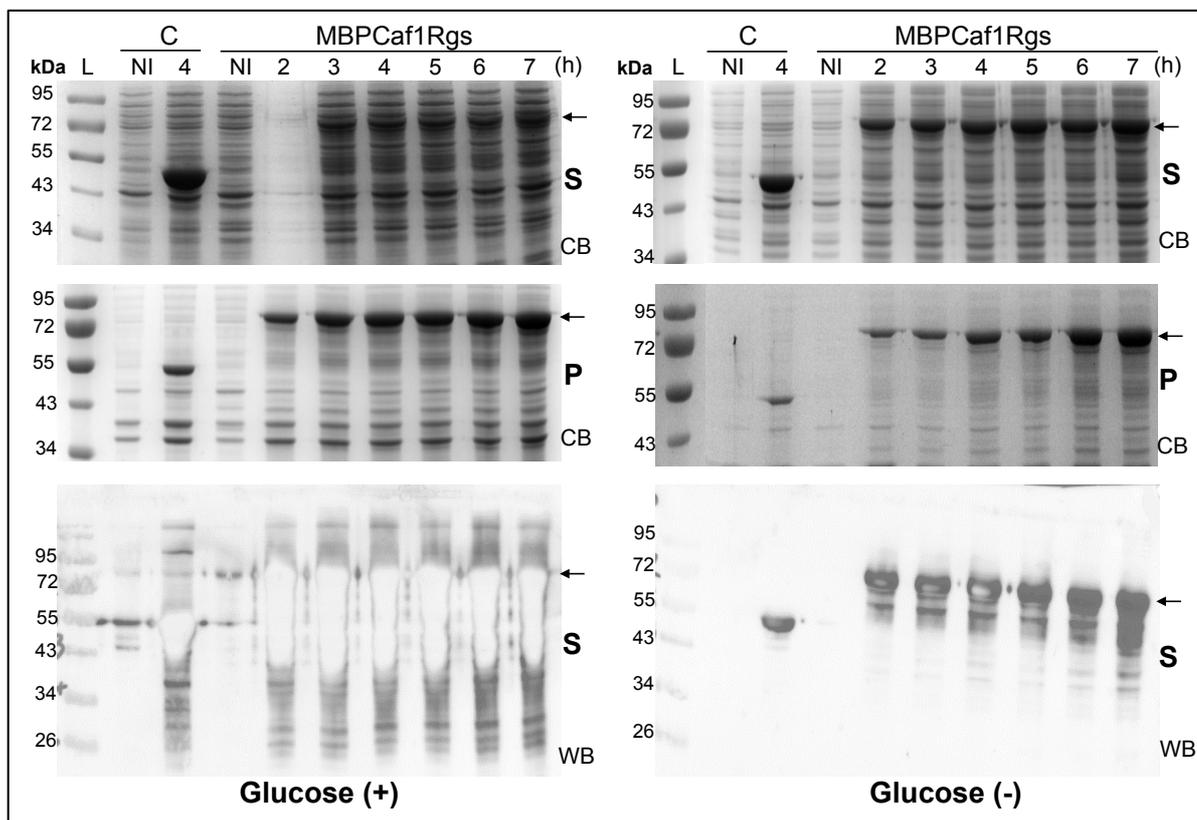


Figure 4.24| **Recovery of soluble MBPCaf1Rgs from *E. coli* K12-ER2508/pMALc2-Caf1R_{gs}.** SDS-PAGE (12.5% acrylamide, CB) and Western blot (WB, on NCM) analysis of MBPCaf1R_{gs}. Cultures were grown at 37°C following induction with 0.3 mM induction ± glucose (1%) with sampling over 2-7 h as indicated. S and P, Soluble and insoluble fractions, prepared from 1 OD unit cells, lysed in 100 µl MBPTrapHP column buffer with protease inhibitor (Roche) followed by centrifugation at 20,000 rpm/15 min at 4°C. NI, non-induced, K12-ER2508/pMALc2-MBPCaf1R_{gs}; C, control with pMALc2x, expresses MBP only (48 kDa). ← arrow, full-length MBPCaf1R_{gs}. MBP in negative control, WB (on NCM) was with 1:10,000 dilutions of both antiMBP (primary) and antiRabbit-HRP (secondary) antibodies.

ii) Maltose affinity purification of MBPCaf1Rgs from *E. coli* K12-ER2508/pMALc2-Caf1R_{gs}

Cell lysate was prepared from 500 ml culture of *E. coli* K12-ER2508/pMALc2-Caf1R_{gs} induced with 0.3 mM IPTG at 25°C for 5 h by sonication in 40 ml of MBPTrapHP column buffer (20 mM Tris-HCl, 200 mM NaCl, 10 mM β-ME, 1 mM EDTA; pH 7.5) supplemented with 1× protease inhibitor (Thermo Scientific, Pierce). Sonicated lysate was ultra-centrifuged (134,877×g/60 min/4°C) and the supernatant was then fractionated on an MBPTrapHP column. The Elution profile from the MBPTrapHP column (**Fig. 4.25a**) indicates a very sharp peak of A₂₈₀ encompassing Fn 2-4, eluting at about 18% of B (1.8 mM maltose). This corresponds to the correct size for MBPCaf1R_{gs} (MWt 78.98 kDa) (**Fig. 4.25b**). The high protein background in the flow through (FT) sample indicates poor binding to column matrix, which could be improved by repeat applications of smaller volumes or using a larger column. Protein concentration in Fn 3 was measured by Nanodrop and found to be 380 µg/ml.

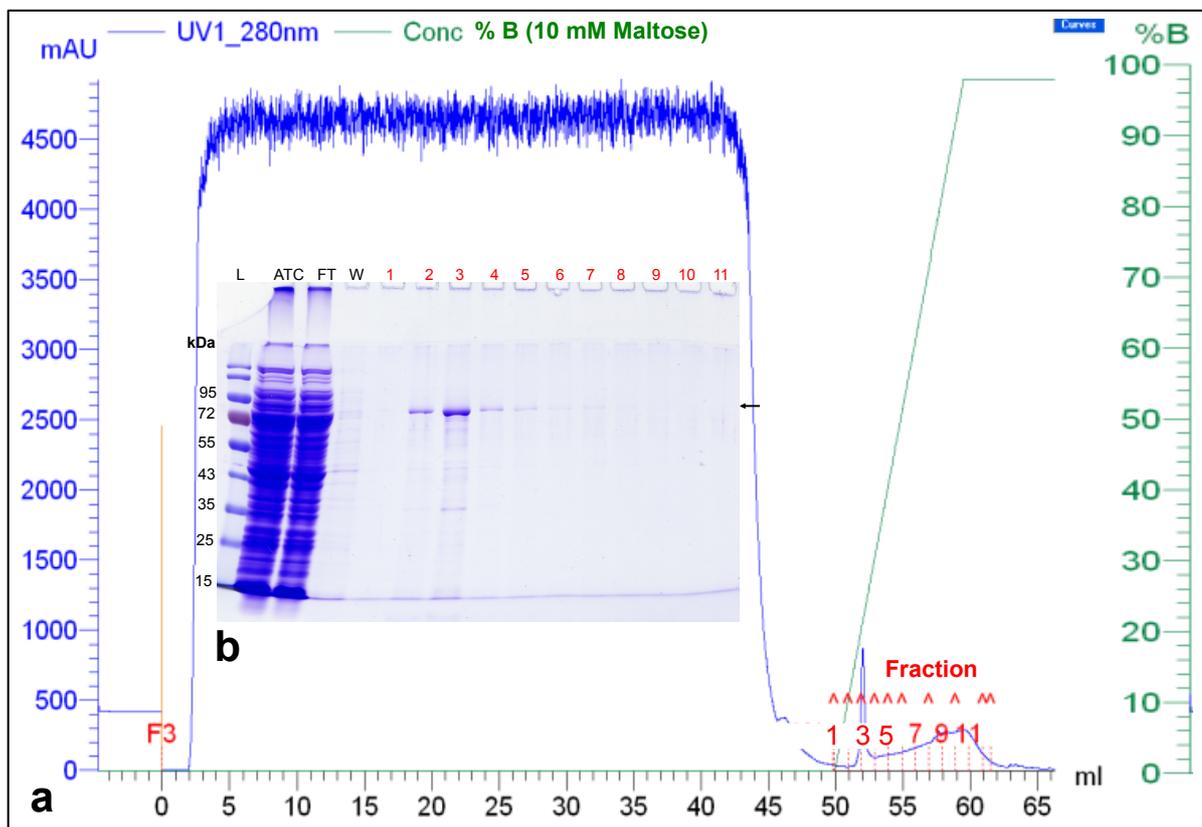


Figure 4.25| Purification of soluble MBPCaf1R_{gs} by maltose affinity chromatography.

a) Elution profile from the MBPTrapHP column. Clarified lysate-supernatant (35 ml) was applied to a MBPTrapHP column (GE Healthcare) and the bound MBPCaf1R_{gs} was eluted with a linear gradient of 0-10 mM maltose (0-100% of B) over 10 ×1 ml fractions as described in section 2.5.4 (ii). **b)** SDS-PAGE (12.5% acrylamide) analysis of the eluted fractions (1-11) along with sample applied to column, flow through (FT) and wash fraction (W). ← arrow indicates expected location and protein band of MBPCaf1R_{gs}. L: Pre-stained protein size marker.

4.4 Conclusion

Caf1R is a member of the AraC/XylS family of transcription regulators. Proteins of this family are naturally expressed at low level and are notoriously difficult to purify in their native state. This is in part due to differences in the pI of their two domains (conserved DBD and highly variant sensing or oligomerisation domain) that can lead to protein aggregation and insolubility (Gallegos et al., 1997; Schleif, 2010). As seen in this study the DBD alone is also insoluble when overexpressed. Despite these problems, some proteins of this family have been successfully overexpressed, purified and characterised including the putative-structural homologs of Caf1R MarA, Rob and XylR proteins of *E. coli* (Jair et al., 1995; Kwon et al., 2000; Ni et al., 2013). In this chapter expression of tagged Caf1R expression was monitored from both the native DNA sequence (*caf1R*) and a synthetic codon optimised gene (*caf1Rgs*). Both native and synthetic gene were subcloned in three expression plasmids, two contain N-terminal His₆-tag, pBADHisA and pET28a⁺ under control of the P_{BAD} and T7 promoter, respectively. The third plasmid, pMALc2x contains a solubility enhancer tag (MBP) at N-terminal under control of the IPTG inducible *ptac* promoter. Expression of hCaf1R from pBADhCaf1R in *E. coli* Top10 resulted in recovery of soluble hCaf1R, but at very low levels only detectable by immunoblotting. Optimising the *caf1R* codon sequence only slightly increased the level of recovery of soluble

hCaf1Rgs. In both cases, induction with 0.02% L-ara for 37°C/4 h was found to be optimum, with primarily soluble protein, no evidence of aggregates but only very low levels of hCaf1R.

To boost the level of expression of His₆-tagged Caf1R, *caf1R* and *caf1Rgs* were subcloned into the overexpression vector pET28a⁺, as pET vectors had been successfully used in the overexpression of putative-structural homologs of Caf1R, MarA (Jair et al., 1995), Rob (Kwon et al., 2000) and XylR (Ni et al., 2013). Surprisingly, levels of recovered hCaf1R^T were no higher from this vector when induced with IPTG. Detectable hCaf1R was primarily recovered in the insoluble fraction and degraded. Decreasing the temperature of incubation and IPTG concentrations did not help. However, reducing the rate of induction from pETCaf1Rgs by inclusion of 1% glucose with IPTG dramatically affected results. Under these conditions high levels of Coomassie blue stainable hCaf1R^Tgs were recovered although this was virtually all insoluble. Relief from IPTG-mediated toxicity by inclusion of glucose is linked to catabolite repression (Deutscher, 2008). Induction at the lower temperatures of 15 and 18°C increased the ratio of soluble to insoluble protein, probably due to reduction in the rate of synthesis and subsequent misfolding of protein (Rosano and Ceccarelli, 2014; Vera et al., 2007), but the majority of hCaf1R^Tgs still aggregated, Expression of the His₆-tagged N-terminal DBD of Caf1R (from pET_hCaf1R_N) was found to be even worse. Virtually all of the protein was recovered in insoluble inclusion bodies and aggregates, indicating that the DBD domain of the regulator is primarily responsible for aggregation. Hence, for production of soluble hCaf1R from pETCaf1Rgs, a higher ratio of soluble to aggregated protein can be obtained with induction at 15°C or higher levels of protein (with a lower ratio of soluble to insoluble) can be obtained with short induction times at 37°C. IMAC isolated and concentrated soluble hCaf1R was primarily large aggregates with a small amount of monomer (40.98 kDa), and possibly dimer (82 kDa) as determined by size exclusion chromatography. AraC/XylS regulators act as monomer, MarA (Rhee et al., 1998) and Rob (Kwon et al., 2000), and dimer, XylR (Jair et al., 1995; Kwon et al., 2000; Ni et al., 2013). Further studies could focus on purification of monomer and dimer of hCaf1R and monitor binding activity of both.

To enhance the solubility of Caf1R, expression of Caf1R was monitored by fusion of Caf1R to MBP. MBP has been shown to enhance the solubility of AraC/XylS regulators and their subsequent characterisation without removing the MBP-tag; examples include SoxS (Fawcett and Wolf, 1994), Rns (Munson and Scott, 1999, 2000) and PchR (Lin et al., 2013; Michel et al., 2005). MBPCaf1R and MBPCaf1R_N was unstable and degraded when expressed from BL21 at both 37°C and 25°C. The NEB recommended strain, *E. coli* K12-ER2508 worked as an excellent host strain for the expression of MBPCaf1R, MBPCaf1R_{gs} and MBPCaf1R_N. This is a *lon* deficient *E. coli* K12 strain that enhances protein solubility of cytosolic recombinant proteins (NEB). High levels of soluble MBPCaf1R were observed from all three versions of MBP-tagged Caf1R constructs using this strain, with no degradation. Although there was still accumulation of insoluble MBPCaf1R, the level of soluble protein recovered from K12-ER2508 equalled the amount of insoluble MBPCaf1R. Importantly, the level of soluble MBPCaf1R represented the major protein in lysed cell supernatants in contrast to the soluble hCaf1R constructs, which were only detectable by immunoblotting of cell lysate supernatants.

Coomassie blue stainable and soluble His₆- and MBP-tagged Caf1R was purified from the respective construct expressing codon optimised *caf1Rgs*. The hCaf1R^{Tgs} was affinity purified from 100 ml culture of BL21(DE3)/pEThCaf1R_{gs}, induced with 0.35 mM IPTG +1% glucose at 37°C for 2.5 h. The hCaf1R^{Tgs} protein was recovered in a flat peak (about 40 µg/ml, roughly 200 µg total). MBPCaf1R_{gs} was purified from 500 ml culture of K12-ER2508/pMALc2-Caf1R_{gs}, induced with 0.3 mM IPTG induction (no glucose) at 25°C for 5 h. MBPCaf1R_{gs} was recovered in a sharp peak from maltose affinity chromatography (about 380 µg/ml, roughly 600-700 µg total MBPCaf1R fusion or 300-350 µg Caf1R). Activity of both fusion proteins is demonstrated in Chapter-5. However, activity was clearly suboptimal and further studies should focus on recovery of fully active protein.

MarA was purified and solubilised from inclusion bodies (Jair et al., 1995) and successfully used to solve the *mar*-MarA cocrystal structure (Rhee et al., 1998). Insoluble hCaf1R^{Tgs} was successfully solubilised from inclusion bodies using urea. Further manipulation of solubilisation conditions and studies to optimise refolding could also be used as a route to obtain functional, soluble hCaf1R for further DNA binding studies, antibody production and even crystallization studies.

Chapter 5

Identification and localisation of *caf* locus promoters and Caf1R binding site(s)

5.1 Introduction

The F1 encoding *caf* locus (5.128 kb) is present on *Y. pestis* specific largest virulence plasmid, pFra (\approx 100 kb). It codes for 4 genes, located in opposite orientations and presumably organised into two or possibly three transcriptional units (**Fig. 5.1**). One transcriptional unit (reverse orientation) contains *caf1R*, a gene encoding a transcriptional regulator (Caf1R) of the AraC/XylS family and the other (forward orientation) includes *caf1M*, *caf1A* and *caf1*, the genes coding for periplasmic chaperone (Caf1M), an outer membrane usher (Caf1A) and F1 structural subunit (Caf1), respectively (Karlyshev et al., 1992). Very little information is available on regulation of this locus. It has been suggested that it may contain two regulatory regions, one the intergenic region (327 bp) between *caf1R* and *caf1M*, which presumably contains divergent promoters, P_R and P_M , and Caf1R binding site(s) (**Fig. 5.1**), responsible for transcription in both directions (Karlyshev et al., 1992). The possibility of a second regulatory region is the intergenic region (80 bp) between *caf1A* and *caf1*, which could influence *caf1* expression, leading to high levels of accumulation of F1, possibly through binding of the Caf1R regulator (Karlyshev et al., 1992; MacIntyre, 2004).

Bacterial RNAP holoenzyme (\sim 460 kDa) is the key determinant for transcription initiation, made up of 6 subunits, 2α , β , β' , ω and σ (Browning and Busby, 2004). The β' is the largest subunit (155.2 kDa), encoded by the *rpoC* gene (Ebright, 2000). This subunit contains the region of the active center essential for RNA synthesis and also contains some determinants for non-sequence-specific interactions with DNA and nascent RNA (Murakami et al., 2002). The rest of the active center essential for RNA synthesis and determinants for non-sequence-specific interactions (with DNA and nascent RNA) is contained within the second-largest subunit, β (150.6 kDa), which is encoded by *rpoB* gene. Third largest subunit of RNAP holoenzyme is α subunit (36.5 kDa) (Ross et al., 2001), which is encoded by *rpoA* gene and is present in 2 copies (α^I and α^{II}) per molecule of RNAP. Each α subunit is made up of two domains, N-terminal domain (α NTD) and a C-terminal domain (α CTD). The α CTD contains determinants for interaction with promoter sequence and thus makes non-sequence-specific interactions at most promoters and sequence-specific interactions at the UP-element containing promoters. It also contains the determinants for interactions with transcription factors or regulators. The smallest subunit of RNAP is ω subunit (10.2 kDa), encoded by *rpoZ* gene. This subunit function as a chaperone and thus facilitates assembly of RNAP and stabilizes assembled RNAP (Mathew and Chatterji, 2006). For binding to the target promoter, RNAP core initially interacts with the transcriptional initiation factor, σ subunit, in order to form RNA polymerase holoenzyme. The σ subunit (factor) decreases the affinity of RNAP for nonspecific DNA interactions and thus enhancing specificity for the target promoters, allowing transcription to begin at the correct site. To initiate transcription, bacterial cell uses different sigma factors under different environmental conditions, which bind the promoters of the target genes depending on the environmental conditions and thus enhance the transcription of those genes (Gruber and Gross, 2003).

All bacterial cells contain a primary or housekeeping σ factor known as σ^{70} (70 kDa), which is encoded by *rpoD* gene and transcribe most genes in the growing cells and thus keeps essential genes and pathways operating (Gruber and Gross, 2003). In *Enterobacteriaceae*, genes recognised by σ^{70} all have similar promoter consensus sequences, containing two hexamer elements, -10 (TATAAT) and -35 (TTGACA), located 10 and 35 nt upstream of the transcription start site (TSS; +1). Other example of σ factors include, σ^{19} (19 kDa), σ^{24} (24 kDa), σ^{28} (28 kDa) σ^{32} (32 kDa) and σ^{38} (38 kDa) encoded by *fecI*, *rpoE*, *rpoF*, *rpoH* and *rpoS* gene. Factor σ^{19} regulates *fec* gene involved in iron transport, σ^{24} is a factor regulating extracytoplasmic/extreme heat stress, σ^{28} regulates gene involved in flagellar synthesis, σ^{32} is turned on when cells are exposed to heat. Genes, whose expression is enhanced by σ^{32} code for heat shock proteins, chaperones, proteases and DNA-repair enzymes. Factor σ^{38} is regulate genes in the starvation/stationary phase of growing cells.

Sequence alignment of RNAP holoenzyme of *Y. pestis* (strain CO92) and *E. coli* (strain K-12) showed a great degree of similarity (Appendix 2a), having high % of amino acid identity in the respective subunits. For examples, RpoA or α subunit (99%), RpoB or β subunit (95%), RpoC or β' subunit (93%), RpoZ or ω subunit (92%) and RpoD or σ^{70} (91%), RpoH or σ^{32} (84%) and RpoS or σ^{38} (92%). In addition, several promoters have been mapped and predicted in *Y. pestis*, containing the -10 and -35 elements quite similar to the σ^{70} based promoters consensus elements of *E. coli*. A tabulated summary of these promoters elements along with the corresponding regulated gene is shown in Appendix 2b. Hence, the aims of this chapter were to localise the core or basic promoter (σ^{70} type) for the *caf1R/M/1* genes and define Caf1R binding within the *caf* locus, by combinatorial approaches of bioinformatics analysis, promoter-*lacZ* fusions and an *in vitro* DNA-binding study of tagged Caf1R.

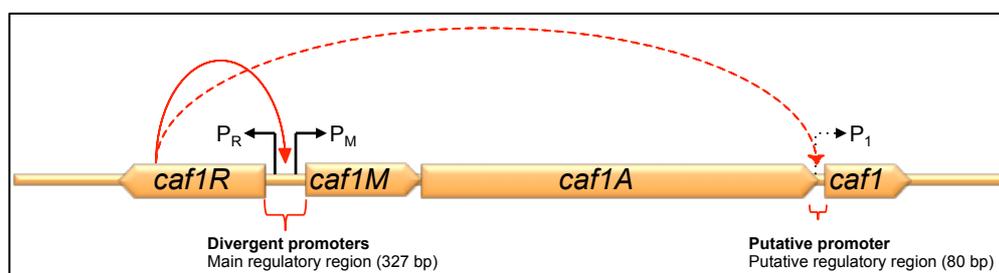


Figure 5.1| **The *caf* locus of *Yersinia pestis*.**

Genes encoding Caf1R regulator, Caf1M periplasmic chaperone, Caf1A outer membrane usher and the F1 subunit (Caf1) are presented according to their relative location and size. The main divergent promoters (P_R and P_M) within the *caf1R-caf1M* intergenic region and an additional putative promoter (P_1), within *caf1A-caf1* intergenic region are indicated by black- solid and dotted arrows, respectively. Proposed binding of Caf1R to the suggested divergent and putative promoters is indicated by a solid and dotted red arrow, respectively.

Results and Discussion

5.2 Analysis of *caf1R-caf1M* and *caf1A-caf1* intergenic regions

Three potential promoters had been predicted previously (Karlyshev et al., 1992) in the intergenic region of *caf1R-caf1M*; one for *caf1R* (P_R^K) and the other two for *caf1M* ($P_M^{K1/2}$). No such promoter had been predicted for the intergenic region of *caf1A-caf1*, suggesting the

intergenic region of *caf1R-caf1M* may be the main regulatory region of the *caf* locus controlling the expression of *caf1R* in one direction and *caf1MA1* in the opposite direction. These predictions raise the questions whether these are the active promoters of the *caf* locus and where exactly Caf1R binds. Alternatively, there may be other promoter(s), which regulate the expression from the *caf* locus. In order to identify potential promoter elements (-10 and -35), Caf1R binding sites and possible Shine-Dalgarno (SD) sequence for *caf1R*, *caf1M* and *caf1* genes, the intergenic region of *caf1R-caf1M* and *caf1A-caf1* (Fig. 5.2) was analysed bioinformatically and visually as described in section 2.11.2-5. The results are discussed in the following sections and properties of predicted promoters are tabulated in Appendix 3.

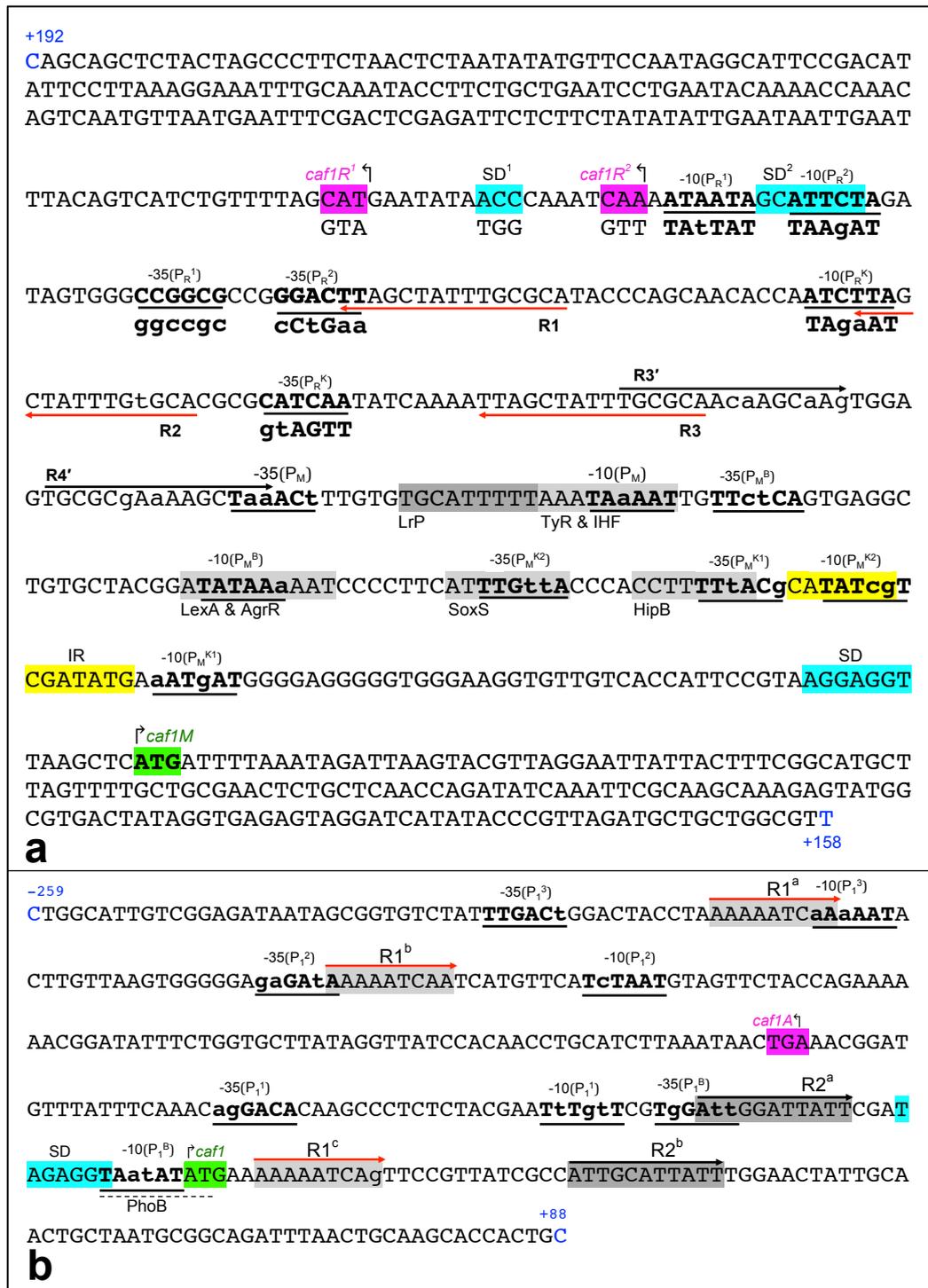


Figure 5.2| Sequence analysis of *caf1R-caf1M* and *caf1A-caf1* intergenic regions of *caf* locus.

a) The *caf* DNA fragment encompassing the complete intergenic region of *caf1R-caf1M* (327 bp) and 192 or 158 bp (w.r.t. ATG start) of coding sequence of *caf1R* (^fM₁-A₇₀ aa) and *caf1M* (M₁-V₅₃ aa), respectively. The two possible start codons, ATG (Met) (Karlyshev et al., 1992) and TTG (Leu or ^fMet) (Parkhill et al., 2001) (CAT and CCA in reverse complement) for *caf1R* and ATG (Met) for *caf1M* are highlighted in pink and green, respectively. Predicted Shine-Dalgarno (SD) sequences for *caf1R* and *caf1M* are highlighted in cyan and numbered SD¹ and SD, respectively. Visually predicted SD sequence for *caf1R* w.r.t. TTG (^fMet) start is also highlighted in cyan and indicated as SD². Predicted and visually identified promoter elements (-10 and -35) are underlined and indicated in bold capital and small (non-consensus) letters. Previously predicted -10 and -35 elements (Karlyshev et al., 1992) are designated as P_R^{K1} and P_M^{K1/2} while BPRM predicted elements are designated as -10(P_M^B) and -35(P_M^B) and visually identified P_R and P_M promoters, (P_R^{1/2}) and (P_M), respectively. Repeated sequences, R1, R2 and R3 (for *caf1R*) and R3' and R4' (for *caf1M*) are shown by red and black arrows, respectively, with variant nucleotides in small letters. BPRM predicted binding site motifs for other transcriptional factors are highlighted in grey and discussed in text (section 5.2.5). An inverted repeat (IR), upstream of -10(P_M^{K1}) is highlighted in yellow. The +192 and +158 in blue at either end show the numbering from the *caf1R* and *caf1M* translation start sites (ATG; M), respectively. **b)** The *caf* DNA fragment encompassing the complete intergenic region of *caf1A-caf1* (80 bp) plus 179 and 88 bp, respectively, from the *caf1A* end (TGA (Z), pink; S₇₇₅-Z₈₃₄ aa) and *caf1* start (ATG, green; M₁-A₃₀ aa). BPRM predicted and visually identified putative P₁ promoter's elements are underlined, shown in bold capital and small (non-consensus) letters and indicated as -10/-35(P₁^B) and -10/-35(P₁^{1/2/3}). Visually predicted SD sequence for *caf1* is highlighted in cyan. Grey highlighted sequences with red and black arrows show repeat sequence motifs R1^{a/b/c} and R2^{a/b}, respectively. Variant nucleotide (G) of R1^c is indicated as g. BPRM predicted PhoB binding site is indicated by dotted underline. The -259 and +88 indicated in blue at either end show the numbering from *caf1* start (ATG).

5.2.1 Identification of potential promoter elements in *caf1R-caf1M* intergenic region

Previously predicted (Karlyshev et al., 1992) -10 and -35 elements of P_R^K and P_M^{K1/2} promoters are indicated as -10/-35(P_{R/M}^{K1/2}) (**Fig. 5.2a**). The intergenic region of *caf1R-caf1M* (327 bp, relative to ATG start of both *caf1R* and *caf1M*) along with first 192 and 158 bp of *caf1R* and *caf1M* genes was initially screened by BPRM (Solovyev, 2011) to identify potential -10 and -35 elements of possible σ^{70} type promoters for *caf1R* and *caf1M*. A single promoter was predicted within this DNA fragment with close proximity to *caf1M*, -104 bp upstream of *caf1M* start codon (ATG) to the beginning of -10 element. Hence this was assigned as P_M^B promoter with predicted -10 (TATAAa) and -35 (TTctCA) elements (**Fig. 5.2a**) and LDF score of 6.38, suggesting an active promoter. The individual score of -10(P_M^B) and -35(P_M^B) element was predicted as 41 and 33, indicating a better prediction of -10(P_M^B) element. In addition, the frequency percentage of nt within -10(P_M^B) element was found comparatively higher than that of -35(P_M^B), suggesting a strong -10 and a weak -35 element with respect to optimum -10 (TATAAT) and -35 (TTGACA) elements of σ^{70} type promoters (**Fig. 5.2a**). This is in good agreement with an optimum promoter prediction where one predicted element shows a higher consensus conservation than the other (Lisser and Margalit, 1993). The distance between the predicted -10(P_M^B) and -35(P_M^B) elements was 18 nt, close to the optimum distance (17 nt) for σ^{70} type promoters of *E. coli* (Lisser and Margalit, 1993).

Following visual analysis (section 2.11.2-5) of both strands of this DNA fragment, three additional promoters were predicted within the *caf1R-caf1M* intergenic region, two upstream of *caf1R*, (P_R¹) and (P_R²) and another for *caf1M*, P_M (**Fig. 5.2a**). The sequence composition of these predicted elements, was TATtAT (-10) and cgccgg (-35) for (P_R¹) and TAgAAT (-10) and aaGtCc (-35) for (P_R²), indicating a strong -10 element with 5/6 consensus nt and a very weak -35 element for both. Similarly, the sequence composition of the predicted P_M promoter, TAAaAT (-10) and TaaAcT (-35), indicated a strong -10 element with 5/6 consensus nt and a reasonable good -35 element with 50% nt identity to the optimum -35 element. The distance between visually

identified -10 and -35 promoter elements was 17 nt, apart from (P_{R^2}) which has 18 nt (**Fig. 5.2**). Among the visually identified promoters, the predicted P_M promoter seems the most promising promoter. The first two and last three nt of its predicted -10 element (TAaAAT) correspond to the % frequency of consensus residues at each position within -10 element ($T_{80}A_{95}T_{45}A_{50}A_{60}T_{96}$) as do the first, fourth and fifth nt of its predicted -35 element, TaaACt ($T_{82}T_{84}G_{78}A_{65}C_{54}A_{48}$) (Lisser and Margalit, 1993).

5.2.2 Identification of potential promoters elements in *caf1A-caf1* intergenic region

The *caf1A-caf1* intergenic region (80 bp) along with 179 and 88 bp from the *caf1A* end (TGA) and *caf1* start (ATG), respectively was also screened by BPROM and then analysed visually for -10 and -35 consensus sequences (**Fig. 5.2b**). BPROM predicted a single promoter, P_1^B within the *caf1A-caf1* intergenic region with an LDF score of 4.25. The individual score of the predicted -10 and -35 elements was 68 and 16, with 17 nt spacer suggesting a strong prediction for -10 element (**Fig. 5.2b**). The nt composition of predicted -10(P_1^B) and -35(P_1^B) elements was TAaAT and TgGAtt, 4/6 and 3/6 consensus nt, respectively for an optimum σ^{70} type promoter. However, the location of the predicted -10(P_1^B) element of this promoter was adjacent to the start codon (ATG) of *caf1*, suggesting it is unlikely to be the real promoter for full-length *caf1*. The possibility was considered that there might be another open reading frame (ORF) downstream, encoding a small fragment of Caf1 controlled by this promoter. Therefore the complete *caf* DNA fragment of pRS*caf1A'-1'-lacZ* promoter-*lacZ* fusion construct was analysed by an ORF finder program (http://www.bioinformatics.org/sms2/orf_find.html) in order to identify any possible ORF(s), downstream, starting with either ATG or commonly used rare but alternate codons, TTG, GTG and CTG (Blattner et al., 1997). A single ORF, with a CTG start codon encoding a 69-amino acids long peptide of 7.817 kDa was predicted (**Fig. 5.3**). There is no recognisable strong SD sequence for this ORF, but if expressed, the BPROM predicted promoter, P_1^B would control the expression of this low-complexity peptide (**Fig. 5.3**).

Visual analysis of the *caf1A-caf1* DNA fragment (**Fig. 5.2b**) showed three additional possible promoters, P_1^{1-3} with 17 nt spacer and predicted -10 and -35 elements (-10(P_1^{1-3}) and -35(P_1^{1-3})) as shown in **Fig. 5.2b** and Appendix 2. The P_1^1 was predicted within the intergenic region of *caf1A-caf1* while P_1^{1-2} were predicted within *caf1A* 3' fragment. The sequence composition of these predicted -10 and -35 elements were TtTgtT -10(P_1^1) and agGACA -35(P_1^1); TcTAAT -10(P_1^2) and gaGAtA -35(P_1^2); aAaAAT -10(P_1^3) and TTGACt -35(P_1^3), indicating 3/6 and 4/6 (for P_1^1); 5/6 and 3/6 (for P_1^2); and 4/6 and 5/6 (for P_1^3) consensus nt (shown in capitals) in the -10 and -35 elements, respectively. Among these three visually identified promoters, P_1^1 might seem to be the logical promoter for full-length *caf1* due to its distance from the *caf1* start codon, ATG (-31 bp upstream from ATG to the beginning of predicted -10(P_1^1) and nt composition of the predicted -10(P_1^1) and -35(P_1^1) elements (**Fig. 5.2b**). Alternatively, the other two promoters, P_1^2 and P_1^3 , predicted within *caf1A* 3' fragment have as good predicted motif-sequence compositions, and could control transcription of a longer 5' UTR on the *caf1* mRNA.

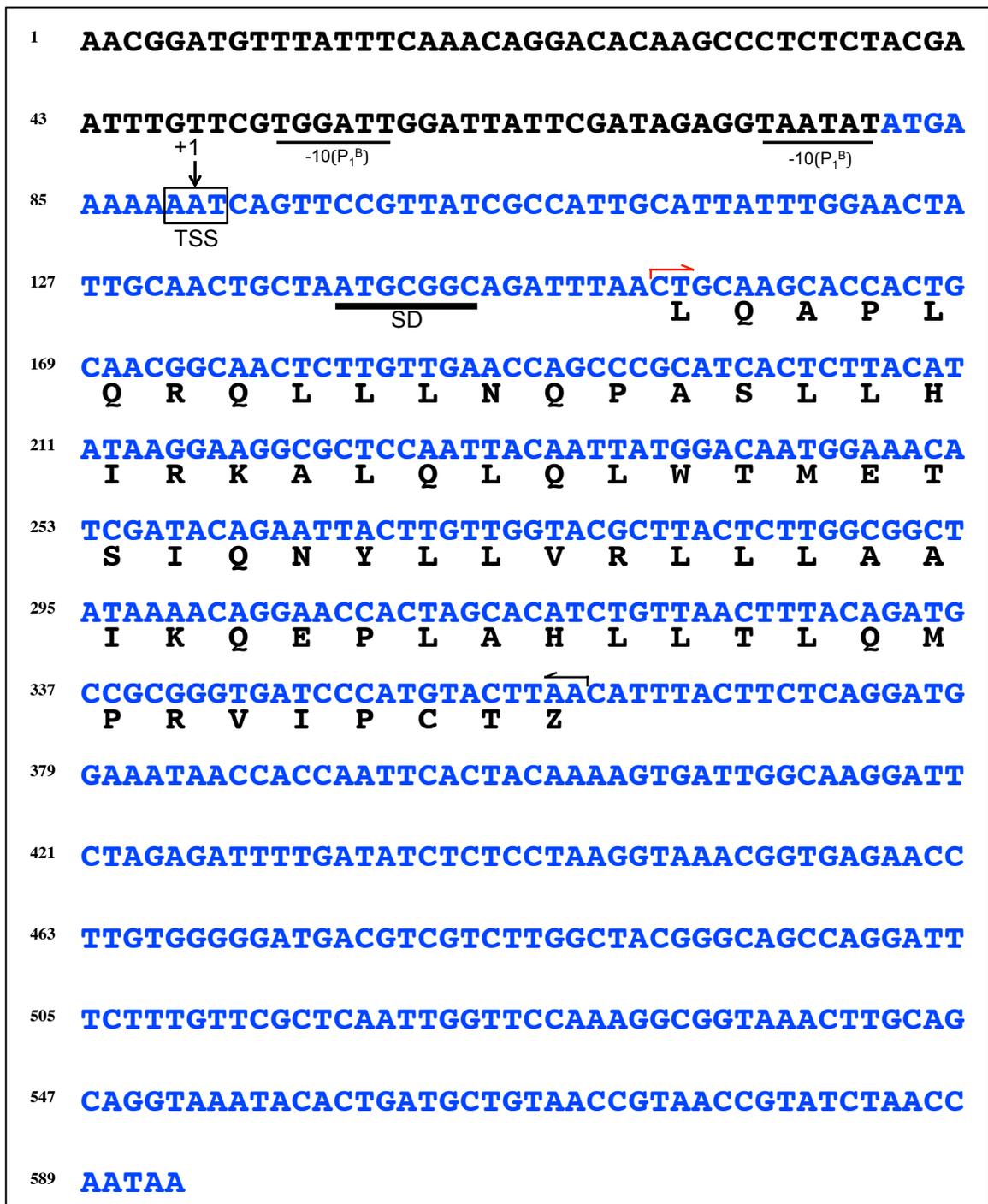


Figure 5.3| Additional predicted ORF within *cafI*.

Nucleotide sequence of the full-length *cafI* gene is indicated in blue text with upstream *cafIA-cafI* intergenic region (in black). The deduced amino acids sequence of predicted ORF is indicated between the red and black arrows. BPROM predicted -10 and -35 elements of P₁^B promoter are under lined. A very weak predicted SD and TSS for this predicted ORF are thick underlined and shown in a box with +1, respectively.

5.2.3 Identification of potential Caf1R-binding sites in *caf1R-caf1M* intergenic region

Using both REPFIND and visual analysis (section 2.11.2-5) of the *caf1R-caf1M* intergenic fragment (used in section 5.2.1), five repeat motifs (R1, R2, R3, R3' and R4') of 15 nt long were identified with few variant nt. Repeat motifs R1, R2 and R3 (red arrows) were identified at 56-70 bp (R1), 89-103 bp (R2) and 122-136 bp, (R3) upstream from ATG start codon of *caf1R*, with 18 nt spacer between each (Fig. 5.2a). Repeat motifs R3' and R4' (black arrows) run in the opposite orientation at 163-177 bp (R4') and 183-197 bp (R3') upstream from the ATG start

codon of *caf1M*, with a 5 nt spacer. There are 14 nt between R3 (*caf1R* motif) and R4' (*caf1M* motif). AraC/XylS family activators often bind immediately upstream of the promoter -35 element (Egan, 2002; Gallegos et al., 1997). This would be consistent with Caf1R binding to R4', immediately upstream and overlapping the -35 element of P_M promoter. Similarly R1 lies immediately upstream and overlaps -35 element of P_{R²} promoter. R3 is located 8 nt upstream of the predicted P_{R^K} -35 element. Repeat motif R3 (upstream of *caf1R*) and R3' (upstream of *caf1M*) overlap each other by 6 nt.

The five potential Caf1R-binding sites or motifs were aligned and compared to predict a consensus-binding motif for Caf1R. They were also compared to the known binding motif of MarA (Rhee et al., 1998) and a published MarA/Rob/SoxS binding consensus (Duval and Lister, 2013) to identify key nt (**Fig. 5.4**). The 15 nt long fragment of these motifs is highly conserved with 9 identical nt and only a few point variants. R3' is the most variant; R1-R3 and R4' share 12 identical nt. The flanking sequences of each repeat shows very poor conservation, although interestingly an A base upstream of R4' and at position one of MarA/Rob/SoxS consensus sequences is invariant (Duval and Lister, 2013; Martin et al., 1999; Wood et al., 1999). A mutagenesis study of this invariant A base in the SoxS consensus demonstrated a two-fold reduction in transcription following substitution with any other base (Griffith and Wolf, 2001). In the *mar*-MarA DNA-protein cocrystal structure (PDB-1bl0) this base was not contacted by any amino acid residue of MarA (Rhee et al., 1998). However, the first eight amino acid residues of MarA are not visible in the cocrystal structure of *mar*-MarA complex and thus it is possible that one or more of these amino acids make base specific contact(s) with this base. This is supported by an NMR study of DNA-MarA complex, where amino acid residues from the N-terminal tail were shown to contact DNA (Dangi et al., 2001).

Caf1R consensus from the aligned sequences of these potential Caf1R-binding motifs were determined as (i) TGCRC^{BS1}RAMWAGCWARD^{BS2}, including all five and (ii) TGCGC^{BS1}RAAWAGCTAAR^{BS2} considering only R1 and R4' (**Fig. 5.4**). In the MarA/Rob/SoxS consensus, the BS1 sequence is RGCAC (R: A/G). The first nt of this BS1 is T in all 5 potential Caf1R-binding motifs whereas the last four nt are GCGC except for R2 which has the same as MarA/Rob/SoxS consensus (GCAC). The predicted BS2 nt of the Caf1R-binding motifs also fits the consensus and is identical over 6 nt to *mar* BS2. The sequence between BS1 and BS2 of the MarA/Rob/SoxS consensus is quite variable (Egan, 2002; Gallegos et al., 1997) but the Caf1R-binding motifs show a greater degree of conservation, primarily a series of A nt.

Often transcriptional regulators bind to DNA as multimeric forms to a sequence motif (Garvie and Wolberger, 2001), which exhibit dyad symmetry. Examples of regulator binding to dyad symmetry sequence include, AraC (Lu et al., 1992), AarP (Macinga et al., 1999), ArsR (Rosenstein et al., 1994) and Fur (Pich et al., 2012). In this context, the most likely binding of Caf1R as a dimer to the identified repeat motifs would be at R1 and R4', which seems to be present in the dyad symmetry (15-14-15) and thus might regulate the expression of *caf1R* and *caf1M* or *caf1MA1* operon. More so, one 15 nt sequence motif represented as 'IR' in **Fig. 5.2a** was identified in the intergenic region of the *caf1R-caf1M*, which seems to be present in the dyad symmetry (7-1-7) and there might be possibility of Caf1R or other regulator(s) binding to this

region. No such repeat sequences with dyad symmetry were identified in rest of the *caf* locus, reinforcing the potential binding site(s) for Caf1R is/are present in the *caf1R-caf1M* intergenic region.

