1194–1202 *Nucleic Acids Research, 1997, Vol. 25, No. 6*

Secondary structure of the 3′ untranslated region of flaviviruses: similarities and differences

Vitali Proutski*, Ernest A. Gould[1] and Edward C. Holmes

The Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK and [1]NERC Institute of Virology and Environmental Microbiology, Mansfield Road, Oxford OX1 3SR, UK

Received December 2, 1996; Revised and Accepted January 31, 1997

ABSTRACT

Genetic algorithm-based RNA secondary structure prediction was used in combination with comparative sequence analysis to construct models of folding for the distal part of the 3′-untranslated region of flaviviruses belonging to four serological groups. Elements of RNA secondary structure that are preserved among all the flaviviruses studied were revealed, despite the high degree of sequence divergence between them. At the same time, structural elements were observed that distinguish members of different serological groups and, in particular, a region of remarkable structural divergence between the tick-borne and mosquito-borne flaviviruses was found. Application of the genetic algorithm also revealed that the 3′-terminus of flaviviral genomic RNA may take on alternative conformations, which are not observed in the 3′-terminus of complementary minus strand RNA. These alternative folding patterns may have roles in the regulation of transcription and translation initiation and in the switch between them.

INTRODUCTION

Flaviviruses share common morphological features of the virion, antigenic determinants and genomic organization. Most of the flaviviruses are arthropod-borne and can be divided into two major groups in this respect: those transmitted by mosquitoes (mosquito-borne) and those where ticks are the main vector (tick-borne). Originally the flaviviruses were classified on the basis of serological relatedness (1,2), a classification which has generally been confirmed by phylogenetic analysis of viral genomic sequences (3,4), although some flaviviruses, for example yellow fever virus (YF), remain unclassified at the serological level (5).

All flaviviruses have a single-stranded RNA genome of positive polarity which is ~11 kb in length. Genomic RNA is the only virus-specific messenger RNA in flavivirus-infected cells and encodes a single open reading frame of ~10 kb. Translation of genomic RNA results in a single large polyprotein precursor, which undergoes proteolytic cleavage and gives rise to 10 mature viral proteins (6,7).

Flanking the coding region are the 5′- and 3′-untranslated regions (UTRs). The 5′-UTR of flaviviruses is relatively short (95–132 bases in length), while the 3′-UTR is usually longer but demonstrates extensive heterogeneity in size and sequence between different viral species and even among different strains within the same species (8–10). This divergence is primarily concentrated within the proximal part of the 3′-UTR following the stop codon, where long deletions, insertions, sequence repeats and even poly(A) stretches have been observed (8,9,11). At the same time, the distal part of the 3′-UTR (~330–400 nt in length) exhibits relatively high sequence identity among different strains of the same viral species and even among the members of serological groups (8–10,12). Interestingly, some strains of tick-borne encephalitis virus (TBE) have very short 3′-UTRs and demonstrate that the distal part of this region alone may be sufficient for the existence of a viable virus (9), suggesting that it may represent a functional core of the flaviviral 3′-UTR where all or most of the important elements in viral translation, replication and assembly are concentrated. Though the precise mechanisms underlying these functions are currently not well understood, some conserved elements of RNA primary and potential secondary structure have been identified. In particular, computer-predicted folding patterns and RNase cleavage experiments have demonstrated that the marginal 3′-terminal nucleotides of all flaviviruses form a long stable hairpin structure (3′-LSH), which preserves its shape despite significant differences in primary structure (8,9,13–17). Such conservation suggests that this structural element has functional importance, a concept supported by the demonstration of a specific interaction between the 3′-LSH of flaviviruses and some host cellular proteins that are thought to be components of the virus replication complex (18) and that it is the secondary structure, which is under selective constraint. Although other conserved RNA sequence motifs have been found within the 3′-UTR of flaviviruses, little is known about the secondary structure upstream of the 3′-LSH. Specifically, while some RNA sequences within the 3′-UTR distinguish mosquito-borne from tick-borne flaviviruses, to date there is no information identifying any structural differences between these two groups.

