Replication enhancer elements within the open reading frame of tick-borne encephalitis virus and their evolution within the Flavivirus genus

Article

Published Version


It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://nar.oxfordjournals.org/content/early/2011/05/27/nar.gkr237.full.pdf?
keytype=ref&ijkey=UI6dZzMlfQmcV2t
To link to this article DOI: http://dx.doi.org/10.1093/nar/gkr237

Publisher: Oxford University Press
All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading’s research outputs online
Replication enhancer elements within the open reading frame of tick-borne encephalitis virus and their evolution within the Flavivirus genus

A. Tuplin1, D. J. Evans1, A. Buckley2, I. M. Jones3, E. A. Gould4 and T. S. Gritsun3,*

1School of Life Sciences, University of Warwick, Coventry, CV4 7AL, 2Cranfield Health, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK, 3School of Biological Sciences, University of Reading, Whiteknights, Reading, RG6 6AJ, UK and 4Unité des Virus Emergents, Faculté de Médecine Timone, 5ème étage Aile Bleu, 27, Bd Jean Moulin, 13385 Marseille, Cedex 05, France

Received February 6, 2011; Revised March 31, 2011; Accepted April 5, 2011

ABSTRACT

We provide experimental evidence of a replication enhancer element (REE) within the capsid gene of tick-borne encephalitis virus (TBEV, genus Flavivirus). Thermodynamic and phylogenetic analyses predicted that the REE folds as a long stable stem–loop (designated SL6), conserved among all tick-borne flaviviruses (TBFV). Homologous sequences and potential base pairing were found in the corresponding regions of mosquito-borne flaviviruses (MBFV). Substitutions were made within a TBEV reverse genetic system and recovered mutants were compared for plaque morphology, single-step replication kinetics and cytopathic effect. The greatest phenotypic changes were observed in mutants with a destabilized stem. Point mutations in the conserved hexanucleotide motif of the terminal loop caused moderate virus attenuation. However, all mutants eventually reached the titre of wild-type virus late post-infection. Thus, although not essential for growth in tissue culture, the SL6 REE acts to up-regulate virus replication. We hypothesize that this modulatory role may be important for TBEV survival in nature, where the virus circulates by non-viraemic transmission between infected and non-infected ticks, during co-feeding on local rodents.

INTRODUCTION

Tick-borne encephalitis virus (TBEV) is a human pathogen that causes about 16 000 human cases of tick-borne encephalitis (TBE) across Europe and Asia annually (1–3). Taxonomically, TBEV is a species within the mammalian tick-borne flaviviruses (mTBFV). Together with the seabird tick-borne flavivirus group (sTBFV), they comprise one ecological group of tick-borne flaviviruses (TBFV) within the genus Flavivirus, family Flaviviridae. Two other ecological groups within the genus Flavivirus are the mosquito-borne flaviviruses (MBFV) and flaviviruses with no-known vector (NKV) (4). A fourth group including Kamiti River virus (KRV) (5), cell fusion agent virus (CFAV) (6) and Culex flavivirus (CuFV) (7) have been isolated only from mosquitoes with no demonstrated capacity to replicate in mammals and are under consideration by the ICTV Committee for classification as ‘probably arthropod-borne viruses’ (PABV).

Flavivirus virions are ∼50-nm particles with a nucleocapsid composed of capsid (C) protein surrounding a positive-sense single-stranded RNA genome of ∼11 kb. The capsid is enclosed in a lipid membrane within which the viral membrane (M) and envelope (E) proteins are embedded. The genome encodes a single polyprotein of approximately 3400 amino acids from which the three structural (C, M and E) and seven non-structural (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) proteins are processed by cellular and viral proteases (8).

Flavivirus genome replication involves synthesis of a negative-sense template strand by the RNA-dependent RNA polymerase (RdRp; NS5<sup>pol</sup>) from which additional genome-sense strands are transcribed. This process is controlled by numerous RNA–RNA and RNA–protein interactions determined by virus RNA sequence motifs and secondary structures, called cis-acting replication

*To whom correspondence should be addressed. Tel: +44 0118 378 6368; Fax +44 (0)118 378 6537; Email: t.s.gritsun@reading.ac.uk
In evolutionary terms, the replication of flavivirus CREs (16). The promoter has been identified as a complex of highly conserved interacting RNA structures formed by four conserved long imperfectly repeated sequences (LSRs), genetic remnants of which are revealed in the MBFV, NKV and PABV groups (17). It has been proposed that the 5'-UTR may have evolved from a trans-terminal duplication of the archival flavivirus 3'-UTR (16).

An additional complexity in flavivirus replication is the presence of replication enhancer elements (REEs) in the 3'-UTR that, while not obligatory for replication of laboratory-maintained viruses, are likely essential for virus circulation and transmission in nature (16,18). Engineered deletions or modifications of the REEs enable the recovery of viable viruses that are attenuated as a result of reduced RNA synthesis (10,19–22). The cumulative effect of several REEs enhances the assembly of the RdRp complex and is probably critical to the survival of flaviviruses in nature (23). The REEs identified for MBFV have become an important target for the development of a live attenuated vaccine for dengue virus (24,25).