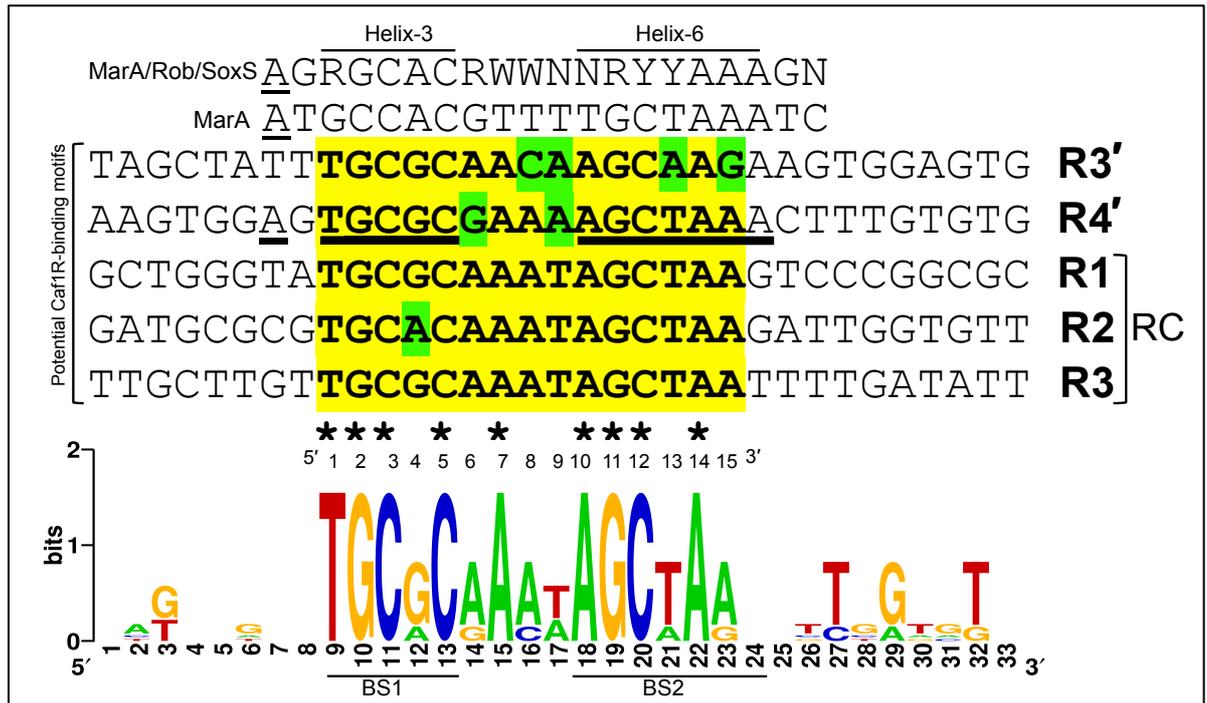


Figure 5.4| **Potential Caf1R-binding motifs, within the *caf1R-caf1M* intergenic region of the *caf* locus.** Sequence alignment of five potential Caf1R-binding repeat motifs (R1, R2, R3, R3' and R4') are highlighted in yellow with green highlight indicating variants. RC, repeat motifs in the reverse complement. Absolutely conserved nt within these repeat motifs are indicated by * and shown in the sequence logo generated using the WEBLOGO program (Crooks et al., 2004). The known binding sequence motif and nt interacting with helix-3 (BS1) and helix-6 (BS2) of MarA in *mar*-MarA complex (Rhee et al., 1998) and the consensus-binding motif for MarA/Rob/SoxS (Duval and Lister, 2013) are shown above the potential Caf1R-binding repeat motifs using the degenerate nucleotide code (**R**: A/G, **Y**: C/T, **W**: A/T and **N**: A/T/G/C). Predicted sequences of R4' interacting with helix-3 (BS1) and helix-6 (BS2) of Caf1R, as well as the corresponding invariant A of MarA/Rob/SoxS consensus are bold underlined.

5.2.4 Repeat sequences within the *caf1A-caf1* intergenic region

Five-repeated sequences with 2 different motifs, R1^{a-c} (9 nt) and R2^{a-b} (11 nt) were identified in this region (**Fig. 5.2b**). R1^a and R1^b are identical over all 9 nt, while R1^c shares eight of these nt, having a variant G base (g) at the end. R2^a and R2^b contain a single nt mismatch. R1^c and R2^b are within the coding sequence of *caf1*, R1^a and R1^b are located within the *caf1A* gene and only R2^a is within the intergenic region. The nt composition of these repeats is unrelated to the repeat sequences identified in the *caf1R-caf1M* intergenic region (**Fig. 5.2a**), suggesting that Caf1R does not bind to this region. Hence, the repeat motifs of *caf1A-caf1* intergenic region could be the remnant of phage or mobile element insertion and some relic of horizontal gene transfer (HGT). In addition, these could contribute to stability of polycistronic *caf* mRNA.

5.2.5 Predicted binding sites for additional transcription factors within the *caf* locus

BPROM screening of *caf1R-caf1M* and *caf1A-caf1* intergenic regions, identified eight binding motifs for additional transcription factors, seven between *caf1R* and *caf1M* and a single one upstream of *caf1* (**Fig. 5.2**). A potential integration host factor (IHF) binding site (AAATAAAAT)

was identified overlapping the predicted -10(P_M) element, IHF binds to the minor groove of DNA in a sequence specific manner, recognising AT rich regions. The primary function of IHF is architectural as it introduces a >160° bend in the DNA and influences transcription (Rice et al., 1996). A single binding motif of LrP (TCGATTTT) and TyrR (TAAATAAA) was indicated upstream of the IHF binding motif (**Fig. 5.2a**). TyrR, a metabolic virulence determinant of *Y. pestis*, essential for extracellular survival and thus plays an important role in enhancing its virulence by regulating 29 genes directly or indirectly (Deng et al., 2015). These include, nine down regulated, 5 (*aroF*, *tyrA*, *aroP*, *aroL*, and *tyrP*) from aromatic-pathways, 2 (*glnL* and *glnG*) of two-component system, encoding sensory histidine kinase and a regulator involved nitrogen assimilation and other 2 (*hdeB* and *hdeD*) of acid-stress. The remaining 20 genes (up regulated, 2.0–4.2-fold) code for proteins of the T3SS (Deng et al., 2015).

In the *caf1A-caf1* intergenic region a single binding motif of PhoB (TAATATAT) was identified, which overlaps the first two nt of the *caf1* start codon (ATG) and the last nt of the predicted SD sequence (**Fig. 5.2b**), suggesting binding of a phosphate regulon transcriptional regulator, PhoB of OmpR family. Prediction of more transcription factor binding sites within the *caf1R-caf1M* intergenic region in comparison to the *caf1A-caf1* intergenic region would be consistent with *caf1R-caf1M* intergenic region being the main regulatory region of the *caf* locus. In addition, *caf* expression is thermo-controlled (discussed in Chapter-6). There is likely to be a complex mechanism of regulation controlling F1 expression.

5.3 Promoter activity within the *caf* locus intergenic regions

Following bioinformatic prediction of promoters, three promoter-*lacZ* fusion constructs, pRScaf1M'-R'-*lacZ*, pRScaf1R'-M'-*lacZ* and pRScaf1R-M'-*lacZ* were used to monitor promoter activity within the intergenic region *caf1R-caf1M*, by monitoring β-galactosidase activity (section 2.3, **Fig. 5.5**). The pRScaf1R'-M'-*lacZ* and pRScaf1R-M'-*lacZ* are constructs in the promoter fusion vector, pRS550, to test for *caf1M* promoter(s); while, pRScaf1M'-R'-*lacZ* is in pRS415 (reverse orientation of insert) to test for *caf1R* promoter(s). An additional construct, pRScaf1A'-1'-*lacZ*, also in pRS550 was used to monitor activity upstream of *caf1*. To test the impact of Caf1R regulator on promoter activity, *caf1R* was supplied either in *cis* orientation (on the same DNA fragment as in the pRScaf1R-M'-*lacZ* construct) or in *trans* by co-transformation with pACYC-R. Activity was routinely monitored from three individual transformants in *E. coli* Top10.

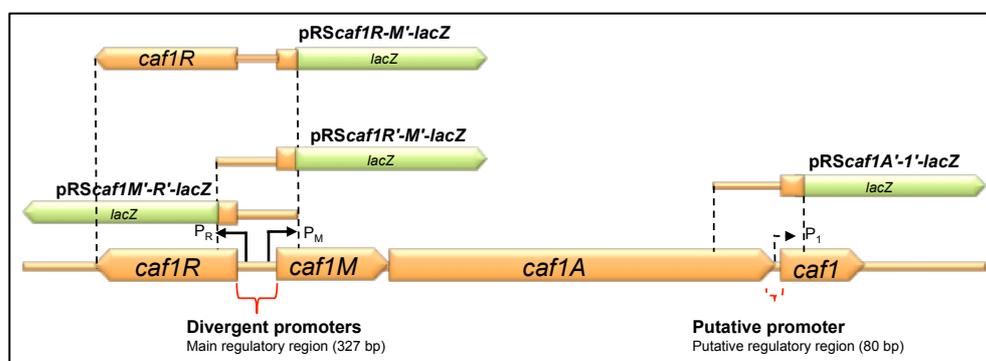


Figure 5.5| **Promoter-*lacZ* fusion constructs of the *caf* locus intergenic regions.**

Fragments of the *caf* locus indicated in orange between dotted lines were PCR-amplified and subcloned upstream of *lacZ* (green) in pRS415 (construct pRScaf1M'-R'-*lacZ*) or pRS550 (all others) to create promoter-

fusion plasmids (Simons et al., 1987). Three constructs, pRScaf1M'-R'-lacZ, pRScaf1R'-M'-lacZ and pRScaf1A'-I'-lacZ were prepared to test promoter activity of DNA sequences upstream of *caf1R*, *caf1M* and *caf1*, respectively. The construct, pRScaf1R'-M'-lacZ was prepared to monitor the impact of Caf1R (encoded in *cis* orientation) on expression from *caf1M* promoter(s).

5.3.1 Promoter activity upstream of *caf1R*

Promoter-fusion construct pRScaf1M'-R'-lacZ, encompassing the *caf1R*-*caf1M* intergenic region, 192 bp of *caf1R* and 158 bp of *caf1M* (**Fig. 5.5**) was used to identify *caf1R* promoter(s) activity. β -galactosidase activity per OD cell density was monitored during growth of *E. coli* Top10 carrying pRScaf1M'-R'-lacZ, pRScaf1M'-R'-lacZ+pACYC-R (expressing Caf1R) or the negative controls, pRS415 and pRScaf1M'-R'-lacZ+pACYC (**Fig. 5.6**). Promoter activity was consistently low throughout growth, in the absence of Caf1R (both *E. coli*/pRScaf1M'-R'-lacZ and *E. coli*/pRScaf1M'-R'-lacZ+pACYC empty vector). In the presence of Caf1R (pRScaf1M'-R'-lacZ+pACYC-R), there was a dramatic increase in the expression of *lacZ* (β -galactosidase) from 2 h throughout the exponential phase of growth. The activity from both tests, plus or minus Caf1R, was maximum in the stationary phase of growth (at 10 h). At 10 h, without Caf1R the maximum activity was 40.30 ± 1.65 units, in cells carrying pRScaf1M'-R'-lacZ plus pACYC. Activity increased to 278.85 ± 0.82 units in the presence of Caf1R (cells carrying both pRScaf1M'-R'-lacZ and pACYC-R plasmids). This suggests Caf1R dependent activation of a promoter within this DNA fragment, with about a 6.9-fold increase in the activity. Activity from pRScaf1M'-R'-lacZ and pRScaf1M'-R'-lacZ+pACYC was almost identical, confirming that the enhanced activity of P_R promoter(s) in the presence of Caf1R is from Caf1R. Comparison of β -galactosidase activity of 26°C-24 h (0 h, inoculum) and 37°C-24 h samples, indicated a substantial increase in activity (**Fig. 5.6**). Without Caf1R (from pRScaf1M'-R'-lacZ+pACYC), the temperature induction was low, 2.9-fold; whereas when complemented with plasmid encoded Caf1R (from pRScaf1M'-R'-lacZ+pACYC-R), induction was about 8.3-fold, suggesting temperature-dependent, Caf1R-mediated activation of P_R promoter(s). This is consistent with thermoinduction of *caf1R*, which is investigated further in Chapter-6. In conclusion, there is weak promoter activity upstream of *caf1R* and this was enhanced substantially in the presence of Caf1R provided in *trans* (from pACYC-R), suggesting that there is a Caf1R-dependent promoter in the *caf1R*-*caf1M* intergenic region controlling expression of *caf1R*. This indicates *caf1R*-mediated autoregulation of transcription within the *caf1R*-*caf1M* intergenic region of the *caf* locus.

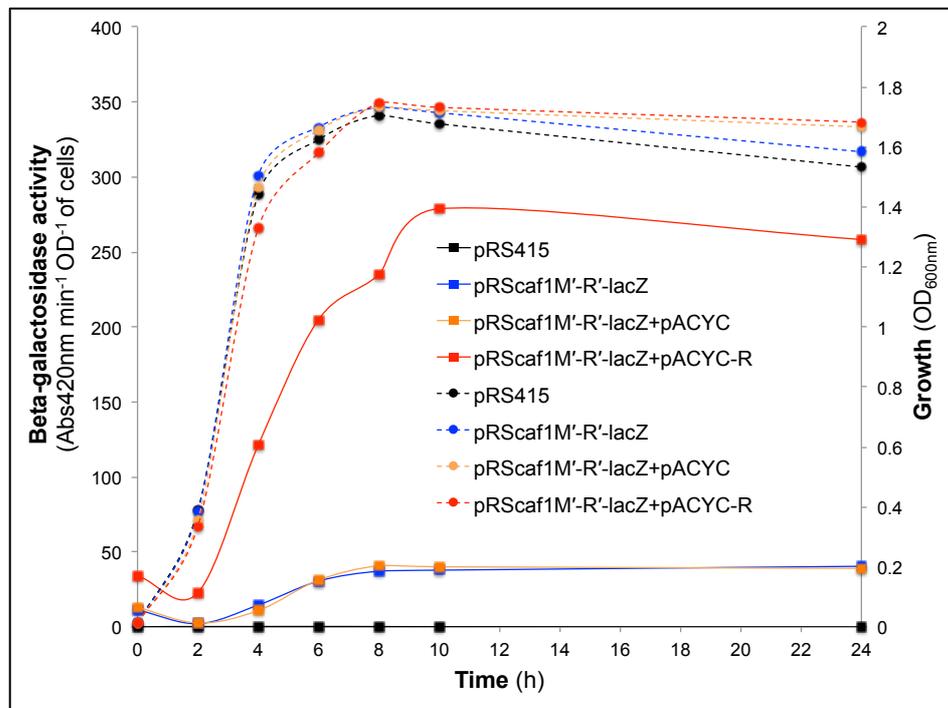


Figure 5.6| **Promoter activity upstream of *caf1R* with and without Caf1R.**

Activity (per OD, solid lines) at 0-24 h in cells of *E. coli* Top10 carrying pRScaf1M'-R'-lacZ (intergenic region alone), pRScaf1M'-R'-lacZ+pACYC-R (intergenic region +Caf1R) or the corresponding negative controls, pRS415 empty plasmid or pRScaf1M'-R'-lacZ+pACYC is presented on the primary Y-axis. Growth scale (OD₆₀₀, dotted lines) is on the secondary Y-axis. Error bars on the activity lines are based on \pm SEM of activity from three separate cultures of each, processed from three individual colonies of each.

5.3.2 Promoter activity upstream of *caf1M*

Promoter activity upstream of *caf1M* was monitored in a similar way using the fusion construct, pRScaf1R'-M'-lacZ. The influence of Caf1R on β -galactosidase activity was monitored using two different *caf1R* encoding constructs, pRScaf1R'-M'-lacZ (*cis* oriented *caf1R*, on the same plasmid) and pRScaf1R'-M'-lacZ+pACYC-R (*trans* oriented *caf1R*, supplied from pACYC-R) (Fig. 5.7).

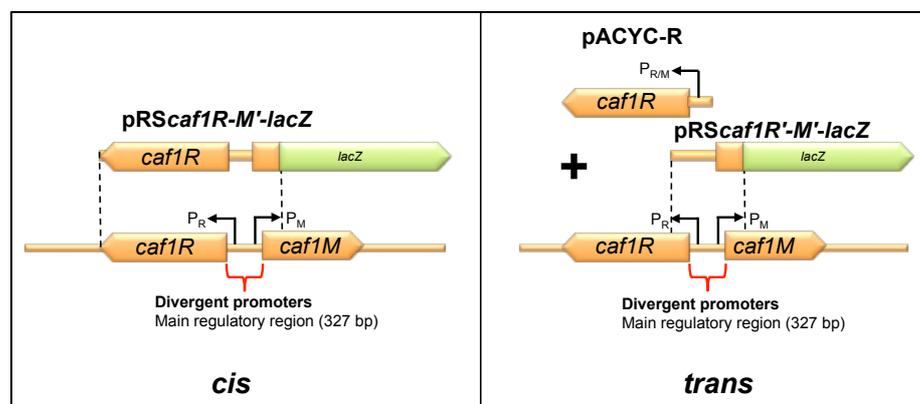


Figure 5.7| **The *cis* and *trans* constructs used to test influence of Caf1R on P_M promoter(s) activity.**

The β -galactosidase activity per OD cell density along with growth (OD_{600}) of test constructs, pRScaf1R'-M'-lacZ (intergenic region alone), pRScaf1R-M'-lacZ (intergenic region plus *cis* oriented *caf1R*), pRScaf1R'-M'-lacZ+pACYC-R (intergenic region plus *caf1R* supplied from pACYC-R) and the corresponding negative controls, pRS550 empty plasmid and pRScaf1R'-M'-lacZ+pACYC co-transformant is presented in **Fig. 5.8**. The activity of *caf1M* promoter(s) in the absence of Caf1R, from pRScaf1R'-M'-lacZ and pRScaf1R'-M'-lacZ+pACYC constructs at 10 h (the time point of maximum activity) was 7.38 ± 0.56 and 8.82 ± 0.65 units, respectively (**Fig. 5.8**), indicating a much lower level of activity compared to the weak activity of the promoter(s) upstream of *caf1R* in the absence of Caf1R (40 ± 0.61 units) (**Fig 5.6**). As with the *caf1R* promoter-lacZ fusion construct, activity from the *caf1M* promoter-lacZ fusion construct, pRScaf1R'-M'-lacZ was enhanced by *trans* expression of Caf1R from pACYC-R. Activity with pRScaf1R'-M'-lacZ+pACYC-R was 283.49 ± 0.33 units at 10 h, indicating about 32-fold Caf1R-dependent activation (**Fig. 5.8**). Almost identical, low-activity from pRScaf1R'-M'-lacZ and pRScaf1R'-M'-lacZ+pACYC confirms that the increased activity from pRScaf1R'-M'-lacZ+pACYC-R is solely from the impact of Caf1R.

In spite of this dramatic activation, when *caf1R* was expressed in the *trans* orientation from pACYC-R or when *caf1R* was expressed from the *cis* orientation in pRScaf1R-M'-lacZ, no significant impact was observed on the activity of P_M promoter(s) (**Fig. 5.8**). Activity was only 3.95 ± 0.07 units (at 10 h), even lower than the activity from pRScaf1R'-M'-lacZ and pRScaf1R'-M'-lacZ+pACYC constructs at 10 h. This seems very strange as the same oriented of *caf1R* and intergenic region is present in pACYCF1. The sequence of *caf1R* and intergenic region has been confirmed (Appendix 4), but prior to speculation, functional β -galactosidase should be confirmed in the pRScaf1R-M'-lacZ construct. If expression from this plasmid carrying *caf1R* in *cis* location is confirmed to be negligible, it may be related to the localised architecture of the DNA. In conclusion, as with the P_R promoter(s), activity of P_M promoter(s) was stimulated by *trans* supply of Caf1R, suggesting Caf1R is an activator with multiple modes of interaction with the *caf1R-caf1M* intergenic DNA.

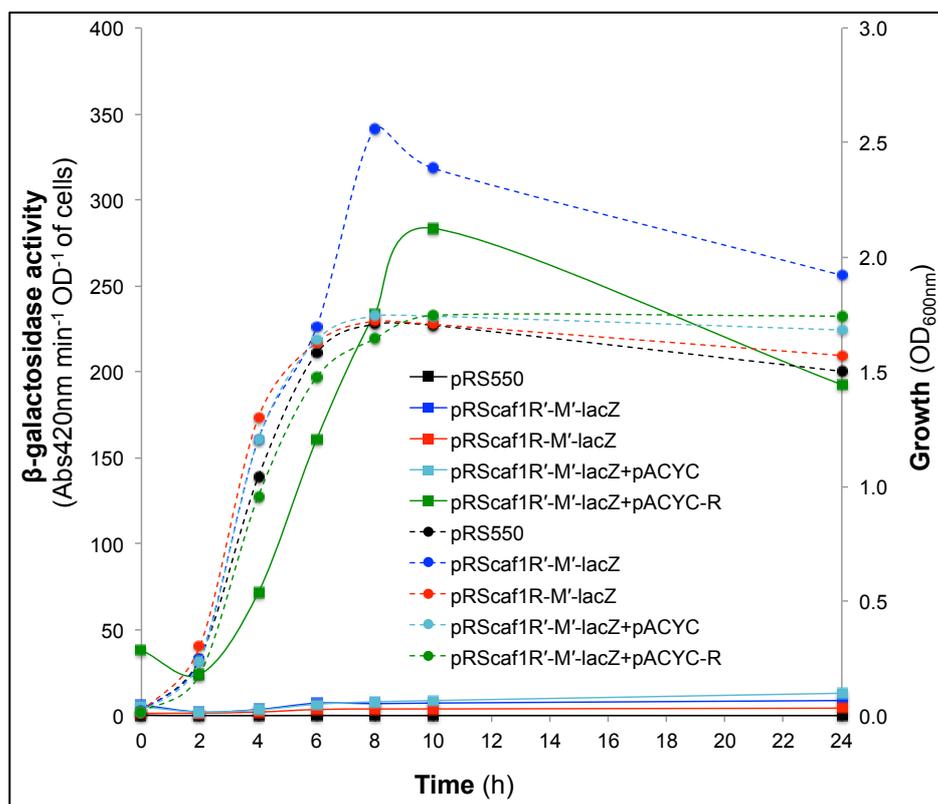


Figure 5.8| **Promoter activity upstream of *caf1M* with and without Caf1R.**

Activity (per OD, solid lines) at 0-24 h in cells of *E. coli* Top10 carrying pRScaf1R'-M'-lacZ (intergenic region alone), pRScaf1R'-M'-lacZ (intergenic region plus *caf1R* on same plasmid), pRScaf1R'-M'-lacZ+pACYC-R (intergenic region +Caf1R) or the corresponding negative controls, pRS550 empty plasmid or pRScaf1R'-M'-lacZ+pACYC is presented on the primary Y-axis. Growth scale (OD_{600nm}, dotted lines) is on the secondary Y-axis. Error bars on the activity lines are based on \pm SEM of activity from three separate cultures of each, processed from three individual colonies of each.

5.3.3 Promoter activity upstream of *caf1*

As above, the activity of predicted *caf1* promoters \pm Caf1R was assayed in *trans* using pRScaf1A'-1'-lacZ and pRScaf1A'-1'-lacZ+pACYC-R constructs along with the corresponding negative controls, pRS550 empty plasmid (pRS550) and pRScaf1A'-1'-lacZ+pACYC. Activity (β -galactosidase activity per OD cell density) and growth (OD₆₀₀) of each construct is presented in **Fig. 5.9**. Activity in the late exponential growth-phase (6 h) and the early stationary phase (8- and 10 h) was 35.01 ± 3.53 (6 h), 39.90 ± 1.98 (8 h) and 41.64 ± 3.28 (10 h) units from *E. coli* Top10/pRScaf1A'-1'-lacZ+pACYC. Activity was only modestly higher in samples with Caf1R, (25.62 ± 0.12 (6 h), 43.35 ± 0.21 (8 h) and 47.80 ± 0.57 (10 h) units) in *E. coli* Top10/pRScaf1A'-1'-lacZ+pACYC-R, indicating no detectable effect of Caf1R on any promoter within the *caf1A-caf1* intergenic fragment tested. This in turn suggests that Caf1R is less likely to bind to *caf1A-caf1* intergenic region. Unlike the very low activity from P_M promoter(s) in the absence of Caf1R, there was definite activity, throughout growth, in presence of *caf1* 5' UTR DNA. At early stationary phase (10 h), activity was 49.82 ± 0.12 units versus 7.38 ± 0.56 units for promoter activity upstream of *caf1M*. Comparison of 26°C-24 h (0 h inoculum) and 37°C-24 h, revealed a temperature upshift from 26 to 37°C of only 2-fold increase in activity \pm Caf1R. Hence, the possibility of temperature-induced transcription effect within the *caf1A-caf1* intergenic region is less likely. In conclusion, promoter activity upstream of *caf1* was low and not activated by Caf1R.

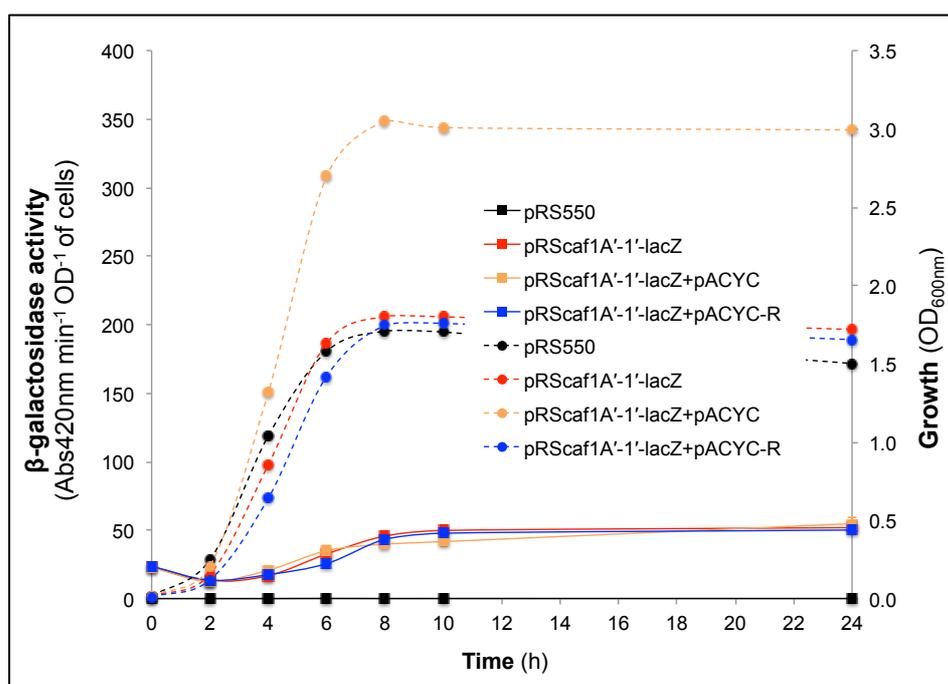


Figure 5.9| **Promoter activity upstream of *cafI* with and without Caf1R.**

Activity (per OD, solid lines) at 0-24 h in cells of *E. coli* Top10 carrying pRScaf1A'-1'-lacZ (intergenic region alone), pRScaf1A'-1'-lacZ+pACYC-R (intergenic region +Caf1R) or the corresponding negative controls, pRS550 empty plasmid or pRScaf1A'-1'-lacZ+pACYC is presented on the primary Y-axis. Growth scale (OD, dotted lines) is on the secondary Y-axis. Error bars on the activity lines are based on \pm SEM of activity from three separate cultures of each, processed from three individual colonies of each.

5.4 Localisation of *caf* promoters and Caf1R-binding site(s)

Following bioinformatic prediction and initial analysis of promoter activity within each complete intergenic region, the next step was to identify and localise which of the predicted promoters is the functional promoter for each gene. Therefore, a series of promoter-*lacZ* fusion constructs were designed for each gene by Infusion cloning technology (section 2.2.8(ii) and 2.3.1). Each series of constructs contained sequential loss of putative promoter and/or repeat motif(s), shown in (Fig 5.2). β -galactosidase activity from all of these promoter-*lacZ* fusion constructs was measured with and without Caf1R (encoded from pACYC-R). The numbering of all promoter-*lacZ* fusion constructs is from the translation start site (ATG) of the corresponding gene.

5.4.1 Localisation of *caf1R* promoter and Caf1R-binding upstream of *caf1R*

Five progressively shorter promoter-*lacZ* fusions, pRScaf1R'-.137+192-lacZ, pRScaf1R'-.114+192-lacZ, pRScaf1R'-.70+192-lacZ, pRScaf1R'-.65+192-lacZ and pRScaf1R'-.46+192-lacZ were created based on predicted promoters and repeats (Fig. 5.10a). The putative P_R promoters, ($P_R^{1/2}$) and (P_R^K) and repeat motifs R1, R2 and R3 (possible Caf1R binding sites) were lost as the constructs became smaller (Fig. 5.10b). The largest construct, pRScaf1R'-.137+192-lacZ contained all the predicted promoters and repeat motifs upstream of *caf1R*. The second largest construct, pRScaf1R'-.114+192-lacZ had lost only repeat R3, while in the third and fourth constructs, pRScaf1R'-.70+192-lacZ and pRScaf1R'-.65+192-lacZ, repeats R3+R2 and also P_R^K promoter had been deleted. The pRScaf1R'-.65+192-lacZ construct was additionally missing half of repeat R1. In the smallest construct,

pRScaf1R'₋₄₆₊₁₉₂-lacZ, all repeat motifs and predicted promoters, P_{R^K} and P_{R²} were deleted hence this construct contained only P_{R¹} predicted promoter (**Fig. 5.10a-b**). In the absence of Caf1R, even the shortest construct (pRScaf1R'₋₄₆₊₁₉₂-lacZ), had promoter activity (**Fig. 5.10c**). However, the pRScaf1R'₋₆₅₊₁₉₂-lacZ construct, which includes P_{R²} promoter, showed a higher level of basal activity, but no Caf1R activation. Inclusion of repeat R1 in construct pRScaf1R'₋₇₀₊₁₉₂-lacZ led to Caf1R activated transcription. As R1 overlaps the -35 element of P_{R²} promoter, this provides good evidence that the P_{R²} promoter is the main Caf1R activated promoter for transcription of *caf1R*, with the ATG start codon as shown (**Fig. 5.10a**). In class II promoters of AraC/XylS family regulators, the transcriptional regulator binding site generally overlaps the -35 motif (Duval and Lister, 2013; Egan, 2002; Gallegos et al., 1997). The pRScaf1R'₋₇₀₊₁₉₂-lacZ, pRScaf1R'₋₁₁₄₊₁₉₂-lacZ and pRScaf1R'₋₁₃₇₊₁₉₂-lacZ constructs contain repeats R1, R1+R2 and R1+R2+R3, respectively. Transcriptional activity of these constructs at 8 h was enhanced 3.8-, 7.6- and 2.5-fold, in the presence of Caf1R (**Fig. 5.10c**). While this would appear to suggest that R2 also is required for maximum Caf1R mediated activation, the construct pRScaf1R'₋₇₀₊₁₉₂-lacZ lacks nt immediately upstream of R1 that may be required for efficient Caf1R binding. This requires further investigation, with *lacZ* fusion construct containing a further 10-15 nt upstream of R1 to ensure optimum binding of Caf1R to R1. The low (2.5-fold) activation from the construct pRScaf1R'₋₁₃₇₊₁₉₂-lacZ is mainly a reflection of higher basal activity in the absence of Caf1R, suggesting P_{R^K} might function as a Caf1R independent promoter. A statistical comparison of the actual β -galactosidase activity from pRScaf1R'₋₁₃₇₊₁₉₂-lacZ and pRScaf1R'₋₁₁₄₊₁₉₂-lacZ in the presence of Caf1R at each time point, identified a statistically decreased level of β -galactosidase production throughout growth from pRScaf1R'₋₁₃₇₊₁₉₂-lacZ (P-value < 0.05 at each time point). This might be an indication of different modes of control in the presence of the 3 repeat sequences, R1, R2 and R3. However, in the presence of the complete intergenic regions in pRScaf1M'-R'-lacZ, full Caf1R mediated activation was restored (7.5-fold activity over basal activity in the absence of Caf1R).

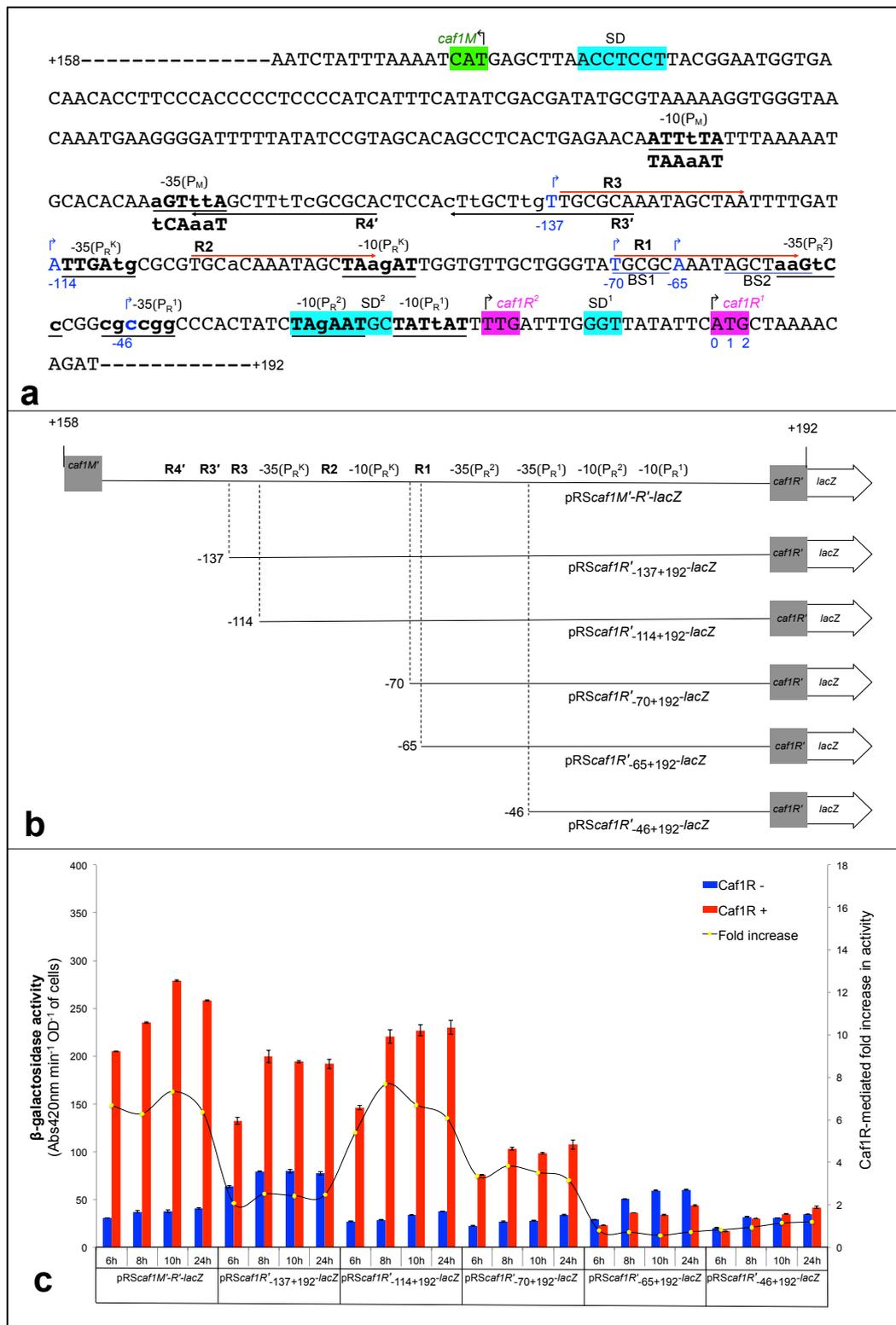


Figure 5.10| Localisation of *caf1R* promoter and Caf1R-binding site(s) upstream of *caf1R*.

a) Annotation of *caf1M-caf1R* intergenic DNA fragment (complement strand of the *caf* locus) as in pRS415*caf1M'-R'-lacZ* construct. Location of the putative promoter elements (-10 and -35) for *caf1R* ($P_R^{1/2}$, P_R^K) and for *caf1M* (P_M) are underlined and indicated in bold; repeat motifs, R1, R2 and R3 (red arrows) and R4' and R3' (black arrows). +158 and +192 indicate corresponding coding sequence of *caf1M* and *caf1R* with respect to ATG (Met) start sites. \uparrow arrows indicate location of deletions, and number of bp (-xx) upstream from the *caf1R* start included in each construct. Start codons and SD sequences of *caf1M* and *caf1R* are shown, with both predicted starts (TTG and ATG) of *caf1R* highlighted. **b)** Deletion strategy and assigned promoter-*lacZ* fusions used to localise functional *caf1R* promoters and Caf1R binding site(s). **c)** β -galactosidase activity (\pm pACYC-R, encoding Caf1R) from promoter-*lacZ* fusions, sampled over 24 h as shown. Activity ($\Delta A_{420}/\text{min}/\text{OD}_{600\text{nm}}$) and Caf1R-mediated fold increase in activity at primary and secondary Y-axis, respectively. Standard error (\pm SEM) of activity was calculated from three individual colonies of each construct. The comparative P-value for activity

of pRScaf1R'.₁₁₄₊₁₉₂-lacZ and pRScaf1R'.₁₃₇₊₁₉₂-lacZ constructs +Caf1R were 0.014, 0.004, and 0.019 for 6 h, 8 h and 10 h respectively. (Calculated using a Paired two-sample t-test, StatPlus software).

5.4.2 Localisation of *caf1M* promoter and Caf1R-binding upstream of *caf1M*

To localise the *caf1M* promoter, six progressively shorter promoter-lacZ fusions, pRScaf1M'.₂₇₀₊₁₅₈-lacZ, pRScaf1M'.₁₉₇₊₁₅₈-lacZ, pRScaf1M'.₁₈₄₊₁₅₈-lacZ, pRScaf1M'.₁₆₉₊₁₅₈-lacZ, pRScaf1M'.₁₄₁₊₁₅₈-lacZ and pRScaf1M'.₁₀₂₊₁₅₈-lacZ (**Fig 5.11**) were created based on the analysis of *caf1M* 5' UTR. These constructs contain the sequential removal of predicted promoters, (P_M^{K1-2}), P_M^B and P_M for *caf1M* and repeat motifs. Briefly, the largest construct, pRScaf1M'.₂₇₀₊₁₅₈-lacZ contains 13 of the 15 nt of R1 and intact R2, R3, R3' and R4' plus all 4 predicted promoters for *caf1M*. The second construct, pRScaf1M'.₁₉₇₊₁₅₈-lacZ has essentially lost all 3 repeats (R1-R3), upstream of *caf1R* but retained R3' and R4', and all predicted promoters for *caf1M*. The third construct, pRScaf1M'.₁₈₄₊₁₅₈-lacZ contained only R4 and all predicted promoters for *caf1M*. The fourth construct, pRScaf1M'.₁₆₉₊₁₅₈-lacZ contained only 7 of the 15 nt of R4' repeat, hence essentially no repeat, but retained all predicted promoters for *caf1M*. Construct number five, pRScaf1M'.₁₄₁₊₁₅₈-lacZ had lost the P_M promoter but retained BPRM predicted P_M^B promoter, while the smallest construct, pRScaf1M'.₁₀₂₊₁₅₈-lacZ contained only the previously assigned promoters (Karlyshev et al., 1992) P_M^{K1-2} for *caf1M*.

The results of the β -galactosidase assays with this series of *caf1M* promoter-lacZ fusion constructs showed a very clear indication of both the functional promoter and the Caf1R binding site (**Fig. 5.11c**). Very low promoter activity (about 2.8-3.9 units) was observed from both pRScaf1M'.₁₀₂₊₁₅₈-lacZ and pRScaf1M'.₁₄₁₊₁₅₈-lacZ constructs, with or without Caf1R. As pRScaf1M'.₁₄₁₊₁₅₈-lacZ contains both previously assigned promoters (Karlyshev et al., 1992), P_M^{K1-2} and the BPRM predicted P_M^B promoter, this suggests that none of these predictions represent an active promoter for *caf1M*. In contrast, pRScaf1M'.₁₆₉₊₁₅₈-lacZ, which contains the visually identified promoter, P_M showed about 10-fold higher activity (27.68 \pm 0.64, 28.97 \pm 0.62, 27.74 \pm 0.08 and 32.55 \pm 0.34 units at 6, 8, 10 and 24 h, respectively) although no Caf1R-mediated activation of transcription was observed. Inclusion of the repeat R4', with an additional 15 bp upstream in the construct pRScaf1M'.₁₈₄₊₁₅₈-lacZ, resulted in clear Caf1R-mediated activation. About 7.5- and 11.4-fold increase in activity was observed at 6 h and 8 h from extracts of cells transformed with pRScaf1M'.₁₈₄₊₁₅₈-lacZ and pACYC-R (Caf1R+) compared to cells transformed with pRScaf1M'.₁₈₄₊₁₅₈-lacZ and pACYC plasmid (Caf1R-). Together these results indicate that the visually identified promoter assigned as P_M is the functional promoter for the transcription of *caf1M*. The results also show that inclusion of 19 bp upstream of the -35 element of this promoter is sufficient for Caf1R mediated activation, consistent with Caf1R binding to the R4' repeat motif. The fact that P_M promoter is the only predicted promoter for *caf1M* associated with a potential Caf1R binding motif is further evidence that this is the functional promoter for Caf1R-mediated activation of transcription of *caf1M*. In contrast to Caf1R-activated transcription at the P_R² promoter (from pRScaf1R'.₇₀₊₁₉₂-lacZ+pACYC-R), Caf1R-activated transcription at P_M promoter (from pRScaf1M'.₁₈₄₊₁₉₂-lacZ+pACYC-R) showed a gradual increase in activity from the exponential (6 h) to the early stationary phase (10 h) followed by slight reduction at 24 h.

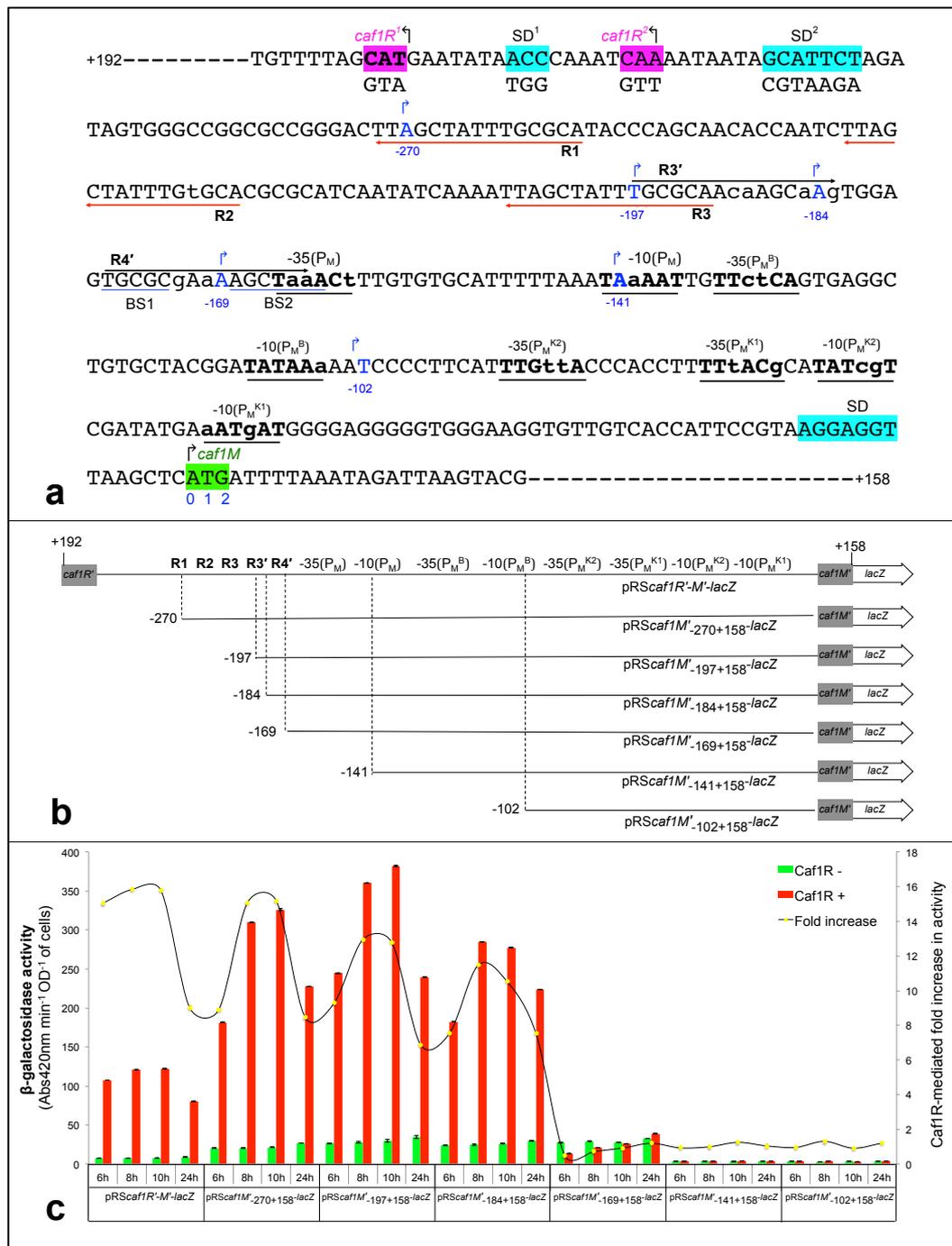


Figure 5.11 | Localisation of *caf1M* promoter and Caf1R binding upstream of *caf1M*.

a) Annotation of the *caf1R-caf1M* intergenic DNA fragment (sense strand) as in construct, pRScaf1R'-M'-lacZ. Location of the putative promoter elements (-10 and -35) of P_M, P_M^B and P_M^{K1/2} are underlined and indicated in bold; repeat motifs, R4' and R3' (black arrows), R1, R2 and R3 (red arrows). +158 and +192 indicate corresponding coding sequence of *caf1M* and *caf1R* with respect to ATG (Met) start sites. \uparrow arrows indicate location of deletions, and number of bp (-xx) upstream from *caf1M* start included in each construct. Start codons and SD sequences of *caf1M* and *caf1R* are shown, with both predicted starts (TTG and ATG) of *caf1R* highlighted. **b**) Deletion strategy and assigned promoter-lacZ fusions constructs, used to localise functional promoters for *caf1M* and Caf1R binding. **c**) β-galactosidase activity (± pACYC-R, encoding Caf1R) from promoter-lacZ fusions, sampled over 24 h as shown. Activity (ΔA420/min/OD_{600 nm}) and Caf1R-mediated fold increase in activity at primary and secondary Y-axis, respectively. Standard error (± SEM) of activity was calculated from three individual colonies of each construct. The comparative P-value for activity of pRScaf1M'-184+158-lacZ and pRScaf1R'-197+158-lacZ constructs +Caf1R were 0.006, 0.004, and 0.001 for 6 h, 8 h and 10 h respectively. (Calculated using a Paired two-sample t-test, StatPlus software).