*To whom correspondence should be addressed. Tel: +44 1865 271272; Fax: +44 1865 310447; Email: vitali.proutski@zoology.oxford.ac.uk

Table 1. GenBank accession nos of sequences used in this analysis

Serological group	Virus	GenBank accession no.
Tick-borne encephalitis	Central European encephalitis (TBE-W)	U27491, U27493, U27494, U27495, U27496, U39292
	Far Eastern encephalitis (TBE-FE)	U27490, U27492
	Powassan (POW)	L06436
Japanese encephalitis	Japanese encephalitis (JE)	D90194, D90195, L48961, M18370, M55506, U14163, U15763
	West Nile (WN)	M12294, L48977
	Murray Valley encephalitis (MVE)	M35172, L48972, L48973, L48974, L48975, L48976
	Kunjin (KUN)	L24512, D00246, L49311, L48978, L48979
Dengue	Dengue type 1 (DEN1)	M87512
	Dengue type 2 (DEN2)	M29095, M19197, M20558, M84727, M84728
	Dengue type 3 (DEN3)	M93130
	Dengue type 4 (DEN4)	M14931
Ungrouped	Yellow fever (YF)	U52423, U52420, U52417, U52414, U52411, U52407, U52401, U52399, U52396, U52390, U21056, U21055, X02807, U17067, U17066

In the present study we constructed models of potential RNA secondary structure for the distal part of the flaviviral 3′-UTR, including the region upstream of the 3′-LSH. These models reveal a number of structural elements that preserve their conformation in most of the flaviviruses, as well as structures that characterize the members of different serological groups. The possible functional implications of some of these structures are discussed.

MATERIALS AND METHODS

Flaviviral 3′-UTR nucleotide sequences

We used complete or partial sequences of the 3′-UTR from both tick-borne and mosquito-borne flaviviruses belonging to three serological groups: Japanese encephalitis (JE), dengue (DEN) and tick-borne encephalitis (TBE) (5). We also analysed the 3′-UTR sequences of wild and vaccine strains of yellow fever virus (YF) (ungrouped according to the classification of Calisher *et al.*; 2) that had previously been used for the construction of a RNA secondary structure model of this virus (Proutski *et al.*, manuscript submitted). This means that all flaviviral 3′-UTR sequences currently available in the GenBank database were utilized. The GenBank accession nos of the sequences used are given in Table 1.

The sequences were grouped according to their serological classification and then aligned using the ClustalW program (19) and subsequently corrected manually. Alignments are available from the authors upon request.

Secondary structure prediction and comparative analysis

The possible folding of the distal part of the 3′-UTR of flaviviruses was predicted using a genetic algorithm (GA) implemented in the STAR program (20,21). The GA has several advantages over the most widely used algorithms for secondary structure prediction, which are based on a search for the minimal free energy state (22). Firstly, the GA simulates the natural folding pathway which takes place during RNA elongation. This not only enables new stems to be added to the growing RNA chain, but also allows structures to be removed at later stages of the simulation if some other pairings are found to be more favourable. The GA also allows the user to follow the folding pathway and reveal the metastable structures that appear during RNA synthesis, which might have some functional importance despite being disrupted at later stages of synthesis (23). Finally, the GA allows prediction of some tertiary interactions.

In this study we performed secondary structure prediction for each sequence and then compared the results of prediction manually and by means of a programing module, CovarSearch (available from the authors on request), which was developed to reveal possible compensatory mutations in the sequences (covariations). The program uses an alignment file as input data and outputs a list of covariant positions within the alignment taking into account only the substitutions that occur in both strands of the possible helix (G-C→G-U substitutions were not counted as covariant). Currently two compensatory mutations are considered to be sufficient to prove the stem region of a hairpin. In this study we revealed a number of elements of secondary structure containing two or more compensatory mutations between several related sequences. Structures with one site of compensatory mutation were also considered preferentialy over those containing no covariant nucleotides. All structures containing compensatory mutation were considered as being under selection pressure and, therefore, functionally important.

Manual comparison of the predicted secondary structures allowed us to reveal and remove from the analysis structural elements that occurred occasionally in individual sequences but retain those structures, although not supported by compensatory mutations, that exhibited similar conformational features among sequences.

Structure prediction and comparative analysis were performed twice. First, whole sequences of the region in question were analysed so that regions with independent structures that did not interfere with the structures from other regions could be defined. Structure prediction and comparison was then repeated for each region to identify possible variants of folding and to construct a general model of secondary structure. Secondary structure models for each sequence in analysis are available from the authors upon request.