The relatively compact nature of the flavivirus genome, together with constraints imposed by the need to replicate in vertebrate and invertebrate hosts, means that additional CRE sequences may be present in parts of the genome other than the non-coding regions. Indeed, RNA secondary structures have been predicted within the coding region of several flaviviruses (26–28). Here, using bioinformatic and reverse genetic analysis we demonstrate that the capsid-encoding region of TBEV contains an REE which we designate SL6 (26,27). Phylogenetic evidence suggests that the MBFV group also contains at least a partial SL6-like structure, though it is absent in the NKV or PABV groups. The significance of these findings in the context of flavivirus evolution and adaptation to transmission is discussed.

MATERIALS AND METHODS

Sequence and structural analysis

Genbank accession numbers for sequences from all four groups of flaviviruses (TBFV, MBFV, NKV and PABV) used for in silico analysis are listed in Supplementary Table S1. RNA nucleotide sequences were aligned using ClustalX (29) and then edited manually. Nucleotide and dinucleotide scans and analysis of suppression of synonymous site variability (SSSV) were determined by mean pair-wise distance comparison at each codon within the ORF using the Simmonics 1.6 package (http://www.picornavirus.org/), as previously described (30). SSSV was calculated only at aligned codon positions in which over 40% of sequence comparisons were synonymous and averaged over a sliding window of 21 codons; consequently, data point are only produced from codon 11.

RNA secondary structures were predicted using the MFold 3.2 and DINAELMT packages (http://mfold.bioinfo.rpi.edu/) with default settings (31,32). Phylogenetically conserved RNA structures were predicted using STRUCTURE_DIST (http://www.picornavirus.org/) to analyze connect files generated using hybrid-ss-min from the UNAFold suite of programs (32).

Cells and viruses

Porcine embryo kidney cells (PS) have been used in experiments with TBEV strain Vasilchenko (Vs) and its infectious clone (pGGVs) as described previously (33–35).

Plasmids and site-directed mutagenesis

The construction of the infectious clone pGGVs for Vs virus and methods of mutagenesis have been described (34,36). Briefly, the pGGVs was subcloned into two plasmids; one, pGGVs660 contained the first 660 nt of the virus genome and the second pGGVs660–10927 included the remainder. Site-directed mutagenesis was accomplished by PCR (details of primers are available on request). Mutated PCR products were cloned into the pGGVs660 between MluI and EcoRI sites followed by sequencing.

Recovery of viruses from infectious clone

The recovery of virus from the two plasmids representing the infectious clone has been described previously (34–36). Briefly, plasmid pGGVs660 (or mutated derivatives) was digested with PsAI, dephosphorylated with Shrimp Alkaline Phosphatase (SAP; USB) and, after heat-inactivation of SAP, digested with AgeI. Similarly, pGGVs660–10927 was digested with NotI, dephosphorylated and then digested with AgeI. The excised linker DNA fragments from pGGVs660 and pGGVs660–10927 were removed using MicroSpin™ S-400 columns (Pharmacia Biotech) and ligated at the AgeI site generating full-length cDNA which was linearized with SmaI and used as a template for SP6 transcription (34). In vitro-synthesized RNA was inoculated intracerebrally into suckling mice to recover the mutant viruses which were not passaged further prior to phenotype evaluation (35). Recovered viruses were amplified by RT–PCR between nucleotides 1–940 (5'-UTR-C-prM region of the TBEV genome) and 10206–10927 (3'-UTR), and sequenced to validate the presence of the introduced mutations and to exclude extraneous mutations at the 5'-UTR and 3'-UTR (36).

Analysis of virus replication cycle and cytopathic effect

For growth curves, monolayers of PS cells in 96-well plates were infected with viruses at a multiplicity of infection (moi) of 1 PFU/cell, in quadruplicates. The inoculum (30 µl) was removed after 1 h, the monolayer
washed thoroughly and replaced with 200 μl of media containing 2% serum. Media (10 μl) was collected at different time-points (8, 12, 16 and 24 h post-infection) and stored frozen at −70°C, before virus quantification by plaque assay. For cytopathic effect (cpe), PS cell monolayers were infected in 96-well plates at an moi of 1 PFU/cell, in quadruplicates, and stained with naphthalene black after 72 h.

**Statistical analysis**

Statistical analysis was performed on the data obtained from the virus growth curve studies and the evaluation of cpe in PS cells. For growth curves, the data were plotted to include the standard error of the mean (SEM) for each data set. At any given time point divergence by at least 2 SD from the mean, between wild-type and mutant viruses, was taken as significant. Measurement of cpe was done visually by three independent evaluators in a ‘blind’ manner. The cpe of viruses were estimated on a scale of 1–4 corresponding to 20–40, 40–60, 60–80 and 80–100% of monolayer destruction following microscopic examination. The interevaluator consistency error was verified using F-test which revealed no one evaluation was significantly different from that of the others.