5.4.3 Confirmation of P_M promoter by mutagenesis in pACYCF1

The predicted -10 element (TAAaAT) of the P_M promoter was substituted by GCcGGC in the pACYCF1 plasmid (carries the entire WT *caf* locus), using SDM and the Pcaf1M_2f and Pcaf1M_2r mutagenic primers as described in section 2.2.4(iv)(a). Plasmid size was monitored and the GCCGGC substitution was confirmed in plasmids from three of the *E. coli* DH5 α transformants by sequencing the complete *caf1R-caf1M* intergenic region. Plasmids were assigned the name Δ P_M_pACYCF1. Following confirmation, F1 expression was tested from all three transformants (1-3) and compared to the level of expression from three transformants (1-3) of *E. coli* DH5 α /pACYCF1 (wild type) (**Fig. 5.12**). F1 expression was monitored directly in whole cells (section 2.4.1), after 4 and 6 h thermo-induction (**Fig. 5.12**). In contrast to the high level of F1 produced in all three transformants of *E. coli* DH5 α /pACYCF1, transformants of DH5 α / Δ P_M_pACYCF1 produced no detectable F1 (Caf1). While this is a drastic mutation that may have other effects including loss of IHF binding, the results are consistent with the finding that the visually identified promoter, P_M, is a strong Caf1R activated promoter, controlling the high level of F1 expression from the *caf* locus.

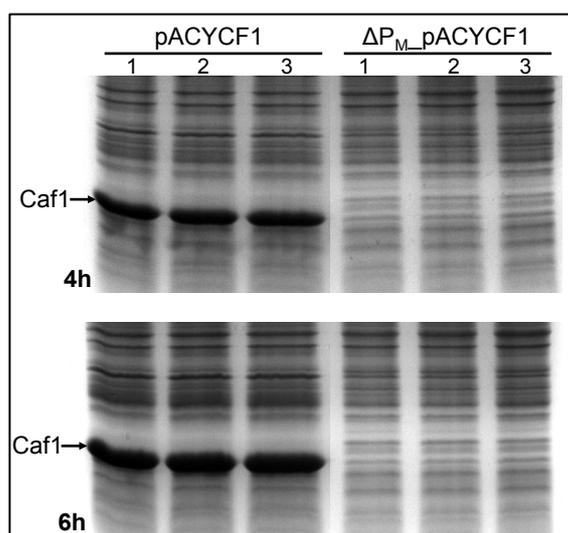


Figure 5.12| **Loss of F1 production on mutation of P_M promoter.**

Samples from three individual transformants of both wild type, DH5 α /pACYCF1 and DH5 α / Δ P_M_pACYCF1 mutant were harvested after 4 and 6 h thermo-induction at 37°C. Equivalent of 0.04 OD units of each sample of heat denatured cells (97.5°C for 15 min) were resolved on 16% acrylamide gels followed by CB staining.

5.4.4 Localisation of promoter(s) upstream of *caf1*

Although there was no indication of Caf1R-binding in the *caf1A-caf1* intergenic region in either the bioinformatics analysis (section 5.2) or with the transcriptional fusion study of cells transformed with pRScaf1A'-1'-lacZ and pACYC-R (section 5.3.3), but there was evidence of some promoter activity in this intergenic region (section 5.3.3). Therefore, six deletion constructs (**Fig. 5.13a-b**) were designed within this region to localise any potential *caf1* promoter. Four of these constructs, pRScaf1'-181+88-lacZ, pRScaf1'-158+88-lacZ, pRScaf1'-61+88-lacZ and pRScaf1'-38+88-lacZ contained the first 88 bp of *caf1* (from ATG start) and progressively larger fragments of *caf1A-caf1* intergenic region with or without bases from *caf1A* 3' terminus (indicated as -xxx). The other two constructs, pRScaf1'-259-61-lacZ and pRScaf1'-259-22-lacZ were missing the first 61 or 22 bp upstream and coding bases of *caf1*. They contained the rest of the

caf1 5' UTR upstream from -61 and -22 bp, respectively, plus the same 3' fragment of *caf1A* (179 bp), present in the pRScaf1A'-1'-lacZ construct. Thus the first construct, pRScaf1'-181+88-lacZ encompassed the DNA fragment containing repeat motifs, R1^c, R2^a and R2^b and -10 and -35 elements of P₁¹, P₁² and P₁^B promoters. The second construct, pRScaf1'-158+88-lacZ contained R1^b, R1^c, R2^a and R2^b repeat motifs and, P₁¹ and P₁^B predicted promoters plus the -10 consensus of P₁². The third construct, pRScaf1'-61+88-lacZ also contained R1^c, R2^a and R2^b repeat motifs, and P₁¹ and P₁^B predicted promoters. The fourth construct, pRScaf1'-38+88-lacZ contained R1^c, R2^a and R2^b repeat motifs, the BPRM predicted promoter, P₁^B plus the -10 element of predicted promoter, P₁¹. The other two constructs lacked promoter P₁^B but contained predicted P₁³ that was absent from the other constructs. pRScaf1'-259-61-lacZ contained R1^a and R1^b repeats motifs and P₁³ and P₁² predicted promoters. pRScaf1'-259-22-lacZ contains R1^a, R1^b, part of R2^a repeat motifs and P₁³, P₁², P₁¹ and the -35 element of P₁^B predicted promoter.

Negligible activity was observed with inclusion of promoters, P₁¹⁻³ in constructs pRScaf1'-259-61-lacZ and pRScaf1'-259-22-lacZ (**Fig. 5.13c**). Thus, P₁¹⁻³ do not appear to be functional promoters for transcription of *caf1*. In contrast, all other constructs from the smallest (pRScaf1'-38+88-lacZ) upwards showed a low level of promoter activity in the exponential (6 h, 23.6 ± 0.05 units) and stationary (at 8 h, 36.6 ± 0.07 units) growth phase (**Fig. 5.13c**). This is consistent with P₁^B being the sole active promoter in this region. However as raised earlier, the location of this promoter adjacent to the start of *caf1* means that it cannot be the promoter responsible for transcription of full-length *caf1*. The possibility remains that it might function as a promoter for the putative small ORF encoding a short peptide of 69-amino acids (**Fig. 5.3**). Caf1R had negligible effect on the level of transcription from this part of the *caf* locus.

In conclusion, no promoter was identified between the beginning of *caf1* and the last 179 bp of *caf1A* that could be responsible for transcription of full-length *caf1* and there was no evidence to support Caf1R binding within this region. BPRM predicted a strong promoter (strongest within the *caf* locus) at 192 bp downstream from the beginning of *caf1A*. The LDF score of this promoter is 6.68 having -10 (TAgAcT; score 62) and -35 (TTaAtg; score 29) and a 21 nucleotide spacer. Activity of this promoter was not tested. While it could in theory promote additional transcription of *caf1*, the transcript would have a leader sequence of 2375 bp.

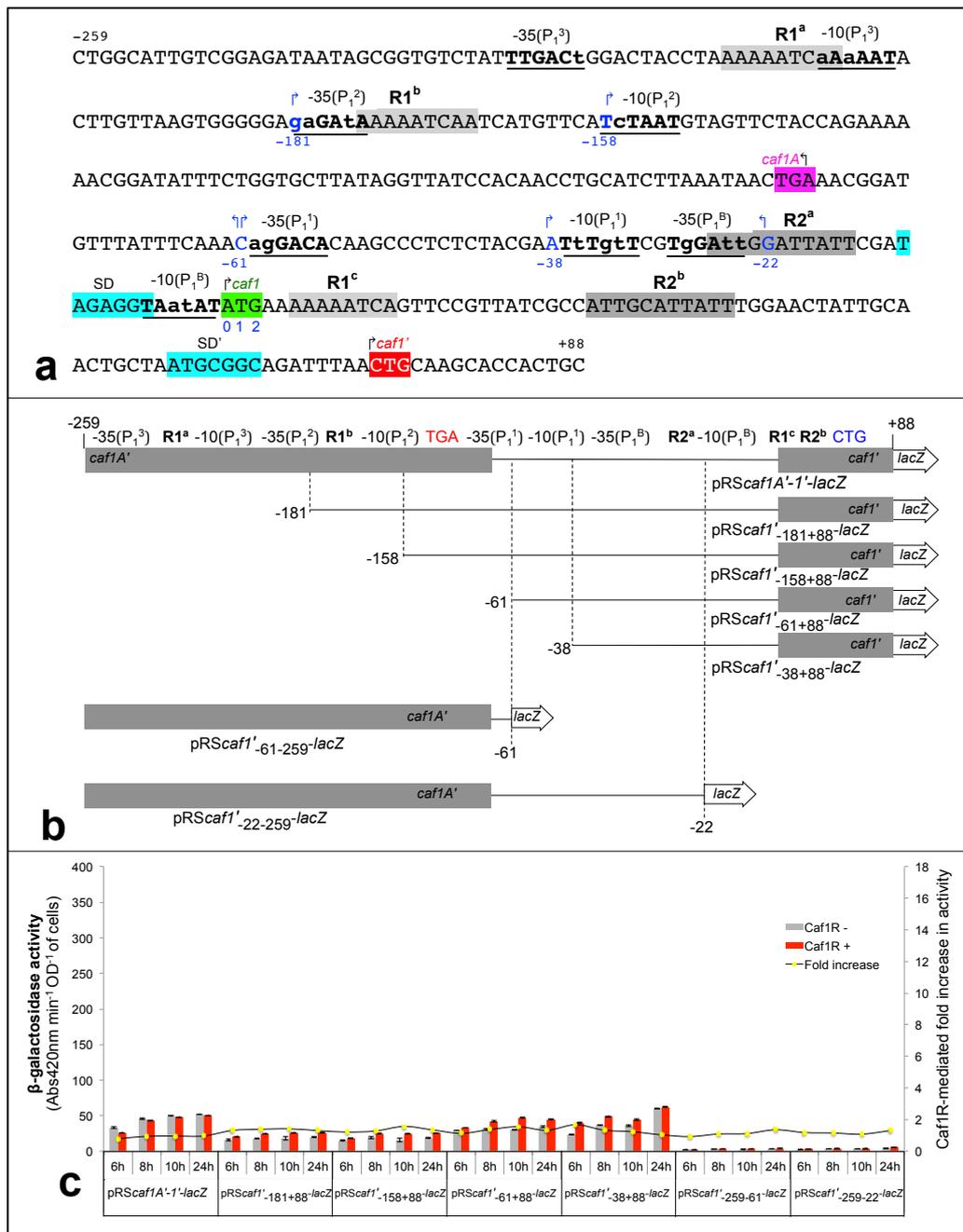


Figure 5.13 | Localisation of *caf1* promoter and lack of effect of Caf1R.

a) Annotation of *caf1A-caf1* intergenic DNA as assigned in section 5.2.2 and included in pRScaf1A'-1'-*lacZ*. Location of the putative promoter elements (-10 and -35) for P₁¹⁻³ and P₁^B are underlined and indicated in bold, repeat motifs, R1^{a-c} and R2^{a-b} are shaded in grey. -259 and +88, numbering of residues with respect to the A of *caf1* ATG start codon (green). TGA stop codon of *caf1A* is in magenta. ↑, arrows indicate start of inserts in fusion constructs with sequence XX to +88; ↓, arrows indicate end of inserts in fusion constructs with insert -269 to XX. **b)** Deletion strategy and assigned promoter-*lacZ* fusions, used to localise functional promoters for *caf1* and to test the impact of Caf1R. **c)** β-galactosidase activity (± pACYC-R, encoding Caf1R) from promoter-*lacZ* fusions, sampled over 24 h as shown. Activity (ΔA420/min/OD_{600 nm}) and Caf1R-mediated fold increase in activity at primary and secondary Y-axis, respectively. Standard error (± SEM) of activity was calculated from three individual colonies of each construct.

5.5 *In vitro* confirmation of Caf1R binding to the R4' repeat, upstream of P_M

Following *in vivo* confirmation of P_M as the main Caf1R-activated promoter within the *caf* locus, an *in vitro* protein-DNA binding approach, electrophoretic mobility shift assay (EMSA) (section 2.10), was used to decipher tagged Caf1R-P_M interactions. Freshly prepared cell lysates containing hCaf1R^{Tgs}, MBPCaf1R_{gs} or MBPCaf1R_N as well as purified protein were used to monitor binding of Caf1R to oligonucleotides containing the putative Caf1R binding repeat R4'. The P_M promoter specific and non-specific DNA probes were selected from the *caf1R-caf1M* intergenic region. The specific biotinylated probe (F2-B) was designed based on results of the promoter fusion assay (Fig. 5.11) and the Caf1R consensus sequence (Fig. 5.4). It contained 47 bp from -140 to -186 bp upstream of *caf1M* start codon, ATG and included the -35 element of P_M promoter and the R4' repeat motif. The non-specific biotinylated probe (F1-B), also 47 bp was selected from -56 to -102 bp upstream of *caf1M* start codon, ATG and includes no recognisable predicted Caf1R binding sequence (Fig. 5.14).

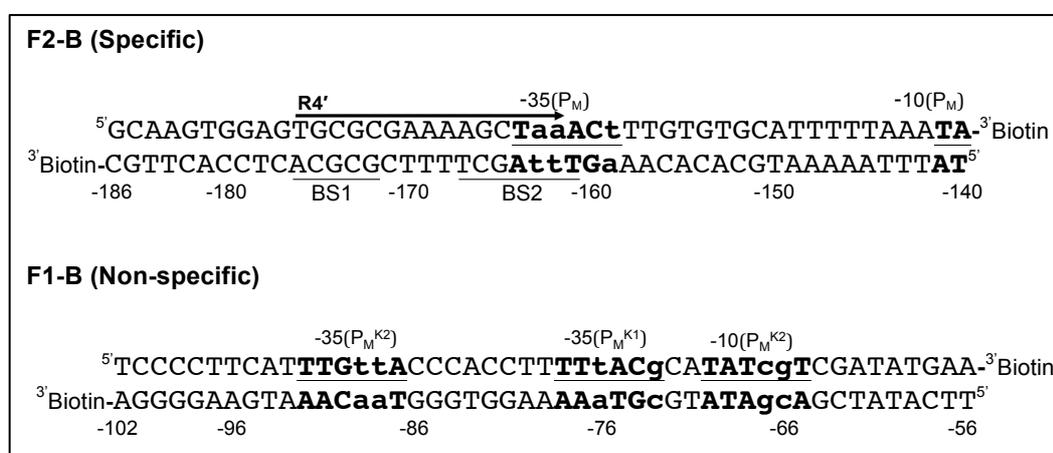


Figure 5.14| **Biotin tagged specific (F2-B) and non-specific (F1-B) EMSA probes.**

F2-B (47 bp), contained predicted Caf1R-binding R4' repeat motif (arrow) plus part of the downstream P_M promoter (bold and underlined). BS1 and BS2 (underlined), putative binding site for Caf1R helix-3 and helix-6 according to *mar*-MarA complex (Rhee et al., 1998). F1-B (47 bp) contains predicted but non-functional P_M^{K1} and P_M^{K2} promoters upstream of *caf1M*. Small bold letters indicate non-consensus nucleotides in the predicted promoter elements. Numbers underneath of F1-B and F2-B are relative to ATG start codon of *caf1M*. See Figs. 5.11 and 5.13 for confirmation of Caf1R dependent functional and non-functional promoters. F2-B and F1-B, biotin labeled probes; F2 and F1, corresponding unlabeled oligonucleotides.

5.5.1 EMSAs with hCaf1R^{Tgs}

i) *In vitro* binding of wild type and mutated hCaf1R^T from cell lysates

Initial binding studies were performed using freshly prepared cell lysates (4 and 6 μl) from *E. coli* BL21(DE3) expressing wild type hCaf1R^{Tgs} (from pEThCaf1Rgs), mutated hCaf1R^{Tgs}_{E98G} (from pEThCaf1Rgs_{E98G}) or no Caf1R (negative control, from pET28a⁺) together with the test (F2-B) or control (F1-B) biotin-labeled probes (5-fmol each). Binding reactions were set up as described in section 2.10.4 and subjected to EMSA as detailed in section 2.10.5. A prominent shift in migration of the biotinylated-F2-B DNA probe was observed with both 4 and 6 μl lysates containing wild type hCaf1R^{Tgs} (Fig. 5.15a, lanes 3 and 7) indicating formation of a hCaf1R^T-F2-B complex (B). The 6 μl lysate produced a stronger shifted band than the 4 μl sample, consistent with dependence of the interaction on Caf1R concentration. A competition reaction, with 100-fold molar excess unlabeled F2 probe (500 fmol) in the binding reaction of wild type

hCaf1R^{Tgs} lysate and F2-B probe showed no shift in the labelled F2-B probe (**Fig. 5.15a, lane 5 and 9**), providing evidence for specificity of binding. Specific binding of hCaf1R^{Tgs} to the test probe, F2-B was confirmed by the very faint shifted band of F1-B (control probe) with 6 μ l hCaf1R^{Tgs} lysate (**Fig. 5.15a, lane 10, WT**). As expected no shift in F2-B was observed from the reactions containing no lysate (NL, lane 1). This and the absence of any shift with lysates obtained from cells carrying empty pET28a⁺ vector (C) (**Fig. 5.15a, lane 2, 6 and 12**) supports identity of the shifted band as a complex of hCaf1R^{Tgs}-F2-B rather than a specific interaction of this DNA sequence with some other cell factor. The very faint complex observed with 4 μ l lysate containing the mutated regulator, hCaf1R^{TgsE98G} (**Fig. 5.15a, lane 4, E98G**) further supports the identity of the complex as hCaf1R^{Tgs}-F2-B.

Thus this initial study with cell lysates provided strong evidence that wild type hCaf1R^{Tgs} binds with high specificity to the F2-B probe, which includes the Caf1R binding consensus sequence. The very weak activity of hCaf1R^{TgsE98G} mutant is interesting, as this would explain loss of transcriptional activation by Caf1R possessing this G98 residue. However, it could also be explained by instability of the mutated protein. Therefore, an immunoblot was performed using fresh lysates along with a repeat of the EMSA (**Fig. 5.15b-c**). It was observed that in this EMSA, hCaf1R^{Tgs}-F2-B interaction resulted in two shifted bands, B1 and B2, in contrast to the single shift detected earlier (**Fig. 5.15b**). Two different shifted complexes have been attributed by others to monomeric and dimeric protein-DNA complexes (Carolina and Celso, 2012; Elliot et al., 2003). This is seen with AraC/XylS-type regulators that function as both activator and repressor (Hollander et al., 2011; Ruiz et al., 2003). A low level of complex only at position B2 was detected with lysate from the hCaf1R^{TgsE98G} mutant. The western blot analysis of these lysate samples detected both hCaf1R^{TgsE98G} and hCaf1R^{Tgs} (**Fig. 5.15c**). The lower level of hCaf1R^{TgsE98G} may suggest a greater instability of hCaf1R^{TgsE98G}, but probably not sufficient to account for the differences in binding activity. To differentiate the impact of the E98G mutation on DNA binding versus regulator stability, this would be best addressed by comparing specific binding activity of both purified proteins.

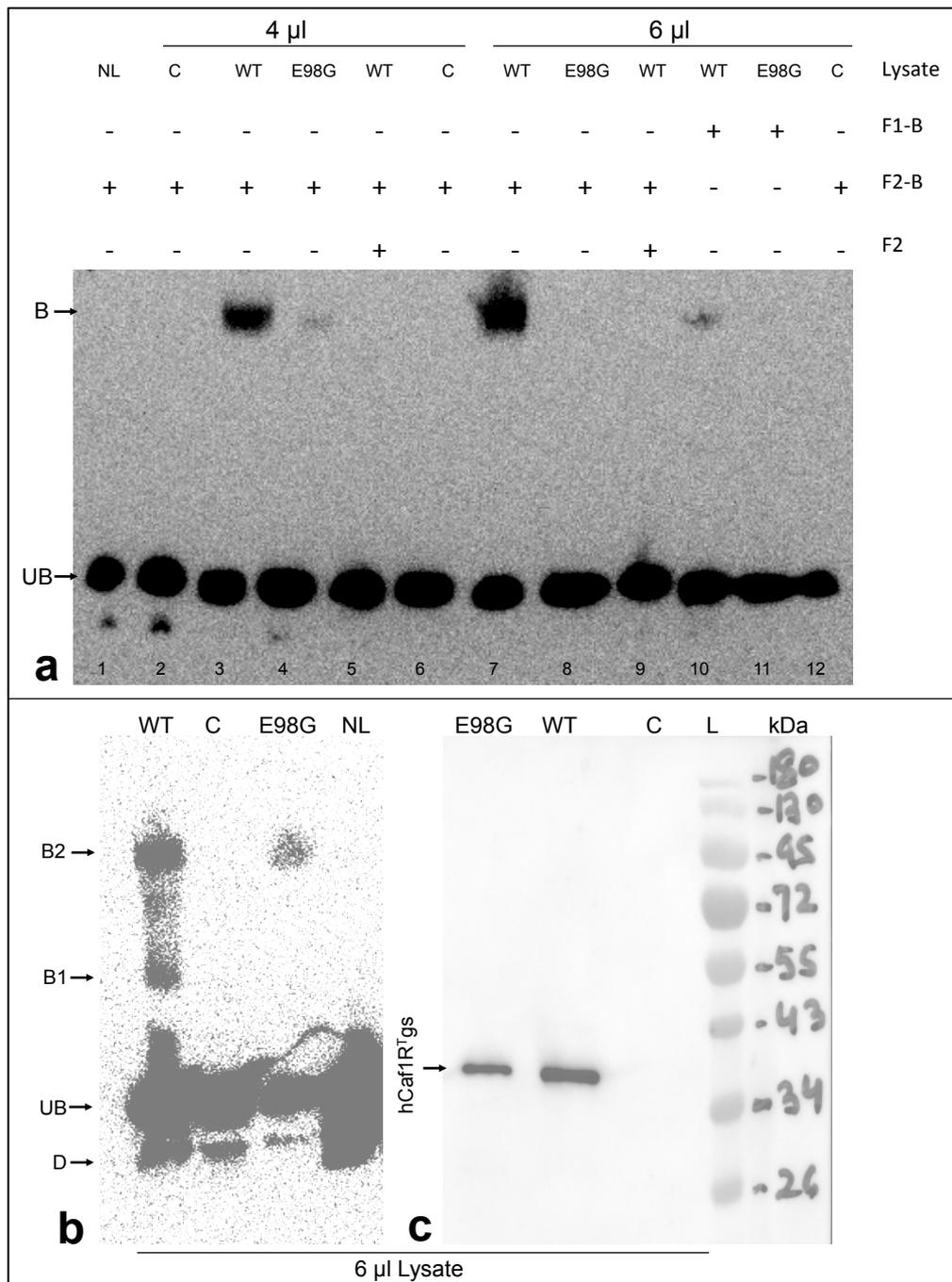


Figure 5.15| EMSA with cell lysates of wild type and mutated hCaf1R^{Tgs}.

a) EMSA blot. Lysates (4 or 6 μ l) of BL21(DE3)/pET_{hCaf1R}gs (WT), pET_{hCaf1R}gs_{E98G} (E98G) or /pET28a⁺ (C, vector control) were incubated with 5-fmol biotinylated DNA probe. Reactions were analysed by DNA retardation gel electrophoresis and visualised using Streptavidin-HRP, as in methods, section 2.10. **b**) Repeat EMSA analysis with hCaf1R^{Tgs} and hCaf1R^{Tgs}_{E98G} (from fresh 6 μ l cell lysate) using F2-B probe (10-fmol). **c**) Immunoblot of fresh lysates (6 μ l) used in (b) with AntiHis-HRP monoclonal antibody (1:10,000) to detect His₆-tagged Caf1R. Probes: F2-B (test, includes proposed Caf1R binding R4' repeat); F1-B (control, *caf* DNA without Caf1R-binding consensus); F2, non-biotinylated test (specific) probe. B/B1/B2: Retarded hCaf1R^{Tgs}-F2-B complexes, D: degradation product and UB: unbound biotinylated probe. NL: no lysate. kDa: SDS-PAGE molecular weight markers. Arrow indicates hCaf1R^{Tgs}, at about 38 kDa.

ii) *In vitro* binding of IMAC purified hCaf1R^T_{gs} to Caf1R specific F2-B probe

To confirm hCaf1R^T_{gs} binding to the R4' DNA probe (F2-B) fresh IMAC purified hCaf1R^T_{gs} (**Fig. 4.21, Fn-10, 40 µg/ml**) was concentrated and dialysed against 20 mM Tris, 300 mM NaCl, pH 8.0 and the final preparation (245 µg/ml, 6308.7 pmol/ml) stored at 4 °C on ice. This was tested in EMSA (section 2.10) using 2.5 to 50 pmol of protein and 10-fmol of F2-B probe. Again 2 band shifts (protein-F2-B complexes), B1 and B2, were observed (**Fig. 5.16**), indicating that purified hCaf1R^T_{gs} may bind as monomer and dimer. However, binding was irrespective of varying protein concentrations and the unbound (UB) F2-B DNA was prominent, suggesting that hCaf1R^T_{gs} had not bound optimally to F2-B probe.

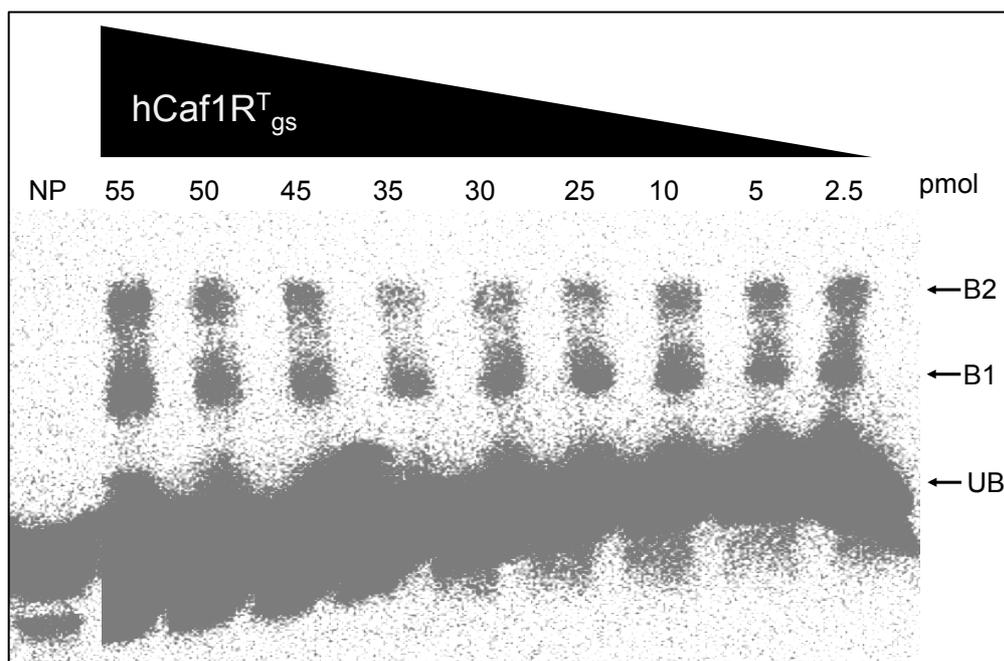


Figure 5.16| *In vitro* binding of IMAC purified hCaf1R^T_{gs} to Caf1R specific F2-B probe. EMSA blot using purified hCaf1R^T_{gs} (6.3 pmol/µl in 20 mM Tris, 300 mM NaCl, pH 8.0) and F2-B biotinylated probe. Increasing amounts of hCaf1R^T_{gs} (2.5-55 pmol) as indicated, were incubated with the Caf1R specific F2-B probe (10-fmol), non-specific competitor DNA Poly dI.dC (50 ng), 1× reaction-binding buffer (10 mM Tris, 50 mM KCl and 1 mM DTT; pH 7.5) in a final reaction volume of 20 µl (section 2.10). B1 and B2: hCaf1R^T_{gs}-F2-B complexes. UB: Unbound, F2-B probe and NP: negative control with no Caf1R protein.

5.5.2 EMSAs with MBPCaf1R and MBPCaf1R_N

His₆-tagged Caf1R consistently bound specifically to the Caf1R specific F2-B DNA fragment, but binding of isolated protein was not optimal. Much higher levels of recombinant Caf1R were recovered with MBPCaf1R_{gs} fusions (Chapter-4). Therefore, MBPCaf1R_{gs} and MBP fused to the N-terminal DBD of Caf1R (MBPCaf1R_N) also were tested for DNA binding capacity.

i) *In vitro* binding of MBPCaf1R_{gs} and MBPCaf1R_N from cell lysate

Cell lysates of *E. coli* K12-ER2508 carrying pMALc2-MBPCaf1R_{gs}, pMALc2-MBPCaf1R_N or pMALc2x empty vector were prepared and subjected to EMSA using 6 µl of corresponding cell lysates and 10-fmol of F2-B probe as described in section 2.10. A prominent shift of protein-DNA complex (B) was observed from the lysates of both pMALc2-MBPCaf1R_{gs} (LMBP-R) and

pMALc2-MBPCaf1R_N (^LMBP-R_N) (**Fig. 5.17a**). The absence of any shifted band in the negative control, cell lysate containing MBP alone from pMALc2x (^LMBP), confirmed that the shifted F2-B band (B) is due to interaction of Caf1R/Caf1R_N with F2-B. Both MBPCaf1R_{gs} and MBPCaf1R_N resulted in an intense band of the shifted complex, suggesting that the N-terminal domain of Caf1R alone is sufficient to mediate DNA binding. The ratio of F2-B in complex (B) to unbound F2-B (UB) was much larger with the MBP fusions (MBPCaf1R_{gs} and MBPCaf1R_N) than assays using hCaf1R^{Tgs} lysates. This can most likely be attributed to the fact that there is substantially more MBPCaf1R_{gs} and MBPCaf1R_N in these lysates than hCaf1R^T. MBPCaf1R_{gs} and MBPCaf1R_N fusions represented the strongest CB stainable band cell lysates (**Fig 5.17b**), whereas hCaf1R^T was only visible following immunoblotting of cell lysates (**Fig 5.15c**).

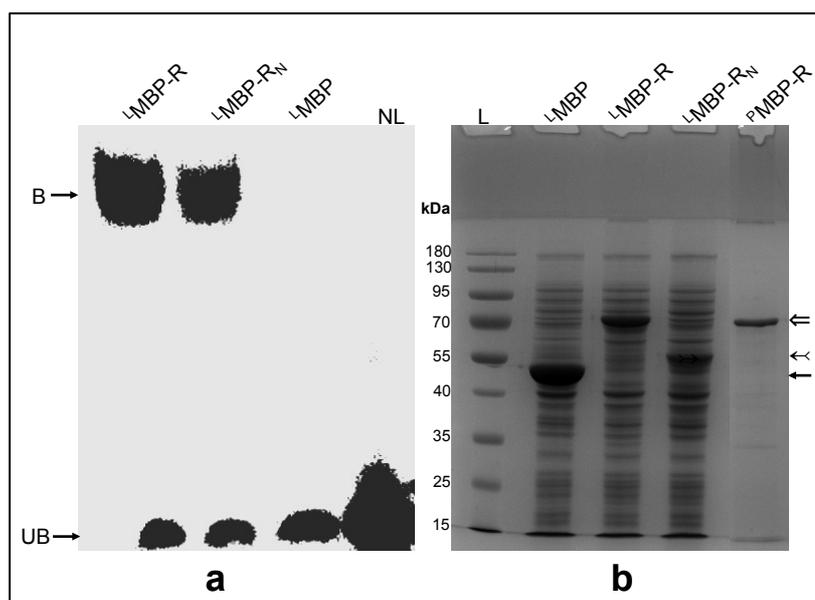


Figure 5.17| **EMSA with cell lysates of MBPCaf1R and MBPCaf1R_N.**

a) EMSA blot. **b)** CB stained SDS-PAGE gel from cell lysates of *E. coli* K12-ER2508 carrying pMALc2-MBPCaf1R_{gs} (^LMBP-R, full length Caf1R), pMALc2-MBPCaf1R_N (^LMBP-R_N, Caf1R N-terminus BDB) or pMALc2x (^LMBP, vector control), induced with 0.3 mM IPTG for 2.5 h at 37°C. For EMSA, 6 µl of cell lysate, was tested with 10-fmol biotinylated F2-B probe as described, section 2.10. NL: no lysate, unbound F2-B control. B: retarded protein-DNA complex from MBP-R/R_N- and F2-B probe. UB: unbound F2-B probe. For SDS-PAGE analysis, 6 µl of corresponding cell lysates or 100 pmol of purified/dialysed MBPCaf1R_{gs} (^PMBP-R) was mixed in 5 mM DTT supplemented SDS-PAGE sample buffer (4×), heat denatured (97.5°C-15 min) and resolved on 12.5% polyacrylamide containing gel followed by CB staining. MBPCaf1R_{gs} (78.98 kDa) and MBPCaf1R_N (57.35 kDa) are indicated by ⇒ and → arrow, respectively. MBP alone (42.956 kDa) is indicated by → arrow. Identity of MBP and MBPCaf1R/R_N fusions was confirmed by Western blotting using anti-MBP antibody (data not shown).

ii) *In vitro* binding of purified MBPCaf1R_{gs} to Caf1R specific F2-B probe

As for purified wild type hCaf1R^{Tgs}, purified MBPCaf1R_{gs} (**380 µg/ml, Fig. 4.25, Fn-3**) was concentrated and dialysed against 20 mM Tris, 300 mM NaCl, pH 8.0 and the final preparation (740 µg/ml or 9366.8 pmol/ml) stored at 4°C on ice. MBPCaf1R_{gs} (2.5, 5.0, 10, 25, 30, 35, 45 and 50 pmol) was tested with 10-fmol F2-B probe in EMSA (section 2.10). Despite the very promising results with cell lysates, poor binding was observed with the purified MBP fusion

proteins irrespective of protein concentration (**Fig. 5.18**). Consistent with purified wild type hCaf1R^{Tgs}, two faint shifted complexes, B1 and B2 could be detected.

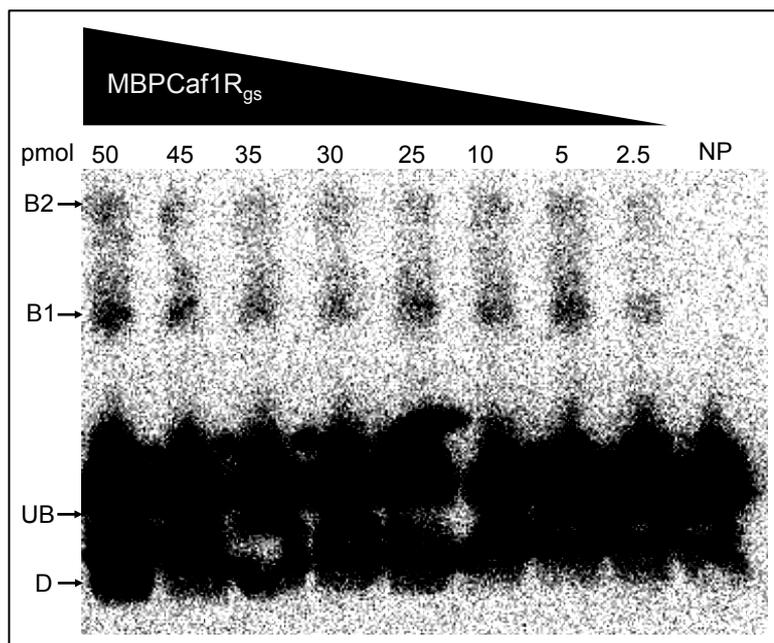


Figure 5.18| *In vitro* binding of purified MBPCaf1R_{gs} to Caf1R specific F2-B probe.

EMSA blot using purified and dialysed MBPCaf1R_{gs} (9.3 pmol/μl) with F2-B probe, as in section 2.10. B1 and B2: MBPCaf1R_{gs}-F2-B shifted complexes. UB: Unbound F2-B probe, NP: negative unbound DNA probe control, no protein and D: degradation product.

5.5.3 Concentration-dependent binding of purchased hCaf1R^P to F2-B probe

In order to establish the concentration dependent effect of tagged Caf1R binding to the F2-B DNA probe, EMSA was repeated using purchased hCaf1R (hCaf1R^P) (MyBiosource, USA). This protein was provided at a concentration of 500 μg/ml in a storage buffer containing 20 mM Tris-HCl, 500 mM NaCl, 50% Glycerol; pH 8.0, stored at -20°C. This preparation of hCaf1R^P was shown to be free from large aggregates by centrifugation (16,000×g/40 min/4°C). Prior to setting up EMSA binding reactions, this protein was also dialysed, as above and then used at a broader range of protein (0.10, 0.25, 0.5, 25 and 50 pmol) and double the amount of F2-B probe (20-fmol). Earlier studies used purified hCaf1R^{Tgs} and MBPCaf1R_{gs} (2.5, 5.0, 10, 25, 30, 35, 45, 50 and 55 pmol) and F2-B probe (10-fmol). Following EMSA (section 2.10), an equal amount of both B1 and B2 complex were again seen, but unlike the previous assay, concentration dependent binding of hCaf1R^P to the F2-B probe was obtained as can be seen from the gradual increase in intensity of B1 and B2 complexes from low (0.1 pmol) to high (50 pmol) concentrations of hCaf1R^P (**Fig. 5.19**). Strangely, as protein concentration increased the intensity of unbound probe at the bottom also appear to increase. The reason for this increase in intensity of unbound probe is unclear, but might be related to increasing salt concentrations in each reaction, as the amount of added protein increased.

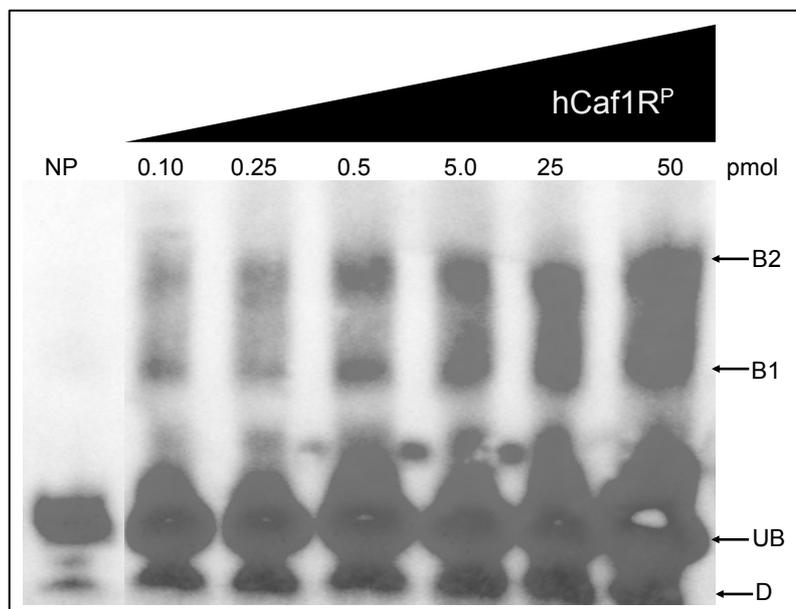


Figure 5.19| **Concentration-dependent binding of purchased hCaf1R^P to Caf1R specific F2-B probe.** EMSA blot with purchased/dialyzed hCaf1R^P (12.8 pmol/μl) and F2-B probe. Increasing amounts of hCaf1R^P (0.1-50 pmol) as indicated were mixed with 1× reaction-binding buffer (10 mM Tris, 50 mM KCl and 1 mM DTT; pH 7.5), non-specific competitor DNA Poly dI.dC (50 ng) and biotinylated F2-B specific probe (20-fmol) in a final reaction volume of 20 μl and subjected to EMSA (section 2.10). B1 and B2: hCaf1R^P-F2-B shifted complexes, possibly representing monomer (B1) and dimers (B2) of HCaf1R. UB: Unbound F2-B probe and NP: negative control with no protein and D: Degradation product.

5.6 Conclusion

The results presented in this chapter show that the *caf1R-caf1M* intergenic region (327 bp) is the main regulatory region of the *caf* locus and provides evidence for identification of Caf1R activated promoters for both *caf1R* and *caf1M*. Five potential Caf1R binding sites of 15 nt were identified in this region, three, R1-R3 upstream of *caf1R* and two, R3' and R4' upstream *caf1M*. A Caf1R-dependent promoter, P_M was shown to activate the transcription of *caf1M*, *in vivo*, using *lacZ*-promoter fusions. The basal expression of *caf1M* from P_M, in the absence of Caf1R was very low suggesting that P_M is more tightly controlled than that of promoters upstream to *caf1R*. The fusion study also provided preliminary evidence of Caf1R-binding upstream of P_M to the R4' repeat, which overlaps the -35 promoter element by 3 nt. Overlap of the Caf1R-binding site with the -35 element categorizes the *caf* P_M promoter as a class II activated promoter of AraC/XylS family regulators (Duval and Lister, 2013; Egan, 2002; Gallegos et al., 1997). With class II promoters, RNA polymerase interacts with the DNA binding activator stabilizing the DNA-regulator interaction (Martin et al., 2002). Binding of Caf1R to R4' repeat was confirmed by *in vitro* EMSA studies using cell lysates containing hCaf1R^{Tgs} or MBPCaf1R_{gs}. In the EMSA studies, Caf1R-binding to the labelled R4' repeat DNA fragment was more efficient with cell lysates than with the isolated proteins. This may suggest that the purified versions of tagged Caf1R lack some important cell factors (lost during purification), essential for optimal binding. For example, RNA polymerase may be required for optimal binding of regulator to the cognate DNA, as suggested by others (Duval and Lister, 2013; Martin et al., 2002). Alternately, recombinant Caf1R may require optimal buffer conditions to maintain reduced status and solubility during purification. Optimisation of the EMSA would permit detailed quantitative comparison of binding to compare specificity of binding to different sequences. The impact of R3' on Caf1R mediated activation is

unclear. This sequence is less conserved than the other 4 binding sites. The promoter-*lacZ* fusion assay indicated slight enhancement of activation in the presence of both R3' and R4' binding sites. Whether this is an indication of dimer binding could be investigated.

Comparison of R4' with the other 4 repeats within the *caf1R-caf1M* intergenic region permitted identification of a potential Caf1R binding consensus motif, 5'TGCRCRAMWAGCWARD3' (underlined nt indicate BS1 and BS2). Within BS1 (TGCRC), both C nt were absolutely conserved when compared to the MarA/Rob/SoxS consensus (Duval and Lister, 2013) and the G varied only in MarA. Within BS2, MarA (Rhee et al., 1998) and Caf1R binding sites were identical over 6 nt (GCTAAA). In the *mar*-MarA crystal structure (Rhee et al., 1998) nt of both binding sites are involved in specific recognition and binding of helix-3 and 6 amino acids. This is considered in more depth in Chapter-7. The higher level of conservation of residues in BS2 compared to BS1 has been attributed to a shared specificity of binding in BS2 among this subfamily of AraC/XylS-like regulators and possibly greater specificity of BS1 binding by members within the family.

Promoter localisation identified no active promoter upstream of *caf1*, responsible for transcription of full-length *caf1*. The results were more consistent with polycistronic transcription of the 3 genes *caf1MA1*. For *caf1R* transcription, a promoter assigned as P_R² was identified as the main Caf1R-dependent promoter for the divergent transcription of *caf1R*. This could also be classified as a class II Caf1R activated promoter. The R1 repeat motif, also lies just upstream of the -35 promoter element, overlapping it by 2 nt. While transcription of *caf1R* is clearly autoregulated, the significance of the remaining 2 Caf1R binding sites (R2 and R3) is not clear at this stage and requires further study. R2 is located between the -10 and -35 elements of P_R^K promoter. The statistically significant decrease in β-galactosidase activity upon inclusion of repeats R2 and R3 might suggest that as levels of Caf1R increase, Caf1R acts as a repressor at R2 and/or R3. In addition, inclusion of an extended sequence (up to R3) encompassing the P_R^K sequence provided an indication of a basal level of Caf1R independent transcription from P_R^K.

Conclusively, two class II promoters, P_R² and P_M were identified, which activate divergent expression within the *caf* locus on Caf1R binding. Caf1R-P_R² autoregulates its own transcription and Caf1R-P_M controls transcription of *caf1MA1*. The role of P_R^K promoter in transcription remained to be clarified.

Chapter 6

Thermoregulation in the *caf* locus of *Y.pestis*

6.1 Introduction

Following entry of *Y. pestis* into the animal host, the host body temperature acts as a decisive signal and a range of temperature dependent virulence factors are expressed, which help in the survival of bacteria in the host. F1 is one such factor, which is expressed at a high-level in the host (Straley and Perry, 1995). An early DNA microarray study of global changes in gene expression in *Y. pestis* on temperature transition from 26 to 37°C reported upregulation of all genes within the *caf* locus, *caf1R* (encodes regulator), *caf1M* (encodes periplasmic chaperone, *caf1A* (encodes outer membrane usher) and *caf1* (encodes F1 subunit) (Motin et al., 2004). The level of *caf1* alone was increased about 100-fold at 10 hours, and this was the highest fold increase in transcript levels within the entire genome. A second study also reported upregulation of the *caf* genes in LB on temperature shift from 28 to 37°C. This study reported a 22.7-fold increase for *caf1A* and 41.5-fold increase for *caf1*, but only 4.3-fold increase for *caf1M* and 2.3-fold increase for *caf1R* (Chauvaux et al., 2007). In the same year, one other transcriptomic study confirmed a higher level of expression of the *caf* genes at 37°C compared to growth at 26°C, after just 1 hour shift in temperature. However, the fold-increase was similar for all genes 3.16 (*caf1R*), 4.30 (*caf1M*), 3.28 (*caf1A*) and 2.96 (*caf1*) (Han et al., 2007). The increased level of *caf1* transcript relative to the level of *caf1M* and *caf1A* with time would be consistent with high stability specifically of the *caf1* transcript or an additional promoter. Recently a global gene expression study of *Y. pestis* in macrophages revealed a gradual increase in the *caf* genes expression over 8 hours (Fukuto et al., 2010). A protein microarray study containing 218 antigens of *Y. pestis* showed much higher F1 antibody response in a pooled plague serum which was adsorbed by *Y. pestis* grown at 37°C than with *Y. pestis* grown at 26°C. This is again consistent with overexpression of F1 from temperature upshift from 26 to 37°C (Li et al., 2011). Hence these studies all confirm that expression of F1 is highly upregulated at 37°C. The mechanism of this thermoregulation within the *caf* locus has never been studied. Thermoregulation of recombinant F1 in heterologous systems such as in *E. coli* (Simpson et al., 1990) and in *Salmonella* (Cao et al., 2012; Titball et al., 1997) provides good evidence that thermoregulation of *Y. pestis* F1 expression can be studied in *E. coli*.

A recent review has given a detailed account of both transcriptional and post-transcriptional mechanisms of thermoregulation in *Yersinia* (Steinmann and Dersch, 2013). Examples include temperature-induced changes in DNA topology and changes in stability of DNA binding proteins (transcriptional level) and thermoregulation *via* thermosensitive RNA secondary structure (post-transcriptional level). An example of transcriptional based thermoregulation in *Yersinia* occurs *via* concerted interplay of H-NS, YmoA and the DNA topology. H-NS binds to DNA specifically as dimers with a preference for AT rich regions with a bend in the DNA (Amit et al., 2003; Dame et al., 2000; Spurio et al., 1997). YmoA (18 kDa) is a small thermosensitive protein that interacts with H-NS and thus enhances downregulation of transcription (Madrid et al., 2007). At higher temperatures (37°C), YmoA is rapidly degraded by host proteases (ClpP and Lon) and the ability of H-NS to bind its target sequence is decreased. This relieves the repression, permitting transcription. A well-characterised example of thermoregulation at both the transcriptional and translational level is expression of *lcrF*, the gene encoding a transcriptional activator (AraC/XylS family) of the *Yersinia* Type 3 secretion system (Bohme et

al., 2012; Hoe and Goguen, 1993). At ambient temperature (25°C), the YmoA regulator binds downstream of the transcription start site (TSS) and thus represses *lcrF* transcription. Upon *Y. pestis* entry into the mammalian host, at the higher temperature YmoA is rapidly degraded permitting transcription of *lcrF* (Bohme et al., 2012). At the post-transcriptional level, translation of the *lcrF* transcript is blocked by a two-stem loop RNA structure, located within the *yscW-lcrF* intergenic region (Bohme et al., 2012; Hoe and Goguen, 1993). This RNA secondary structure sequesters the RBS (AGGA) by annealing with a consecutive 4U sequence and prevents access of the ribosomes at lower temperature. On temperature increase, during infection, structural changes in this RNA secondary structure (melting of the stem-loop structure), allows access to ribosomes, synthesis of LcrF regulator and consequently induction of all LcrF-dependent virulence genes.

In general, bacterial pathogens often use these type of thermosensitive RNA secondary structures to synthesis virulence factors only when required and respond to changes in temperature (Righetti and Narberhaus, 2014). These structures are usually located within the 5' untranslated region (UTR) of the transcript of virulence genes and are known as RNA thermometers (RNATs) (Righetti and Narberhaus, 2014). These RNATs modulate the expression of virulence genes in response to temperature changes and provide a control mechanism that is fast, energy saving and bypasses the requirement for expression of transcription factors. These RNATs generate a base-paired structure that includes the ribosomal binding site (Quade et al.) and/or translation initiation codon (AUG) at the lower temperature (**Fig. 6.1**), thus blocking translation. Following temperature upshift to 37°C, these RNATs melt in a zipper-like mechanism and liberate the RBS and/or AUG initiation codon which is then free to form the translation initiation complex with ribosomes, as required for translation (**Fig. 6.1**).

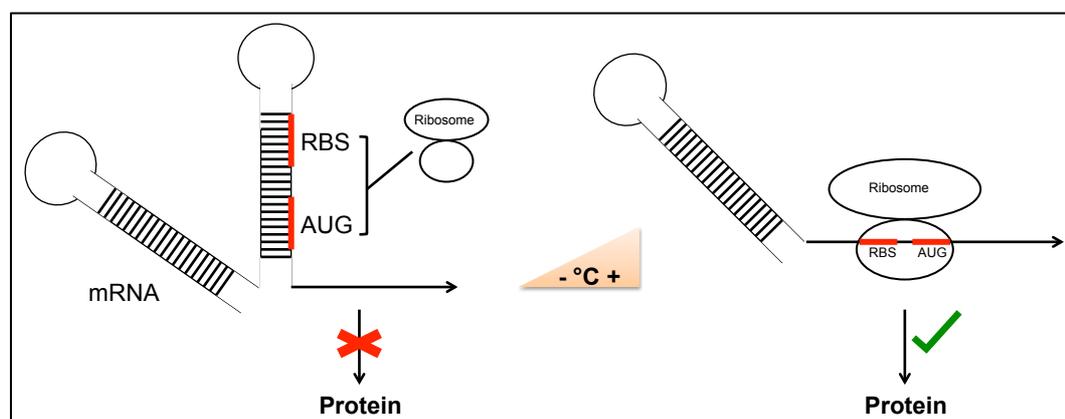


Figure 6.1| **General principle of RNAT mediated control of protein expression.**

At ambient temperature RNAT masks the RBS and the translation start site (AUG) by base pairing within a secondary structure. The increased temperature modulates the structure of RNAT and liberates RBS and AUG for subsequent translation. Picture adapted from (Righetti and Narberhaus, 2014).

In view of the recognised temperature dependent regulation of expression of the *caf* genes, it is reasonable to hypothesize that post-transcriptional control *via* a thermosensitive RNA secondary structure(s) might contribute to control of expression of F1. Therefore, the aims of this chapter were specifically to localise the site(s) within the *caf* locus that is/are involved in

thermoregulation and to explore potential RNATs upstream of *caf1R* and *caf1M* that could contribute to thermoregulation of F1 expression.

Results and Discussion

6.2 Thermoinduction of F1 expression through Caf1R

As Chapter-5 provides the indirect evidence that Caf1R activates F1 expression by binding to the 5' UTR of both *caf1R* and *caf1M*. The promoter-*lacZ* fusions and EMSA identified binding to the R4' repeat, located upstream of the P_M promoter. Fusions also identified autoregulation of Caf1R and binding most likely at the R1 repeat adjacent to the P_R² promoter. As a first step to investigate thermoinduction of F1 expression, the impact on *caf1R* expression was tested. A complementation assay was used that could test if the 5' UTR of *caf1R* contributes to this temperature dependent shift in expression of F1. Two complementing combinations of constructs used were pACYCF1+pBADHisA (complete *caf* locus + pBADHisA vector) and pACYC-MA1+pBADhCaf1R (*cafMA1* operon, without regulator + P_{BAD} controlled hCaf1R) (**Fig. 6.2**). The hypothesis was that Caf1R production from the *caf* locus within pACYCF1 is controlled by a thermosensitive RNAT, which may be located at 5' UTR of *caf1R*, but that expression of Caf1R from pBADhCaf1R would not be thermosensitive because of its regulation by temperature independent P_{BAD} promoter (L-ara inducible). There are many reports of successful expression from pBADHis vectors at lower temperatures (Bohme et al., 2012; Guzman et al., 1995; Hoe and Goguen, 1993; Klinkert et al., 2012; Weber et al., 2014). In both pACYCF1 and pACYC-MA1, transcription of the *caf* operon, MA1 is controlled by binding of Caf1R to repeat 4 (R4'), upstream of *caf1M* (Chapter-5). Thus, in both cases *cafMA1* genes are only expressed and F1 only accumulates when Caf1R regulator is expressed. Hence, expression of Caf1R could be monitored by production/loss of F1 assembly.

F1 expression was monitored in whole cells of *E. coli* Top10, carrying either pACYCF1+pBADHisA or pACYC-MA1+pBADhCaf1R at 26, 28 and 30°C. On analysis of whole cells carrying pACYCF1, there was little F1 after 9 h at 26°C and F1 production was induced following temperature upshift, to 28 and 30°C (**Fig. 6.2a**). This was consistent with results from others in the laboratory where only low levels (blottable) of F1 were detected at 26°C (data not shown). In contrast, in cells carrying pACYC-MA1+pBADhCaf1R, F1 was expressed at roughly the same level at all 3 temperatures tested, following induction with L-arabinose (**Fig. 6.2b**). Immunoblotting of whole cells carrying pACYC-MA1+pBADhCaf1R confirmed that hCaf1R was expressed at all 3 temperatures from pBADhCaf1R (**Fig. 6.2c**). As pACYC-MA1 contains the complete 5' UTR for *caf1M*, absence of thermo-control when pACYC-MA1 is complemented with pBADhCaf1R is consistent with thermo-control *via* the 5' UTR of *caf1R*.

An effect of growth temperature on the β-galactosidase activity of promoter-*lacZ* fusions, described in Chapter-5, had also been noted. Therefore the results from Chapter-5 were analysed in more detail to compare activity of 24 h samples grown at 26°C with those grown at 37°C (**Fig 6.3**). Each construct contains the *caf1R* promoters, start codon and 192 bp of *caf1R* (relative to ATG start codon) fused upstream of the promoterless *lacZ* gene (with its own RBS and start codon). Activity of the *caf1R* promoter fusion constructs, pRScaf1R'.₇₀₊₁₉₂-*lacZ*

(contains P_{R^2} promoter) and pRScaf1R'-114+192-lacZ (contains P_{R^K} promoter) was increased 2.5-fold and 3.3-fold following the temperature shift from 24°C to 37°C. This was in contrast to a fusion construct containing the *caf1A-caf1* intergenic region where there was no apparent effect of temperature. Thus the increase in β -galactosidase activity following growth at 37°C is consistent with temperature induced transcription of *caf1R*. Transcription of *caf1R* is autoregulated, therefore the impact of temperature in the presence of complementing pACYC-R encoded Caf1R was also compared. The effect was amplified substantially due to expression of pACYC-R encoded Caf1R in a temperature dependent fashion. In the presence of Caf1R the constructs pRScaf1R'-70+192-lacZ and pRScaf1R'-114+192-lacZ showed a 7.5 and 12.7-fold higher activity, respectively, following growth temperature shift from 26 to 37°C (**Fig. 6.3**). This might be explained by additional post-transcriptional control of expression from the *caf1R* transcript encoded on pACYC-R.

The P_M promoter was also activated by Caf1R, which bound at R4' repeat, immediately upstream of P_M (Chapter-5; **Fig. 5.11**). Therefore, β -galactosidase activity of the *caf1M* promoter fusion construct, pRScaf1M'-184+158-lacZ (contains P_M promoter and complete R4' repeat) was also reanalysed in the context of temperature shift. Transformants carrying pACYC-R together with pRScaf1M'-184+158-lacZ already showed an increase in activity even at the lower temperature of 26°C. This may reflect sensitivity of P_M to very low levels of Caf1R. Following growth temperature shift from 26 to 37°C there was a 4.0-fold induction of activity (presence of Caf1R) whereas in the absence of Caf1R, the temperature-mediated increase in activity was 3.6-fold. These results would be consistent with expression of *caf1M* being strictly dependent on activation by Caf1R and the temperature requirement being for expression of *caf1R* (from pACYC-R; controlled by native 5' UTR). Hence interpretation of this data reinforces the 5' UTR of *caf1R* appears to be a site of thermoinduction within the *caf* locus. Expression of Caf1R provides an indication of thermoregulation of basal levels of transcription of both *caf1R* and *caf1M* in the absence of Caf1R. The large impact on the presence of *caf1R* might readily be explained by an RNAT within the 5' UTR of *caf1R* affecting translation from pACYC-R.

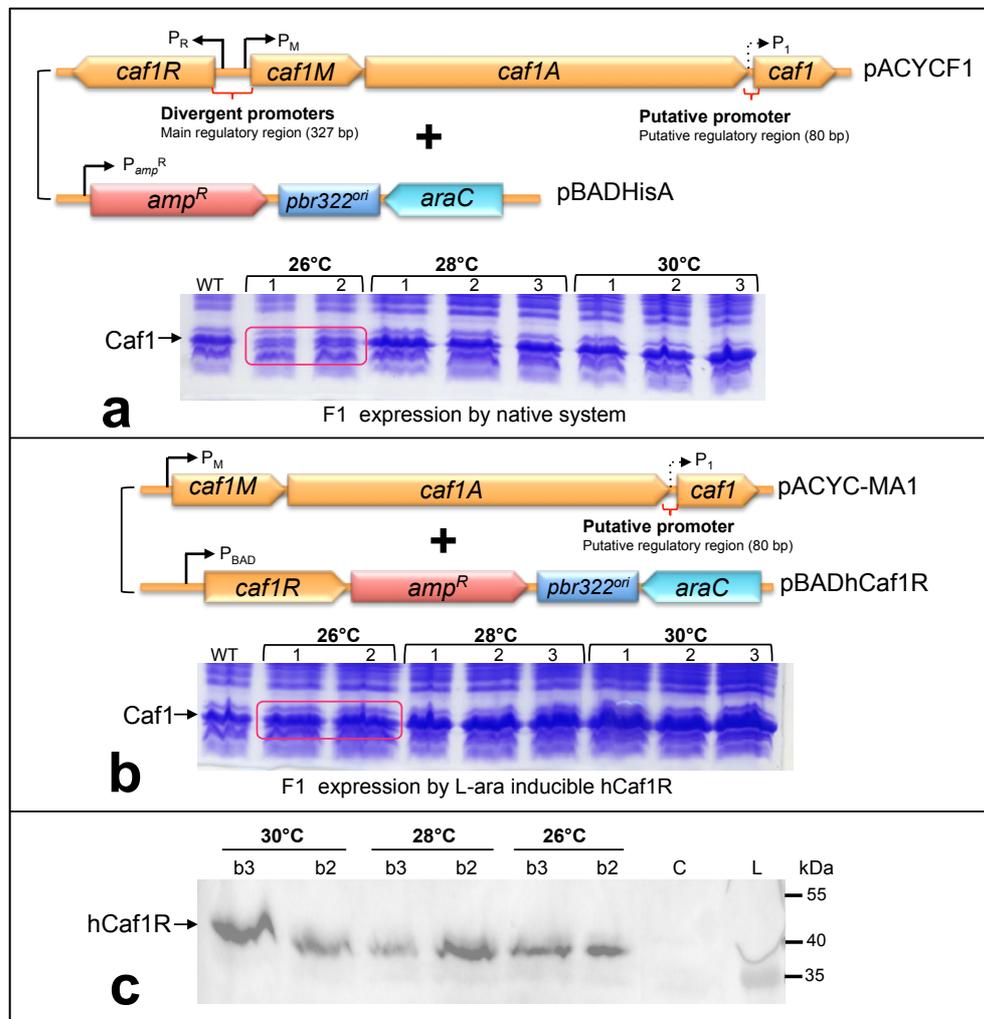


Figure 6.2| **Dependency of thermoregulation of F1 on *caf1R* 5' UTR.**

a) *caf1R* expressed from native 5' UTR in pACYCF1 b) *caf1R* expression controlled by pBADhCaf1R UTR. Schematic presentations show the plasmid combinations used in both tests to monitor the influence of *caf1R* 5' UTR on thermoinduction of F1 expression. *E. coli* Top10, two or three different transformants co-transformed with both plasmids were grown (with 0.02% L-ara for (b)) at 26, 28 and 30°C for 9 h. F1 production was monitored by analysing whole cells (0.07 OD units) by SDS PAGE (16% acrylamide). WT, positive control F1, from whole cell (Top10/pACYCF1), recovered after 37°C/4 h. c) Immunoblot of samples shown in (b) with anti-His mAb (1:10,000). L, pre-stained protein ladder and C, negative control of hCaf1R (Top10/pBADHisA).

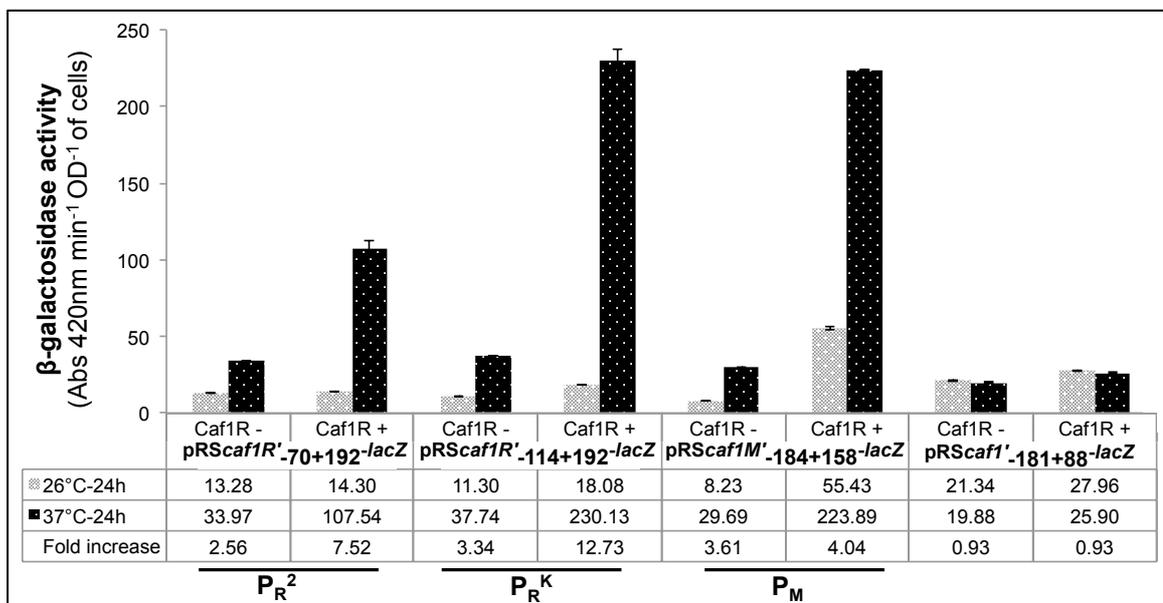


Figure 6.3| **Caf1R-mediated thermoinduction at P_{R^2} , P_{R^K} and P_M promoters.**

β -galactosidase activity per OD following growth at 26 and 37°C for 24 h, from promoter-*lacZ* fusion constructs, pRScaf1R'.70+192-lacZ, pRScaf1R'.114+192-lacZ and pRScaf1M'.184+158-lacZ, which contain promoters P_{R^2} , P_{R^K} and P_M , as indicated. Caf1R+ and Caf1R-, complemented with pACYC-R or pACYC vector. pRScaf1'.181+88-lacZ contains complete *caf1A-caf1* intergenic region (80 bp), 101 bp from 3' and 88 bp of *caf1* and is included for comparison. Standard error on the activity bars is from \pm SEM of the activity from three individual co-transformants of each. Note: Data presented are re-analysed data from Chapter-5; Figs. 5.10, 5.11 and 5.13.

6.3 Prediction of RNATs within the 5' UTR of *caf1R* and *caf1M*

In order to predict RNATs at the 5' UTR of *caf1R* and *caf1M*, initially the corresponding mRNA was assigned by predicting the transcriptional start site (TSS, +1 site) for the *caf1R* and *caf1M* or *caf1MA1* transcript. This was defined as 10 nt downstream from the first residue of the -10 element (Ruff et al., 2015). Promoter fusion studies had identified P_{R^2} and possibly also P_{R^K} as the promoter for *caf1R* and P_M as the promoter for *caf1M* and possibly a single *caf1MA1* transcript (Fig. 6.4). The sequences encompassing a few residues upstream from +1 to a tail of the first 1-10 coding amino acids of *caf1R* and *caf1M* (Fig. 6.4) were analysed at 26, 28, 30 and 37°C by the RNA folding program, Mfold version 2.3. Mfold is based on the algorithm of minimum free energy (ΔG) of decomposition or destabilization (Zuker, 2003). The selection criteria for inclusion of 1-10 codons of the translated mRNA was based on studies from previously identified bacterial RNATs (Bohme et al., 2012; Kouse et al., 2013; Loh et al., 2013; Matsunaga et al., 2013; Righetti and Narberhaus, 2014; Weber et al., 2014). In each case the most stable RNA secondary structure with lowest free energy (most negative ΔG value), was selected and then analysed for the extent of nucleotide pairing at the predicted SD sequences (RBSs) and translation initiation start codons.

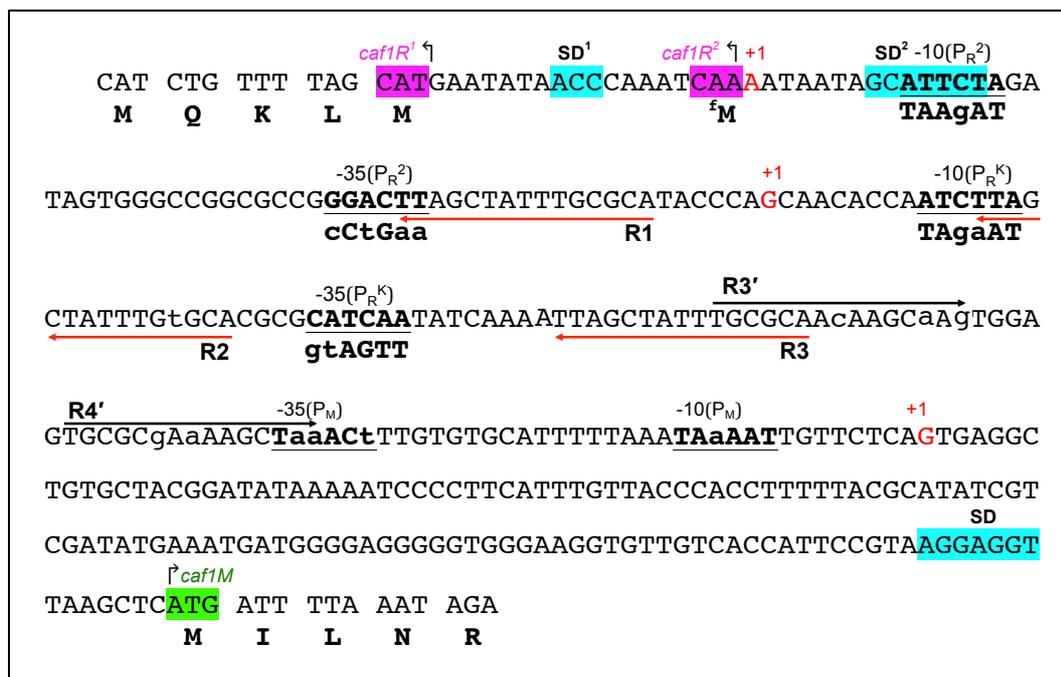


Figure 6.4| DNA fragment used to predict RNATs at 5' UTR of *caf1R* and *caf1M*.

Predicted +1 transcriptional start sites, corresponding to -10 elements of P_{R^2} , P_{R^K} and P_M promoters are indicated. Only the first 5 amino acids of both Caf1M and Caf1R (with ATG start codon) are shown. fM indicates the alternate predicted start site for *caf1R*. Identification of promoters and other key features are described in Chapter-5.

6.3.1 Predicted RNA secondary structure within 5' UTR of *caf1R*

As the 5' UTR from P_{R^2} promoter is only 19 nt, RNATs were predicted by inclusion of ten, five and one codon of *caf1R*. For P_{R^K} which has a longer 5' UTR (78 nt), only the first five codons from the translated sequence were included. For the mRNA corresponding to P_{R^2} , the predicted structures all showed an identical hairpin stem-loop structure at each temperature tested, with a less stable structure (less negative ΔG values), upon increasing temperature (**Fig. 6.5a (i-iii)**). Upon inclusion of only a single *caf1R* codon (AUG), an identical short weak stem-loop (from U11-A23 nt) was predicted that traps only the putative RBS¹ (**Fig. 6.5a (i)**). ΔG for these structures varied from -2.11 at 26°C to -1.80 at 37°C. Upon inclusion of the first five and ten codons of *caf1R*, the resulting stem-loops became gradually longer and more stable with increasingly negative ΔG values (**Fig. 6.5a (i-ii)**). Structures predicted on inclusion of the first 5 codons of *caf1R* trapped the RBS¹ and START¹ within the loop region (**Fig. 6.5a (ii)**). On inclusion of the first 10 codons, RBS¹ and START¹ were trapped in an unstable internal loop and a bulge, respectively. Hence, all predicted structures from a P_{R^2} transcript are in a CLOSED conformation (trapping of both RBS and AUG) at each temperature tested.

If P_{R^K} is used for the transcription of *caf1R*, the transcript includes both potential start codons (START² (UUG) and START¹ (AUG)) and each respective RBS. Structures of the 5' UTR of this transcript (59 and 77 nt with respect to START² and START¹) were more complex (**Fig. 6.5a (i-iv)**). Three stem-loops encompassing nt G3-C33, G37-U74 and nt C80-G95 were predicted at 26, 28 and 30°C. At 37°C the third stem-loop encompassing nt C80-G95 including START¹ was completely melted, forming a linear single stranded OPEN conformation. Notably, this third stem-loop was dependent on the presence of the first 5 codons of *caf1R*. If this stem-loop exists

it would suggest thermosensing within the *caf1R* coding sequence. More so, within the second stem-loop, at each temperature the predicted RBS¹, GGU was paired with CCG and located in a stable stem (G-C paired environment). However, RBS² (AGAAUGC) and START² (UUG) spanned the highly unstable external loop and internal bulge, respectively. First four nt of RBS² paired with 4U, as in a classic 4U RNATs (Grosso-Becera et al., 2015). Interestingly, in this transcript from P_{R^K}, RBS¹ is preceded by a potential RNase E endoribonuclease cleavage site (GAUUU) consistent with the *E. coli* consensus G/AAUUA/U (Ehretsmann et al., 1992). This may indicate particular sensitivity of this region to degradation driving the kinetics to an unstructured 5' UTR.

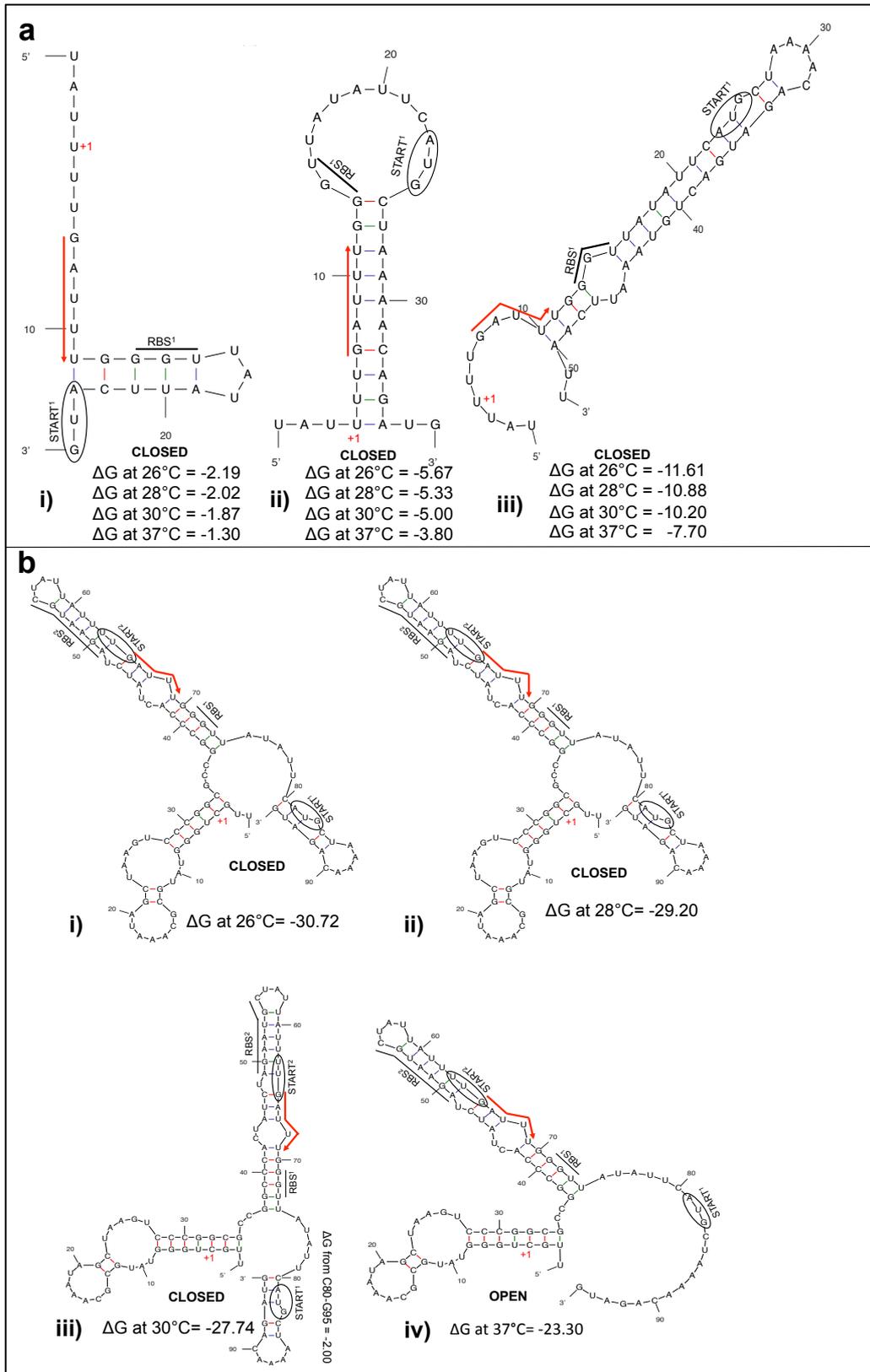


Figure 6.5| Predicted RNATs at 5' UTR of *cafIR*.

The Mfold program (Zuker, 2003) was used for the prediction of temperature modulated RNA secondary structures (RNATs) within the 5' UTR of *cafIR* transcript from P_R^2 promoter (a) and P_R^K promoter (b). The structure with the lowest predicted free energy (ΔG ; kcal/mole) (indicated) at each temperature is shown. +1, predicted transcription start site (TSS) from each promoter. The two different predicted start codons for *cafIR* (AUG (START¹) and UUG (START²)) are encircled and corresponding RBSs (RBS¹ and RBS²) are underlined. Red arrows indicate a potential RNase E cleavage site. Prediction of P_R^2 structure (a) was with inclusion of one (i), five (ii) and ten (iii) codons of *cafIR* (relative to AUG). RNATs corresponding to message from P_R^K were predicted only by inclusion of first five codons of *cafIR*. OPEN and CLOSED indicate accessibility of the START and RBS site. ΔG values for all segments of each structure are tabulated in Appendix 5 and 6.

6.3.2 Predicted RNA secondary structure within 5' UTR of *caf1M*

The 5' UTR of the transcript corresponding to P_M promoter is quite long (128 nt). Potential RNA secondary structures within this sequence plus the first 5 codons of *caf1M* showed four small stem-loops encompassed nt G6-U14, G17-C30, G96-C107 and U109-G120, connected by an unstable internal stem (**Fig. 6.6**). A fifth much longer structure (C44-G93) contained one particularly stable stack of 11 bp separated by a small internal loop from a second stack of 9 bp with a pentagonal bulge, one non-paired pair and a classic 4 nt hairpin. This structure does not encompass either the RBS or AUG and hence is likely to be more important in mRNA stability than in thermoregulation. Structures predicted at 26, 28, 30°C (**Fig. 6.6a-c**) were identical while a dramatic change was observed between 30 and 37°C. The internal stem, trapping the START codon, melted to an OPEN conformation (**Fig. 6.6d**). However, at all temperatures the putative RBS (AGGAGGU) was trapped in a stem-loop. This long *caf1M* transcript (128 nt) might contribute to the stability of the mRNA, although the single RNase E site (GAUUU) following the AUG start codon might indicate instability.

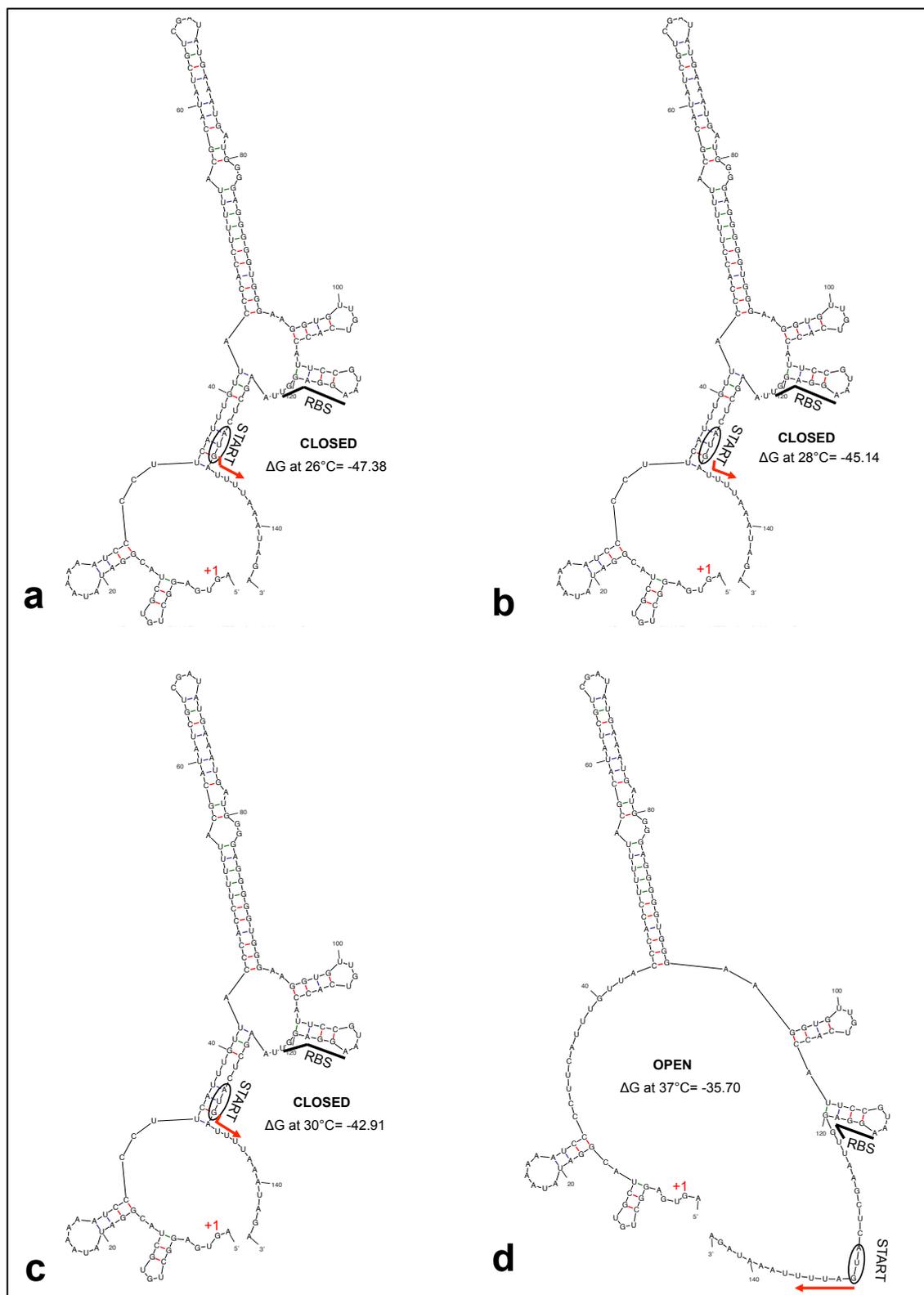


Figure 6.6| **Predicted RNATs at 5' UTR of *cafIM*.**

The Mfold program (Zuker, 2003) was used for the prediction of the RNA secondary structures (RNATs) at 26 (a), 28 (b), 30 (c) and 37°C (d). The mRNA fragment (message) was selected from the predicted +1 transcription start site (indicated in red), corresponding predicted P_M promoter of *cafIM* or *cafIMAI*. The putative RBS and translation initiation codon (AUG) for *cafIM* are underlined and encircled, respectively. Red arrows indicate RNase E cleavage site. RNATs at each selected temperature were predicted by inclusion of the first five codons of *cafIM*. Conformation (OPEN or CLOSED) relative to translation initiation codon (AUG; START) and RBS site is indicated. The most probable prediction with lowest free energy (ΔG ; kcal/mole) at each temperature is shown. ΔG for all segments of each structure is tabulated in Appendix 7.

6.4 Conclusion

The results presented in this chapter support the hypothesis that thermosensing within the *caf* locus is linked to the expression of *caf1R*. This was initially shown by complementation of *caf1MA1* operon with P_{BAD} regulated hCaf1R. Thermosensing was lost when Caf1R was expressed from pBADhCaf1R, F1 was produced even at 26°C. The fact that F1 was produced means that the P_M promoter must be functional at 26°C as the Caf1M chaperone and the Caf1A usher are required for F1 assembly. This is in contrast to loss of F1 production at 26°C when expressed from the native *caf1R* promoter (in pACYCF1). Some evidence of thermoregulation of transcription of *caf1R* and *caf1M* was indicated by reanalysis of the data for β -galactosidase activity from the promoter-*lacZ* fusion study, comparing the 26°C/24 h inoculum with 37°C/24 h samples where a 2-3 fold increase in activity with the 37°C samples in the absence of Caf1R was noted. This was in contrast to fusion of the *caf1A-caf1* intergenic region which showed no enhanced activity with the 37°C samples. In the presence of pACYC-R encoded Caf1R, a 7.5-fold and 12.7-fold thermoinduction of β -galactosidase was obtained with P_{R^2} , P_{R^K} promoters, respectively. This marked increase could be explained by an additional temperature dependent impact on translation of *caf1R* from pACYC-R. While *caf1M* transcription was also dependent on pACYC-R encoded Caf1R, moderate levels of transcription from pRScaf1M'-184+158-*lacZ* (P_M) +pACYC-R at 26°C resulted in only a 4.0-fold increase at 37°C. This presumably reflects high sensitivity of P_M to Caf1R. Together these results provide strong evidence for thermoregulation of the *caf1MA1* operon *via* expression of Caf1R, with possibilities for contributions of both transcriptional and translational control.

One common mechanism of thermoregulation of virulence factors from bacteria that live in different environments is translational control *via* thermosensitive structures within the 5' UTR of the transcript (Grosso-Becera et al., 2015; Hoe and Goguen, 1993; Kouse et al., 2013; Matsunaga et al., 2013). Of the 2 potential promoters upstream of *caf1R*, P_{R^2} produces the shortest 5' UTR (19 nt) with a small stem-loop structure encompassing the RBS¹ at both 26 and 37°C and weak stability (ΔG value of -2.19 and -1.30 kcal/mole, respectively). More stable structures were only generated with inclusion of the coding sequence. The RBS in one structure (with 10 codons) was associated with a bulge within the stem structure similar to the RNAT of the PfrA transcriptional activator of *L. monocytogenes* (Loh et al., 2009). Location of the RBS in the external loop, as in the other structure (with 5 codons) would be unusual (Grosso-Becera et al., 2015). Hence, the most likely structure within the P_{R^2} transcript 5' UTR would be of only marginally stable. In contrast to transcript from P_{R^2} , the 5' UTR of the transcript from P_{R^K} (78 nt) includes both predicted start codons and the corresponding RBSs. Analysis of the 5' UTR produced the same basic structure with respect to RBS¹, RBS² and UUG start, with and without additional coding sequence. Only the AUG start codon paired with additional 15 nt of the coding sequence forming an unstable structure that was lost between 30 and 37°C, to form a linear single stranded OPEN conformation with respect to this start codon. RBS1 formed a stable G-C rich base-pairing stem (ΔG of -10.10 kcal/mole). In contrast, the RBS² was located in a less stable stem (ΔG of -4.20 kcal/mole) including four consecutive A/G-U pairs, typical of the 4U family of RNAT as seen in other AraC/XylS family regulators such as LcrF regulator of *Y. pestis* (Bohme et al., 2012; Hoe and Goguen, 1993) and ToxT of *V. cholera* (Weber et al., 2014). Although, the 4U

paired RBS was still at the same position at 37°C it spanned a highly unstable internal bulge and an external loop. This suggests that this region readily melts *in vivo* at 37°C with associated destabilisation of the RBS1 pairing sequence and melting of the entire structure to an OPEN conformation accessible for translation.

RNA secondary structure of the *caf1M* 5' UTR (128 nt) was longer and more complex than that of *caf1R*. It included a relatively unstable internal stem structure connecting two large internal loops. This stem, which encompassed the AUG start codon, melted between 30 and 37°C. The RBS was located within a relatively stable stem-loop structure (ΔG of -4.9 kcal/mole) with a 4 nt paired stem and a 4 nt loop. This may also indicate some thermosensing within the *caf1M* transcript, although this was not evident from the complementation study using P_{BAD} regulated *caf1R*. A long stable G-C rich stem structure was predicted within the *caf1M* 5' UTR that might relate to mRNA stability.

In conclusion, the evidence so far is consistent with expression of *caf1R* being the primary site for thermoregulation of F1 production. This study has not directly addressed factors that might influence thermoregulation of transcription such as stability of additional DNA binding proteins as seen with *Yersinia* YmoA (Bohme et al., 2012) and its *E. coli* homolog Hha (Madrid et al., 2002). The predicted RNA secondary structures presented provide a framework for confirmation of the role of a *caf1R* 5' RNAT in controlling F1 expression. This would be addressed by site-specific mutagenesis targeting non-canonical or non-base pairing residues to increase stability and canonical base-pairing nt to decrease stability of predicted structures. Primary targets would be nt pairing with RBS¹ in the P_{R^2} transcript and both RBS¹ and RBS² in the P_{R^K} transcript.

Chapter 7

General discussion and future directions

7.1 Introduction

This study elucidated the primary regulatory mechanisms controlling the expression of key surface antigen F1 of *Yersinia pestis*. Despite a wealth of knowledge on other aspects of F1, such as its assembly by the CU pathway, and its use in plague diagnostics and vaccine design, understanding the regulation of this key antigen was untouched prior to this study. The *caf* locus is controlled by a divergently linked AraC/XylS family regulator, Caf1R (Karlyshev et al., 1992). CU assembled pili/fimbriae have acquired a diverse range of regulatory mechanisms to control their expression (Clegg et al., 2011). An AraC/XylS family regulator is one such mechanism, but this has received little attention thus far. There are a few examples of AraC/XylS type regulators which ‘globally’ regulate several unlinked CU system such as Rns-mediated regulation of CS1-CS4, CS14, CS17 and CS19 fimbriae of enterotoxigenic *E. coli* (Basturea et al., 2008; Caron et al., 1989; Froehlich et al., 1994; Mahon et al., 2010; Munson et al., 2002; Munson and Scott, 1999, 2000); AggR-mediated expression of Agg3, Agg and AAF fimbriae of enteroaggregative *E. coli* (Dudley et al., 2006; Morin et al., 2013; Nataro et al., 1994) but there has been no in-depth study of a dedicated AraC/XylS family regulator controlling a linked CU system. This study has highlighted key residues involved in DNA-Caf1R regulator interaction, including a bridge interaction likely to be involved in stabilising the correct orientation of BS2 residues. Promoters have been defined and an interesting model for thermo-control *via* expression of *caf1R* is presented. A database survey identified many closely related regulators controlling CU systems within the *Enterobacteriaceae*. From consideration of conserved residues, many features appear to be conserved among this CU family of regulators. This characterisation of Caf1R-mediated regulation of F1 provides the first fundamental model on regulation of CU pili/fimbriae, controlled by a directly linked AraC/XylS type regulator. The most significant and novel aspects of this study are discussed below.

7.2 Caf1R-dependent and independent promoters of the *caf* locus

Five potential Caf1R binding repeat motifs, R1, R2, R3, R3' and R4' (15 nt long) were identified within the *caf1R-caf1M* intergenic region. R1-R3 are equally spaced (18 nt) and located upstream of *caf1R*, while repeats R3' and R4' (separated by 5 nt) run in the opposite orientation upstream of *caf1M*. Transcriptional fusions, designed around these repeat motifs, identified the Caf1R dependent promoters, P_M (for *caf1M*) and P_R² (for *caf1R*). In contrast to earlier speculation (Karlyshev et al., 1992; MacIntyre, 2004), no Caf1R binding site or indication of promoter activity was detected within 246 bp upstream of *caf1* SD sequence. This would suggest that either there is another promoter further upstream leading to an exceptionally long mRNA or that a single polycistronic mRNA, encoding Caf1M, Caf1A and Caf1, is transcribed from the P_M promoter. Transcriptomic studies have reported exceptionally high levels of *caf1* mRNA compared to *caf1M* and *caf1A* mRNA, following several hours at 37°C (Motin et al., 2004). A feasible explanation for this would be differential susceptibility of *caf1M*, *caf1A* and *caf1* mRNA to degradation, leaving highly stable mRNA of *caf1*. This would then contribute to the high level of F1 that accumulates on the cell surface of *Y. pestis* (Amies, 1951; Simpson et al., 1990).