**RESULTS**

**In silico analysis of RNA structures in flaviviruses**

Previous in silico studies have predicted a stable RNA structure designated SL6, in the C protein-encoding region, for a limited number of viruses within the mTBFV subgroup (16, 26–28). Structural RNA elements were also revealed in the C region of some MBFV (28) although their homology to SL6 of TBFV had not been established. Here, we utilized a variety of independent structure prediction methods and a much larger sample of viral sequences to analyze whether or not the SL6-like structure was conserved throughout the entire genus Flavivirus.

**In silico analysis of SL6 in the TBFV subgroup**

It was found that the overall folding of the first 333 nt of TBFV was highly conserved among several members of the mTBFV subgroup (16, 26–28), with six stable SLs (enumerated 1–6 in Figure 1A). This analysis was extended to investigate the conservation of SL6 in the larger group of distantly related mTBFV, sTBFV and KADV (37). A nucleotide alignment of the C region was generated and optimized by the introduction of numerous gaps (Supplementary Figure S1A); it shows that divergent RFV, GGYV and KSIV (distant virus species of the mTBFV) maintained homology in the SL6 region. However, some nucleotide perturbations in the SL6 region were observed between mTBFV, sTBFV (MEAV, TYUV and SREV) and KADV proving that the region between the initiation codon and SL6 had evolved with frame shifts as we previously demonstrated (16). We conducted MFold analysis to investigate the presence of SL6-like structures in the distantly related mTBFV (RFV, GGYV and KSIV), sTBFV and KADV groups (Figure 1B).

Despite sequence divergence, all viruses in the mTBFV group formed similar SL6-like structures when the 333 nt or a longer nucleotide region (up to 1000 nt) was used for MFold analysis (data not shown). The SL6-like folds contained a remarkably high number of co-variant and semi-covariant substitutions which maintained the general conformation across divergent viruses (Figure 1B). The minimum free energy dG of folding for SL6 varied between −32.3 and −17.2 kcal/mol with RFV and LIV/GGYV as extremes in this range. Although KADV had a shorter SL6 compared with other TBFVs, the energy of folding was −17.32 kcal/mol, within the range found for mTBFV.

In comparison to SL6 of the mTBFV, the SL6 of sTBFV was shorter and less stable, with a dG in the range −12.5 to −10.6 kcal/mol (Figure 1B). However, the SL6-like structures of sTBFVs were observed as elements of longer and branched RNA conformations (data not shown).

A smaller terminal loop was revealed in the KFDV/AHFV and KSIV sequences resulting in the formation of the tetraloop U(GCCA)A (Figure 1B). The presence of U:A as a loop-closing base pair has been shown to decrease tetraloop stability considerably; in combination with some intraloop sequences this results in intermolecular tensions that prevent the folded tetraloop from achieving a global thermodynamic minimum (38,39). Thus, despite the MFold-mediated predictions, a tetraloop may not form for KFDV/AHFV and KSIV or at least not be sufficiently stable for biologically significant (RNA–RNA or RNA–protein) interactions.

The conservation of a UGCCAA hexanucleotide motif in the terminal loop of SL6 in all the divergent TBFVs was striking. Both TYUV and KADV showed one substitution in the hexanucleotide UGCCUA; TYUV has also lost the first nucleotide (Supplementary Figure S1A).

In the minus-sense orientation, the conservation of an SL6-like structure was not as robust as in the positive-sense. Although most of the TBFVs formed a structure in the minus-sense RNA, the number of hydrogen bonds, the lengths of the stems and free minimal energy of folding varied significantly even between closely related viruses (data not shown). Consequently, the formation of SL6 is likely to be biologically significant only in the positive-sense RNA.

Structure predictions correlated with evidence for SSSV in TBFV genomes (Figure 2). A remarkable drop in SSSV was observed in the SL6 region between positions 209 and 254 of the Vs sequence. The most extreme drop in variability was observed in a window centred on position 221 within the apical stem of SL6. The levels of SSSV within the remainder of the structural protein-encoding region (positions 295–2435) were higher than the upstream portion. Similarly, high levels of SSSV were observed across the non-structural portion of the genome between positions 322 and 2425 (data not shown).

We excluded the possibility that SSSV in the C-coding region was due to codon bias by analyzing nucleotide composition at each position within the codons.
Figure 1. Conservation of SL6 among the TBFVs. (A) MFold-simulated RNA secondary structure (stable stem loops numerated 1–6) between nucleotides 1–333 of the Vs virus genome. The 5’-CYCL, initiation AUG codon and conserved hexanucleotide UGCCAA are outlined. (B) Comparison of SL6 between TBFV species. Numeration in brackets corresponds to SL6 numbered from the start codon of each virus (abbreviated in Table S1). Free energy dG values of folding are shown in kcal/mol. Covariant and semi-covariant substitutions are underlined on Vs virus.
No unusual variation of G/C or purine/pyrimidine composition was observed at the third codon position or at positions one or two of the codon (not shown). Likewise, we analyzed the dinucleotide composition at all three possible positions. Although there was a general under-representation of CpG and UpA, and over-representation of CpA and UpG, there was no correlation between areas of SSSV and regions of unusual dinucleotide frequencies (data not shown). These results indicated that evolutionary constraints restrict nucleotide variation within the 5'-coding regions of flavivirus genomes.