For transcription of *caf1M* or *caf1MA1* operon, P_M showed a strong dependency on Caf1R. An *in vitro* protein-DNA binding study confirmed binding of Caf1R at R4'. The R4' sequence overlaps this promoter -35 element by 3 nt, and has a 19 nt spacer between the -10 element and predicted R4' BS2, thus classifying P_M as a class II type promoter (Browning and Busby, 2004). This is a common mode of promoter activation by AraC/XylS family regulators (Martin et al., 2002; Martin et al., 1999; Martin and Rosner, 2002). With class II promoters, the regulator initially binds to region 4 of the $\sigma 70$ subunit and the α -CTD of the RNAP and then scans the DNA for subsequent transcription initiation complex formation (Jair et al., 1996a; Jair et al., 1996b; Martin et al., 1999; Shah and Wolf, 2004; Zafar et al., 2011; Zafar et al., 2010). The R3' sequence is located only 5 nt upstream of R4'. A similar short space between 2 binding sites was found to be critical for DNA-bending by 98° and optimal activation by a dimer of XylS (Dominguez-Cuevas et al., 2010). To ensure binding of Caf1R HTH1 and HTH 2 at R4', *caf* DNA must almost certainly bend, as seen in the cocrystal structure of *mar*-MarA (Rhee et al., 1998) and in the *mar* DNA-Caf1R model (Fig. 3.18c). The promoter fusion studies indicated a slight enhancement of P_M activation on inclusion of R3' as well as R4' with P_M promoter. It is possible that *caf* DNA is additionally bent between R3' and R4' as Caf1R binds as a dimer to enhance activation. However, as yet there is no strong evidence of dimerisation of Caf1R and R3' has diverged more than the other repeats. Any involvement of this sequence remains to be elucidated. A model depicting current understanding of Caf1R-mediated activation at P_M promoter is shown in Fig. 7.1.

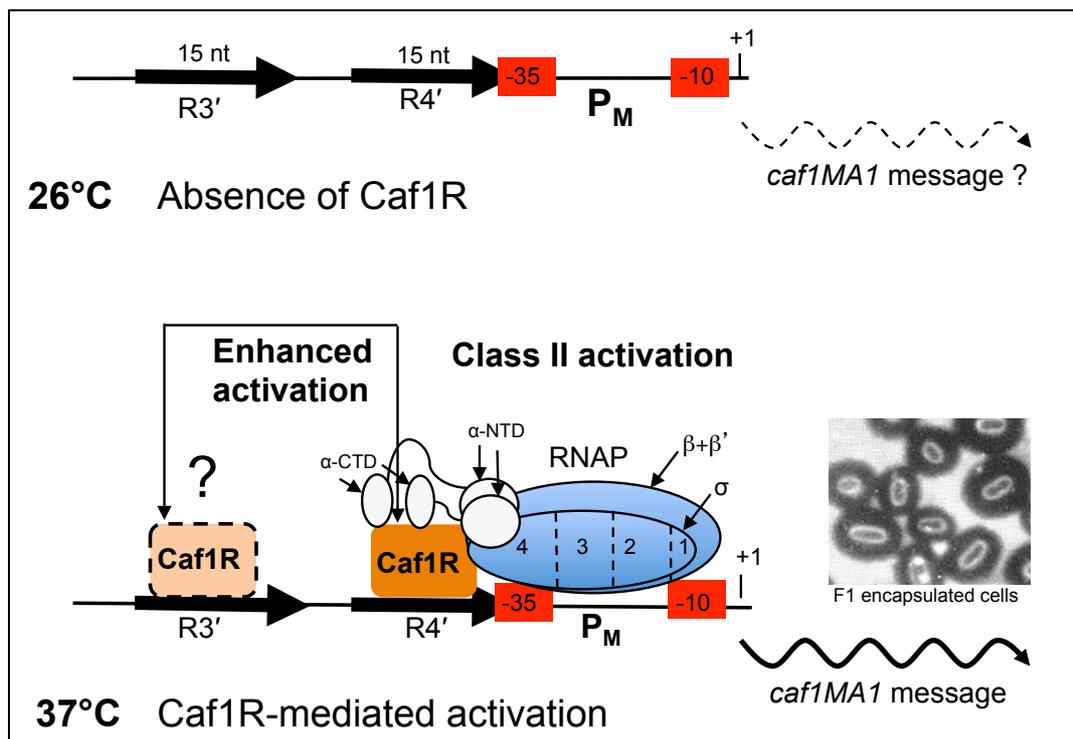


Figure 7.1| **Caf1R-mediated transcription activation at the class II P_M promoter.**

At 26°C, there is little *caf1MA1* message (due to absence of thermo-inducible expression of *caf1R*). Following temperature transition to 37°C, mimicking *Y. pestis* transmission from flea vector to mammalian host, Caf1R regulator is synthesised (see Fig 7.2 for detail). Based on published evidence of RNAP-regulator interactions at class II promoters, it is proposed that Caf1R then interacts with region 4 of the $\sigma 70$ subunit and the α -CTD of RNAP. RNAP-Caf1R complex would then scan the DNA, bind at R4' and activate transcription of *caf1MA1*. This leads to high levels of surface F1, giving the appearance of encapsulated cells. Involvement of R3' in enhancing Caf1R mediated activation is currently speculative. Depiction of class II activation adapted from (Browning and Busby, 2004) for the *caf* locus.

Regulation of *caf1R* expression appears more complicated with 3 repeat sequences, R1-3, each separated by 18 nt. P_{R^2} can also be defined as a class II activated promoter, with the Caf1R binding site overlapping the -35 consensus by 2 nt. P_{R^2} clearly mediates autoactivation of *caf1R*. When Caf1R binds at R1 there is a 3.8-7.8-fold increase in transcription, leading to enhanced production of Caf1R (**Fig 7.2**). Autoregulation is common among these regulators. Examples of autoactivation include Rns (Munson and Scott, 2000), PerA (Porter et al., 2004), MarA (Martin et al., 1996) and AggR (Morin et al., 2010). Unlike *caf1R* these regulators are primarily unlinked to the target operons. Further constructs are required to assess the impact of inclusion of R2 on Caf1R mediated activation at P_{R^2} . However, results from the promoter fusion constructs suggested that a second predicted promoter, P_{R^K} may work at a slightly higher basal level in the absence of Caf1R. The location of the R2 repeat between the P_{R^K} -10 and -35 elements then resembles a classic mode of transcriptional repression (Browning and Busby, 2004; Sanchez et al., 2011). This together with the lower fold activation observed with *lacZ*-fusions containing P_{R^K} , R2 and R3 may reflect Caf1R mediated repression at this P_{R^K} promoter. This may or may not involve a Caf1R dimer. Thus far, from EMSA with cell lysates containing either MBPCaf1R fusion or hCaf1R, effective binding was obtained with a single Caf1R binding repeat (R4').

F1 expression is known to be temperature controlled with expression OFF in the flea and ON in mammalian hosts (Chauvaux et al., 2007; Hinnebusch, 2005; Straley and Perry, 1995; Zhou et al., 2006). Thermo-control of F1 expression is known to be maintained in recombinant plasmids both in *E. coli* and *Salmonella* (Cao et al., 2012; Simpson et al., 1990; Titball et al., 1997). This study has provided the first evidence of how changes in temperature control expression from the *caf* locus. Thermosensing was localised to expression of *caf1R* and a thermosensitive RNA secondary structure (RNAT) was predicted within the transcript from P_{R^K} . This predicted secondary RNA structure trapped both the ribosomal binding site (Quade et al.) and the AUG start codon of Caf1R in two stem-loop structures at lower temperatures of 26°C. Temperature mediated modifications were observed in this predicted structure at 37°C, with melting of the stem-loop encompassing the AUG start codon and destabilisation of the structure encompassing the RBS. Interestingly, an RNase E site was located directly preceding the RBS within a bulge in the stem structure. Cleavage at this site, in a partially unfolded structure, would drive the reaction towards the OPEN structure permitting access of ribosomes for subsequent translation of Caf1R. This model of thermoregulation of *caf1R* is depicted in **Fig 7.2**. In the model, an RNAT structure within the leader of *caf1R* transcript synthesised at 26°C remains untranslated until the temperature increases. At 37°C, on infection of the mammalian host, Caf1R regulator is produced and binds at R4' to activate transcription of *caf1MA1* from the P_M promoter. Caf1R would also bind at R1, autoactivating transcription of *caf1R*. It is hypothesized that Caf1R would also bind at R2 blocking further expression from P_{R^K} . The *caf1R* gene would then only be transcribed from P_{R^2} . These features could be addressed by further *lacZ*-fusions, directed mutagenesis and using primer extension, or RT-PCR to identify temperature dependent leader sequence.

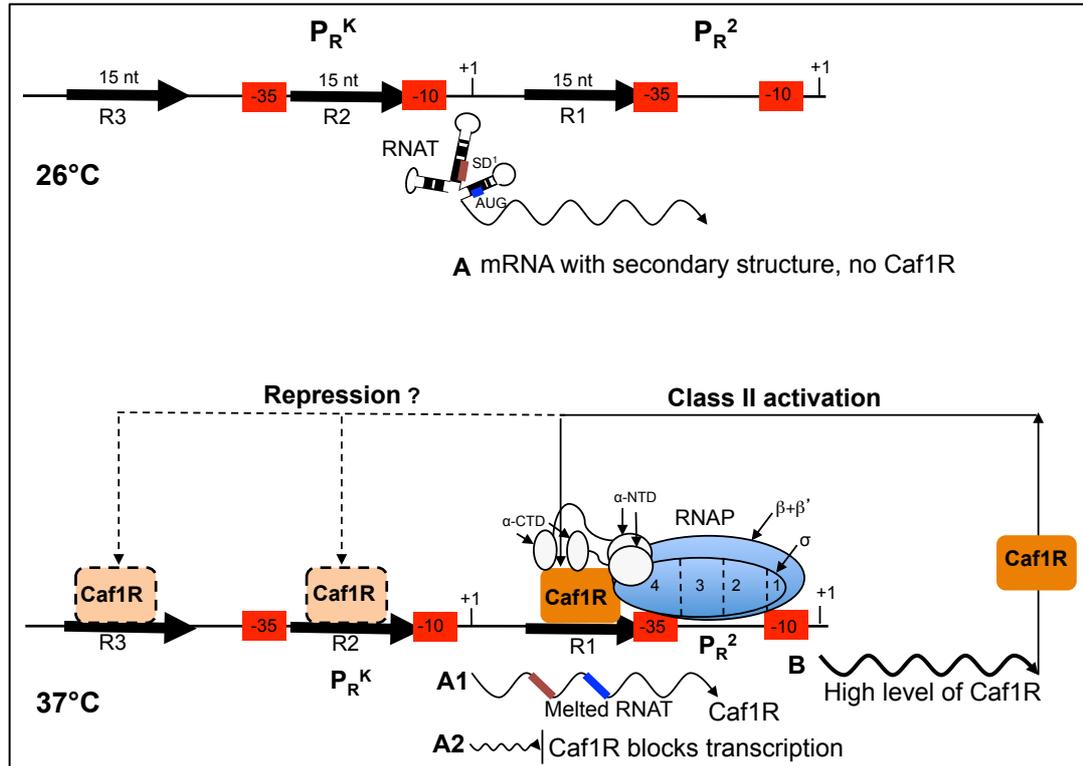


Figure 7.2| **Working model of Caf1R- independent and dependent transcription of *caf1R*.**

At 26°C, *caf1R* mRNA is transcribed and mRNA leader sequence folds to form an RNAT (A), translation of Caf1R is blocked. At 37°C, the RNAT melts (A1) and Caf1R is translated. Low-level transcription from P_{R^K} may continue. Caf1R binds R1, activating transcription from P_{R^2} promoter, in a class II type manner involving binding and optimum promoter RNAP complex formation as described in Fig. 7.1. This leads to a high level of Caf1R (B). A high level of Caf1R may also result in binding to R2 (possibly also involving R3 binding) resulting in repression of transcription (A2). The model of class II activation adapted from (Browning and Busby, 2004) for the *caf* locus.

Activation by AraC/ XylS family regulators, as seen with P_M , is common. In many well-studied cases, binding of a small metabolite is required for activation (Dominguez-Cuevas et al., 2008; Kolin et al., 2008; Lowden et al., 2010; Ni et al., 2013; Schleif, 2010). Caf1R also possesses a C-terminal 'sensing'-like domain but it remains to be seen if this binds anything or is involved in dimerisation. The proposed mode of regulation of *caf1R* is more unusual. Two promoters are proposed; one involved in autoactivation (P_{R^2}) and the other (P_{R^K}) in thermoregulation of basal levels of Caf1R expression at low temperatures and autorepression at 37°C. The best-studied example of transcription repression by AraC/XylS family regulators is that mediated by AraC itself. When an AraC dimer binds to operator site O1 it blocks self-transcription (Ogden et al., 1980). In addition, it represses transcription of *araBAD* via a DNA-looping mechanism (Schleif, 2000, 2010) involving binding of AraC dimer to two distant sites, O2 and I1, which are 171 nt apart. The working model presented for regulation by Caf1R is completely different. No additional distal Caf1R binding site is present, R1-R4' are 80 nt apart. It remains to be seen if other CU systems are controlled in a similar manner proposed here for Caf1R.

7.3 Caf1R residues involved in DNA binding

The Caf1R binding motif (R4') showed a high degree of similarity with the MarA binding motif (Rhee et al., 1998) (**Fig. 7.3**). In BS1 (binds helix-3, within HTH1), only 2 of 5 bp are identical between the 2 sequences, whereas in BS2 (binds helix-6, within HTH2), 6 of the 7 bp are identical. This is consistent with the fact that bases in BS1 contribute to regulator specificity (Egan, 2002; Gallegos et al., 1997). The high level of identity at BS2 reflects more conserved DNA-protein interactions within this subfamily of AraC/XylS regulators. This is also consistent with poor amino acid conservation within the regulator HTH1 motif in comparison to a much higher level of conservation within HTH2. DNA-Caf1R interactions had been identified using the *mar* DNA-Caf1R model. Having confirmed binding of Caf1R to R4' sequence, the *mar* DNA nt could be replaced for *caf* DNA as shown in **Fig. 7.3**. In comparison with *mar*-MarA interactions, the key helix-6 residues of Caf1R (Q93 and R97) and MarA (Q92 and R96) are involved in virtually identical BS2 base interactions as shown for Caf1R in **Fig. 7.3**. MarA-Q92/ Caf1R-Q93 interacts with T-7 and T-8 bases while MarA-R96/Caf1R-R97 showed extensive hydrogen bonds with A-9, G-10 and G-40 bases. The only exception was the Caf1R-R97 interaction with A-39, which corresponds to MarA-R96 interaction with T-39. In the modelled Caf1R-DNA structure, the BS1 interactions are very different from those in the *mar*-MarA cocrystal structure. Caf1R-R43 was the only residue with specific base interactions. It interacts with the highly conserved nt, G-31 and C-32 in BS1. Y44 interacts with G-29 at the border of the consensus sequence. In contrast, in MarA three residues, R46, W42 and Q45 all interact with nt of BS1. W42, corresponding to Caf1R-R43, interacts with T-18, C-31 and C-32 (van der Waals interactions), Q45 with G-17 and T-18 (van der Waals interactions) while R46 forms H-bonds with G-20, G-30 and C-31 (Rhee et al., 1998). These differences could be explained by the very different residues in helix-3 of both regulators, which in turn could contribute to differences in specificity.

Alanine scanning mutagenesis of all Caf1R residues involved in DNA binding was used to assess the importance of these residues to Caf1R function. Comparison with published data on mutations in MarA is summarised in Appendix 8. Substitution of Q93 and R97 in Caf1R indicated an equally critical role of both of these invariant residues in Caf1R function. The corresponding nt to Caf1R-R97 in MarA and SoxS was also found to be critical using Ala mutagenesis. MarA_{R96A} had a lower level of transcriptional activity at both class I and class II promoters (0-20% on *fpr*, *zwf* and *micF*; 21-40% on *mar*; 41-60% on *fumC*) (Gillette et al., 2000). SoxS_{R90A} had an even greater effect, particularly on *zwf* and *micF* promoters. It generated only 2% and 38% activity at the class I promoters, *zwf* and *fpr*, respectively, and 45% and 8% activity at the class II promoters, *fumC* and *micF* (Griffith and Wolf, 2002). DNA-protein mobility shift assays also highlighted the critical role of this BS2 Arg residue in both MarA and SoxS for stable DNA binding, consistent with the extensive nt interaction (Gillette et al., 2000; Griffith and Wolf, 2002; Rhee et al., 1998). The impact of loss of the conserved Gln residue was less pronounced. With MarA_{Q92A}, 61-80% transcriptional activation remained at both class I (*zwf*) and class II (*fumC*) promoters (Gillette et al., 2000). With SoxS_{Q86A} there was no impact at any of the promoters tested (Griffith and Wolf, 2002).

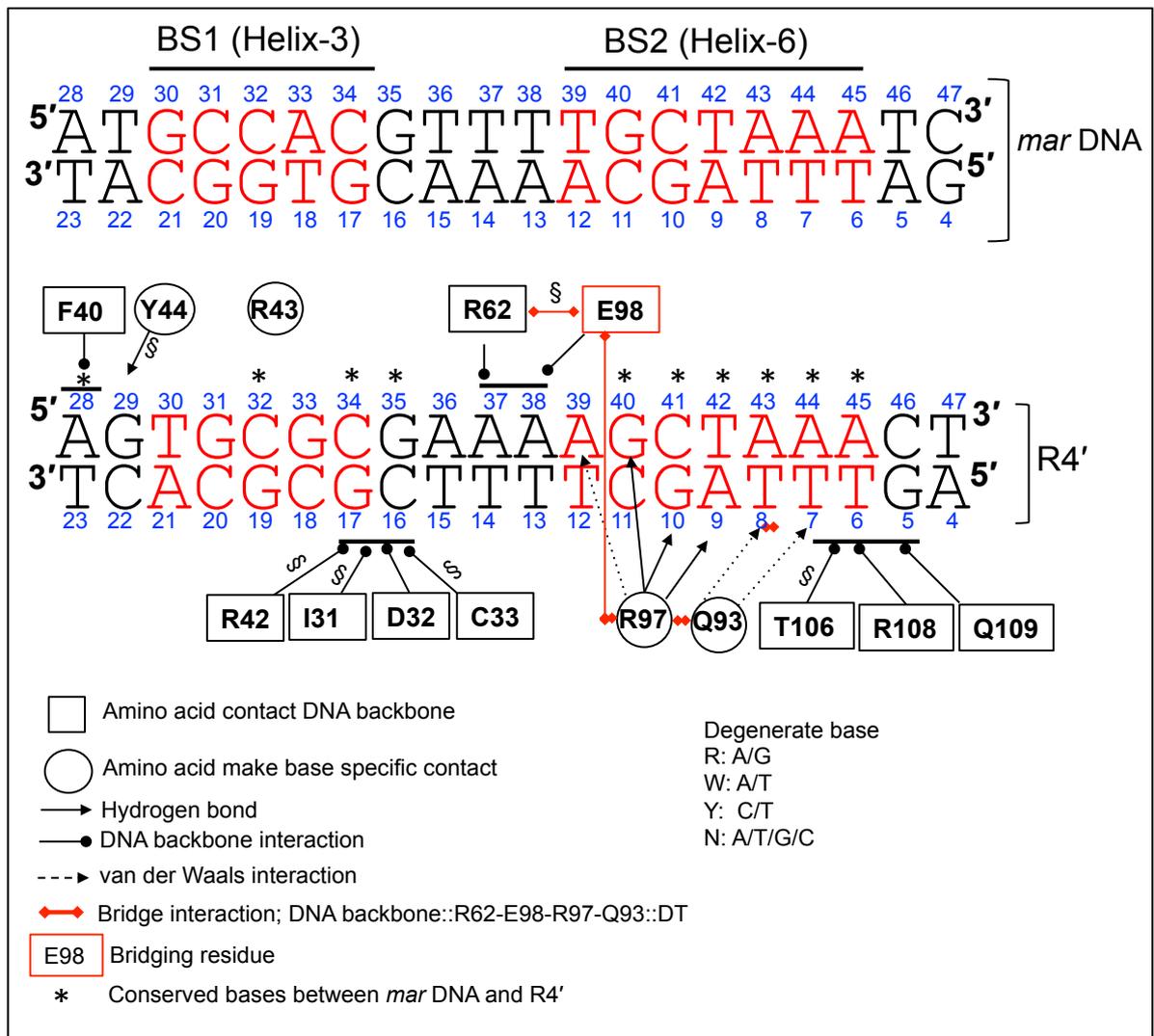


Figure 7.3| Modelled interactions of Caf1R residues at R4' repeat.

R4' nt corresponding to *mar* BS1 and BS2 are numbered according to the *mar* DNA in the *mar*-MarA crystal structure (Rhee et al., 1998). BS1 and BS2 represent DNA residues involved in binding helix-3 and 6 of MarA regulator. The corresponding residues in *caf* R4' sequence (red) are shown. Highlighted amino acids are Caf1R residues found to be involved in DNA binding in the *mar* DNA-Caf1R model (see Fig. 3.23). These amino acids are assumed to interact with the same corresponding position in the *caf* DNA. DNA-protein interactions identified in the *mar* DNA-Caf1R model that are different from those in the *mar*-MarA cocrystal structure are indicated by §.

Ala substitution of Caf1R- R43 and Y44 had much less of an impact. F1 production was reduced by only 25-35%. This suggests a lesser role of these residues in stabilizing the DNA-protein interactions. In contrast, BS1 interacting residues of MarA and SoxS were found to be more critical (Gillette et al., 2000; Griffith and Wolf, 2002). Alanine substitution of W42 in MarA (corresponding to R43) resulted in a drastic reduction in the transcription activity (0-20%) at class I promoters (*fpr* and *zwf*) and had a moderate effect on class II promoters (*fumC* and *micF*) (Gillette et al., 2000). Similarly, Ala substitution of MarA- H43 produced a severe effect (0-20% of WT) on both class I and class II promoters, except class I *mar* promoter, on which wild type phenotype was observed (Gillette et al., 2000).

The DNA backbone interacting residues also play an important role. Key residues are Caf1R-131, D32, C33, F40 and R108, which correspond to MarA- L30, E31, K32, Y39 and H107. The most pronounced effect was seen with F40A mutation in Caf1R, Y39A mutation in MarA (Gillette et al., 2000) and Y33A in SoxS (Griffith and Wolf, 2002). In all cases, regulator activity was virtually abolished. Griffith and Wolf suggested that the Y39 residue of MarA might interact with highly conserved (invariant) A base (at position 28 in **Fig. 7.3**) (Griffith and Wolf, 2002). A similar conclusion was made from an NMR study of *mar*-MarA-RNAP complex (PDB-1xs9), where *mar*-MarA contacts were shown from the N-terminal residues of MarA (Dangi et al., 2001) that are absent in the cocrystal structure of *mar*-MarA complex (Rhee et al., 1998). At the other end of the regulator, close to BS2, Caf1R-R108 was found to be critical as was SoxS-S101 (Griffith and Wolf, 2002), suggesting that that Caf1R-R108 together with F40 may be critical in anchoring both ends of the protein-DNA interaction. However, Ala substitution of the corresponding residue in MarA, H107, had no impact (Gillette et al., 2000).

A spontaneous mutation (E98G), within Caf1R-DBD virtually abolished F1 assembly and loss of Caf1R binding to *caf* DNA in an *in vitro* EMSA binding. Strikingly, the analogous position of Glu98 was identical in all of the analysed regulators linked to the CU system. But in non-CU regulators such as Rob, MarA and SoxS this position was highly variable, containing Thr, Ala and Val, respectively. DNA-Caf1R modelling provided evidence that E98 stabilizes the correct protein-DNA interaction through formation of a bridge between the BS2 recognition helix-6 and the helix-4 residue R62. R62 and helix-4 then binds to the DNA backbone and connects BS1 and BS2. R62 and E98 interact with an interatomic distance of 3.2Å. Upon E98G substitution, the proposed bridge interaction was completely lost (no interaction between R62 and G98). Moreover, although the modelled E98G mutant showed no disruption of the R97 and Q93/BS2 interactions, the small size and low helix propensity of G98 may disturb DNA interactions of helix-6. This could explain the severe defect with Caf1R_{E98G}. This bridge-type interaction was further validated by site-specific mutagenesis of R62 (R62A/S) and E98 (E98T/A/K) with which significant reduction in F1 assembly and loss of the modelled bridge was observed. R62 is an invariant residue, R61 in MarA. In MarA, R61 also interacts with the DNA backbone in the *mar*-MarA cocrystal structure (Rhee et al., 1998). However, Ala substitution of MarA-R61 had variable results. There was no defect in transcription from the class II promoters *fumC* and *micF*, but a marked reduction in transcriptional activity at the class I promoter *fpr* (21-40% of WT) (Gillette et al., 2000). Mutation of SoxS gave similar results to Caf1R. A SoxS with an R55A mutation resulted in 40-80% reduction in transcriptional activity with all promoters tested (*zwf*, *fpr*, *micF* and *fumC*) (Griffith and Wolf, 2002). There was less impact in other regulators carrying an Ala substitution for the variable residue corresponding to Caf1R-E98. In contrast to Caf1R_{E98A} (58% of WT F1), MarA T97A showed 81-100% activation with class I promoters and 61-81% activation with class II promoters and SoxS V91A had only a minor effect on *fumC* promoter (Gillette et al., 2000; Griffith and Wolf, 2002). No T97 mediated bridge interaction was observed in the *mar*-MarA cocrystal structure (Rhee et al., 1998), suggesting the modelled Caf1R-DNA bridge interaction may be a special feature of CU-associated regulators, which contain invariant R62 and E98 residues of this subfamily. Ultimately, high resolution structural studies would be required to valid these interactions.

7.4 Future directions

Initial key aspects requiring attention include confirmation of the model of thermoregulation at the *caf1R* promoter and optimisation of EMSA. This would permit a detailed analysis of the significance of each repeat motif and consideration of whether Caf1R functions as a monomer or dimer. This could include mapping of transcription start site (TSS) for *caf1R*, *caf1M* and *caf1* transcript. The question of whether the C-terminal domain of Caf1R functions as a sensor domain is interesting, particularly in view of the fact that the closest relative appears to be involved in carbohydrate transport. This could be addressed using a pull-down assay and mass spectrometry. Another possible role of the C-terminal domain is for dimerisation as with AraC (Soisson et al., 1997), XylR (Ni et al., 2013) and XylS (Dominguez-Cuevas et al., 2010) and Rns (Mahon et al., 2012) regulators. All the constructs and optimised expression conditions for recombinant tagged Caf1R are now available. Availability of purified tagged Caf1R would be very useful for further EMSA study, preparation of anti-Caf1R antibody and structural analysis.

This study has focused on the role of Caf1R in regulation of the *caf* locus. This represents a good model for studying the CU-system controlled by a dedicated regulator. There will be many other factors controlling expression of the *caf* locus. To fully understand regulation of F1 expression it will be important to map the role of other cellular factors onto the proposed model. Caf1R is a dedicated regulator linked to the *caf1MA1* operon. Many AraC/XylS like regulators function as global regulators, however, most of these are unlinked to the target operons. Examples include, Rob, MarA and SoxS (Ariza et al., 1995; Barbosa and Levy, 2000; Duval and Lister, 2013; Jair et al., 1996a), Rns (Basturea et al., 2008; Caron et al., 1989; Froehlich et al., 1994; Mahon et al., 2010; Munson et al., 2002; Munson and Scott, 1999, 2000) and AggR (Dudley et al., 2006; Morin et al., 2013; Nataro et al., 1994). Virulence of *Y. pestis* is still not well understood and regulators involved in global regulation may be critically important. Hence, despite the fact that Caf1R appears to be a dedicated regulator, it is still important to review any possibility of global regulation. A first step toward this would be searching *Y. pestis* genome for Caf1R consensus binding sites.

In vaccine design, including anti-plague vaccines, there is a lot of interest in development of attenuated vaccines and heterologous expression of F1 (Derbise et al., 2012; Feodorova and Motin, 2012). In this type of vaccine it is essential to retain optimum expression of F1. For this reason a full understanding of factors controlling expression of F1 is essential.

Appendices

1. Transcriptional fusion vectors used to construct promoter-*lacZ* fusions.

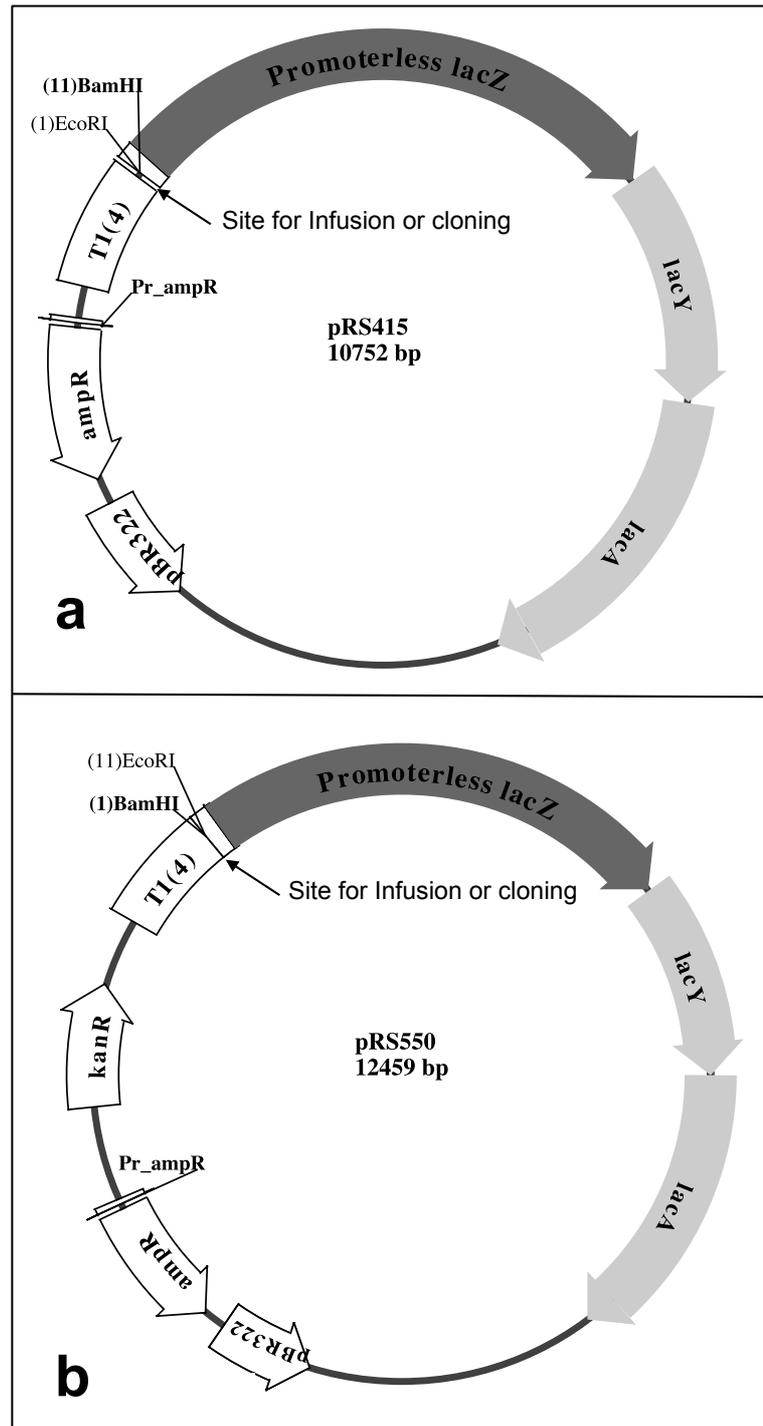


Figure A.1| **Transcriptional fusion vectors used to construct promoter-*lacZ* fusions.**

a) pRS415 used to design constructs for identification of promoters for *cafIR*. **b)** pRS550 used to design constructs for identification of promoters for *cafIM* and *cafI*. Key features of each vector are indicated along with reverse location of *EcoRI* and *BamHI* cloning site upstream of the promoterless *lacZ* gene. The 100 bp including SD sequence (AGGA) between the *lacZ* ATG start codon and multiple cloning sites are the same in both plasmids. This is in contrast to the available repository sequence (<http://www.mimg.ucla.edu/faculty/simons/vectors/lac-based/lac-based.htm>) of these vectors, where pRS550 missing a G nt at position 23 (relative to *BamHI* site) within a sequence, ATCCxGACA; x is missing G. T1(4), four consecutive sequences of *rrnB* terminator (Simons et al., 1987).

2a. Similarity between RNAP holoenzyme of *Y. pestis* CO92 and *E. coli* K-12.

RpoA or α subunit						
Score	Expect	Method	Identities	Positives	Gaps	Frame
659 bits(1701)	0.0()	Compositional matrix adjust.	327/329(99%)	328/329(99%)	0/329(0%)	
Features:						
Query	1	MQGSVTEFLKPRLDIEQVSSSTHAKVTLEPLERGFHTLGNALRRILLSSMPGCAVTEVE				60
Sbjct	1	MQGSVTEFLKPRLDIEQVSSSTHAKVTLEPLERGFHTLGNALRRILLSSMPGCAVTEVE				60
Query	61	IDGVLHEYSTKEGVQEDILEILLNLKGLAVRVQKDEVILTNNKSGIGPVTAADITHDGD				120
Sbjct	61	IDGVLHEYSTKEGVQEDILEILLNLKGLAVRVQKDEVILTNNKSGIGPVTAADITHDGD				120
Query	121	VEIVKPOHVICHLDENASINMRIKVGORGRGVVPASARIHSEEDERPIGRLLVDACYSVP				180
Sbjct	121	VEIVKPOHVICHLDENASINMRIKVGORGRGVVPASTRIHSEEDERPIGRLLVDACYSVP				180
Query	181	ERIAYNVEAARVEQRTDLDKLVIEEMTNGTIDPEEAIRRAATILAEQLEAFVDRDVRQP				240
Sbjct	181	ERIAYNVEAARVEQRTDLDKLVIEEMTNGTIDPEEAIRRAATILAEQLEAFVDRDVRQP				240
Query	241	EVKKEKPEPDPILLRPVDDLELTVRSANCLKAEAIHYIGDLVQRTVEVELLKPNNLKKSL				300
Sbjct	241	EVKKEKPEPDPILLRPVDDLELTVRSANCLKAEAIHYIGDLVQRTVEVELLKPNNLKKSL				300
Query	301	TEIKDVLASRGLSLGMRLNWPPIADE	329			
Sbjct	301	TEIKDVLASRGLSLGMRLNWPPIADE	329			
RpoB or β subunit						
Score	Expect	Method	Identities	Positives	Gaps	Frame
2620 bits(6792)	0.0()	Compositional matrix adjust.	1270/1342(95%)	1321/1342(98%)	0/1342(0%)	
Features:						
Query	1	MVYSYTEKKRIRKDFGKRPQVLDIPYLLSIQLDSFOKFIQDPEGGHGLEAFAFRSVPFIQ				60
Sbjct	1	MVYSYTEKKRIRKDFGKRPQVLDIPYLLSIQLDSFOKFIQDPEGGHGLEAFAFRSVPFIQ				60
Query	61	SYSGNSELOQVSYRLEGEVDFVKECOIRCVTYSAPLRVKLRVLYEREAEPTVKDIKEQ				120
Sbjct	61	SYSGNSELOQVSYRLEGEVDFVKECOIRCVTYSAPLRVKLRVLYEREAEPTVKDIKEQ				120
Query	121	EVYMGELPLMTENGTFVINGTFRVIVSQLRHSPGVFFSDSKGKTHSSGKVLNARIIPYR				180
Sbjct	121	EVYMGELPLMTENGTFVINGTFRVIVSQLRHSPGVFFSDSKGKTHSSGKVLNARIIPYR				180
Query	181	GSWLDPEFDPKDNLFVRIDRRRKLPAIILRALNFTTAQIILLDFPEKVVFEIRDNKLQME				240
Sbjct	181	GSWLDPEFDPKDNLFVRIDRRRKLPAIILRALNFTTAQIILLDFPEKVVFEIRDNKLQME				240
Query	241	LVPERLRGETASFDIEANGKVVYVEKARRITARHIRELEKDDIRIEVPEYIAGKVVAKD				300
Sbjct	241	LVPERLRGETASFDIEANGKVVYVEKARRITARHIRELEKDDIRIEVPEYIAGKVVAKD				300
Query	301	YVDASTGELICANMELSLDILLAKLSQAGHKQIETLFTNDLDHGAYISETLRVDPDRL				360
Sbjct	301	YIDESTGELICANMELSLDILLAKLSQSGHKRIETLFTNDLDHGPIYISETLRVDPDRL				360
Query	361	SALVEIYRMMRPGPEPTREAAENLFENLFFSEDRYDLSAVGRMKFNRSLLRDEIEGSGIL				420
Sbjct	361	SALVEIYRMMRPGPEPTREAAE+LFENLFFSEDRYDLSAVGRMKFNRSLLRDEIEGSGIL				420
Query	421	SKEDI+TEVMKKLIDIRNGGEVDDIDHLDGNRRIRSVGEMAEQFRVGLVVRERAVKERLS				480
Sbjct	421	SKDDIDVMMKKLIDIRNGGEVDDIDHLDGNRRIRSVGEMAEQFRVGLVVRERAVKERLS				480
Query	481	LGDLDTLMPQDMINAKPISAAVKEFFGSSQLSQFMDQNNPLSEITHKRRISALGPGLTR				540
Sbjct	481	LGDLDTLMPQDMINAKPISAAVKEFFGSSQLSQFMDQNNPLSEITHKRRISALGPGLTR				540
Query	541	ERAGFEVRDVHPHGRVCPVETPEGPNIGLINSLSVYAQTNEYGFLETPYRVRDGVVT				600
Sbjct	541	ERAGFEVRDVHPHGRVCPVETPEGPNIGLINSLSVYAQTNEYGFLETPYRVRDGVVT				600
Query	601	DEINYLSAIEEGNVIAQANSNLDDEGRFLEDLVTCRSKGESSLFSRQVDVMDVSTQOI				660
Sbjct	601	DEIHYLSAIEEGNVIAQANSNLD+EG F+EDLVTCRSKGESSLFSRQVDVMDVSTQOQ				660
Query	661	VSVGASLIPFLEHDDANRALMGANMQRQAVPTLRADKPLVGTGEMERAVAVDSGVT+VARR				720
Sbjct	661	VSVGASLIPFLEHDDANRALMGANMQRQAVPTLRADKPLVGTGEMERAVAVDSGVT+VARR				720
Query	721	GGTVQVVDASRIKVNEDEMHPGEAGIDIYNLTKYTRSNQNTCINQMPCVLGEPIERG				780
Sbjct	721	GGVVQVVDASRIKVNEDEMHPGEAGIDIYNLTKYTRSNQNTCINQMPCVLGEPIERG				780
Query	781	DVLADGPSTDLGELALGQNMRFVMPWNGYFNEDSILVSEVVQEDRFTTHIQELACVS				840
Sbjct	781	DVLADGPSTDLGELALGQNMRFVMPWNGYFNEDSILVSEVVQEDRFTTHIQELACVS				840
Query	841	RDTKLGPEEITADIPNVGEAALSCLDESIGVYIGAEVTGGDILVGVKTPKGETOLTPEEK				900
Sbjct	841	RDTKLGPEEITADIPNVGEAALSCLDESIGVYIGAEVTGGDILVGVKTPKGETOLTPEEK				900
Query	901	LLRAIFGEKASDVKSSLRVPGVSGTVIDVQVFTRDGVEKDKRALEIEEMQLKQAKKDL				960
Query	961	TEELQILEAGLFAIRIHAVLVSGGIEAEKLSKLPREWRLELGLTDEKQNOLEQLAEQYDE				1020
Sbjct	961	SEELQILEAGLFSRIRAVLVAGGVEAEKLDKLPDRWLELGLTDEKQNOLEQLAEQYDE				1020
Query	1021	MKSEFEKMDAKRRKITQCDLAPGVKIVKVVYLAVKRQIQPCDKMAGRHNKGVISKIN				1080
Sbjct	1021	LKHEFEKLEAKRRKITQCDLAPGVKIVKVVYLAVKRRIQPCDKMAGRHNKGVISKIN				1080
Query	1081	PIEDMPYDENGTPVDIVLNLPGVPSRMNIGQILETHLGMAAKIGIKINAMLKQEEVAK				1140
Sbjct	1081	PIEDMPYDENGTPVDIVLNLPGVPSRMNIGQILETHLGMAAKIGIKINAMLKQEEVAK				1140
Query	1141	LREFIQAYDLGDNVCKVDLSFFDDEVLRLAENLKGMP+IATPVFDGAPEKIEKELLO				1200
Sbjct	1141	LREFIQAYDLGADVRQVDLSFFDDEVLRLAENLKGMP+IATPVFDGAPEKIEKELLL				1200
Query	1201	LGGLPTSGQITLFDGRTGEQFERQVTVGYMYMLKLNHLVDDKMHARSTGSYSLVTOQPLG				1260
Sbjct	1201	LGGLPTSGQITLFDGRTGEQFERVTVGYMYMLKLNHLVDDKMHARSTGSYSLVTOQPLG				1260
Query	1261	GKAQFGGQRFGEMEVWALEAYGAAATLQEMLTVKSDDVNGRTKMYKNIVDGDMPEFGMP				1320
Sbjct	1261	GKAQFGGQRFGEMEVWALEAYGAAATLQEMLTVKSDDVNGRTKMYKNIVDGNHMQEPGMP				1320
Query	1321	ESFNVLKKEIRSLGINIELEEE	1342			
Sbjct	1321	ESFNVLKKEIRSLGINIELEDE	1342			

RpoC or β' subunit

Score	Expect	Method	Identities	Positives	Gaps	Frame
2680 bits(6948)	0.0()	Compositional matrix adjust.	1304/1404(93%)	1358/1404(96%)	0/1404(0%)	
Features:						
Query	1	MKDLLKFLKAQTKTEEFDAIKIALASPD	MIRSWSFGEVKKPETINVRTFKPERDGLFCAR	60		
Sbjct	1	MKDLLKFLKAQTKTEEFDAIKIALASPD	MIRSWSFGEVKKPETINVRTFKPERDGLFCAR	60		
Query	61	IFGPVKDYECCLCGYKRLKHRGVI	CEKCGVEVTQTKVRRERMCHIELASPTAHIWFLKSL	120		
Sbjct	61	IFGPVKDYECCLCGYKRLKHRGVI	CEKCGVEVTQTKVRRERMCHIELASPTAHIWFLKSL	120		
Query	121	PSRIGLLLDMPLRDIERVLYFESYVV	IEGGMTNLERRQILTEEQYLDAALEEFGEFDDAKM	180		
Sbjct	121	PSRIGLLLDMPLRDIERVLYFESYVV	IEGGMTNLER+QILTEEQYLDAALEEFGEFDDAKM	180		
Query	181	GAEAIQALLKNDLEAECEILREELNET	NETSKRKKLTKRIKLEAFVQSGNKPEWMILT	240		
Sbjct	181	GAEAIQALLK+MDLE ECE LREELNET	NETSKRKKLTKRIKLEAFVQSGNKPEWMILT	240		
Query	241	VLPVLPDLRPLVPLDGGRFATS	DLNDLYRRVINRNRNLKRLDLAAPDIIVRNEKRMQLQ	300		
Sbjct	241	VLPVLPDLRPLVPLDGGRFATS	DLNDLYRRVINRNRNLKRLDLAAPDIIVRNEKRMQLQ	300		
Query	301	EAVDALLDNGRRRAITGSKNRPLKSL	ADMIKKGQGRFRONLLGKRVDSYSGRSVITVGPY	360		
Sbjct	301	EAVDALLDNGRRRAITGSKNRPLKSL	ADMIKKGQGRFRONLLGKRVDSYSGRSVITVGPY	360		
Query	361	LRLHQCGLPKMALELFKPF	IYGKLELRGLATTIKAAKKMVEREEAVVWDLDEVIREHP	420		
Sbjct	361	LRLHQCGLPKMALELFKPF	IYGKLELRGLATTIKAAKKMVEREEAVVWDLDEVIREHP	420		
Query	421	VLLNRAPTLHRLGIQAFEPV	IEGKAIQLHPLVCAAYNADFDGDMAVHVPILTLEAQLEA	480		
Sbjct	421	VLLNRAPTLHRLGIQAFEPV	IEGKAIQLHPLVCAAYNADFDGDMAVHVPILTLEAQLEA	480		
Query	481	RALMMSNNILSPANGPEI	IIVPSQDVVLGLYMTDRDCVNAKGEGMVLTGPKEAERIYRAG	540		
Sbjct	481	RALMMSNNILSPANGPEI	IIVPSQDVVLGLYMTDRDCVNAKGEGMVLTGPKEAERIYRAG	540		
Query	541	LASLHARVKVRITEEIRNTEGES	ITRSTSIDTTVGRAILWMIIVPQGLPYSIVNQPLGKKA	600		
Sbjct	541	LASLHARVKVRITEEIRNTEGES	ITRSTSIDTTVGRAILWMIIVPQGLPYSIVNQPLGKKA	600		
Query	601	ISKMLNTCYRILGLKPTVIFADQ	IMYTGFAAARSASGASVDDMVPEAKAGIEEAEETE	660		
Sbjct	601	ISKMLNTCYRILGLKPTVIFADQ	IMYTGFAAARSASGASVDDMVPEAKAGIEEAEETE	660		
Query	661	VAEIQEQFQSGLVTAGERYNKVI	DIWAAANERVAKAMMDNLSVEDVNRDGVQVQVSN	720		
Sbjct	661	VAEIQEQFQSGLVTAGERYNKVI	DIWAAANERVAKAMMDNLSVEDVNRDGVQVQVSN	720		
Query	721	SIFMMADSGARSAQIRQLAGMR	LMAKPDGSI IETPTANFREGLNVLQYFISTHGAR	780		
Sbjct	721	SIFMMADSGARSAQIRQLAGMR	LMAKPDGSI IETPTANFREGLNVLQYFISTHGAR	780		
Query	781	KGLADTALKTANSGYLTRRLVDV	AQDLVVTEDDCGTHNGIIVMTPVIEGGDVKEPLDRVLL	840		
Sbjct	781	KGLADTALKTANSGYLTRRLVDV	AQDLVVTEDDCGTHNGIIVMTPVIEGGDVKEPLDRVLL	840		
Query	841	GRVTAEEVVKPGSADILVPRNTL	LLEKWCDDLEENSVDSVVKRSVSCETDFGVCANCYG	900		
Sbjct	841	GRVTAEEVVKPGSADILVPRNTL	LLEKWCDDLEENSVDSVVKRSVSCETDFGVCANCYG	900		
Query	901	RDLARGHINKGEAVGVIAAQS	I GEPGTQMTMRTFHHGGAASRAAAESSIQVKNKGS	960		
Sbjct	901	RDLARGHINKGEAVGVIAAQS	I GEPGTQMTMRTFHHGGAASRAAAESSIQVKNKGS	960		
Query	961	SNVKFVNAAGKLVITSRNTE	KLIDEFGRTKESYKVPYGAVMAGDGAEOGGETVANW	1020		
Sbjct	961	SNVKFVNAAGKLVITSRNTE	KLIDEFGRTKESYKVPYGAVMAGDGAEOGGETVANW	1020		
Query	1021	DPHIMPVVTEVSGFIRFADMV	DGQTIITRQDELTLGSLVLDVLSAERTGSKDLRPAKI	1080		
Sbjct	1021	DPHIMPVVTEVSGFIRFADMV	DGQTIITRQDELTLGSLVLDVLSAERTGSKDLRPAKI	1080		
Query	1081	VDAKGNVLIPIGTDMPAQYFL	PGKAIQVLEDGIQIAGDTLARIPQESSGTDITGGLPR	1140		
Sbjct	1081	VDAKGNVLIPIGTDMPAQYFL	PGKAIQVLEDG+QI +GDTLARIPQES GTKDTITGGLPR	1140		
Query	1141	VADLFEARRPKPAALAEISGI	ISFGKETKGRRLVISPDLGSDAYEEMIIPKWRQLNVFE	1200		
Sbjct	1141	VADLFEARRPKPAALAEISGI	ISFGKETKGRRLVITPVDGSDPYEEMIIPKWRQLNVFE	1200		
Query	1201	GEVVERGDVSDGPEPHDILRL	RGVHAVTRYITNEVQEVYRLQGVKINDKHIEIVRQM	1260		
Sbjct	1201	GEVVERGDVSDGPEPHDILRL	RGVHAVTRYI NEVQ+VYRLQGVKINDKHIEIVRQM	1260		
Query	1261	LRKGTIVDAGSTDFLEGEQA	EMSRVKIANRKLAEKGIEATFTRDLLGITKASLATESFI	1320		
Sbjct	1261	LRKGTIVDAGSTDFLEGEQA	EMSRVKIANRKLAEKGIEATFTRDLLGITKASLATESFI	1320		
Query	1321	SAASFQETTRVLTEAAVAGKR	DELRLGKENVIVGRLIPAGTGYAYHQRMRRAQGEAPV	1380		
Sbjct	1321	SAASFQETTRVLTEAAVAGKR	DELRLGKENVIVGRLIPAGTGYAYHQRMRRAQGEAPV	1380		
Query	1381	VPOVSADAEATANLAELLNAG	FNN 1404			
Sbjct	1381	POV+AAA+LAELLNAG G +	APQVTAEDASASLAELLNAGLGG 1404			

RpoZ or ω subunit

Score	Expect	Method	Identities	Positives	Gaps	Frame
131 bits(330)	1e-45()	Compositional matrix adjust.	84/91(92%)	87/91(95%)	0/91(0%)	
Features:						
Query	1	MARVTVQDAVEKIGNRFDLVLVAARRARQIQSGGKDALVPEENDKVTVIALREIEEGLIT				60
Sbjct	1	MARVTVQDAVEKIGNRFDLVLVAARRARQ+O GGD LVPEENDK TVIALREIEEGLI				60
Query	61	NQILLVRRERQEQEQAAEIQAQVTAIAEGRR	91			
Sbjct	61	NQILLVRRERQEQEQAAELQAQVTAIAEGRR	91			

RpoD or $\sigma 70$ factor

Score	Expect	Method	Identities	Positives	Gaps	Frame
1106 bits(2861)	0.0()	Compositional matrix adjust.	562/615(91%)	587/615(95%)	5/615(0%)	
Features:						
Query	1	MEQNPOSQKLLVTRGKEQGYLYAEVNDHLPEDIVDSQIEDIIQMINDMGIQVLEEAP				60
Sbjct	1	MEQNPOSQKLLVTRGKEQGYLYAEVNDHLPEDIVDSQIEDIIQMINDMGIQV+EEAP				60
Query	61	DADDLMLAENTTDDDDAAAAQVLSVSESEIGRTDPVRMYMREMGTVELLTREGEID				120
Sbjct	61	DADDLMLAENT D D AAAAAQVLSVSESEIGRTDPVRMYMREMGTVELLTREGEID				118
Query	121	IAKRIEDGINQVQCSVAEYPEAITYLLEQYDRVEAGESRLSDLTGTFVDPNAEEDIAPTA				180
Sbjct	119	IAKRIEDGINQVQCSVAEYPEAITYLLEQYDRVEA E+RLSDLTGTFVDPNAEED+APTA				178
Query	181	THVGSLS---TEEMDEDEDEDEDDAEDDNSIDPELARQKFSDLREQYENARMEIKKN				237
Sbjct	179	THVGSLSQEDLDDEDEDEDEDDSDADDNSIDPELAREKFAELRAQYVVRTDTIKAK				238
Query	238	GRNHANAEEILKLVSEVFKQFRLVVKQFDYLVNMRAMMDRVRTQERIMKLCVEQCKMP				297
Sbjct	239	GRSHATAQEEILKLVSEVFKQFRLVVKQFDYLVNSMRVMDRVRTQERIMKLCVEQCKMP				298
Query	298	KKNFVTLFSSNETSDTFWNAAVAMKPKWSEKLDVSEVQRSLOKLRQIEEETGLTIEQV				357
Sbjct	299	KKNFITLFTGNETSDFWNAAIAMNKPWSEKLDVSEVHRALQKQIEEETGLTIEQV				358
Query	358	KDINRRMSIGAKARRAKKEMVEANLRLVSI AKKYTNRGLQFLDLIQEIGNGLMKAVDK				417
Sbjct	359	KDINRRMSIGAKARRAKKEMVEANLRLVSI AKKYTNRGLQFLDLIQEIGNGLMKAVDK				418
Query	418	FEYRRGYKFSTYATWWIRQAITRSIADQARTIRIPVHMIEITINKLNRISRQMLQEMGREP				477
Sbjct	419	FEYRRGYKFSTYATWWIRQAITRSIADQARTIRIPVHMIEITINKLNRISRQMLQEMGREP				478
Query	478	TPEELAEERMLMPEDKIRKVLKIAKEPISMETPIGDDEDSHLDGPIEDTTLELPDLSATSE				537
Sbjct	479	TPEELAEERMLMPEDKIRKVLKIAKEPISMETPIGDDEDSHLDGPIEDTTLELPDLSATTE				538
Query	538	SLRSATHDVLAGLTAREAKVLRMRFGIDMNTDHTLEEVGKQFDVTRERIRQIEAKALRKL				597
Sbjct	539	SLRAATHDVLAGLTAREAKVLRMRFGIDMNTDYTLEEVGKQFDVTRERIRQIEAKALRKL				598
Query	598	RHPSRSEVLSRSLDD	612			
Sbjct	599	RHPSRSEVLSRSLDD	613			

RpoH or σ32 factor						
Score	Expect	Method	Identities	Positives	Gaps	Frame
483 bits(1243) 2e-177() Compositional matrix adjust. 238/285(84%) 265/285(92%) 1/285(0%)						
Features:						
Query	1	MTKEMQTLALVPQGSLEAYIRAANAYPMLTAEERELAERLHYQGDLAGAAKQLILSHLRF				60
Sbjct	1	MT +MQ+LAL P G+L++YIRAANA+PML+A+EER LAE+LHY GDL AAK LILSHLRF				60
Query	61	VAHVARNYSYGLPQADLIQEGNIGLMAKAVRRFNPEVGVRLVSFAVHWIKAEIHEYVLRN				120
Sbjct	61	V H+ARNY+GYGLPQADLIQEGNIGLMAKAVRRFNPEVGVRLVSFAVHWIKAEIHEYVLRN				120
Query	121	WRIVKVATTKAQRKLFNLRKTKQRLGWFNQDEVELVAKELGVTSKDVREMSRMSAQDM				180
Sbjct	121	WRIVKVATTKAQRKLFNLRKTKQRLGWFNQDEVE+VA+ELGVTSKDVREMSRMA+AQDM				180
Query	181	TFDPSDDEARDQGFMAPVLYLQDKTSDFADGIEEDNWDNSHAADKLSYALEGLDERSQHI				240
Sbjct	181	TFD S D+ D Q MAPVLYLQDK+S+FADGIE+DNW+ AA++L+ A++GLDERSQ I				239
Query	241	IRARWLDDENKSTLQELADQYGVSAERVROLEKNAMKKLRMAIEA				285
Sbjct	240	IRARWLD++NKSTLQELAD+YGVSAERVROLEKNAMKKLR AIEA				284
RpoS or σ38 factor						
Score	Expect	Method	Identities	Positives	Gaps	Frame
615 bits(1587) 0.0() Compositional matrix adjust. 307/332(92%) 324/332(97%) 2/332(0%)						
Features:						
Query	1	MSQSTLKVNELHEDADFDENSTETEIFDEKALVDDEPTESELADDELLAQGVTRVLDAT				60
Sbjct	1	MSQ+TLKV++L+EDR+FDEN E +FDEKALV+ EP+ ++L+ ++ELL+QC TQRVLDAT				58
Query	61	QYLGEIGYSPLLTAEEEVYFARRALRGDVSRRRMIESNLRVVKIARRYSNRGLALLD				120
Sbjct	59	QYLGEIGYSPLLTAEEEVYFARRALRGDVSRRRMIESNLRVVKIARRYGNRGLALLD				118
Query	121	LIEEGNLGLIRAVEKFDPERGFRFSTYATWWIRQTIERAIMNQTRTIRLPIHIVKELNVY				180
Sbjct	119	LIEEGNLGLIRAVEKFDPERGFRFSTYATWWIRQTIERAIMNQTRTIRLPIHIVKELNVY				178
Query	181	LRTARELSHKLDHEPSAEIEAEQLDKPVDDVSRMLRLNERITSVDTPLGGDSEKALLDIL				240
Sbjct	179	LRTARELSHKLDHEPSAEIEAEQLDKPVDDVSRMLRLNERITSVDTPLGGDSEKALLDIL				238
Query	241	SDENENGPEDTTQDDDMKQSIKWLFLFELNAKQREVLARRFGLLYEAATLEDVGREIGLT				300
Sbjct	239	+DE ENGPEDTTQDDDMKQSIKWLFLFELNAKQREVLARRFGLLYEAATLEDVGREIGLT				298
Query	301	RERVROIQVEGLRRLREILQTOGLSIEALPRE				332
Sbjct	299	RERVROIQVEGLRRLREILQTOGLNIEALPRE				330

Figure A.2a| **Similarity between RNAP holoenzyme of *Y. pestis* CO92 and *E. coli* K-12.**

The amino acid sequence of the corresponding subunits and σ factors of RNAP holoenzyme of *Y. pestis* CO92 and *E. coli* K-12 were retrieved from NCBI with UniProtKB/Swiss-Prot accession numbers, Q8ZJ87 (RpoA or α subunit), Q8ZAP5 (RpoB or β subunit), A0A0B6NVW2 (RpoC or β' subunit), A0A0B6NRU2 (RpoZ or ω subunit), A0A0B6NX87 (RpoD or σ 70 factor), A0A0B6NRK5 (RpoH or σ 32 factor), A0A0B6NVW7 (RpoS or σ 38 factor); *Y. pestis* CO92 and P0A7Z4 (RpoA or α subunit), P0A8V2 (RpoB or β subunit), P0A8T7 (RpoC or β' subunit), P0A800 (RpoZ or ω subunit), P00579 (RpoD or σ 70 factor), P0AGB3 (RpoH or σ 32 factor), P13445 (RpoS or σ 38 factor); *E. coli* K-12 and were analysed by pairwise sequence alignment using NCBP BLASTp service with default parameters. Query, *Y. pestis* and Sbjct, *E. coli*.

2b. Promoters elements confirmed and predicted for some *Y. pestis* genes.

Gene	-10 TATAAT	-35 TTGACA	Spacer (nt)	Reference
<i>yopN</i>	TATAAT	TTGgCA	21	(King et al., 2013); Regulated by LcrF/VirF
<i>lcrG</i>	TAgAAT	cTGACA	21	
<i>yscN</i>	TAaAAT	cTacaA	21	
<i>yscB</i>	TAaAAT	TTtAaA	21	
<i>pla</i>	TtTAAT	acGAgA	19	(Kim et al., 2007); Regulated by CRP
<i>fyu</i>	TATcAT	TaGcac	16	(Gao et al., 2008); Regulated by Fur and YbtA
<i>irp2</i>	TATtAT	TaGcat	15	
<i>ybtA</i>	ataAta	TgGcgt	17	
<i>pst</i>	TAattT	TTGACA	17	(Rakin et al., 1996); Regulated by bacteriocin
<i>mgtC</i>	TATAcT	TTtACA	17	(Li et al., 2008); Regulated by PhoP
<i>katA</i>	TAcAAc	TTGctg	17	
<i>sodB</i>	TATggT	TcGAtc	17	
<i>sodC</i>	TAaAAc	TTatCc	17	
<i>uspB</i>	TAcAcT	agGctA	17	
<i>uspA</i>	TATAcT	TcaACc	17	
<i>pagP</i>	TATAga	TTcAac	17	
<i>oppA</i>	TATcAc	TTGACc	17	
<i>metJ</i>	TAaAAT	TTGAgc	17	
<i>furR</i>	TtagAT	TTGctc	17	
<i>phoP</i>	TATcgT	TTtAtA	17	
<i>slyA</i>	TATtAT	TaaAgA	17	
<i>astC</i>	TtatAT	gTGgCA	17	

Figure A.2b| **Promoter elements confirmed (light) and predicted (dark) for some *Y. pestis* genes.** Consensus and non-consensus nt in each -10 and -35 are indicated by small and capital letters, respectively.

3. The *caf* locus promoters predicted upstream of *caf1R*, *caf1M* and *caf1*.

Promoter and location relative to ATG start	-10 element consensus and frequency % at each position (T ₈₀ A ₉₅ T ₄₅ A ₅₀ A ₆₀ T ₉₆)	-35 element consensus and frequency % at each position (T ₈₂ T ₈₄ G ₇₈ A ₆₅ C ₅₄ A ₄₈)	Spacer (nt)	Source
Upstream of <i>caf1R</i>				
P _R ¹ (-20 to -48)	TATtAT (5)	cgccgg (0)	17	Visual
P _R ² (-28 to -57)	TAgAAT (5)	aaGtCc (2)	18	Visual
P _R ^K (-86 to -113)	TAagAT (4)	TTGAtg (4)	16	(Karlyshev et al., 1992)
Upstream of <i>caf1M</i>				
P _M ^{K1} (-51 to -78)	aATgAT (3)	TTtACg (4)	16	(Karlyshev et al., 1992)
P _M ^{K2} (-65 to -92)	TATcgT (4)	TTGtTA (4)	16	(Karlyshev et al., 1992)
P _M ^B (-105 to -134) LDF score 6.38	TATAAa (5) 41	TTctCA (4) 33	18	BPROM
P _M (-137 to -165)	TAaAAT (5)	TaaACt (3)	17	Visual
Upstream of <i>caf1</i>				
P ₁ ^B (-1 to -29) LDF score 4.25	TAatAT (4) 68	TgGAtt (3) 16	17	BPROM
P ₁ ¹ (-32 to -60)	TtTgtT (3)	agGACA (4)	17	Visual
P ₁ ² (-153 to -181)	TcTAAT (5)	gaGAtA (3)	17	Visual
P ₁ ³ (-199 to -227)	aAaAAT (4)	TTGAct (5)	17	Visual

Figure A.3| The *caf* locus promoters predicted upstream of *caf1R*, *caf1M* and *caf1*.

Capital and small letters in both -10 and -35 elements represent the corresponding consensus and non-consensus nucleotides (nt) at the respective position. Number in bracket (x), at the end of each -10 and -35 element indicates the total number of consensus nt out of 6 nt for each element. (-xx to -xx) represent the upstream location (in bp) of each promoter relative to ATG start codon of the corresponding gene. LDF score of BPROM predicted promoters along with individual score of both -10 and -35 elements is shown underneath. BROM identified a strongest promoter within *caf1A* (at +192 to +219 bp, relative to ATG start) with LDF score 6.68, -10 (TAgAcT; score 62), -35 (TTaAtg; score 29) and 21 nt spacer.

4. Sequencing result of pRScaf1R-caf1M'-lacZ.

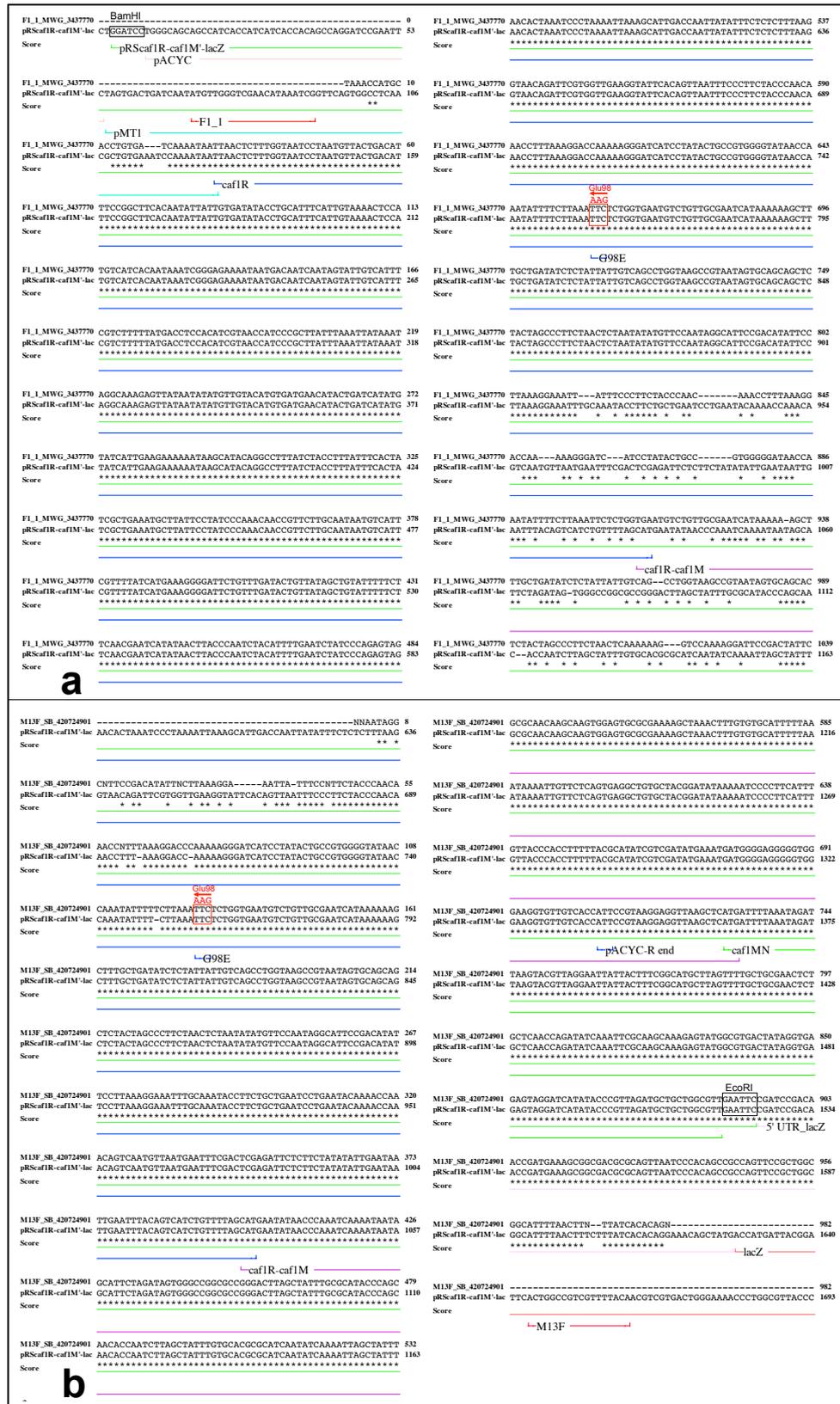


Figure A.4| Sequencing result of pRScaf1R-caf1M'-lacZ.

Pairwise sequence alignment of designed construct, pRScaf1R-caf1M'-lacZ (annotated) with retrieved sequences from sequencing primers, F1_1, forward (a) and M13F, reverse (b). Repaired Glu98 residue (from E98G/K) is shown in the red boxes and annotated. Cloning sites (BamHI and EcoRI) on either side are in-boxed (black). Sequence accession number from MWG, Germany (for F1_1) and Source Bioscience (SB), Oxford (for M13F) is indicated on left side of each sequence. Alignment was generated using DNA dynamo via ClustalW interface.

5. ΔG of the predicted RNATs at 5' UTR of *caf1R* relative to transcript from P_R^2 .

RNAT_PR2_26C_Met $\Delta G = -2.19$			PR2_28C_Met $\Delta G = -2.02$		
Structural element	δG	Information	Structural element	δG	Information
External loop	-0.87	12 ss bases & 1 closing helices.	External loop	-0.86	12 ss bases & 1 closing helices.
Stack	-2.11	External closing pair is U ₁₁ -A ₂₃	Stack	-2.05	External closing pair is U ₁₁ -A ₂₃
Stack	-1.54	External closing pair is G ₁₂ -C ₂₂	Stack	-1.51	External closing pair is G ₁₂ -C ₂₂
Stack	-0.51	External closing pair is G ₁₃ -U ₂₁	Stack	-0.49	External closing pair is G ₁₃ -U ₂₁
Stack	-1.11	External closing pair is G ₁₄ -U ₂₀	Stack	-1.09	External closing pair is G ₁₄ -U ₂₀
Helix	-5.27	5 base pairs.	Helix	-5.14	5 base pairs.
Hairpin loop	3.95	Closing pair is U ₁₅ -A ₁₉	Hairpin loop	3.98	Closing pair is U ₁₅ -A ₁₉
RNAT_PR2_30C_Met $\Delta G = -1.87$			RNAT_PR2_37C_Met $\Delta G = -1.30$		
Structural element	δG	Information	Structural element	δG	Information
External loop	-0.85	12 ss bases & 1 closing helices.	External loop	-0.80	12 ss bases & 1 closing helices.
Stack	-2.00	External closing pair is U ₁₁ -A ₂₃	Stack	-1.80	External closing pair is U ₁₁ -A ₂₃
Stack	-1.49	External closing pair is G ₁₂ -C ₂₂	Stack	-1.40	External closing pair is G ₁₂ -C ₂₂
Stack	-0.47	External closing pair is G ₁₃ -U ₂₁	Stack	-0.40	External closing pair is G ₁₃ -U ₂₁
Stack	-1.07	External closing pair is G ₁₄ -U ₂₀	Stack	-1.00	External closing pair is G ₁₄ -U ₂₀
Helix	-5.03	5 base pairs.	Helix	-4.60	5 base pairs.
Hairpin loop	4.01	Closing pair is U ₁₅ -A ₁₉	Hairpin loop	4.10	Closing pair is U ₁₅ -A ₁₉

i)

RNAT_PR2_26C_5 codons $\Delta G = -5.67$			RNAT_PR2_28C_5 codons $\Delta G = -5.33$		
Structural element	δG	Information	Structural element	δG	Information
External loop	-0.87	5 ss bases & 1 closing helices.	External loop	-0.86	5 ss bases & 1 closing helices.
Stack	-1.19	External closing pair is U ₄ -A ₃₅	Stack	-1.17	External closing pair is U ₄ -A ₃₅
Stack	-0.61	External closing pair is U ₅ -G ₃₄	Stack	-0.59	External closing pair is U ₅ -G ₃₄
Stack	-2.11	External closing pair is U ₆ -A ₃₃	Stack	-2.05	External closing pair is U ₆ -A ₃₃
Helix	-3.91	4 base pairs.	Helix	-3.81	4 base pairs.
Interior loop	0.77	External closing pair is G ₇ -C ₃₂	Interior loop	0.78	External closing pair is G ₇ -C ₃₂
Stack	-1.10	External closing pair is U ₉ -A ₃₀	Stack	-1.07	External closing pair is U ₉ -A ₃₀
Stack	-1.10	External closing pair is U ₁₀ -A ₂₉	Stack	-1.07	External closing pair is U ₁₀ -A ₂₉
Stack	-0.91	External closing pair is U ₁₁ -A ₂₈	Stack	-0.89	External closing pair is U ₁₁ -A ₂₈
Stack	-2.28	External closing pair is G ₁₂ -U ₂₇	Stack	-2.21	External closing pair is G ₁₂ -U ₂₇
Helix	-5.39	5 base pairs.	Helix	-5.24	5 base pairs.
Hairpin loop	3.73	Closing pair is G ₁₃ -C ₂₆	Hairpin loop	3.80	Closing pair is G ₁₃ -C ₂₆
RNAT_PR2_30C_5 codons $\Delta G = -5.00$			RNAT_PR2_37C_5 codons $\Delta G = -3.80$		
Structural element	δG	Information	Structural element	δG	Information
External loop	-0.85	5 ss bases & 1 closing helices.	External loop	-0.80	5 ss bases & 1 closing helices.
Stack	-1.16	External closing pair is U ₄ -A ₃₅	Stack	-1.10	External closing pair is U ₄ -A ₃₅
Stack	-0.57	External closing pair is U ₅ -G ₃₄	Stack	-0.50	External closing pair is U ₅ -G ₃₄
Stack	-2.00	External closing pair is U ₆ -A ₃₃	Stack	-1.80	External closing pair is U ₆ -A ₃₃
Helix	-3.73	4 base pairs.	Helix	-3.40	4 base pairs.
Interior loop	0.78	External closing pair is G ₇ -C ₃₂	Interior loop	0.80	External closing pair is G ₇ -C ₃₂
Stack	-1.03	External closing pair is U ₉ -A ₃₀	Stack	-0.90	External closing pair is U ₉ -A ₃₀
Stack	-1.03	External closing pair is U ₁₀ -A ₂₉	Stack	-0.90	External closing pair is U ₁₀ -A ₂₉
Stack	-0.87	External closing pair is U ₁₁ -A ₂₈	Stack	-0.80	External closing pair is U ₁₁ -A ₂₈
Stack	-2.14	External closing pair is G ₁₂ -U ₂₇	Stack	-1.90	External closing pair is G ₁₂ -U ₂₇
Helix	-5.07	5 base pairs.	Helix	-4.50	5 base pairs.
Hairpin loop	3.87	Closing pair is G ₁₃ -C ₂₆	Hairpin loop	4.10	Closing pair is G ₁₃ -C ₂₆

ii)

RNAT_PR2_26C_10 codons $\Delta G = -11.61$			RNAT_PR2_28C_10 codons $\Delta G = -10.88$		
Structural element	δG	Information	Structural element	δG	Information
External loop	-0.87	11 ss bases & 1 closing helices.	External loop	-0.86	11 ss bases & 1 closing helices.
Stack	-1.10	External closing pair is U ₁₀ -A ₅₀	Stack	-1.07	External closing pair is U ₁₀ -A ₅₀
Stack	-2.11	External closing pair is U ₁₁ -A ₄₉	Stack	-2.05	External closing pair is U ₁₁ -A ₄₉
Stack	-1.54	External closing pair is G ₁₂ -C ₄₈	Stack	-1.51	External closing pair is G ₁₂ -C ₄₈
Helix	-4.75	4 base pairs.	Helix	-4.63	4 base pairs.
Interior loop	1.18	External closing pair is G ₁₃ -U ₄₇	Interior loop	1.28	External closing pair is G ₁₃ -U ₄₇
Stack	-1.10	External closing pair is U ₁₅ -A ₄₄	Stack	-1.07	External closing pair is U ₁₅ -A ₄₄
Stack	-1.35	External closing pair is U ₁₆ -A ₄₃	Stack	-1.30	External closing pair is U ₁₆ -A ₄₃
Stack	-1.11	External closing pair is A ₁₇ -U ₄₂	Stack	-1.09	External closing pair is A ₁₇ -U ₄₂
Stack	-0.91	External closing pair is U ₁₈ -G ₄₁	Stack	-0.89	External closing pair is U ₁₈ -G ₄₁
Helix	-4.47	5 base pairs.	Helix	-4.35	5 base pairs.
Interior loop	0.77	External closing pair is A ₁₉ -U ₄₀	Interior loop	0.78	External closing pair is A ₁₉ -U ₄₀
Stack	-2.69	External closing pair is U ₂₁ -A ₃₈	Stack	-2.62	External closing pair is U ₂₁ -A ₃₈
Stack	-2.11	External closing pair is C ₂₂ -G ₃₇	Stack	-2.05	External closing pair is C ₂₂ -G ₃₇
Stack	-1.07	External closing pair is A ₂₃ -U ₃₆	Stack	-1.04	External closing pair is A ₂₃ -U ₃₆
Helix	-5.87	4 base pairs.	Helix	-5.71	4 base pairs.
Bulge loop	1.07	External closing pair is U ₂₄ -A ₃₅	Bulge loop	1.17	External closing pair is U ₂₄ -A ₃₅
Stack	-1.91	External closing pair is C ₂₆ -G ₃₄	Stack	-1.87	External closing pair is C ₂₆ -G ₃₄
Helix	-1.91	2 base pairs.	Helix	-1.87	2 base pairs.
Hairpin loop	3.24	Closing pair is U ₂₇ -A ₃₃	Hairpin loop	3.31	Closing pair is U ₂₇ -A ₃₃

RNAT_PR2_30C_10 codons $\Delta G = -10.20$			RNAT_PR2_37C_10 codons $\Delta G = -7.70$		
Structural element	δG	Information	Structural element	δG	Information
External loop	-0.85	11 ss bases & 1 closing helices.	External loop	-0.80	11 ss bases & 1 closing helices.
Stack	-1.03	External closing pair is U ₁₀ -A ₅₀	Stack	-0.90	External closing pair is U ₁₀ -A ₅₀
Stack	-2.00	External closing pair is U ₁₁ -A ₄₉	Stack	-1.80	External closing pair is U ₁₁ -A ₄₉
Stack	-1.49	External closing pair is G ₁₂ -C ₄₈	Stack	-1.40	External closing pair is G ₁₂ -C ₄₈
Helix	-4.52	4 base pairs.	Helix	-4.10	4 base pairs.
Interior loop	1.37	External closing pair is G ₁₃ -U ₄₇	Interior loop	1.70	External closing pair is G ₁₃ -U ₄₇
Stack	-1.03	External closing pair is U ₁₅ -A ₄₄	Stack	-0.90	External closing pair is U ₁₅ -A ₄₄
Stack	-1.26	External closing pair is U ₁₆ -A ₄₃	Stack	-1.10	External closing pair is U ₁₆ -A ₄₃
Stack	-1.07	External closing pair is A ₁₇ -U ₄₂	Stack	-1.00	External closing pair is A ₁₇ -U ₄₂
Stack	-0.87	External closing pair is U ₁₈ -G ₄₁	Stack	-0.80	External closing pair is U ₁₈ -G ₄₁
Helix	-4.23	5 base pairs.	Helix	-3.80	5 base pairs.
Interior loop	0.78	External closing pair is A ₁₉ -U ₄₀	Interior loop	0.80	External closing pair is A ₁₉ -U ₄₀
Stack	-2.55	External closing pair is U ₂₁ -A ₃₈	Stack	-2.30	External closing pair is U ₂₁ -A ₃₈
Stack	-2.00	External closing pair is C ₂₂ -G ₃₇	Stack	-1.80	External closing pair is C ₂₂ -G ₃₇
Stack	-1.01	External closing pair is A ₂₃ -U ₃₆	Stack	-0.90	External closing pair is A ₂₃ -U ₃₆
Helix	-5.56	4 base pairs.	Helix	-5.00	4 base pairs.
Bulge loop	1.26	External closing pair is U ₂₄ -A ₃₅	Bulge loop	1.60	External closing pair is U ₂₄ -A ₃₅
Stack	-1.83	External closing pair is C ₂₆ -G ₃₄	Stack	-1.70	External closing pair is C ₂₆ -G ₃₄
Helix	-1.83	2 base pairs.	Helix	-1.70	2 base pairs.
Hairpin loop	3.38	Closing pair is U ₂₇ -A ₃₃	Hairpin loop	3.60	Closing pair is U ₂₇ -A ₃₃

iii)

Figure A.5 | ΔG of the predicted RNATs at 5' UTR of *caf1R* relative to transcript from P_R^2 . Using Mfold 2.3 program (Zuker, 2003) RNA secondary structures were predicted by inclusion of Caf1R first codon (i), first 5 codons (ii) and first 10 codons (iii) at 26, 28, 30 and 37°C. The most probable prediction with lowest free energy (ΔG ; kcal/mole) at each temperature is indicated.

6. ΔG of the predicted RNATs at 5' UTR of *caf1R* relative to transcript from P_R^K.

RNAT_PRK_26C_5 codons $\Delta G = -30.72$			RNAT_PRK_28C_5 codons $\Delta G = -29.20$		
Structural element	δG	Information	Structural element	δG	Information
External loop	-3.63	10 ss bases & 3 closing helices.	External loop	-3.56	10 ss bases & 3 closing helices.
Stack	-2.11	External closing pair is C ₈₀ -G ₉₅	Stack	-2.05	External closing pair is C ₈₀ -G ₉₅
Stack	-1.07	External closing pair is A ₈₁ -U ₉₄	Stack	-1.04	External closing pair is A ₈₁ -U ₉₄
Helix	-3.18	3 base pairs.	Helix	-3.09	3 base pairs.
Bulge loop	1.07	External closing pair is U ₈₂ -A ₉₃	Bulge loop	1.17	External closing pair is U ₈₂ -A ₉₃
Stack	-1.91	External closing pair is C ₈₄ -G ₉₂	Stack	-1.87	External closing pair is C ₈₄ -G ₉₂
Helix	-1.91	2 base pairs.	Helix	-1.87	2 base pairs.
Hairpin loop	3.24	Closing pair is U ₈₅ -A ₉₁	Hairpin loop	3.31	Closing pair is U ₈₅ -A ₉₁
Stack	-0.51	External closing pair is G ₃₇ -U ₇₄	Stack	-0.49	External closing pair is G ₃₇ -U ₇₄
Stack	-2.37	External closing pair is G ₃₈ -U ₇₃	Stack	-2.32	External closing pair is G ₃₈ -U ₇₃
Stack	-3.23	External closing pair is C ₃₉ -G ₇₂	Stack	-3.17	External closing pair is C ₃₉ -G ₇₂
Stack	-3.23	External closing pair is C ₄₀ -G ₇₁	Stack	-3.17	External closing pair is C ₄₀ -G ₇₁
Stack	-2.11	External closing pair is C ₄₁ -G ₇₀	Stack	-2.05	External closing pair is C ₄₁ -G ₇₀
Helix	-11.45	6 base pairs.	Helix	-11.20	6 base pairs.
Interior loop	2.18	External closing pair is A ₄₂ -U ₆₉	Interior loop	2.22	External closing pair is A ₄₂ -U ₆₉
Stack	-1.07	External closing pair is A ₄₅ -U ₆₇	Stack	-1.04	External closing pair is A ₄₅ -U ₆₇
Stack	-2.69	External closing pair is U ₄₆ -A ₆₆	Stack	-2.62	External closing pair is U ₄₆ -A ₆₆
Helix	-3.76	3 base pairs.	Helix	-3.66	3 base pairs.
Bulge loop	1.65	External closing pair is C ₄₇ -G ₆₅	Bulge loop	1.74	External closing pair is C ₄₇ -G ₆₅
Stack	-0.61	External closing pair is A ₄₉ -U ₆₄	Stack	-0.59	External closing pair is A ₄₉ -U ₆₄
Stack	-1.19	External closing pair is G ₅₀ -U ₆₃	Stack	-1.17	External closing pair is G ₅₀ -U ₆₃
Stack	-1.10	External closing pair is A ₅₁ -U ₆₂	Stack	-1.07	External closing pair is A ₅₁ -U ₆₂
Stack	-1.07	External closing pair is A ₅₂ -U ₆₁	Stack	-1.04	External closing pair is A ₅₂ -U ₆₁
Stack	-0.91	External closing pair is U ₅₃ -A ₆₀	Stack	-0.89	External closing pair is U ₅₃ -A ₆₀
Helix	-4.88	6 base pairs.	Helix	-4.76	6 base pairs.
Hairpin loop	3.97	Closing pair is G ₅₄ -U ₅₉	Hairpin loop	4.01	Closing pair is G ₅₄ -U ₅₉
Stack	-3.78	External closing pair is G ₃ -C ₃₃	Stack	-3.71	External closing pair is G ₃ -C ₃₃
Stack	-2.28	External closing pair is C ₄ -G ₃₂	Stack	-2.21	External closing pair is C ₄ -G ₃₂
Stack	-1.22	External closing pair is U ₅ -G ₃₁	Stack	-1.22	External closing pair is U ₅ -G ₃₁
Stack	-3.23	External closing pair is G ₆ -C ₃₀	Stack	-3.17	External closing pair is G ₆ -C ₃₀
Stack	-3.23	External closing pair is G ₇ -C ₂₉	Stack	-3.17	External closing pair is G ₇ -C ₂₉
Helix	-13.74	6 base pairs.	Helix	-13.48	6 base pairs.
Interior loop	1.04	External closing pair is G ₈ -C ₂₈	Interior loop	1.14	External closing pair is G ₈ -C ₂₈
Stack	-3.78	External closing pair is G ₁₂ -C ₂₂	Stack	-3.71	External closing pair is G ₁₂ -C ₂₂
Helix	-3.78	2 base pairs.	Helix	-3.71	2 base pairs.
Hairpin loop	2.46	Closing pair is C ₁₃ -G ₂₁	Hairpin loop	2.54	Closing pair is C ₁₃ -G ₂₁

i)

ii)

RNAT_PRK_30C_5 codons $\Delta G = -27.74$			RNAT_PRK_37C_5 codons $\Delta G = -23.30$		
Structural element	δG	Information	Structural element	δG	Information
External loop	-2.30	8 ss bases & 3 closing helices.	External loop	-2.10	24 ss bases & 2 closing helices.
Stack	-2.00	External closing pair is C ₈₀ -G ₉₅	Stack	-0.40	External closing pair is G ₃₇ -U ₇₄
Stack	-1.01	External closing pair is A ₈₁ -U ₉₄	Stack	-2.10	External closing pair is G ₃₈ -U ₇₃
Helix	-3.01	3 base pairs.	Stack	-2.90	External closing pair is C ₃₉ -G ₇₂
Bulge loop	1.26	External closing pair is U ₈₂ -A ₉₃	Stack	-2.90	External closing pair is C ₄₀ -G ₇₁
Stack	-1.83	External closing pair is C ₈₄ -G ₉₂	Stack	-1.80	External closing pair is C ₄₁ -G ₇₀
Helix	-1.83	2 base pairs.	Helix	-10.10	6 base pairs.
Hairpin loop	3.38	Closing pair is U ₈₅ -A ₉₁	Interior loop	2.40	External closing pair is A ₄₂ -U ₆₉
Stack	-0.47	External closing pair is G ₃₇ -U ₇₄	Stack	-0.90	External closing pair is A ₄₅ -U ₆₇
Stack	-2.27	External closing pair is G ₃₈ -U ₇₃	Stack	-2.30	External closing pair is U ₄₆ -A ₆₆
Stack	-3.11	External closing pair is C ₃₉ -G ₇₂	Helix	-3.20	3 base pairs.
Stack	-3.11	External closing pair is C ₄₀ -G ₇₁	Bulge loop	2.10	External closing pair is C ₄₇ -G ₆₅
Stack	-2.00	External closing pair is C ₄₁ -G ₇₀	Stack	-0.50	External closing pair is A ₄₉ -U ₆₄
Helix	-10.96	6 base pairs.	Stack	-1.10	External closing pair is G ₅₀ -U ₆₃
Interior loop	2.26	External closing pair is A ₄₂ -U ₆₉	Stack	-0.90	External closing pair is A ₅₁ -U ₆₂
Stack	-1.01	External closing pair is A ₄₅ -U ₆₇	Stack	-0.90	External closing pair is A ₅₂ -U ₆₁
Stack	-2.55	External closing pair is U ₄₆ -A ₆₆	Stack	-0.80	External closing pair is U ₅₃ -A ₆₀
Helix	-3.56	3 base pairs.	Helix	-4.20	6 base pairs.
Bulge loop	1.81	External closing pair is C ₄₇ -G ₆₅	Hairpin loop	4.20	Closing pair is G ₅₄ -U ₅₉
Stack	-0.57	External closing pair is A ₄₉ -U ₆₄	Stack	-1.20	External closing pair is U ₂ -G ₃₄
Stack	-1.16	External closing pair is G ₅₀ -U ₆₃	Stack	-3.40	External closing pair is G ₃ -C ₃₃
Stack	-1.03	External closing pair is A ₅₁ -U ₆₂	Stack	-1.90	External closing pair is C ₄ -G ₃₂
Stack	-1.01	External closing pair is A ₅₂ -U ₆₁	Stack	-1.20	External closing pair is U ₅ -G ₃₁
Stack	-0.87	External closing pair is U ₅₃ -A ₆₀	Stack	-2.90	External closing pair is G ₆ -C ₃₀
Hairpin loop	4.05	Closing pair is G ₅₄ -U ₅₉	Stack	-2.90	External closing pair is G ₇ -C ₂₉
Stack	-1.21	External closing pair is U ₂ -G ₃₄	Helix	-13.50	7 base pairs.
Stack	-3.64	External closing pair is G ₃ -C ₃₃	Interior loop	1.60	External closing pair is G ₈ -C ₂₈
Stack	-2.14	External closing pair is C ₄ -G ₃₂	Stack	-3.40	External closing pair is G ₁₂ -C ₂₂
Stack	-1.21	External closing pair is U ₅ -G ₃₁	Helix	-3.40	2 base pairs.
Stack	-3.11	External closing pair is G ₆ -C ₃₀	Hairpin loop	2.90	Closing pair is C ₁₃ -G ₂₁
Stack	-3.11	External closing pair is G ₇ -C ₂₉			
Helix	-14.42	7 base pairs.			
Interior loop	1.23	External closing pair is G ₈ -C ₂₈			
Stack	-3.64	External closing pair is G ₁₂ -C ₂₂			
Helix	-3.64	2 base pairs.			
Hairpin loop	2.63	Closing pair is C ₁₃ -G ₂₁			

iv)

iii)

Figure A.6 | ΔG of the predicted RNATs at 5' UTR of *caf1R* relative to transcript from P_R^K. Using Mfold 2.3 program (Zuker, 2003) RNA secondary structures were predicted by inclusion of Caf1R first 5 codons at 26°C (i), 28°C (ii), 30°C (iii) and 37°C (iv). The most probable prediction with lowest free energy (ΔG ; kcal/mole) at each temperature is indicated.

7. ΔG of the predicted RNATs at 5' UTR of *caf1M* relative to transcript from P_M .