The phylogenetic conservation of thermodynamically stable RNA structures across all TBFV group ORFs was further analyzed using the program STRUCTURE_DIST (Figure 2) (40). This method quantifies phylogenetically concordant structures predicted using the widely accepted MFold or UNAFold algorithms, which can then be aligned and overlaid with SSSV results (31,32). Analysis of the entire ORF showed the most striking evidence for conserved base-pairing between the initiation codon at position 133 and position 318, after which a large drop in the frequency of base-paired nucleotides was observed. Within this region SL6 was predicted to be the most significant structure, with conserved pairing between 209 and 254 centred on a region with a conserved lack of base pairing between positions 228 to 236, representing the unpaired apical loop of SL6. The base-paired stem of SL6 contained conserved short single-stranded regions between positions 218–220 and nucleotides 244 and 245 consistent with the unpaired bulge, either side of the paired stem. This corresponds exactly to the position and structure of SL6 predicted by MFold (Figure 1 and Supplementary Figure S1A).

**In silico prediction of SL6 in the MBFV group**

An annotated nucleotide alignment of the C-coding region between TBFV and three MBFV groups (JEV, DENV, and others) was analyzed using the program STRUCTURE_DIST (Figure 2) (40). This method quantifies phylogenetically concordant structures predicted using the widely accepted MFold or UNAFold algorithms, which can then be aligned and overlaid with SSSV results (31,32). Analysis of the entire ORF showed the most striking evidence for conserved base-pairing between the initiation codon at position 133 and position 318, after which a large drop in the frequency of base-paired nucleotides was observed. Within this region SL6 was predicted to be the most significant structure, with conserved pairing between 209 and 254 centred on a region with a conserved lack of base pairing between positions 228 to 236, representing the unpaired apical loop of SL6. The base-paired stem of SL6 contained conserved short single-stranded regions between positions 218–220 and nucleotides 244 and 245 consistent with the unpaired bulge, either side of the paired stem. This corresponds exactly to the position and structure of SL6 predicted by MFold (Figure 1 and Supplementary Figure S1A)
Figure 2. Phylogenetic conservation of RNA structures in the TBFV and MBFV groups across the 5'-UTR and structural protein coding regions. Black lines represent SSSV for TBFV, DENV, JEV and YFV. Gray filled bars represent phylogenetic conservation of base pairing from a STRUCTURE_DIST pair-wise comparison of connect files generated by Mfold.
and YFV) was constructed based on a previously presented alignment (16) but modified to include newly sequenced distantly related mTBFV, sTBFV and KADV isolates (Supplementary Figure S1A). The C protein TBFV/MBFV alignment (available on request) was used to anchor the divergent nucleotide sequences. The annotations include the 5'-CYCL of MBFV, an 8 nt long cyclisation domain highly conserved between all MBFVs (16). The 5'-CYCL interacts with a complementary sequence 3'-CYCL in the 3'-UTR to form a dsRNA panhandle, a vital element of the replication promoter that initiates viral RNA synthesis (16). For the TBFV the 21-nt long 5'-CYCL is located in the 5'UTR (i.e. outside the alignment in Supplementary Figure S1A; highlighted in Figure 1A). The 5'-CYCL for MBFV mapped to the capsid gene and, among the TBFV, aligns optimally with a region that is identified only in TUYV (Supplementary Figure S1A).

Nucleotide sequence homology was observed between the TBFVs and MBFVs particularly in the SL6 region of some JEV group viruses. For example, WNV was observed to share both the stem and loop sequences of some JEV group viruses. For example, WNV was the TBFVs and MBFVs particularly in the SL6 region of TBFV. This alignment was further annotated with available sequences, no thermodynamically stable RNA structures were observed in the region corresponding to the TBFV SL6 region. However, an SL6-like structure (Supplementary Figure S1A and Figure 2) was predicted in all MBFV groups, although it was less stable in the DENV group, particularly the DENV3 serotype.

In silico prediction of SL6 in the NKV and PABV groups

In contrast to TBFV and MBFV, the NKV and PABV groups are not arboviruses and their replication is limited to only one natural host, i.e. rodents/bats (NKV) or mosquitoes (PABV). The high nucleotide divergence (Supplementary Figure S1B and S1C) and limited number of complete published sequences for members of the NKV and PABV groups precluded the use of both phylogenetic and thermodynamic approaches to RNA structure prediction. When MFold analysis was performed with available sequences, no thermodynamically stable RNA structures were observed in the region corresponding to the TBFV SL6 region. However, an SL6-like structure, with a similar apical loop CCAA motif was observed in KRV (PABV), upstream of the analogous TBFV SL6.