RNAT_PM_26C_5 codons
 $\Delta G = -47.38$

Structural element	δG	Information
External loop	-3.61	21 ss bases & 3 closing helices.
Stack	-2.69	External closing pair is U ₃₄ -A ₁₃₃
Stack	-2.11	External closing pair is C ₃₅ -G ₁₃₂
Stack	-1.07	External closing pair is A ₃₆ -U ₁₃₁
Helix	-5.87	4 base pairs.
Interior loop	1.00	External closing pair is U ₃₇ -A ₁₃₀
Stack	-2.37	External closing pair is G ₄₀ -C ₁₂₇
Stack	-0.61	External closing pair is U ₄₁ -G ₁₂₆
Helix	-2.98	3 base pairs.
Multi-loop	2.35	External closing pair is U ₄₂ -A ₁₂₅ 8 ss bases & 4 closing helices.
Stack	-0.61	External closing pair is U ₁₀₉ -G ₁₂₀
Stack	-2.69	External closing pair is U ₁₁₀ -A ₁₁₉
Stack	-3.23	External closing pair is C ₁₁₁ -G ₁₁₈
Helix	-6.53	4 base pairs.
Hairpin loop	0.30	Closing pair is C ₁₁₂ -G ₁₁₇
Stack	-3.23	External closing pair is G ₉₆ -C ₁₀₇
Stack	-2.39	External closing pair is G ₉₇ -C ₁₀₆
Stack	-2.11	External closing pair is U ₉₈ -A ₁₀₅
Helix	-7.73	4 base pairs.
Hairpin loop	3.85	Closing pair is G ₉₉ -C ₁₀₄
Stack	-3.23	External closing pair is C ₄₄ -G ₉₃
Stack	-3.23	External closing pair is C ₄₅ -G ₉₂
Stack	-2.11	External closing pair is C ₄₆ -G ₉₁
Stack	-2.39	External closing pair is A ₄₇ -U ₉₀
Stack	-3.23	External closing pair is C ₄₈ -G ₈₉
Stack	-2.28	External closing pair is C ₄₉ -G ₈₈
Stack	-0.51	External closing pair is U ₅₀ -G ₈₇
Stack	-0.51	External closing pair is U ₅₁ -G ₈₆
Stack	-0.61	External closing pair is U ₅₂ -G ₈₅
Stack	-1.19	External closing pair is U ₅₃ -A ₈₄
Helix	-19.29	11 base pairs.
Interior loop	-0.11	External closing pair is U ₅₄ -G ₈₃
Stack	-1.22	External closing pair is C ₅₆ -G ₈₀
Helix	-1.22	2 base pairs.
Bulge loop	1.39	External closing pair is G ₅₇ -U ₇₉
Stack	-2.11	External closing pair is C ₅₈ -G ₇₇
Stack	-1.07	External closing pair is A ₅₉ -U ₇₆
Helix	-3.18	3 base pairs.
Interior loop	0.77	External closing pair is U ₆₀ -A ₇₅
Stack	-2.69	External closing pair is U ₆₂ -A ₇₃
Stack	-1.22	External closing pair is C ₆₃ -G ₇₂
Stack	-1.11	External closing pair is G ₆₄ -U ₇₁
Helix	-5.02	4 base pairs.
Hairpin loop	4.18	Closing pair is U ₆₅ -A ₇₀
Stack	-3.23	External closing pair is G ₁₇ -C ₃₀
Stack	-2.69	External closing pair is G ₁₈ -C ₂₉
Stack	-1.07	External closing pair is A ₁₉ -U ₂₈
Helix	-6.99	4 base pairs.
Hairpin loop	3.42	Closing pair is U ₂₀ -A ₂₇
Stack	-2.28	External closing pair is G ₆ -U ₁₄
Stack	-3.78	External closing pair is G ₇ -C ₁₃
Helix	-6.06	3 base pairs.
Hairpin loop	3.95	Closing pair is C ₈ -G ₁₂

i)

RNAT_PM_28C_5 codons
 $\Delta G = -45.14$

Structural element	δG	Information
External loop	-3.56	21 ss bases & 3 closing helices.
Stack	-2.62	External closing pair is U ₃₄ -A ₁₃₃
Stack	-2.05	External closing pair is C ₃₅ -G ₁₃₂
Stack	-1.04	External closing pair is A ₃₆ -U ₁₃₁
Helix	-5.71	4 base pairs.
Interior loop	1.07	External closing pair is U ₃₇ -A ₁₃₀
Stack	-2.32	External closing pair is G ₄₀ -C ₁₂₇
Stack	-0.59	External closing pair is U ₄₁ -G ₁₂₆
Helix	-2.91	3 base pairs.
Multi-loop	2.53	External closing pair is U ₄₂ -A ₁₂₅ 8 ss bases & 4 closing helices.
Stack	-0.59	External closing pair is U ₁₀₉ -G ₁₂₀
Stack	-2.62	External closing pair is U ₁₁₀ -A ₁₁₉
Stack	-3.17	External closing pair is C ₁₁₁ -G ₁₁₈
Helix	-6.38	4 base pairs.
Hairpin loop	0.39	Closing pair is C ₁₁₂ -G ₁₁₇
Stack	-3.17	External closing pair is G ₉₆ -C ₁₀₇
Stack	-2.34	External closing pair is G ₉₇ -C ₁₀₆
Stack	-2.05	External closing pair is U ₉₈ -A ₁₀₅
Helix	-7.56	4 base pairs.
Hairpin loop	3.91	Closing pair is G ₉₉ -C ₁₀₄
Stack	-3.17	External closing pair is C ₄₄ -G ₉₃
Stack	-3.17	External closing pair is C ₄₅ -G ₉₂
Stack	-2.05	External closing pair is C ₄₆ -G ₉₁
Stack	-2.34	External closing pair is A ₄₇ -U ₉₀
Stack	-3.17	External closing pair is C ₄₈ -G ₈₉
Stack	-2.21	External closing pair is C ₄₉ -G ₈₈
Stack	-0.49	External closing pair is U ₅₀ -G ₈₇
Stack	-0.49	External closing pair is U ₅₁ -G ₈₆
Stack	-0.59	External closing pair is U ₅₂ -G ₈₅
Stack	-1.17	External closing pair is U ₅₃ -A ₈₄

ii)

Helix	-18.85	11 base pairs.
Interior loop	0.00	External closing pair is U ₅₄ -G ₈₃
Stack	-1.22	External closing pair is C ₅₆ -G ₈₀
Helix	-1.22	2 base pairs.
Bulge loop	1.47	External closing pair is G ₅₇ -U ₇₉
Stack	-2.05	External closing pair is C ₅₈ -G ₇₇
Stack	-1.04	External closing pair is A ₅₉ -U ₇₆
Helix	-3.09	3 base pairs.
Interior loop	0.78	External closing pair is U ₆₀ -A ₇₅
Stack	-2.62	External closing pair is U ₆₂ -A ₇₃
Stack	-1.22	External closing pair is C ₆₃ -G ₇₂
Stack	-1.09	External closing pair is G ₆₄ -U ₇₁
Helix	-4.93	4 base pairs.
Hairpin loop	4.22	Closing pair is U ₆₅ -A ₇₀
Stack	-3.17	External closing pair is G ₁₇ -C ₃₀
Stack	-2.62	External closing pair is G ₁₈ -C ₂₉
Stack	-1.04	External closing pair is A ₁₉ -U ₂₈
Helix	-6.83	4 base pairs.
Hairpin loop	3.47	Closing pair is U ₂₀ -A ₂₇
Stack	-2.21	External closing pair is G ₆ -U ₁₄
Stack	-3.71	External closing pair is G ₇ -C ₁₃
Helix	-5.92	3 base pairs.
Hairpin loop	3.98	Closing pair is C ₈ -G ₁₂

RNAT_PM_30C_5 codons
 $\Delta G = -42.91$

Structural element	δG	Information
External loop	-3.50	21 ss bases & 3 closing helices.
Stack	-2.55	External closing pair is U ₃₄ -A ₁₃₃
Stack	-2.00	External closing pair is C ₃₅ -G ₁₃₂
Stack	-1.01	External closing pair is A ₃₆ -U ₁₃₁
Helix	-5.56	4 base pairs.
Interior loop	1.14	External closing pair is U ₃₇ -A ₁₃₀
Stack	-2.27	External closing pair is G ₄₀ -C ₁₂₇
Stack	-0.57	External closing pair is U ₄₁ -G ₁₂₆
Helix	-2.84	3 base pairs.
Multi-loop	2.71	External closing pair is U ₄₂ -A ₁₂₅ 8 ss bases & 4 closing helices.
Stack	-0.57	External closing pair is U ₁₀₉ -G ₁₂₀
Stack	-2.55	External closing pair is U ₁₁₀ -A ₁₁₉
Stack	-3.11	External closing pair is C ₁₁₁ -G ₁₁₈
Helix	-6.23	4 base pairs.
Hairpin loop	0.48	Closing pair is C ₁₁₂ -G ₁₁₇
Stack	-3.11	External closing pair is G ₉₆ -C ₁₀₇
Stack	-2.28	External closing pair is G ₉₇ -C ₁₀₆
Stack	-2.00	External closing pair is U ₉₈ -A ₁₀₅
Helix	-7.39	4 base pairs.
Hairpin loop	3.98	Closing pair is G ₉₉ -C ₁₀₄
Stack	-3.11	External closing pair is C ₄₄ -G ₉₃
Stack	-3.11	External closing pair is C ₄₅ -G ₉₂
Stack	-2.00	External closing pair is C ₄₆ -G ₉₁
Stack	-2.28	External closing pair is A ₄₇ -U ₉₀
Stack	-3.11	External closing pair is C ₄₈ -G ₈₉
Stack	-2.14	External closing pair is C ₄₉ -G ₈₈
Stack	-0.47	External closing pair is U ₅₀ -G ₈₇
Stack	-0.47	External closing pair is U ₅₁ -G ₈₆
Stack	-0.57	External closing pair is U ₅₂ -G ₈₅
Stack	-1.16	External closing pair is U ₅₃ -A ₈₄

iii)

Helix	-18.42	11 base pairs.
Interior loop	0.11	External closing pair is U ₅₄ -G ₈₃
Stack	-1.21	External closing pair is C ₅₆ -G ₈₀
Helix	-1.21	2 base pairs.
Bulge loop	1.54	External closing pair is G ₅₇ -U ₇₉
Stack	-2.00	External closing pair is C ₅₈ -G ₇₇
Stack	-1.01	External closing pair is A ₅₉ -U ₇₆
Helix	-3.01	3 base pairs.
Interior loop	0.78	External closing pair is U ₆₀ -A ₇₅
Stack	-2.55	External closing pair is U ₆₂ -A ₇₃
Stack	-1.21	External closing pair is C ₆₃ -G ₇₂
Stack	-1.07	External closing pair is G ₆₄ -U ₇₁
Helix	-4.83	4 base pairs.
Hairpin loop	4.26	Closing pair is U ₆₅ -A ₇₀
Stack	-3.11	External closing pair is G ₁₇ -C ₃₀
Stack	-2.55	External closing pair is G ₁₈ -C ₂₉
Stack	-1.01	External closing pair is A ₁₉ -U ₂₈
Helix	-6.67	4 base pairs.
Hairpin loop	3.52	Closing pair is U ₂₀ -A ₂₇
Stack	-2.14	External closing pair is G ₆ -U ₁₄
Stack	-3.64	External closing pair is G ₇ -C ₁₃
Helix	-5.78	3 base pairs.
Hairpin loop	4.01	Closing pair is C ₈ -G ₁₂

RNAT_PM_37C_5 codons $\Delta G = -35.70$					
Structural element	δG	Information	Structural element	δG	Information
External loop	-6.80	47 ss bases & 5 closing helices.	Helix	-1.20	2 base pairs.
Stack	-0.50	External closing pair is U ₁₀₉ -G ₁₂₀	Bulge loop	1.80	External closing pair is G ₅₇ -U ₇₉
Stack	-2.30	External closing pair is U ₁₁₀ -A ₁₁₉	Stack	-1.80	External closing pair is C ₅₈ -G ₇₇
Stack	-2.90	External closing pair is C ₁₁₁ -G ₁₁₈	Stack	-0.90	External closing pair is A ₅₉ -U ₇₆
Helix	-5.70	4 base pairs.	Helix	-2.70	3 base pairs.
Hairpin loop	0.80	Closing pair is C ₁₁₂ -G ₁₁₇	Interior loop	0.80	External closing pair is U ₆₀ -A ₇₅
Stack	-2.90	External closing pair is G ₉₆ -C ₁₀₇	Stack	-2.30	External closing pair is U ₆₂ -A ₇₃
Stack	-2.10	External closing pair is G ₉₇ -C ₁₀₆	Stack	-1.20	External closing pair is C ₆₃ -G ₇₂
Stack	-1.80	External closing pair is U ₉₈ -A ₁₀₅	Stack	-1.00	External closing pair is G ₆₄ -U ₇₁
Helix	-6.80	4 base pairs.	Helix	-4.50	4 base pairs.
Hairpin loop	4.20	Closing pair is G ₉₉ -C ₁₀₄	Hairpin loop	4.40	Closing pair is U ₆₅ -A ₇₀
Stack	-2.90	External closing pair is C ₄₄ -G ₉₃	Stack	-2.90	External closing pair is G ₁₇ -C ₃₀
Stack	-2.90	External closing pair is C ₄₅ -G ₉₂	Stack	-2.30	External closing pair is G ₁₈ -C ₂₉
Stack	-1.80	External closing pair is C ₄₆ -G ₉₁	Stack	-0.90	External closing pair is A ₁₉ -U ₂₈
Stack	-2.10	External closing pair is A ₄₇ -U ₉₀	Helix	-6.10	4 base pairs.
Stack	-2.90	External closing pair is C ₄₈ -G ₈₉	Hairpin loop	3.70	Closing pair is U ₂₀ -A ₂₇
Stack	-1.90	External closing pair is C ₄₉ -G ₈₈	Stack	-1.90	External closing pair is G ₆ -U ₁₄
Stack	-0.40	External closing pair is U ₅₀ -G ₈₇	Stack	-3.40	External closing pair is G ₇ -C ₁₃
Stack	-0.40	External closing pair is U ₅₁ -G ₈₆	Helix	-5.30	3 base pairs.
Stack	-0.50	External closing pair is U ₅₂ -G ₈₅	Hairpin loop	4.10	Closing pair is C ₈ -G ₁₂
Stack	-1.10	External closing pair is U ₅₃ -A ₈₄			
Helix	-16.90	11 base pairs.			
Interior loop	0.50	External closing pair is U ₅₄ -G ₈₃			
Stack	-1.20	External closing pair is C ₅₆ -G ₈₀			

Figure A.7 | ΔG of the predicted RNATs at 5' UTR of *caf1M* relative to transcript from P_M. Using Mfold 2.3 program (Zuker, 2003) RNA secondary structures were predicted by inclusion of Caf1M first 5 codons at 26°C (i), 28°C (ii), 30°C (iii) and 37°C (iv). The most probable prediction with lowest free energy (ΔG ; kcal/mole) at each temperature is indicated.

8. Comparison of Caf1R-DBD mutations with Ala scanning of MarA

pACYCF1-R mutant		F1 recovery (%) at 37°C				MarA mutant	Transcriptional activity at promoter					Role	Location
		4 h		6 h			Class I			Class II			
							<i>mar</i>	<i>fpr</i>	<i>zwf</i>	<i>fumC</i>	<i>micF</i>		
WT		100 ± 1.32		100 ± 1.57									
WT		100 ± 5.33		100 ± 3.43									
I31A		33.92 ± 1.31		73.22 ± 1.16		L30A							Helix-2
D32A		61.25 ± 4.40		86.76 ± 1.83		E31A						P	Helix-2
C33A		59.50 ± 7.84		96.20 ± 9.44		K32A							Helix-2
F40A		4.18 ± 0.33		5.53 ± 0.06		Y39A						P	Loop 2-3
R42A		143.87 ± 3.51		85.19 ± 1.77		K41A							Helix-3
R43A		79.48 ± 4.02		66.35 ± 0.66		W42A						V (C,C)	Helix-3
Y44A		72.06 ± 1.39		76.44 ± 2.34		H43A						P	Helix-3
R62A		76.33 ± 3.66		96.55 ± 1.27		R61A							Helix-4
R62S		58.85 ± 0.20		92.97 ± 2.94									
Q93A		6.99 ± 0.36		33.90 ± 1.77		Q92A						V (T, T)	Helix-6
R97A		4.79 ± 1.83		86.53 ± 4.78		R96A						H (G,G,A), V (T)	Helix-6
E98G		13.05 ± 0.62		14.16 ± 0.32		T97A						P	Helix-6
E98A		58.85 ± 2.15		95.24 ± 3.24									
E98T		67.38 ± 1.57		73.12 ± 2.44									
E98K		00 ± 00		00 ± 00									
T106A		125.12 ± 1.48		112.73 ± 1.84		P105A							Loop 6-7
R108A		57.22 ± 2.17		49.05 ± 0.63		H107A						P	Helix-7
Q109A		121.25 ± 5.73		112.74 ± 5.94		K108A						P	Helix-7
WT or >100%	96-100%	76-95%	46-75%	26-45%	0-25%		81-100%	61-80%	41-60%	21-40%	0-20%		

Figure A.8| **Site-specific mutagenesis of Caf1R-DBD with respect to MarA.**

F1 recovery from Caf1R-DBD mutants (in pACYCF1) and relatedness of mutagenesis targeted Caf1R-DBD residues with *mar*-MarA complex (Rhee et al., 1998) and alanine-scanning results of MarA (Gillette et al., 2000) is indicated. Interaction with DNA backbone (P), base specific hydrogen bond (H) and Van der Waals interactions (V) are indicated with corresponding *mar* DNA base in parentheses, (). Dark and light brown-shaded mutants indicate, F1 recovery from whole cell and cell surface with respective wild type (WT).

References

- Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., and Carniel, E. (1999). *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* *96*, 14043-14048.
- Alekshun, M.N., Levy, S.B., Mealy, T.R., Seaton, B.A., and Head, J.F. (2001). The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. *Nat Struct Biol* *8*, 710-714.
- Amabile-Cuevas, C.F., and Demple, B. (1991). Molecular characterization of the soxRS genes of *Escherichia coli*: two genes control a superoxide stress regulon. *Nucleic acids research* *19*, 4479-4484.
- Amies, C.R. (1951). The envelope substance of *Pasteurella pestis*. *Br J Exp Pathol* *32*, 259-273.
- Amit, R., Oppenheim, A.B., and Stavans, J. (2003). Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophys J* *84*, 2467-2473.
- Andersen, K.R., Leksa, N.C., and Schwartz, T.U. (2013). Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* *81*, 1857-1861.
- Andrews, G.P., Heath, D.G., Anderson, G.W., Jr., Welkos, S.L., and Friedlander, A.M. (1996). Fraction 1 capsular antigen (F1) purification from *Yersinia pestis* CO92 and from an *Escherichia coli* recombinant strain and efficacy against lethal plague challenge. *Infection and immunity* *64*, 2180-2187.
- Ariza, R.R., Li, Z., Ringstad, N., and Demple, B. (1995). Activation of multiple antibiotic resistance and binding of stress-inducible promoters by *Escherichia coli* Rob protein. *Journal of bacteriology* *177*, 1655-1661.
- Associated Press in Antananarivo, M. (2013). Bubonic plague outbreak kills 32 in Madagascar. *The Guardian, United Kingdom*. <http://www.theguardian.com/world/2013/dec/20/bubonic-plague-outbreak-deaths-madagascar>.
- Azam, T.A., Hiraga, S., and Ishihama, A. (2000). Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells* *5*, 613-626.
- Baga, M., Goransson, M., Normark, S., and Uhlin, B.E. (1985). Transcriptional activation of a pap pilus virulence operon from uropathogenic *Escherichia coli*. *The EMBO journal* *4*, 3887-3893.
- Baldwin, R.L. (1986). Temperature dependence of the hydrophobic interaction in protein folding. *Proceedings of the National Academy of Sciences of the United States of America* *83*, 8069-8072.
- Barbosa, T.M., and Levy, S.B. (2000). Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. *Journal of bacteriology* *182*, 3467-3474.
- Bartra, S.S., Styer, K.L., O'Bryant, D.M., Nilles, M.L., Hinnebusch, B.J., Aballay, A., and Plano, G.V. (2008). Resistance of *Yersinia pestis* to complement-dependent killing is mediated by the Ail outer membrane protein. *Infection and immunity* *76*, 612-622.
- Basturea, G.N., Boderio, M.D., Moreno, M.E., and Munson, G.P. (2008). Residues near the amino terminus of Rns are essential for positive autoregulation and DNA binding. *Journal of bacteriology* *190*, 2279-2285.
- BBC, U.K. (2013). Madagascar bubonic plague warning. <http://www.bbc.co.uk/news/world-africa-24461474>.
- Benedek, O., Nagy, G., and Emody, L. (2004). Intracellular signalling and cytoskeletal rearrangement involved in *Yersinia pestis* plasminogen activator (Pla) mediated HeLa cell invasion. *Microb Pathog* *37*, 47-54.
- Bi, Y., Du, Z., Han, Y., Guo, Z., Tan, Y., Zhu, Z., and Yang, R. (2009). *Yersinia pestis* and host macrophages: immunodeficiency of mouse macrophages induced by YscW. *Immunology* *128*, e406-417.

- Bianucci, R., Rahalison, L., Ferroglia, E., Massa, E.R., and Signoli, M. (2007). A rapid diagnostic test for plague detects *Yersinia pestis* F1 antigen in ancient human remains. *C R Biol* 330, 747-754.
- Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453-1462.
- Blomfield, I.C. (2001). The regulation of pap and type 1 fimbriation in *Escherichia coli*. *Adv Microb Physiol* 45, 1-49.
- Bodero, M.D., Harden, E.A., and Munson, G.P. (2008). Transcriptional regulation of subclass 5b fimbriae. *BMC Microbiol* 8, 180.
- Bohme, K., Steinmann, R., Kortmann, J., Seekircher, S., Heroven, A.K., Berger, E., Pisano, F., Thiermann, T., Wolf-Watz, H., Narberhaus, F., *et al.* (2012). Concerted actions of a thermo-labile regulator and a unique intergenic RNA thermosensor control *Yersinia* virulence. *PLoS Pathog* 8, e1002518.
- Bondos, S.E., and Bicknell, A. (2003). Detection and prevention of protein aggregation before, during, and after purification. *Analytical biochemistry* 316, 223-231.
- Botelho, G. (2015). Second case of plague-another Yosemite visitor-investigated in California (California, USA: CNN).
- Braaten, B.A., Nou, X., Kaltenbach, L.S., and Low, D.A. (1994). Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*. *Cell* 76, 577-588.
- Brinkman, A.B., Ettema, T.J., de Vos, W.M., and van der Oost, J. (2003). The Lrp family of transcriptional regulators. *Molecular microbiology* 48, 287-294.
- Brown, N.L., Stoyanov, J.V., Kidd, S.P., and Hobman, J.L. (2003). The MerR family of transcriptional regulators. *FEMS microbiology reviews* 27, 145-163.
- Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. *Nat Rev Microbiol* 2, 57-65.
- Burns, L.S., Smith, S.G., and Dorman, C.J. (2000). Interaction of the FimB integrase with the fimS invertible DNA element in *Escherichia coli* in vivo and in vitro. *Journal of bacteriology* 182, 2953-2959.
- Busby, S., and Ebright, R.H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79, 743-746.
- Busch, A., and Waksman, G. (2012). Chaperone-usher pathways: diversity and pilus assembly mechanism. *Philos Trans R Soc Lond B Biol Sci* 367, 1112-1122.
- Bustos, S.A., and Schleif, R.F. (1993). Functional domains of the AraC protein. *Proceedings of the National Academy of Sciences of the United States of America* 90, 5638-5642.
- Butler, T. (2009). Plague into the 21st century. *Clin Infect Dis* 49, 736-742.
- Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Molecular cell* 9, 527-539.
- Cantey, J.R., Blake, R.K., Williford, J.R., and Moseley, S.L. (1999). Characterization of the *Escherichia coli* AF/R1 pilus operon: novel genes necessary for transcriptional regulation and for pilus-mediated adherence. *Infection and immunity* 67, 2292-2298.
- Cao, L., Lim, T., Jun, S., Thornburg, T., Avci, R., and Yang, X. (2012). Vulnerabilities in *Yersinia pestis* caf operon are unveiled by a *Salmonella* vector. *PLoS One* 7, e36283.

- Carolina, A., and Celso, C. (2012). Electrophoretic Mobility Shift Assay: Analysing Protein - Nucleic Acid Interactions, Gel Electrophoresis - Advanced Techniques, Dr. Samesh Magdeldin (Edn), ISBN: 978-953-51-0457-5, InTech, Available from: <http://cdn.intechopen.com/pdfs-wm/34920.pdf>
- Caron, J., Coffield, L.M., and Scott, J.R. (1989). A plasmid-encoded regulatory gene, *rns*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* *86*, 963-967.
- Caron, J., and Scott, J.R. (1990). A *rns*-like regulatory gene for colonization factor antigen I (CFA/I) that controls expression of CFA/I pilin. *Infection and immunity* *58*, 874-878.
- Cathelyn, J.S., Crosby, S.D., Lathem, W.W., Goldman, W.E., and Miller, V.L. (2006). *RovA*, a global regulator of *Yersinia pestis*, specifically required for bubonic plague. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 13514-13519.
- CDC (2012). Ecology and Transmission of Plague. <http://www.cdc.gov/plague/transmission/>.
- CDC (2015). Plague: Resources for Clinicians. <http://www.cdc.gov/plague/healthcare/clinicians.html>.
- Chanteau, S., Rahalison, L., Ratsitorahina, M., Mahafaly, Rasolomaharo, M., Boisier, P., O'Brien, T., Aldrich, J., Keleher, A., Morgan, C., *et al.* (2000). Early diagnosis of bubonic plague using F1 antigen capture ELISA assay and rapid immunogold dipstick. *Int J Med Microbiol* *290*, 279-283.
- Chapman, D.A., Zavalov, A.V., Chernovskaya, T.V., Karlyshev, A.V., Zav'yalova, G.A., Vasiliev, A.M., Dudich, I.V., Abramov, V.M., Zav'yalov, V.P., and MacIntyre, S. (1999). Structural and functional significance of the FGL sequence of the periplasmic chaperone Caf1M of *Yersinia pestis*. *Journal of bacteriology* *181*, 2422-2429.
- Chauvaux, S., Rosso, M.L., Frangeul, L., Lacroix, C., Labarre, L., Schiavo, A., Marceau, M., Dillies, M.A., Foulon, J., Coppee, J.Y., *et al.* (2007). Transcriptome analysis of *Yersinia pestis* in human plasma: an approach for discovering bacterial genes involved in septicemic plague. *Microbiology* *153*, 3112-3124.
- Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S.J., and Knight, S.D. (1999). X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* *285*, 1061-1066.
- Chouikha, I., and Hinnebusch, B.J. (2012). *Yersinia*-flea interactions and the evolution of the arthropod-borne transmission route of plague. *Curr Opin Microbiol* *15*, 239-246.
- Chubiz, L.M., Glekas, G.D., and Rao, C.V. (2012). Transcriptional cross talk within the *mar-sox-rob* regulon in *Escherichia coli* is limited to the *rob* and *marRAB* operons. *Journal of bacteriology* *194*, 4867-4875.
- Clegg, S., Wilson, J., and Johnson, J. (2011). More than one way to control hair growth: regulatory mechanisms in enterobacteria that affect fimbriae assembled by the chaperone/usher pathway. *Journal of bacteriology* *193*, 2081-2088.
- Clouthier, S.C., Muller, K.H., Doran, J.L., Collinson, S.K., and Kay, W.W. (1993). Characterization of three fimbrial genes, *sefABC*, of *Salmonella enteritidis*. *Journal of bacteriology* *175*, 2523-2533.
- Cohn, S.K. (2003). *The Black Death Transformed: Disease and Culture in Early Renaissance Europe*, 1 edn (Hodder Arnold publication).
- Cole, S.T., and Buchrieser, C. (2001). Bacterial genomics. A plague o' both your hosts. *Nature* *413*, 467, 469-470.
- Costa, T.R., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., and Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* *13*, 343-359.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome research* *14*, 1188-1190.

- Daber, R., Stayrook, S., Rosenberg, A., and Lewis, M. (2007). Structural analysis of lac repressor bound to allosteric effectors. *Journal of molecular biology* 370, 609-619.
- Dame, R.T., Wyman, C., and Goosen, N. (2000). H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic acids research* 28, 3504-3510.
- Dangi, B., Gronenborn, A.M., Rosner, J.L., and Martin, R.G. (2004). Versatility of the carboxy-terminal domain of the alpha subunit of RNA polymerase in transcriptional activation: use of the DNA contact site as a protein contact site for MarA. *Molecular microbiology* 54, 45-59.
- Dangi, B., Pelupessey, P., Martin, R.G., Rosner, J.L., Louis, J.M., and Gronenborn, A.M. (2001). Structure and dynamics of MarA-DNA complexes: an NMR investigation. *Journal of molecular biology* 314, 113-127.
- Davis, K.J., Fritz, D.L., Pitt, M.L., Welkos, S.L., Worsham, P.L., and Friedlander, A.M. (1996). Pathology of experimental pneumonic plague produced by fraction 1-positive and fraction 1-negative *Yersinia pestis* in African green monkeys (*Cercopithecus aethiops*). *Arch Pathol Lab Med* 120, 156-163.
- De Haan, L.A., Willshaw, G.A., van der Zeijst, B.A., and Gaastra, W. (1991). The nucleotide sequence of a regulatory gene present on a plasmid in an enterotoxigenic *Escherichia coli* strain of serotype O167:H5. *FEMS microbiology letters* 67, 341-346.
- deHaseh, P.L., Zupancic, M.L., and Record, M.T., Jr. (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *Journal of bacteriology* 180, 3019-3025.
- Del Prete, G., Santi, L., Andrianaivoarimanana, V., Amedei, A., Domarle, O., MM, D.E., Arntzen, C.J., Rahalison, L., and Mason, H.S. (2009). Plant-derived recombinant F1, V, and F1-V fusion antigens of *Yersinia pestis* activate human cells of the innate and adaptive immune system. *Int J Immunopathol Pharmacol* 22, 133-143.
- Demple, B. (1996). Redox signaling and gene control in the *Escherichia coli* soxRS oxidative stress regulon--a review. *Gene* 179, 53-57.
- Deng, Z., Liu, Z., He, J., Wang, J., Yan, Y., Wang, X., Cui, Y., Bi, Y., Du, Z., Song, Y., *et al.* (2015). TyrR, the regulator of aromatic amino acid metabolism, is required for mice infection of *Yersinia pestis*. *Front Microbiol* 6, 1-6.
- Derbise, A., Cerda Marin, A., Ave, P., Blisnick, T., Huerre, M., Carniel, E., and Demeure, C.E. (2012). An encapsulated *Yersinia pseudotuberculosis* is a highly efficient vaccine against pneumonic plague. *PLoS Negl Trop Dis* 6, e1528.
- Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* 11, 87-93.
- Di Yu, X., Dubnovitsky, A., Pudney, A.F., Macintyre, S., Knight, S.D., and Zavialov, A.V. (2012). Allosteric mechanism controls traffic in the chaperone/usher pathway. *Structure* 20, 1861-1871.
- Dominguez-Cuevas, P., Marin, P., Busby, S., Ramos, J.L., and Marques, S. (2008). Roles of effectors in XylS-dependent transcription activation: intramolecular domain derepression and DNA binding. *Journal of bacteriology* 190, 3118-3128.
- Dominguez-Cuevas, P., Ramos, J.L., and Marques, S. (2010). Sequential XylS-CTD binding to the Pm promoter induces DNA bending prior to activation. *Journal of bacteriology* 192, 2682-2690.
- Donato, G.M., and Kawula, T.H. (1999). Phenotypic analysis of random hns mutations differentiate DNA-binding activity from properties of fimA promoter inversion modulation and bacterial motility. *Journal of bacteriology* 181, 941-948.
- Donato, G.M., Lelivelt, M.J., and Kawula, T.H. (1997). Promoter-specific repression of fimB expression by the *Escherichia coli* nucleoid-associated protein H-NS. *Journal of bacteriology* 179, 6618-6625.

- Dorman, C.J., and Higgins, C.F. (1987). Fimbrial phase variation in *Escherichia coli*: dependence on integration host factor and homologies with other site-specific recombinases. *Journal of bacteriology* *169*, 3840-3843.
- Dorman, C.J., and Ni Bhriain, N. (1992). Thermal regulation of *fimA*, the *Escherichia coli* gene coding for the type 1 fimbrial subunit protein. *FEMS microbiology letters* *78*, 125-130.
- Dove, S.L., Darst, S.A., and Hochschild, A. (2003). Region 4 of sigma as a target for transcription regulation. *Molecular microbiology* *48*, 863-874.
- Drozhdov, I.G., Anisimov, A.P., Samoilova, S.V., Yezhov, I.N., Yeregin, S.A., Karlyshev, A.V., Krasilnikova, V.M., and Kravchenko, V.I. (1995). Virulent non-capsulate *Yersinia pestis* variants constructed by insertion mutagenesis. *J Med Microbiol* *42*, 264-268.
- Drummelsmith, J., and Whitfield, C. (2000). Translocation of group 1 capsular polysaccharide to the surface of *Escherichia coli* requires a multimeric complex in the outer membrane. *The EMBO journal* *19*, 57-66.
- Du, Y., Rosqvist, R., and Forsberg, A. (2002). Role of fraction 1 antigen of *Yersinia pestis* in inhibition of phagocytosis. *Infection and immunity* *70*, 1453-1460.
- Dubnovitsky, A.P., Duck, Z., Kersley, J.E., Hard, T., MacIntyre, S., and Knight, S.D. (2010). Conserved hydrophobic clusters on the surface of the Caf1A usher C-terminal domain are important for F1 antigen assembly. *Journal of molecular biology* *403*, 243-259.
- Dudley, E.G., Thomson, N.R., Parkhill, J., Morin, N.P., and Nataro, J.P. (2006). Proteomic and microarray characterization of the AggR regulon identifies a *pheU* pathogenicity island in enteroaggregative *Escherichia coli*. *Molecular microbiology* *61*, 1267-1282.
- Duval, V., and Lister, I.M. (2013). MarA, SoxS and Rob of - Global regulators of multidrug resistance, virulence and stress response. *Int J Biotechnol Wellness Ind* *2*, 101-124.
- Ebright, R.H. (1993). Transcription activation at Class I CAP-dependent promoters. *Molecular microbiology* *8*, 797-802.
- Ebright, R.H. (2000). RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *Journal of molecular biology* *304*, 687-698.
- Egan, S.M. (2002). Growing repertoire of AraC/XylS activators. *Journal of bacteriology* *184*, 5529-5532.
- Ehretsmann, C.P., Carpousis, A.J., and Krisch, H.M. (1992). Specificity of *Escherichia coli* endoribonuclease RNase E: in vivo and in vitro analysis of mutants in a bacteriophage T4 mRNA processing site. *Genes Dev* *6*, 149-159.
- Eidam, O., Dworkowski, F.S., Glockshuber, R., Grutter, M.G., and Capitani, G. (2008). Crystal structure of the ternary FimC-FimF(t)-FimD(N) complex indicates conserved pilus chaperone-subunit complex recognition by the usher FimD. *FEBS Lett* *582*, 651-655.
- Eisenstein, B.I., Sweet, D.S., Vaughn, V., and Friedman, D.I. (1987). Integration host factor is required for the DNA inversion that controls phase variation in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* *84*, 6506-6510.
- El-Labany, S., Sohanpal, B.K., Lahooti, M., Akerman, R., and Blomfield, I.C. (2003). Distant cis-active sequences and sialic acid control the expression of *fimB* in *Escherichia coli* K-12. *Molecular microbiology* *49*, 1109-1118.
- Elliot, M.A., Locke, T.R., Galibois, C.M., and Leskiw, B.K. (2003). BldD from *Streptomyces coelicolor* is a non-essential global regulator that binds its own promoter as a dimer. *FEMS microbiology letters* *225*, 35-40.
- Evans, D.G., Silver, R.P., Evans, D.J., Jr., Chase, D.G., and Gorbach, S.L. (1975). Plasmid-controlled colonization factor associated with virulence in *Escherichia coli* enterotoxigenic for humans. *Infection and immunity* *12*, 656-667.

- Falkow, S. (1988). Molecular Koch's postulates applied to microbial pathogenicity. *Rev Infect Dis* 10 *Suppl* 2, S274-276.
- Fawcett, W.P., and Wolf, R.E., Jr. (1994). Purification of a MalE-SoxS fusion protein and identification of the control sites of *Escherichia coli* superoxide-inducible genes. *Molecular microbiology* 14, 669-679.
- Feodorova, V.A., and Corbel, M.J. (2009). Prospects for new plague vaccines. *Expert Rev Vaccines* 8, 1721-1738.
- Feodorova, V.A., and Motin, V.L. (2012). Plague vaccines: current developments and future perspectives. *Emerg Microbes Infect* 1, e36.
- Fernandez, L.A., and Berenguer, J. (2000). Secretion and assembly of regular surface structures in Gram-negative bacteria. *FEMS microbiology reviews* 24, 21-44.
- Filippov, A.A., Solodovnikov, N.S., Kookleva, L.M., and Protsenko, O.A. (1990). Plasmid content in *Yersinia pestis* strains of different origin. *FEMS microbiology letters* 55, 45-48.
- Ford, B., Rego, A.T., Ragan, T.J., Pinkner, J., Dodson, K., Driscoll, P.C., Hultgren, S., and Waksman, G. (2010). Structural homology between the C-terminal domain of the PapC usher and its plug. *Journal of bacteriology* 192, 1824-1831.
- Forman, S., Wulff, C.R., Myers-Morales, T., Cowan, C., Perry, R.D., and Straley, S.C. (2008). yadBC of *Yersinia pestis*, a new virulence determinant for bubonic plague. *Infection and immunity* 76, 578-587.
- Friedlander, A.M., Welkos, S.L., Worsham, P.L., Andrews, G.P., Heath, D.G., Anderson, G.W., Jr., Pitt, M.L., Estep, J., and Davis, K. (1995). Relationship between virulence and immunity as revealed in recent studies of the F1 capsule of *Yersinia pestis*. *Clin Infect Dis* 21 *Suppl* 2, S178-181.
- Froehlich, B., Husmann, L., Caron, J., and Scott, J.R. (1994). Regulation of rns, a positive regulatory factor for pili of enterotoxigenic *Escherichia coli*. *Journal of bacteriology* 176, 5385-5392.
- Fukuto, H.S., Svetlanov, A., Palmer, L.E., Karzai, A.W., and Bliska, J.B. (2010). Global gene expression profiling of *Yersinia pestis* replicating inside macrophages reveals the roles of a putative stress-induced operon in regulating type III secretion and intracellular cell division. *Infection and immunity* 78, 3700-3715.
- Gaastra, W., and Svennerholm, A.M. (1996). Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol* 4, 444-452.
- Galimand, M., Guiyoule, A., Gerbaud, G., Rasoamanana, B., Chanteau, S., Carniel, E., and Courvalin, P. (1997). Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *The New England journal of medicine* 337, 677-680.
- Gallegos, M.T., Marques, S., and Ramos, J.L. (1996). Expression of the TOL plasmid xylS gene in *Pseudomonas putida* occurs from a alpha 70-dependent promoter or from alpha 70- and alpha 54-dependent tandem promoters according to the compound used for growth. *Journal of bacteriology* 178, 2356-2361.
- Gallegos, M.T., Michan, C., and Ramos, J.L. (1993). The XylS/AraC family of regulators. *Nucleic acids research* 21, 807-810.
- Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K., and Ramos, J.L. (1997). Arac/XylS family of transcriptional regulators. *Microbiology and molecular biology reviews* : MMBR 61, 393-410.
- Gally, D.L., Bogan, J.A., Eisenstein, B.I., and Blomfield, I.C. (1993). Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in *Escherichia coli* K-12: effects of temperature and media. *Journal of bacteriology* 175, 6186-6193.
- Galyov, E.E., Karlishev, A.V., Chernovskaya, T.V., Dolgikh, D.A., Smirnov, O., Volkovoy, K.I., Abramov, V.M., and Zav'yalov, V.P. (1991). Expression of the envelope antigen F1 of *Yersinia pestis* is mediated by the

product of caf1M gene having homology with the chaperone protein PapD of Escherichia coli. FEBS Lett 286, 79-82.

Galyov, E.E., Smirnov, O., Karlishchev, A.V., Volkovoy, K.I., Denesyuk, A.I., Nazimov, I.V., Rubtsov, K.S., Abramov, V.M., Dalvadyanz, S.M., and Zav'yalov, V.P. (1990). Nucleotide sequence of the Yersinia pestis gene encoding F1 antigen and the primary structure of the protein. Putative T and B cell epitopes. FEBS Lett 277, 230-232.

Gao, H., Zhou, D., Li, Y., Guo, Z., Han, Y., Song, Y., Zhai, J., Du, Z., Wang, X., Lu, J., *et al.* (2008). The iron-responsive Fur regulon in Yersinia pestis. Journal of bacteriology 190, 3063-3075.

Garvie, C.W., and Wolberger, C. (2001). Recognition of specific DNA sequences. Molecular cell 8, 937-946.

Gillette, W.K., Martin, R.G., and Rosner, J.L. (2000). Probing the Escherichia coli transcriptional activator MarA using alanine-scanning mutagenesis: residues important for DNA binding and activation. Journal of molecular biology 299, 1245-1255.

Goransson, M., Forsman, P., Nilsson, P., and Uhlin, B.E. (1989). Upstream activating sequences that are shared by two divergently transcribed operons mediate cAMP-CRP regulation of pilus-adhesin in Escherichia coli. Molecular microbiology 3, 1557-1565.

Goransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelsson, K., and Uhlin, B.E. (1990). Transcriptional silencing and thermoregulation of gene expression in Escherichia coli. Nature 344, 682-685.

Goransson, M., and Uhlin, B.E. (1984). Environmental temperature regulates transcription of a virulence pili operon in E. coli. The EMBO journal 3, 2885-2888.

Gourse, R.L., Ross, W., and Gaal, T. (2000). UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. Molecular microbiology 37, 687-695.

Grewal, H.M., Valvatne, H., Bhan, M.K., van Dijk, L., Gaastra, W., and Sommerfelt, H. (1997). A new putative fimbrial colonization factor, CS19, of human enterotoxigenic Escherichia coli. Infection and immunity 65, 507-513.

Griffith, K.L., Fitzpatrick, M.M., Keen, E.F., 3rd, and Wolf, R.E., Jr. (2009). Two functions of the C-terminal domain of Escherichia coli Rob: mediating "sequestration-dispersal" as a novel off-on switch for regulating Rob's activity as a transcription activator and preventing degradation of Rob by Lon protease. Journal of molecular biology 388, 415-430.

Griffith, K.L., Shah, I.M., and Wolf, R.E., Jr. (2004). Proteolytic degradation of Escherichia coli transcription activators SoxS and MarA as the mechanism for reversing the induction of the superoxide (SoxRS) and multiple antibiotic resistance (Mar) regulons. Molecular microbiology 51, 1801-1816.

Griffith, K.L., and Wolf, R.E., Jr. (2001). Systematic mutagenesis of the DNA binding sites for SoxS in the Escherichia coli zwf and fpr promoters: identifying nucleotides required for DNA binding and transcription activation. Molecular microbiology 40, 1141-1154.

Griffith, K.L., and Wolf, R.E., Jr. (2002). A comprehensive alanine scanning mutagenesis of the Escherichia coli transcriptional activator SoxS: identifying amino acids important for DNA binding and transcription activation. Journal of molecular biology 322, 237-257.

Grosso-Becera, M.V., Servin-Gonzalez, L., and Soberon-Chavez, G. (2015). RNA structures are involved in the thermoregulation of bacterial virulence-associated traits. Trends Microbiol 23, 509-518.

Gruber, T.M., and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol 57, 441-466.

Guiyoule, A., Gerbaud, G., Buchrieser, C., Galimand, M., Rahalison, L., Chanteau, S., Courvalin, P., and Carniel, E. (2001). Transferable plasmid-mediated resistance to streptomycin in a clinical isolate of Yersinia pestis. Emerg Infect Dis 7, 43-48.

- Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology* *177*, 4121-4130.
- Hachler, H., Cohen, S.P., and Levy, S.B. (1991). *marA*, a regulated locus which controls expression of chromosomal multiple antibiotic resistance in *Escherichia coli*. *Journal of bacteriology* *173*, 5532-5538.
- Haensch, S., Bianucci, R., Signoli, M., Rajerison, M., Schultz, M., Kacki, S., Vermunt, M., Weston, D.A., Hurst, D., Achtman, M., *et al.* (2010). Distinct clones of *Yersinia pestis* caused the black death. *PLoS Pathog* *6*, e1001134.
- Halliwell, C.M., Morgan, G., Ou, C.P., and Cass, A.E. (2001). Introduction of a (poly)histidine tag in L-lactate dehydrogenase produces a mixture of active and inactive molecules. *Analytical biochemistry* *295*, 257-261.
- Han, Y., Qiu, J., Guo, Z., Gao, H., Song, Y., Zhou, D., and Yang, R. (2007). Comparative transcriptomics in *Yersinia pestis*: a global view of environmental modulation of gene expression. *BMC Microbiol* *7*, 96.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology* *166*, 557-580.
- Hanahan, D., Jessee, J., and Bloom, F.R. (1991). Plasmid transformation of *Escherichia coli* and other bacteria. *Methods in enzymology* *204*, 63-113.
- Hancock, K. (2001). Purification of (His)₆-tagged recombinant proteins expressed as inclusion bodies in *E. coli* using a Ni²⁺ charged HiTrap Chelating HP column. In *Life Science News*, Amersham Biosciences, pp. 1-3.
- Hart, M.K., Saviolakis, G.A., Welkos, S.L., and House, R.V. (2012). Advanced Development of the rF1V and rBV A/B Vaccines: Progress and Challenges. *Adv Prev Med* *2012*, 731604.
- Hartig, S.M. (2013). Basic image analysis and manipulation in ImageJ. *Curr Protoc Mol Biol Chapter 14*, Unit 14.15.
- Hazen, T.H., Sahl, J.W., Redman, J.C., Morris, C.R., Daugherty, S.C., Chibucos, M.C., Sengamalay, N.A., Fraser-Liggett, C.M., Steinsland, H., Whittam, T.S., *et al.* (2012). Draft genome sequences of the diarrheagenic *Escherichia coli* collection. *Journal of bacteriology* *194*, 3026-3027.
- Heath, D.G., Anderson, G.W., Jr., Mauro, J.M., Welkos, S.L., Andrews, G.P., Adamovicz, J., and Friedlander, A.M. (1998). Protection against experimental bubonic and pneumonic plague by a recombinant capsular F1-V antigen fusion protein vaccine. *Vaccine* *16*, 1131-1137.
- Hernday, A., Krabbe, M., Braaten, B., and Low, D. (2002). Self-perpetuating epigenetic pili switches in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* *99 Suppl 4*, 16470-16476.
- Hidalgo, E., Leautaud, V., and Demple, B. (1998). The redox-regulated SoxR protein acts from a single DNA site as a repressor and an allosteric activator. *The EMBO journal* *17*, 2629-2636.
- Hinnebusch, B.J. (2003). Transmission factors: *Yersinia pestis* genes required to infect the flea vector of plague. *Advances in experimental medicine and biology* *529*, 55-62.
- Hinnebusch, B.J. (2005). The evolution of flea-borne transmission in *Yersinia pestis*. *Curr Issues Mol Biol* *7*, 197-212.
- Hoe, N.P., and Goguen, J.D. (1993). Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. *Journal of bacteriology* *175*, 7901-7909.
- Holden, N.J., Totsika, M., Mahler, E., Roe, A.J., Catherwood, K., Lindner, K., Dobrindt, U., and Gally, D.L. (2006). Demonstration of regulatory cross-talk between P fimbriae and type 1 fimbriae in uropathogenic *Escherichia coli*. *Microbiology* *152*, 1143-1153.