Experimental evidence supporting the predicted structure lSL6

Strategy of mutagenesis on stem–loop 6. Initial design of mutations focused on synonymous codon positions. However, in all but a few instances, this was limited due to the distinctive sequence organization of the apical loop and base paired stem. The first and third codons of the conserved MPN tripeptide (loop region) are limited in respect of variation; M could not be changed and N has two possible silent variations both of which are outside the apical loop (Figure 3). Consequently, when mutating the terminal loop sequence UGCCAAAU, silent substitutions immediately either side of the region of maximum SSSV suggesting that they are less conserved than DV3 (Supplementary Figure S1A). STRUCTURE_DIST also predicts the formation of the DV2 and DV3 but not the SL6-like structure (Supplementary Figure S1A and Figure 2). However, a truncated SL6-like structure was predicted to form in all DENV serotypes, albeit at a sub-optimal energy level, when the SL6-like region was folded independently from neighboring regions that form more stable overlapping structures (Supplementary Figure S1A). Taken together, these data indicate that the DENV SL6-like structure was the least stable conformation among the MBFVs, potentially preventing its prediction by statistical approaches used here and elsewhere (28). Despite this, the short-stem region of putative DENV SL6-like structures is highly conserved within the DENV group (DENV serotypes 1–4) and also between DENV and JEV (Supplementary Figure S1A) suggesting that a linear or conformational signal at this location might have some functionality.

A similar restriction in SSSV was observed in the YFV C-coding region, with maximum SSSV corresponding to the ALIDOT-predicted structure YF4 (Figure 2) (28). Among the MBFVs, only the YFV SL6-like structure was predicted by both thermodynamic and phylogenetic methods. In summary, a proximally truncated SL6-like structure was predicted in all MBFV groups, although it was less stable in the DENV group, particularly the DENV3 serotype.
could only be introduced into the P codon. Similar difficulties were encountered with mutagenesis of the stem, in which the vast majority of possible synonymous and non-synonymous mutations resulted in no significant conformational changes. The MFold-simulated folding of numerous SL6-mutants revealed a high level of evolutionary protection of SL6 against spontaneous single mutations (not shown) and provides additional evidence for the maintenance of SL6 functionality.

In order to resolve the difficulties with design of mutations, three different approaches were adopted (Figure 3). First, we introduced all possible silent substitutions, to target the conserved hexanucleotide and the stem (mutants C12, C13, C14, C16 and C33). Second, we introduced mutations (C10, C15, C17, C19 and C34) that mimicked ‘natural’ amino acid substitutions observed in this region of other mTBFV spp. Third, as a control for mutations that changed amino acids we also introduced compensatory substitutions encoding the same mutated amino acids while restoring the SL6 structure. Accordingly, mutations R32, S31, N28, V39, V39 and P28 were designed as controls for non-synonymous mutants C22, C23, C27 and C34 (Figure 3).

The predicted impact of each substitution (Figure 3) on the secondary structure of SL6 is shown in Figure 4. The plaque characteristics, cpe and growth dynamics of each mutant compared with those of original pGGVs virus (Table 1 and Figure 5). Single-step growth curves revealed differences of ~1 log10 between the mutants early after infection (12–16 h p.i.) which were reproducible and statistically significant (Figure 5).

To exclude the effect of spontaneous mutations in the 5′- and 3′-UTRs which contain TBFV promoter and enhancer elements (16) that might compensate for the effect of the SL6-mutations, rescued virus was not passaged prior to phenotype evaluation and key regions of the genome (1–940 and 10206–10927) were sequenced following recovery of each SL6-mutated virus. Only the intended substitutions were present, with no reversions or other compensatory mutations were observed. The effect of each mutation (reduction from large wild-type plaques to medium, small or pin pointed) was scored if the SL6-mutated strain contained >90% of plaques with the altered morphology. The presence of a minor plaque population (between 1 and 10%) was considered as the inevitable result of the variation inherent

---

**Figure 3.** Site-directed mutagenesis of SL6. Nucleotide (A) and amino acid (B) alignments of SL6 sequences within C gene of mTBFV and mutants produced from the infectious clone of TBEV. Viruses are specified in Supplementary Table S1. Sequence of Vs virus is on the top; nucleotide and amino acid substitutions are shown by letters. The conserved hexanucleotide UGCCAA is underlined. SL6 region is located between nucleotides 209 and 254 and amino acids 26–41 of the Vs virus genome. Unpaired terminal loop of SL6 is highlighted by a gray box.
in all RNA viruses, the consequence of a high error rate in the virus RdRp (41).

**Sequence changes in the apical loop of SL6.** In mutants C12, C13, C14, C16 and C19 substitutions within the apical loop changed the nucleotide sequence without altering the overall conformation (Figures 3 and 4). Four of these mutations C12, C13, C14 and C19 were introduced into the conserved hexanucleotide UGCCAA. Silent mutations C12, C13 and C14 changed plaque morphology; the C13 and C14 mutants that contain purine-to-pyrimidine substitutions also showed reduced growth characteristics and cpe (Table 1 and Figure 5). Silent substitution C16, located outside the conserved hexanucleotide in SL6, did not affect the virus phenotype. Two purines were changed for two pyrimidines in mutant C19, one in the conserved hexanucleotide. This mutant was highly attenuated in cell culture producing no cpe and a small turbid plaque phenotype (Figures 3–5 and Table 1). These two purine–pyrimidine substitutions resulted in the amino acid substitution M33L that mimicked the corresponding natural amino acid in KFDV and AHFV (Figure 3). Nevertheless, M33L, mutant C15, produced by different nucleotide
substitutions had only a moderate effect on virus replication (below). Therefore, the biological consequences of mutation C19 may be attributed, at least partially, to the nucleotide substitutions.

Conformational changes to the apical loop of SL6. Mutations C10, C11, C15, C21, C22 and C23 changed the shape of the loop and base-paired stem within SL6 (Figure 4). Replication of mutant C10 (with an enlarged loop and shortened stem) and C17 (restored wild type conformation due to the second compensatory mutation) was delayed in the early stage of the infection cycle; C17 caused slightly reduced cpe but the plaque morphology of both was equivalent to that of the parental pGGVs virus (Figure 5 and Table 1). The minor phenotypic changes resulting from these mutations could be explained by the accompanying amino acid substitutions N35→K and Q32→P imitating POWV (Figure 3). However, a silent substitution A234→G that also enlarged the apical loop of mutant C11 (Figure 4)

Table 1. Affect of mutations within the SL6 on TBEV phenotype

<table>
<thead>
<tr>
<th>Virus</th>
<th>pGGVs</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
<th>C14</th>
<th>C15</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT/AA*</td>
<td>Wild type</td>
<td>T227G/ N35K</td>
<td>T227C</td>
<td>A234G</td>
<td>A234C</td>
<td>A234T</td>
<td>A234T/T235K</td>
</tr>
<tr>
<td>Plaque phenotype</td>
<td>Large/bright</td>
<td>Large/bright</td>
<td>Large/bright</td>
<td>Small/bright</td>
<td>Small/bright</td>
<td>Medium/turbid</td>
<td>Medium/turbid</td>
</tr>
<tr>
<td>Plaque morphology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CPE</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus</td>
<td>C16</td>
<td>C17</td>
<td>C19</td>
<td>C21</td>
<td>C22</td>
<td>R32</td>
<td>C23</td>
</tr>
<tr>
<td>Plaque phenotype</td>
<td>Large/bright</td>
<td>Large/bright</td>
<td>Pinpoint/turbid</td>
<td>Medium/turbid</td>
<td>Medium, turbid</td>
<td>Large, bright</td>
<td>Pinpoint, turbid</td>
</tr>
<tr>
<td>Plaque morphology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CPE</th>
<th>4</th>
<th>3</th>
<th>1</th>
<th>3</th>
<th>3</th>
<th>4</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus</td>
<td>S31</td>
<td>C27</td>
<td>N31V9</td>
<td>V9</td>
<td>C33</td>
<td>C34</td>
<td>P38</td>
</tr>
<tr>
<td>NT/AA*</td>
<td>G223A/ N31S T224G; A236C; C241T; C245G</td>
<td>C216A/ Q22N A215T</td>
<td>C216A/ Q22N T241G/L245V G246T; A215C</td>
<td>T241G/L245V G246T</td>
<td>C253A G255G</td>
<td>A215C/ Q22P A216C; T241G; C246C</td>
<td>A215C/ Q22P</td>
</tr>
<tr>
<td>Plaque phenotype</td>
<td>Large, bright</td>
<td>Pinpoint, turbid</td>
<td>Large, bright</td>
<td>Large, bright</td>
<td>Pinpoint turbid</td>
<td>Pinpoint turbid</td>
<td>Large, bright</td>
</tr>
<tr>
<td>Plaque morphology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CPE</th>
<th>4</th>
<th>2</th>
<th>4</th>
<th>4</th>
<th>1</th>
<th>1</th>
<th>4</th>
</tr>
</thead>
</table>

Plaque size for each mutant was defined as large (5–6 mm), medium (3–4 mm), small (1–2 mm) or pinpointed (>1 mm). Some plaques, in comparison with parent Vs virus, were described as turbid. The cpe produced by each mutant in comparison to the wild-type virus was evaluated on a scale of 0–4 where 0 indicates no cpe and 4 is maximum cpe (i.e. 80% cell lysis as observed for the control pGGVs virus) in five repeated experiments, each in quadruplicates. NT/AA* - Nucleotide/amino acid substitutions.
caused similar biological effects; it did not affect virus plaque size or level of cpe, but reduced virus replication rate early after infection (Figure 5 and Table 1).