- Holden, N.J., Uhlin, B.E., and Gally, D.L. (2001). PapB paralogues and their effect on the phase variation of type 1 fimbriae in *Escherichia coli*. *Molecular microbiology* *42*, 319-330.
- Hollander, A., Mercante, A.D., Shafer, W.M., and Cornelissen, C.N. (2011). The iron-repressed, AraC-like regulator MpeR activates expression of fetA in *Neisseria gonorrhoeae*. *Infection and immunity* *79*, 4764-4776.
- Hu, P., Elliott, J., McCready, P., Skowronski, E., Garnes, J., Kobayashi, A., Brubaker, R.R., and Garcia, E. (1998). Structural organization of virulence-associated plasmids of *Yersinia pestis*. *Journal of bacteriology* *180*, 5192-5202.
- Huang, X.Z., and Lindler, L.E. (2004). The pH 6 antigen is an antiphagocytic factor produced by *Yersinia pestis* independent of *Yersinia* outer proteins and capsule antigen. *Infection and immunity* *72*, 7212-7219.
- Huang, Y., Smith, B.S., Chen, L.X., Baxter, R.H., and Deisenhofer, J. (2009). Insights into pilus assembly and secretion from the structure and functional characterization of usher PapC. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 7403-7407.
- Hung, D.L., Knight, S.D., Woods, R.M., Pinkner, J.S., and Hultgren, S.J. (1996). Molecular basis of two subfamilies of immunoglobulin-like chaperones. *The EMBO journal* *15*, 3792-3805.
- Hung, D.L., Raivio, T.L., Jones, C.H., Silhavy, T.J., and Hultgren, S.J. (2001). Cpx signaling pathway monitors biogenesis and affects assembly and expression of P pili. *The EMBO journal* *20*, 1508-1518.
- Ibarra, J.A., Perez-Rueda, E., Segovia, L., and Puente, J.L. (2008). The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors. *Genetica* *133*, 65-76.
- Inglesby, T.V., Dennis, D.T., Henderson, D.A., Bartlett, J.G., Ascher, M.S., Eitzen, E., Fine, A.D., Friedlander, A.M., Hauer, J., Koerner, J.F., *et al.* (2000). Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *Jama* *283*, 2281-2290.
- Iriarte, M., and Cornelis, G.R. (1995). MyfF, an element of the network regulating the synthesis of fibrillae in *Yersinia enterocolitica*. *Journal of bacteriology* *177*, 738-744.
- Jair, K.W., Fawcett, W.P., Fujita, N., Ishihama, A., and Wolf, R.E., Jr. (1996a). Ambidextrous transcriptional activation by SoxS: requirement for the C-terminal domain of the RNA polymerase alpha subunit in a subset of *Escherichia coli* superoxide-inducible genes. *Molecular microbiology* *19*, 307-317.
- Jair, K.W., Martin, R.G., Rosner, J.L., Fujita, N., Ishihama, A., and Wolf, R.E., Jr. (1995). Purification and regulatory properties of MarA protein, a transcriptional activator of *Escherichia coli* multiple antibiotic and superoxide resistance promoters. *Journal of bacteriology* *177*, 7100-7104.
- Jair, K.W., Yu, X., Skarstad, K., Thony, B., Fujita, N., Ishihama, A., and Wolf, R.E., Jr. (1996b). Transcriptional activation of promoters of the superoxide and multiple antibiotic resistance regulons by Rob, a binding protein of the *Escherichia coli* origin of chromosomal replication. *Journal of bacteriology* *178*, 2507-2513.
- Karlyshev, A.V., Galyov, E.E., Abramov, V.M., and Zav'yalov, V.P. (1992). Caf1R gene and its role in the regulation of capsule formation of *Y. pestis*. *FEBS Lett* *305*, 37-40.
- Ke, Y., Chen, Z., and Yang, R. (2013). *Yersinia pestis*: mechanisms of entry into and resistance to the host cell. *Front Cell Infect Microbiol* *3*, 106.
- Kersley, J.E., Zavialov, A.V., Moslehi, E., Knight, S.D., and MacIntyre, S. (2003). Mutagenesis elucidates the assembly pathway and structure of *Yersinia pestis* F1 polymer. *Advances in experimental medicine and biology* *529*, 113-116.
- Kim, T.J., Chauhan, S., Motin, V.L., Goh, E.B., Igo, M.M., and Young, G.M. (2007). Direct transcriptional control of the plasminogen activator gene of *Yersinia pestis* by the cyclic AMP receptor protein. *Journal of bacteriology* *189*, 8890-8900.

- King, J.M., Schesser Bartra, S., Plano, G., and Yahr, T.L. (2013). ExsA and LcrF recognize similar consensus binding sites, but differences in their oligomeric state influence interactions with promoter DNA. *Journal of bacteriology* *195*, 5639-5650.
- Klaasen, P., and de Graaf, F.K. (1990). Characterization of FapR, a positive regulator of expression of the 987P operon in enterotoxigenic *Escherichia coli*. *Molecular microbiology* *4*, 1779-1783.
- Klinkert, B., Cimdins, A., Gaubig, L.C., Rossmann, J., Aschke-Sonnenborn, U., and Narberhaus, F. (2012). Thermogenetic tools to monitor temperature-dependent gene expression in bacteria. *J Biotechnol* *160*, 55-63.
- Knight, S.D. (2007). Structure and assembly of *Yersinia pestis* F1 antigen. *Advances in experimental medicine and biology* *603*, 74-87.
- Kolin, A., Balasubramaniam, V., Skredenske, J.M., Wickstrum, J.R., and Egan, S.M. (2008). Differences in the mechanism of the allosteric l-rhamnose responses of the AraC/XylS family transcription activators RhaS and RhaR. *Molecular microbiology* *68*, 448-461.
- Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S.A. (2000). A structural model of transcription elongation. *Science* *289*, 619-625.
- Kouse, A.B., Righetti, F., Kortmann, J., Narberhaus, F., and Murphy, E.R. (2013). RNA-mediated thermoregulation of iron-acquisition genes in *Shigella dysenteriae* and pathogenic *Escherichia coli*. *PLoS One* *8*, e63781.
- Kukkonen, M., and Korhonen, T.K. (2004). The omptin family of enterobacterial surface proteases/adhesins: from housekeeping in *Escherichia coli* to systemic spread of *Yersinia pestis*. *Int J Med Microbiol* *294*, 7-14.
- Kwon, H.J., Bennik, M.H., Demple, B., and Ellenberger, T. (2000). Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. *Nat Struct Biol* *7*, 424-430.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* *227*, 680-685.
- Lahteenmaki, K., Kukkonen, M., and Korhonen, T.K. (2001). The Pla surface protease/adhesin of *Yersinia pestis* mediates bacterial invasion into human endothelial cells. *FEBS Lett* *504*, 69-72.
- Ledent, P., Duez, C., Vanhove, M., Lejeune, A., Fonze, E., Charlier, P., Rhazi-Filali, F., Thamm, I., Guillaume, G., Samyn, B., *et al.* (1997). Unexpected influence of a C-terminal-fused His-tag on the processing of an enzyme and on the kinetic and folding parameters. *FEBS Lett* *413*, 194-196.
- Lee, S., Nam, D., Jung, J.Y., Oh, M.K., Sang, B.I., and Mitchell, R.J. (2012). Identification of *Escherichia coli* biomarkers responsive to various lignin-hydrolysate compounds. *Bioresour Technol* *114*, 450-456.
- Li, B., Tan, Y., Guo, J., Cui, B., Wang, Z., Wang, H., Zhou, L., Guo, Z., Zhu, Z., Du, Z., *et al.* (2011). Use of protein microarray to identify gene expression changes of *Yersinia pestis* at different temperatures. *Can J Microbiol* *57*, 287-294.
- Li, Y., Gao, H., Qin, L., Li, B., Han, Y., Guo, Z., Song, Y., Zhai, J., Du, Z., Wang, X., *et al.* (2008). Identification and characterization of PhoP regulon members in *Yersinia pestis* biovar *Microtus*. *BMC Genomics* *9*, 143.
- Lindler, L.E., Plano, G.V., Burland, V., Mayhew, G.F., and Blattner, F.R. (1998). Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. *Infection and immunity* *66*, 5731-5742.
- Lisser, S., and Margalit, H. (1993). Compilation of *E. coli* mRNA promoter sequences. *Nucleic acids research* *21*, 1507-1516.
- Livingstone, C.D., and Barton, G.J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci* *9*, 745-756.

- Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009). A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell* *139*, 770-779.
- Loh, E., Kugelberg, E., Tracy, A., Zhang, Q., Gollan, B., Ewles, H., Chalmers, R., Pelicic, V., and Tang, C.M. (2013). Temperature triggers immune evasion by *Neisseria meningitidis*. *Nature* *502*, 237-240.
- Lowden, M.J., Skorupski, K., Pellegrini, M., Chiorazzo, M.G., Taylor, R.K., and Kull, F.J. (2010). Structure of *Vibrio cholerae* ToxT reveals a mechanism for fatty acid regulation of virulence genes. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 2860-2865.
- Lu, Y., Flaherty, C., and Hendrickson, W. (1992). AraC protein contacts asymmetric sites in the *Escherichia coli* araFGH promoter. *The Journal of biological chemistry* *267*, 24848-24857.
- Lukaszewski, R.A., Kenny, D.J., Taylor, R., Rees, D.G., Hartley, M.G., and Oyston, P.C. (2005). Pathogenesis of *Yersinia pestis* infection in BALB/c mice: effects on host macrophages and neutrophils. *Infection and immunity* *73*, 7142-7150.
- Macinga, D.R., Paradise, M.R., Parojcic, M.M., and Rather, P.N. (1999). Activation of the 2'-N-acetyltransferase gene [aac(2')-Ia] in *Providencia stuartii* by an interaction of AarP with the promoter region. *Antimicrobial agents and chemotherapy* *43*, 1769-1772.
- MacIntyre, S., Knight, S.D. and Fooks, L.J. (2004). Structure, assembly and applications of the polymeric F1 antigen of *Yersinia pestis*. In *Yersinia: Molecular and Cellular Biology* ed Carniel, E and Hinnenbusch, BJ Norfolk, UK: Horizon Bioscience, 363-407.
- MacIntyre, S., Zyrianova, I.M., Chernovskaya, T.V., Leonard, M., Rudenko, E.G., Zav'Yalov, V.P., and Chapman, D.A. (2001). An extended hydrophobic interactive surface of *Yersinia pestis* Caf1M chaperone is essential for subunit binding and F1 capsule assembly. *Molecular microbiology* *39*, 12-25.
- Madrid, C., Balsalobre, C., Garcia, J., and Juarez, A. (2007). The novel Hha/YmoA family of nucleoid-associated proteins: use of structural mimicry to modulate the activity of the H-NS family of proteins. *Molecular microbiology* *63*, 7-14.
- Madrid, C., Nieto, J.M., and Juarez, A. (2002). Role of the Hha/YmoA family of proteins in the thermoregulation of the expression of virulence factors. *Int J Med Microbiol* *291*, 425-432.
- Mahon, V., Fagan, R.P., and Smith, S.G. (2012). Snap denaturation reveals dimerization by AraC-like protein Rns. *Biochimie* *94*, 2058-2061.
- Mahon, V., Smyth, C.J., and Smith, S.G. (2010). Mutagenesis of the Rns regulator of enterotoxigenic *Escherichia coli* reveals roles for a linker sequence and two helix-turn-helix motifs. *Microbiology* *156*, 2796-2806.
- Makhatadze, G.I., and Privalov, P.L. (1995). Energetics of protein structure. *Adv Protein Chem* *47*, 307-425.
- Marketon, M.M., DePaolo, R.W., DeBord, K.L., Jabri, B., and Schneewind, O. (2005). Plague bacteria target immune cells during infection. *Science* *309*, 1739-1741.
- Martin, R.G., Gillette, W.K., Martin, N.I., and Rosner, J.L. (2002). Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in *Escherichia coli*. *Molecular microbiology* *43*, 355-370.
- Martin, R.G., Gillette, W.K., Rhee, S., and Rosner, J.L. (1999). Structural requirements for marbox function in transcriptional activation of mar/sox/rob regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. *Molecular microbiology* *34*, 431-441.
- Martin, R.G., Jair, K.W., Wolf, R.E., Jr., and Rosner, J.L. (1996). Autoactivation of the marRAB multiple antibiotic resistance operon by the MarA transcriptional activator in *Escherichia coli*. *Journal of bacteriology* *178*, 2216-2223.

- Martin, R.G., and Rosner, J.L. (1995). Binding of purified multiple antibiotic-resistance repressor protein (MarR) to mar operator sequences. *Proceedings of the National Academy of Sciences of the United States of America* *92*, 5456-5460.
- Martin, R.G., and Rosner, J.L. (1997). Fis, an accessorial factor for transcriptional activation of the mar (multiple antibiotic resistance) promoter of *Escherichia coli* in the presence of the activator MarA, SoxS, or Rob. *Journal of bacteriology* *179*, 7410-7419.
- Martin, R.G., and Rosner, J.L. (2001). The AraC transcriptional activators. *Curr Opin Microbiol* *4*, 132-137.
- Martin, R.G., and Rosner, J.L. (2002). Genomics of the marA/soxS/rob regulon of *Escherichia coli*: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. *Molecular microbiology* *44*, 1611-1624.
- Martinez-Antonio, A., and Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* *6*, 482-489.
- Mathew, R., and Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol* *14*, 450-455.
- Matsunaga, J., Schlax, P.J., and Haake, D.A. (2013). Role for cis-acting RNA sequences in the temperature-dependent expression of the multiadhesive lig proteins in *Leptospira interrogans*. *Journal of bacteriology* *195*, 5092-5101.
- McGuffin, L.J., Atkins, J.D., Salehe, B.R., Shuid, A.N., and Roche, D.B. (2015). IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic acids research* *43*, 169-173.
- McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* *16*, 404-405.
- McMurry, L.M., and Levy, S.B. (2010). Evidence that regulatory protein MarA of *Escherichia coli* represses rob by steric hindrance. *Journal of bacteriology* *192*, 3977-3982.
- Merrick, M.J. (1993). In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). *Molecular microbiology* *10*, 903-909.
- Michan, C., Machado, M., and Pueyo, C. (2002). SoxRS down-regulation of rob transcription. *Journal of bacteriology* *184*, 4733-4738.
- Miller, J., Williamson, E.D., Lakey, J.H., Pearce, M.J., Jones, S.M., and Titball, R.W. (1998). Macromolecular organisation of recombinant *Yersinia pestis* F1 antigen and the effect of structure on immunogenicity. *FEMS Immunol Med Microbiol* *21*, 213-221.
- Miller, V.L., Beer, K.B., Heusipp, G., Young, B.M., and Wachtel, M.R. (2001). Identification of regions of Ail required for the invasion and serum resistance phenotypes. *Molecular microbiology* *41*, 1053-1062.
- Morelli, G., Song, Y., Mazzoni, C.J., Eppinger, M., Roumagnac, P., Wagner, D.M., Feldkamp, M., Kusecek, B., Vogler, A.J., Li, Y., *et al.* (2010). *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* *42*, 1140-1143.
- Morin, N., Santiago, A.E., Ernst, R.K., Guillot, S.J., and Nataro, J.P. (2013). Characterization of the AggR regulon in enteroaggregative *Escherichia coli*. *Infection and immunity* *81*, 122-132.
- Morin, N., Tirling, C., Ivison, S.M., Kaur, A.P., Nataro, J.P., and Steiner, T.S. (2010). Autoactivation of the AggR regulator of enteroaggregative *Escherichia coli* in vitro and in vivo. *FEMS Immunol Med Microbiol* *58*, 344-355.
- Motin, V.L., Georgescu, A.M., Fitch, J.P., Gu, P.P., Nelson, D.O., Mabery, S.L., Garnham, J.B., Sokhansanj, B.A., Ott, L.L., Coleman, M.A., *et al.* (2004). Temporal global changes in gene expression during temperature transition in *Yersinia pestis*. *Journal of bacteriology* *186*, 6298-6305.

- Motin, V.L., Nakajima, R., Smirnov, G.B., and Brubaker, R.R. (1994). Passive immunity to yersiniae mediated by anti-recombinant V antigen and protein A-V antigen fusion peptide. *Infection and immunity* *62*, 4192-4201.
- Muller-Hill, B. (1998). Some repressors of bacterial transcription. *Curr Opin Microbiol* *1*, 145-151.
- Munson, G.P., Holcomb, L.G., Alexander, H.L., and Scott, J.R. (2002). In vitro identification of Rns-regulated genes. *Journal of bacteriology* *184*, 1196-1199.
- Munson, G.P., and Scott, J.R. (1999). Binding site recognition by Rns, a virulence regulator in the AraC family. *Journal of bacteriology* *181*, 2110-2117.
- Munson, G.P., and Scott, J.R. (2000). Rns, a virulence regulator within the AraC family, requires binding sites upstream and downstream of its own promoter to function as an activator. *Molecular microbiology* *36*, 1391-1402.
- Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* *296*, 1285-1290.
- Murphy, J.R., Donini, S., and Kappock, T.J. (2015). An active site-tail interaction in the structure of hexahistidine-tagged *Thermoplasma acidophilum* citrate synthase. *Acta Crystallographica Section F: Structural Biology Communications* *71(10)*, 1292-1299.
- Mycroft, Z. (2011). The role of the Caf1A outer membrane usher in the production of F1 surface fibres of *Yersinia pestis* In School of Biological Sciences (University of Reading, UK).
- Nataro, J.P., Yikang, D., Yingkang, D., and Walker, K. (1994). AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli*. *Journal of bacteriology* *176*, 4691-4699.
- Ni, L., Tonthat, N.K., Chinnam, N., and Schumacher, M.A. (2013). Structures of the *Escherichia coli* transcription activator and regulator of diauxie, XylR: an AraC DNA-binding family member with a LacI/GalR ligand-binding domain. *Nucleic acids research* *41*, 1998-2008.
- Ninh, A. (2011). First Case of Bubonic Plague in 2011 Appears in New Mexico. In *TIME (TIME)*.
- Noteborn, C.H.S.a.M.H.M. (1988). Formation of Soluble Recombinant Proteins in *Escherichia coli* is Favored by Lower Growth Temperature. *Nature Biotechnology* *6*, 291 - 294.
- Nou, X., Braaten, B., Kaltenbach, L., and Low, D.A. (1995). Differential binding of Lrp to two sets of pap DNA binding sites mediated by Pap I regulates Pap phase variation in *Escherichia coli*. *The EMBO journal* *14*, 5785-5797.
- Nou, X., Skinner, B., Braaten, B., Blyn, L., Hirsch, D., and Low, D. (1993). Regulation of pyelonephritis-associated pili phase-variation in *Escherichia coli*: binding of the PapI and the Lrp regulatory proteins is controlled by DNA methylation. *Molecular microbiology* *7*, 545-553.
- Nuccio, S.P., and Baumber, A.J. (2007). Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiology and molecular biology reviews* : *MMBR* *71*, 551-575.
- Nunoshiba, T., Hidalgo, E., Amabile Cuevas, C.F., and Demple, B. (1992). Two-stage control of an oxidative stress regulon: the *Escherichia coli* SoxR protein triggers redox-inducible expression of the soxS regulatory gene. *Journal of bacteriology* *174*, 6054-6060.
- Nunoshiba, T., Hidalgo, E., Li, Z., and Demple, B. (1993). Negative autoregulation by the *Escherichia coli* SoxS protein: a dampening mechanism for the soxRS redox stress response. *Journal of bacteriology* *175*, 7492-7494.
- Ogden, S., Haggerty, D., Stoner, C.M., Kolodrubetz, D., and Schleif, R. (1980). The *Escherichia coli* L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. *Proceedings of the National Academy of Sciences of the United States of America* *77*, 3346-3350.

- Okusu, H., Ma, D., and Nikaido, H. (1996). AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. *Journal of bacteriology* *178*, 306-308.
- Olsen, P.B., and Klemm, P. (1994). Localization of promoters in the *fim* gene cluster and the effect of H-NS on the transcription of *fimB* and *fimE*. *FEMS microbiology letters* *116*, 95-100.
- Olsen, P.B., Schembri, M.A., Gally, D.L., and Klemm, P. (1998). Differential temperature modulation by H-NS of the *fimB* and *fimE* recombinase genes which control the orientation of the type 1 fimbrial phase switch. *FEMS microbiology letters* *162*, 17-23.
- Otto, K., and Silhavy, T.J. (2002). Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. *Proceedings of the National Academy of Sciences of the United States of America* *99*, 2287-2292.
- Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., *et al.* (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* *413*, 523-527.
- Paul, B.J., Berkmen, M.B., and Gourse, R.L. (2005). DksA potentiates direct activation of amino acid promoters by ppGpp. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 7823-7828.
- Perez-Rueda, E., and Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic acids research* *28*, 1838-1847.
- Perez-Rueda, E., Tenorio-Salgado, S., Huerta-Saquero, A., Balderas-Martinez, Y.I., and Moreno-Hagelsieb, G. (2015). The functional landscape bound to the transcription factors of *Escherichia coli* K-12. *Comput Biol Chem* *58*, 93-103.
- Perry, R.D., and Fetherston, J.D. (1997). *Yersinia pestis*--etiologic agent of plague. *Clin Microbiol Rev* *10*, 35-66.
- Peruski, L.F., Jr., Kay, B.A., El-Yazeed, R.A., El-Etr, S.H., Cravioto, A., Wierzbza, T.F., Rao, M., El-Ghorab, N., Shaheen, H., Khalil, S.B., *et al.* (1999). Phenotypic diversity of enterotoxigenic *Escherichia coli* strains from a community-based study of pediatric diarrhea in periurban Egypt. *Journal of clinical microbiology* *37*, 2974-2978.
- Pettersson, J., Holmstrom, A., Hill, J., Leary, S., Frithz-Lindsten, E., von Euler-Matell, A., Carlsson, E., Titball, R., Forsberg, A., and Wolf-Watz, H. (1999). The V-antigen of *Yersinia* is surface exposed before target cell contact and involved in virulence protein translocation. *Molecular microbiology* *32*, 961-976.
- Phan, G., Remaut, H., Wang, T., Allen, W.J., Pirker, K.F., Lebedev, A., Henderson, N.S., Geibel, S., Volkan, E., Yan, J., *et al.* (2011). Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature* *474*, 49-53.
- Pich, O.Q., Carpenter, B.M., Gilbreath, J.J., and Merrell, D.S. (2012). Detailed analysis of *Helicobacter pylori* Fur-regulated promoters reveals a Fur box core sequence and novel Fur-regulated genes. *Molecular microbiology* *84*, 921-941.
- Pichel, M., Binsztein, N., and Viboud, G. (2000). CS22, a novel human enterotoxigenic *Escherichia coli* adhesin, is related to CS15. *Infection and immunity* *68*, 3280-3285.
- Pilonieta, M.C., Boderio, M.D., and Munson, G.P. (2007). CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic *Escherichia coli*. *Journal of bacteriology* *189*, 5060-5067.
- Plano, G.V., and Schesser, K. (2013). The *Yersinia pestis* type III secretion system: expression, assembly and role in the evasion of host defenses. *Immunol Res* *57*, 237-245.
- Plumbridge, J. (2002). Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc. *Curr Opin Microbiol* *5*, 187-193.

- Pomposiello, P.J., Bennik, M.H., and Demple, B. (2001). Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. *Journal of bacteriology* *183*, 3890-3902.
- Porter, M.E., Mitchell, P., Roe, A.J., Free, A., Smith, D.G., and Gally, D.L. (2004). Direct and indirect transcriptional activation of virulence genes by an AraC-like protein, PerA from enteropathogenic *Escherichia coli*. *Molecular microbiology* *54*, 1117-1133.
- Prentice, M.B., James, K.D., Parkhill, J., Baker, S.G., Stevens, K., Simmonds, M.N., Mungall, K.L., Churcher, C., Oyston, P.C., Titball, R.W., *et al.* (2001). *Yersinia pestis* pFra shows biovar-specific differences and recent common ancestry with a *Salmonella enterica* serovar Typhi plasmid. *Journal of bacteriology* *183*, 2586-2594.
- Pujol, C., and Bliska, J.B. (2005). Turning *Yersinia* pathogenesis outside in: subversion of macrophage function by intracellular yersiniae. *Clin Immunol* *114*, 216-226.
- Quade, N., Mendonca, C., Herbst, K., Heroven, A.K., Ritter, C., Heinz, D.W., and Dersch, P. (2012). Structural basis for intrinsic thermosensing by the master virulence regulator RovA of *Yersinia*. *The Journal of biological chemistry* *287*, 35796-35803.
- Quenee, L.E., Ciletti, N., Berube, B., Krausz, T., Elli, D., Hermanas, T., and Schneewind, O. (2011). Plague in Guinea pigs and its prevention by subunit vaccines. *Am J Pathol* *178*, 1689-1700.
- Quenee, L.E., Cornelius, C.A., Ciletti, N.A., Elli, D., and Schneewind, O. (2008). *Yersinia pestis* caf1 variants and the limits of plague vaccine protection. *Infection and immunity* *76*, 2025-2036.
- Quinn, B. (2012). Oregon man recovering from rare case of the plague. *The Guardian*, United Kingdom. <http://www.theguardian.com/world/2012/jul/18/oregon-man-case-plague>.
- Rakin, A., Boolgakowa, E., and Heesemann, J. (1996). Structural and functional organization of the *Yersinia pestis* bacteriocin pesticin gene cluster. *Microbiology* *142 (Pt 12)*, 3415-3424.
- Reed, W.P., Palmer, D.L., Williams, R.C., Jr., and Kisch, A.L. (1970). Bubonic plague in the Southwestern United States. A review of recent experience. *Medicine (Baltimore)* *49*, 465-486.
- Remaut, H., Tang, C., Henderson, N.S., Pinkner, J.S., Wang, T., Hultgren, S.J., Thanassi, D.G., Waksman, G., and Li, H. (2008). Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell* *133*, 640-652.
- Reuter, S., Connor, T.R., Barquist, L., Walker, D., Feltwell, T., Harris, S.R., Fookes, M., Hall, M.E., Petty, N.K., Fuchs, T.M., *et al.* (2014). Parallel independent evolution of pathogenicity within the genus *Yersinia*. *Proceedings of the National Academy of Sciences of the United States of America* *111*, 6768-6773.
- Rhee, S., Martin, R.G., Rosner, J.L., and Davies, D.R. (1998). A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proceedings of the National Academy of Sciences of the United States of America* *95*, 10413-10418.
- Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* *87*, 1295-1306.
- Righetti, F., and Narberhaus, F. (2014). How to find RNA thermometers. *Front Cell Infect Microbiol* *4*, 1-6.
- Robichon, C., Luo, J., Causey, T.B., Benner, J.S., and Samuelson, J.C. (2011). Engineering *Escherichia coli* BL21(DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Appl Environ Microbiol* *77*, 4634-4646.
- Roche, D.B., Buenavista, M.T., Tetchner, S.J., and McGuffin, L.J. (2011). The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic acids research* *39*, W171-176.

- Rocke, T.E., Smith, S., Marinari, P., Kreeger, J., Enama, J.T., and Powell, B.S. (2008). Vaccination with F1-V fusion protein protects black-footed ferrets (*Mustela nigripes*) against plague upon oral challenge with *Yersinia pestis*. *J Wildl Dis* 44, 1-7.
- Rodgers, M.E., Holder, N.D., Dirla, S., and Schleif, R. (2009). Functional modes of the regulatory arm of AraC. *Proteins* 74, 81-91.
- Rojas, T.C., Maluta, R.P., Parizzi, L.P., Koenigkan, L.V., Yang, J., Yu, J., Pereira, G.A., and Dias da Silveira, W. (2013). Genome Sequences of Avian Pathogenic *Escherichia coli* Strains Isolated from Brazilian Commercial Poultry. *Genome Announc* 1, e0011013.
- Rollins, S.E., Rollins, S.M., and Ryan, E.T. (2003). *Yersinia pestis* and the plague. *Am J Clin Pathol* 119 *Suppl*, S78-85.
- Rosano, G.L., and Ceccarelli, E.A. (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microb Cell Fact* 8, 41.
- Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* 5, 172.
- Rosenberg, E.Y., Bertenthal, D., Nilles, M.L., Bertrand, K.P., and Nikaido, H. (2003). Bile salts and fatty acids induce the expression of *Escherichia coli* AcrAB multidrug efflux pump through their interaction with Rob regulatory protein. *Molecular microbiology* 48, 1609-1619.
- Rosenstein, R., Nikoleit, K., and Gotz, F. (1994). Binding of ArsR, the repressor of the *Staphylococcus xylosus* (pSX267) arsenic resistance operon to a sequence with dyad symmetry within the ars promoter. *Molecular & general genetics : MGG* 242, 566-572.
- Rosner, J.L., Dangi, B., Gronenborn, A.M., and Martin, R.G. (2002). Posttranscriptional activation of the transcriptional activator Rob by dipyrindyl in *Escherichia coli*. *Journal of bacteriology* 184, 1407-1416.
- Ross, W., Ernst, A., and Gourse, R.L. (2001). Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev* 15, 491-506.
- Ruff, E.F., Record, M.T., Jr., and Artsimovitch, I. (2015). Initial events in bacterial transcription initiation. *Biomolecules* 5, 1035-1062.
- Ruiz, R., Marques, S., and Ramos, J.L. (2003). Leucines 193 and 194 at the N-terminal domain of the XylS protein, the positive transcriptional regulator of the TOL meta-cleavage pathway, are involved in dimerization. *Journal of bacteriology* 185, 3036-3041.
- Runco, L.M., Myrczek, S., Bliska, J.B., and Thanassi, D.G. (2008). Biogenesis of the fraction 1 capsule and analysis of the ultrastructure of *Yersinia pestis*. *Journal of bacteriology* 190, 3381-3385.
- Sakai, T., Sasakawa, C., Makino, S., and Yoshikawa, M. (1986). DNA sequence and product analysis of the virF locus responsible for congo red binding and cell invasion in *Shigella flexneri* 2a. *Infection and immunity* 54, 395-402.
- Sakai, T., Sasakawa, C., and Yoshikawa, M. (1988). Expression of four virulence antigens of *Shigella flexneri* is positively regulated at the transcriptional level by the 30 kiloDalton virF protein. *Molecular microbiology* 2, 589-597.
- Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T., and Shendure, J. (2015). Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome research* 25, 119-128.
- Sambrook, J., and Russell, D.W. (2001). *Molecular cloning - a laboratory manual*, 3rd edn (Cold Spring Harbor Laboratory Press).
- Sanchez, A., Osborne, M.L., Friedman, L.J., Kondev, J., and Gelles, J. (2011). Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *The EMBO journal* 30, 3940-3946.

- Sauer, F.G., Remaut, H., Hultgren, S.J., and Waksman, G. (2004). Fiber assembly by the chaperone-usher pathway. *Biochim Biophys Acta* *1694*, 259-267.
- Saviola, B., Seabold, R., and Schleif, R.F. (1998). Arm-domain interactions in AraC. *Journal of molecular biology* *278*, 539-548.
- Scaletsky, I.C., Michalski, J., Torres, A.G., Dulguer, M.V., and Kaper, J.B. (2005). Identification and characterization of the locus for diffuse adherence, which encodes a novel afimbrial adhesin found in atypical enteropathogenic *Escherichia coli*. *Infection and immunity* *73*, 4753-4765.
- Schellman, J.A. (1997). Temperature, stability, and the hydrophobic interaction. *Biophys J* *73*, 2960-2964.
- Schleif, R. (2000). Regulation of the L-arabinose operon of *Escherichia coli*. *Trends in genetics : TIG* *16*, 559-565.
- Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS microbiology reviews* *34*, 779-796.
- Schneiders, T., and Levy, S.B. (2006). MarA-mediated transcriptional repression of the rob promoter. *The Journal of biological chemistry* *281*, 10049-10055.
- Schuller, A., Slater, A.W., Norambuena, T., Cifuentes, J.J., Almonacid, L.I., and Melo, F. (2012). Computer-based annotation of putative AraC/XylS-family transcription factors of known structure but unknown function. *J Biomed Biotechnol* *2012*, 103132.
- Schwan, W.R. (2011). Regulation of genes in uropathogenic. *World J Clin Infect Dis* *1*, 17-25.
- Schwan, W.R., Lee, J.L., Lenard, F.A., Matthews, B.T., and Beck, M.T. (2002). Osmolarity and pH growth conditions regulate fim gene transcription and type 1 pilus expression in uropathogenic *Escherichia coli*. *Infection and immunity* *70*, 1391-1402.
- Schwan, W.R., Seifert, H.S., and Duncan, J.L. (1994). Analysis of the fimB promoter region involved in type 1 pilus phase variation in *Escherichia coli*. *Molecular & general genetics : MGG* *242*, 623-630.
- Sebbane, F., Jarrett, C., Gardner, D., Long, D., and Hinnebusch, B.J. (2009). The *Yersinia pestis* caf1M1A1 fimbrial capsule operon promotes transmission by flea bite in a mouse model of bubonic plague. *Infection and immunity* *77*, 1222-1229.
- Servin, A.L. (2005). Pathogenesis of Afa/Dr diffusely adhering *Escherichia coli*. *Clin Microbiol Rev* *18*, 264-292.
- Sha, J., Agar, S.L., Baze, W.B., Olano, J.P., Fadl, A.A., Erova, T.E., Wang, S., Foltz, S.M., Suarez, G., Motin, V.L., *et al.* (2008). Braun lipoprotein (Lpp) contributes to virulence of yersiniae: potential role of Lpp in inducing bubonic and pneumonic plague. *Infection and immunity* *76*, 1390-1409.
- Shah, I.M., and Wolf, R.E., Jr. (2004). Novel protein-protein interaction between *Escherichia coli* SoxS and the DNA binding determinant of the RNA polymerase alpha subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. *Journal of molecular biology* *343*, 513-532.
- Sheridan, S.D., Opel, M.L., and Hatfield, G.W. (2001). Activation and repression of transcription initiation by a distant DNA structural transition. *Molecular microbiology* *40*, 684-690.
- Shin, M., Kang, S., Hyun, S.J., Fujita, N., Ishihama, A., Valentin-Hansen, P., and Choy, H.E. (2001). Repression of deoP2 in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *The EMBO journal* *20*, 5392-5399.
- Simons, R.W., Houman, F., and Kleckner, N. (1987). Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene* *53*, 85-96.
- Simpson, W.J., Thomas, R.E., and Schwan, T.G. (1990). Recombinant capsular antigen (fraction 1) from *Yersinia pestis* induces a protective antibody response in BALB/c mice. *Am J Trop Med Hyg* *43*, 389-396.

- Skurnik, M., Peippo, A., and Ercela, E. (2000). Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O:1b. *Molecular microbiology* *37*, 316-330.
- Smith, S.G., and Dorman, C.J. (1999). Functional analysis of the FimE integrase of *Escherichia coli* K-12: isolation of mutant derivatives with altered DNA inversion preferences. *Molecular microbiology* *34*, 965-979.
- Sohanpal, B.K., El-Labany, S., Lahooti, M., Plumbridge, J.A., and Blomfield, I.C. (2004). Integrated regulatory responses of fimB to N-acetylneuraminic (sialic) acid and GlcNAc in *Escherichia coli* K-12. *Proceedings of the National Academy of Sciences of the United States of America* *101*, 16322-16327.
- Sohanpal, B.K., Friar, S., Roobol, J., Plumbridge, J.A., and Blomfield, I.C. (2007). Multiple co-regulatory elements and IHF are necessary for the control of fimB expression in response to sialic acid and N-acetylglucosamine in *Escherichia coli* K-12. *Molecular microbiology* *63*, 1223-1236.
- Soisson, S.M., MacDougall-Shackleton, B., Schleif, R., and Wolberger, C. (1997). The 1.6 Å crystal structure of the AraC sugar-binding and dimerization domain complexed with D-fucose. *Journal of molecular biology* *273*, 226-237.
- Solovyev, V., and Salamov, A., (2011). Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies* R.W. Li, ed. (Nova Science), pp. 61-78.
- Song, Y., Tong, Z., Wang, J., Wang, L., Guo, Z., Han, Y., Zhang, J., Pei, D., Zhou, D., Qin, H., *et al.* (2004). Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res* *11*, 179-197.
- Soto, G.E., and Hultgren, S.J. (1999). Bacterial adhesins: common themes and variations in architecture and assembly. *Journal of bacteriology* *181*, 1059-1071.
- Spears, P.A., Schauer, D., and Orndorff, P.E. (1986). Metastable regulation of type 1 piliation in *Escherichia coli* and isolation and characterization of a phenotypically stable mutant. *Journal of bacteriology* *168*, 179-185.
- Spurio, R., Falconi, M., Brandi, A., Pon, C.L., and Gualerzi, C.O. (1997). The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *The EMBO journal* *16*, 1795-1805.
- Starmer, J., Stomp, A., Vouk, M., and Bitzer, D. (2006). Predicting Shine-Dalgarno sequence locations exposes genome annotation errors. *PLoS computational biology* *2*, e57.
- Steinmann, R., and Dersch, P. (2013). Thermosensing to adjust bacterial virulence in a fluctuating environment. *Future Microbiol* *8*, 85-105.
- Stenseth, N.C., Atshabar, B.B., Begon, M., Belmain, S.R., Bertherat, E., Carniel, E., Gage, K.L., Leirs, H., and Rahalison, L. (2008). Plague: past, present, and future. *PLoS Med* *5*, e3.
- Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal transduction. *Annual review of biochemistry* *69*, 183-215.
- Straley, S.C., and Perry, R.D. (1995). Environmental modulation of gene expression and pathogenesis in *Yersinia*. *Trends Microbiol* *3*, 310-317.
- Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology* *189*, 113-130.
- Taliaferro, L.P., Keen, E.F., 3rd, Sanchez-Alberola, N., and Wolf, R.E., Jr. (2012). Transcription activation by *Escherichia coli* Rob at class II promoters: protein-protein interactions between Rob's N-terminal domain and the sigma(70) subunit of RNA polymerase. *Journal of molecular biology* *419*, 139-157.

- Tanaka, T., Horii, T., Shibayama, K., Sato, K., Ohsuka, S., Arakawa, Y., Yamaki, K., Takagi, K., and Ohta, M. (1997). RobA-induced multiple antibiotic resistance largely depends on the activation of the AcrAB efflux. *Microbiol Immunol* *41*, 697-702.
- Thanassi, D.G., Bliska, J.B., and Christie, P.J. (2012). Surface organelles assembled by secretion systems of Gram-negative bacteria: diversity in structure and function. *FEMS microbiology reviews* *36*, 1046-1082.
- Thanassi, D.G., Saulino, E.T., and Hultgren, S.J. (1998). The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Curr Opin Microbiol* *1*, 223-231.
- Titball, R.W., Howells, A.M., Oyston, P.C., and Williamson, E.D. (1997). Expression of the *Yersinia pestis* capsular antigen (F1 antigen) on the surface of an *aroA* mutant of *Salmonella typhimurium* induces high levels of protection against plague. *Infection and immunity* *65*, 1926-1930.
- Titball, R.W., and Williamson, E.D. (2001). Vaccination against bubonic and pneumonic plague. *Vaccine* *19*, 4175-4184.
- Tobe, T., Schoolnik, G.K., Sohel, I., Bustamante, V.H., and Puente, J.L. (1996). Cloning and characterization of *bfpTVW*, genes required for the transcriptional activation of *bfpA* in enteropathogenic *Escherichia coli*. *Molecular microbiology* *21*, 963-975.
- Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America* *76*, 4350-4354.
- Tsujikawa, L., Tsodikov, O.V., and deHaseth, P.L. (2002). Interaction of RNA polymerase with forked DNA: evidence for two kinetically significant intermediates on the pathway to the final complex. *Proceedings of the National Academy of Sciences of the United States of America* *99*, 3493-3498.
- Valentin-Hansen, P., Sogaard-Andersen, L., and Pedersen, H. (1996). A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Molecular microbiology* *20*, 461-466.
- Vasina, J.A., and Baneyx, F. (1997). Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli* *cspA* and *tac* promoter systems. *Protein expression and purification* *9*, 211-218.
- Vera, A., Gonzalez-Montalban, N., Aris, A., and Villaverde, A. (2007). The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol Bioeng* *96*, 1101-1106.
- Vinue, L., McMurry, L.M., and Levy, S.B. (2013). The 216-bp *marB* gene of the *marRAB* operon in *Escherichia coli* encodes a periplasmic protein which reduces the transcription rate of *marA*. *FEMS microbiology letters* *345*, 49-55.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* *25*, 1189-1191.
- Weber, G.G., Kortmann, J., Narberhaus, F., and Klose, K.E. (2014). RNA thermometer controls temperature-dependent virulence factor expression in *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America* *111*, 14241-14246.
- Welch, T.J., Fricke, W.F., McDermott, P.F., White, D.G., Rosso, M.L., Rasko, D.A., Mammel, M.K., Eppinger, M., Rosovitz, M.J., Wagner, D., *et al.* (2007). Multiple antimicrobial resistance in plague: an emerging public health risk. *PLoS One* *2*, e309.
- Weyand, N.J., Braaten, B.A., van der Woude, M., Tucker, J., and Low, D.A. (2001). The essential role of the promoter-proximal subunit of CAP in *pap* phase variation: Lrp- and helical phase-dependent activation of *papBA* transcription by CAP from -215. *Molecular microbiology* *39*, 1504-1522.

- White-Ziegler, C.A., Angus Hill, M.L., Braaten, B.A., van der Woude, M.W., and Low, D.A. (1998). Thermoregulation of *Escherichia coli* pap transcription: H-NS is a temperature-dependent DNA methylation blocking factor. *Molecular microbiology* 28, 1121-1137.
- White-Ziegler, C.A., Villapakkam, A., Ronaszeki, K., and Young, S. (2000). H-NS controls pap and daa fimbrial transcription in *Escherichia coli* in response to multiple environmental cues. *Journal of bacteriology* 182, 6391-6400.
- Wood, T.I., Griffith, K.L., Fawcett, W.P., Jair, K.W., Schneider, T.D., and Wolf, R.E., Jr. (1999). Interdependence of the position and orientation of SoxS binding sites in the transcriptional activation of the class I subset of *Escherichia coli* superoxide-inducible promoters. *Molecular microbiology* 34, 414-430.
- Wosten, M.M. (1998). Eubacterial sigma-factors. *FEMS microbiology reviews* 22, 127-150.
- Wu, J., and Filutowicz, M. (1999). Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. *Acta Biochim Pol* 46, 591-599.
- Xia, Y., Gally, D., Forsman-Semb, K., and Uhlin, B.E. (2000). Regulatory cross-talk between adhesin operons in *Escherichia coli*: inhibition of type 1 fimbriae expression by the PapB protein. *The EMBO journal* 19, 1450-1457.
- Xie, Y., Yao, Y., Kolisnychenko, V., Teng, C.H., and Kim, K.S. (2006). HbiF regulates type 1 fimbriation independently of FimB and FimE. *Infection and immunity* 74, 4039-4047.
- Yang, Y., and Isberg, R.R. (1997). Transcriptional regulation of the *Yersinia pseudotuberculosis* pH6 antigen adhesin by two envelope-associated components. *Molecular microbiology* 24, 499-510.
- Yu, X., Visweswaran, G.R., Duck, Z., Marupakula, S., MacIntyre, S., Knight, S.D., and Zavialov, A.V. (2009). Caf1A usher possesses a Caf1 subunit-like domain that is crucial for Caf1 fibre secretion. *The Biochemical journal* 418, 541-551.
- Yu, X.D., Fooks, L.J., Moslehi-Mohebi, E., Tischenko, V.M., Askarieh, G., Knight, S.D., Macintyre, S., and Zavialov, A.V. (2012). Large is fast, small is tight: determinants of speed and affinity in subunit capture by a periplasmic chaperone. *Journal of molecular biology* 417, 294-308.
- Zafar, M.A., Sanchez-Alberola, N., and Wolf, R.E., Jr. (2011). Genetic evidence for a novel interaction between transcriptional activator SoxS and region 4 of the sigma(70) subunit of RNA polymerase at class II SoxS-dependent promoters in *Escherichia coli*. *Journal of molecular biology* 407, 333-353.
- Zafar, M.A., Shah, I.M., and Wolf, R.E., Jr. (2010). Protein-protein interactions between sigma(70) region 4 of RNA polymerase and *Escherichia coli* SoxS, a transcription activator that functions by the prerecruitment mechanism: evidence for "off-DNA" and "on-DNA" interactions. *Journal of molecular biology* 401, 13-32.
- Zav'yalov, V., Zavialov, A., Zav'yalova, G., and Korpela, T. (2010). Adhesive organelles of Gram-negative pathogens assembled with the classical chaperone/usher machinery: structure and function from a clinical standpoint. *FEMS microbiology reviews* 34, 317-378.
- Zavialov, A., Zav'yalova, G., Korpela, T., and Zav'yalov, V. (2007). FGL chaperone-assembled fimbrial polyadhesins: anti-immune armament of Gram-negative bacterial pathogens. *FEMS microbiology reviews* 31, 478-514.
- Zavialov, A.V., Batchikova, N.V., Korpela, T., Petrovskaya, L.E., Korobko, V.G., Kersley, J., MacIntyre, S., and Zav'yalov, V.P. (2001). Secretion of recombinant proteins via the chaperone/usher pathway in *Escherichia coli*. *Appl Environ Microbiol* 67, 1805-1814.
- Zavialov, A.V., Berglund, J., Pudney, A.F., Fooks, L.J., Ibrahim, T.M., MacIntyre, S., and Knight, S.D. (2003). Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* 113, 587-596.

- Zavialov, A.V., Kersley, J., Korpela, T., Zav'yalov, V.P., MacIntyre, S., and Knight, S.D. (2002). Donor strand complementation mechanism in the biogenesis of non-pilus systems. *Molecular microbiology* 45, 983-995.
- Zavialov, A.V., Tischenko, V.M., Fooks, L.J., Brandsdal, B.O., Aqvist, J., Zav'yalov, V.P., Macintyre, S., and Knight, S.D. (2005). Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *The Biochemical journal* 389, 685-694.
- Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids research* 33, 2302-2309.
- Zhang, Y., Wang, L., Fang, N., Qu, S., Tan, Y., Guo, Z., Qiu, J., Zhou, D., and Yang, R. (2013). Reciprocal regulation of pH 6 antigen gene loci by PhoP and RovA in *Yersinia pestis* biovar *Microtus*. *Future Microbiol* 8, 271-280.
- Zhou, D., Han, Y., and Yang, R. (2006). Molecular and physiological insights into plague transmission, virulence and etiology. *Microbes Infect* 8, 273-284.
- Zhou, D., Tong, Z., Song, Y., Han, Y., Pei, D., Pang, X., Zhai, J., Li, M., Cui, B., Qi, Z., *et al.* (2004). Genetics of metabolic variations between *Yersinia pestis* biovars and the proposal of a new biovar, *microtus*. *Journal of bacteriology* 186, 5147-5152.
- Zhou, L., Ying, W., Han, Y., Chen, M., Yan, Y., Li, L., Zhu, Z., Zheng, Z., Jia, W., Yang, R., *et al.* (2012). A proteome reference map and virulence factors analysis of *Yersinia pestis* 91001. *J Proteomics* 75, 894-907.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* 31, 3406-3415.