Mutation C15, which shortened the apical loop (Figure 4), did not affect virus growth but changed the plaque morphology and delayed the development of cpe (Table 1 and Figure 5). The C15 mutation altered the amino acid M33→L, which imitates the KFDV/AHFV group (Figure 3), potentially contributing to the observed biological effect. Mutation C21 that reduced the apical loop size (Figure 4) interfering with exposure of the hexanucleotide, also had a moderate affect on virus growth although the accompanying effect of amino acid substitution Q32→H (Figure 3) cannot be excluded (Table 1 and Figure 5).

Mutant C22 contained three substitutions that considerably increased the size of the apical loop thus shortening the base paired stem. Three nucleotide substitutions present in mutant C23 had the opposite affect in shrinking the apical loop (Figures 4). Both C22 and C23 had altered growth dynamics, plaque morphology and cpe (Table 1 and Figure 5). The nucleotide substitutions of both led to amino acid substitutions Q32→R and V31→S, respectively. To exclude their influence on virus growth, counterpart ‘control’ mutants R32 and S31 were analyzed, with the same amino acid substitutions but without alteration of the SL6 conformation (Figures 3 and 4). Both of these control mutants exhibited wild-type plaque morphology and cpe characteristics (Table 1).

Substitutions in the stem of SL6. Three mutants were designed to investigate the influence of SL6 stem length. Most attempts to design synonymous substitutions had little effect on the stem folding conformation. Only silent
mutant C33 (C253→A and C255→G) exhibited a significantly shortened duplex stem, with a corresponding elevated level of dG folding energy. These positions are highly conserved among the mTBFV (Figure 3) and, as expected, had a profound effect on virus replication; C33 displayed pinpoint plaques, reduced growth characteristics and almost no cpe (Table 1 and Figure 5).

Two other mutants C27 and C34 had shortened stems due to the formation of a large internal bulge (Figure 4), and exhibited profoundly altered biological characteristics (Table 1 and Figure 5). However, C27 and C34 included amino acid substitutions Q28→N and Q28→P, respectively, the latter resembling POWV (Figure 3). To rule out the amino acid change as influential, two mutants were designed as a control for C27; double mutant N28V39 and single mutant V39, neither of which affected SL6 conformation (Figure 4). Similarly mutant P28, a control for C34, contained the same amino acid substitution Q28→P but maintained SL6 conformation (Figures 3 and 4). All three control mutants, N28V39, V39 and P28, displayed wild-type large plaque phenotype and cpe (Table 1).

DISCUSSION
In a previous study using MFold-simulated RNA structures for a limited number of mTBFV species, we predicted the existence of SL6 in the C-coding region of TBFV. A conserved hexanucleotide UGCCAA in the apical loop and compensatory mutations in the duplex stem of the SL6 implied the formation of the stable RNA structure in ORF of the TBFV genomes (26,27). However, contradicting these findings a deletion within the C-coding region, which included SL6, did not prevent recovery of viable, albeit attenuated, TBEV (42). In this study, we employed a variety of complementary phylogenetic and thermodynamic methods to examine the evolutionary conservation of SL6 using a much larger sample of significantly divergent TBFVs, including new members of the mTBFV, sTBFV and Kadam subgroups (37). The viruses in the other ecological groups, namely MBFV, NKV and PABV were also included in this analysis to trace the evolution of SL6 throughout the entire genus Flavivirus. In addition, we used a reverse genetic system (34,36) to engineer TBEV strains with mutated SL6 to reveal the biological significance of this structure.

Thermodynamic and phylogenetic analysis of large sequence data sets indicated that all TBFVs including even the distantly related mTBFV, KADV and sTBFV form an SL6-like structure with an exposed conserved hexanucleotide although the molecular details of the predicted stem–loop varied among the mTBFV, sTBFV and KADV subgroups (Figure 1B). In similar manner, an SL6-like structure has been predicted for MBFV although with less stability in comparison to SL6 in TBFV. Two other flavivirus groups NKV and PABV demonstrated no significant sequence homology with the TBFV SL6 region although the genome of KRV (PABV group) formed a thermodynamically stable structure in close vicinity to the TBFV SL6 with a similar terminal loop motif CCAA (TBFV—UGCCAA) (Supplementary Figure S1).

To test the biological significance of SL6 in the TBFV group we engineered 21 mutant viruses with point mutations that altered the linear sequence of the unpaired apical loop or destabilized the base-paired stem. Substitutions within the conserved hexanucleotide loop down-regulated virus growth kinetics whereas changes in the terminal loop outside the hexanucleotide sequence did not alter the observed phenotype. The most significant changes of virus phenotype resulted from substitutions that distorted the stem of SL6; mutations that influenced the length or stability of the stem resulted in the recovery of viruses that formed small and/or turbid plaques. Increasing or decreasing the size of the apical loop had a minor biological effect on virus replication although this could also be interpreted as an effect of the altered stem length. However, the changes in replication kinetics from all modifications of SL6 were moderate and manifested themselves predominantly during the early stage of the virus replication cycle (Table 1).

Previous analysis of RNA secondary structure across the Flavivirus genus led to the concept of promoter and enhancer elements that initiate assembly of the virus polymerase complex (16–18,23,27,43,44). Enhancers were identified as RNA structures that individually produce only small biological effects on virus replication. However, the significance of enhancers as targets for the attenuation of flaviviruses to engineer live vaccines is evident from the example of dengue virus (24,25). Moreover, sequence and structural conservation of flavivirus enhancers is consistent with a role as key players in virus survival in the natural environment. We previously proposed that the cumulative action of several enhancer elements could contribute significantly to the overall rate of assembly of polymerase complexes, thereby enhancing virus survival across a range of natural hosts (17,18,23,27,43,44). In this respect, the presented experimental data indicate that SL6 belongs to the category of REEs, i.e. RNA structures that accelerate the replication of viruses (45–56). This eliminates the apparent contradictions between extremely high levels of SL6 conservation across divergent TBFV virus species and the redundancy of this element for the replication of laboratory-maintained TBEV strains (45–56). However, the specific mechanism by which SL6 functions to enhance virus replication remains to be elucidated.

It has recently been demonstrated that a short but highly conserved RNA hairpin (sHP) localized in the 3′-UTR of DENV2 RNA regulates the transition from a circular (required for the initiation of RNA replication) to linear RNA form during the progress of viral RNA synthesis (57). The SL6-like structure of MBFV is localized immediately downstream of the 5′-CYCL (i.e. within the capsid gene, Supplementary Figure S1A) suggesting it could also contribute to genome circularization. It is possible that in accord with the 3′ sHP (highly conserved throughout the genus Flavivirus), it contributes to the unpairing of the 5′–3′-CYCL panhandle, to promote RNA elongation on the linear template. In contrast, the
5’-CYCL of TBFV is mapped to the 5’-UTR (i.e. upstream of the capsid gene, Figure 1A) and therefore other tentative functions of the TBFV SL6 are not excluded, such as enhancing virus translation, RNA replication or playing a role in regulation between these processes; the possibility of a kissing-loop enhancer of genome circularization was previously discussed (17,18,23,27,43,44).

The C protein of flaviviruses is highly basic at the N-terminus, specifically binding virus genomic RNA during encapsidation and plausibly acting as an RNA chaperone as shown for other viruses (58). The sequence of SL6 within the C coding region localizes to the junction of the positively charged domain and a following hydrophobic domain that interacts with the virus envelope proteins during assembly (42). It is possible that additional synonymous codon flexibility may be accommodated in this region due to the requirement to conserve the charge or hydrophobic characteristics of the domain, rather than any specific amino acid sequence.

Although our studies provide support for the REE role of SL6 in TBEV it is unclear if SL6-like structures of MBFV act similarly as functionally significant REE. However, the remarkable resemblance of the WNV SL6-like structure to TBFV SL6 suggests that it might serve a similar function, at least in one virus group. However, a final conclusion for the MBFV and also for the more distant NKV or PABV groups is not possible ahead of further functional studies.

Being arboviruses, MBFVs and TBFVs are adapted for transmission between distantly related vertebrate hosts and invertebrate vectors. The requirement to adapt to different molecular environments might result in the evolution of enhancer elements essential for virus replication in one host while being redundant in another. This could explain the contradiction between strict conservation of the different flavivirus enhancers and their apparent redundancy in laboratory systems, which are largely based on mammalian cells (17,18,23,27,43,44). Mutations in SL6 described here have demonstrated its enhancer properties in mammalian cells and it will be interesting to evaluate SL6 enhancer activity in ticks, the major host for maintenance of the TBFV group in the environment (59–61).

In conclusion, bioinformatic analysis demonstrated the presence of a conserved RNA secondary structure in the C coding region of the divergent TBFV group. Disruption of this structure compromised virus replication implying an REE function for SL6. By homology with the TBFVs, SL6-like structures were observed in the genomes of some MBFVs and plausibly indicate a similar role as replication enhancers. Future studies using sub-genomic replicons will allow direct measurement of the influence of these sequences on RNA replication and on interaction with viral and host proteins.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank M. Armesto for technical assistance in the experimental part of the project and Prof. Peter Simmonds for technical assistance in the bioinformatics and critical reading of the manuscript.

FUNDING
The Biotechnology and Biological Sciences Research Council grant BBS/B/00697, the Sixth European Union Framework grant (VIZIER EU Contract LSHG-CT-2004-511960); Medical Research Council grant (ID 87652 to T.S.G. and I.M.J.); M.R.C. support (to D.J.E.). Funding for open access charge: Medical Research Council grant (ID 87652)

Conflict of interest statement. None declared.

REFERENCES


a cis-acting replication element within the poliovirus coding region. J. Virol., 74, 4590–4600.