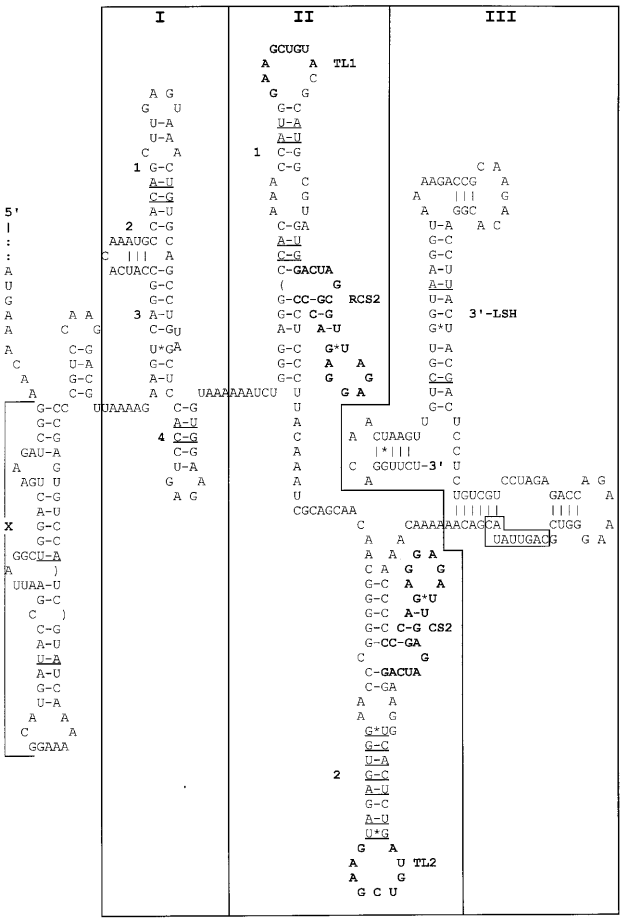


Figure 1. Proposed secondary structure for sequences of the 3′-UTR of DEN viruses. The sequence used for construction of the secondary structure model is from the DEN2 virus Jamaica strain (GenBank accession no. M20558), although very similar structures are found in all DEN viruses. The conserved sequence motifs (CS2, RCS2, TL1 and TL2) are shown in bold. The covariant nucleotides are underlined. The putative cyclization sequence is boxed.

Figure 2. Proposed secondary structure for sequences of the 3′-UTR of viruses from the JE serogroup. The sequence used for construction of the secondary structure model is from the JaOArS982 strain of JE virus (GenBank accession no. M18370), although very similar structures are found in all JE serogroup viruses. The conservative sequence motifs (CS2, RCS2 and TL2) are shown in bold. Other regions and nucleotides of importance are denoted as in Figure 1.

RESULTS

Three structural regions could be defined within the distal part of the flaviviral 3′-UTR, each containing independent elements of secondary structure that are not likely to interfere with the elements from other regions (Figs 1–4).

Region I

Region I of all flaviviruses can form a long hairpin with a branching stem–loop structure or, in few cases, a bulge–loop on the 5′-side of the main hairpin. A number of compensatory mutations within this structure were found among the sequences belonging to each serological group. It should be noted that the sequences of flaviviruses exhibit a high level of divergence in this region, even between different strains of the same viral species. This divergence is reflected in the fact that there are structural variations of region I (Figs 1–4), although for all the flaviviruses analysed we found a structural similarity which greatly exceeded what might be expected given the level of sequence variation. The top stem–loop structure (I-1) displayed the highest conservation among different viruses, while the lateral stem–loop (I-2), whilst showing variation in length, is also present in a majority of viruses, though it is unlikely to exist in some DEN viruses which are more likely to have a bulge–loop.

Following the long hairpin in region I is a relatively short hairpin (I-4), which can be formed in all flaviviruses except those from the TBE serological group. The length and sequence of this structure vary greatly. For YF and DEN viruses, the structure contains covariant substitutions which can be considered as evidence for the existence of this structure.

Upstream of region I in most flaviviruses is a stable hairpin structure (structure X in Figs 1–4). For YF viruses, this structure was short and invariant among all strains. In the case of viruses from the DEN, JE and TBE serological groups this structure was much longer, formed by relatively divergent sequences and contained several compensatory mutations. However, strains RK1424 and 132 of TBE virus (GenBank accession nos U27496 and U27490 respectively) cannot form this structure, since this part of their sequence is replaced by a poly(A) stretch. We presume, therefore, that although this structural element may have a certain function, it is of less importance, so that a viable virus can be formed without this hairpin.

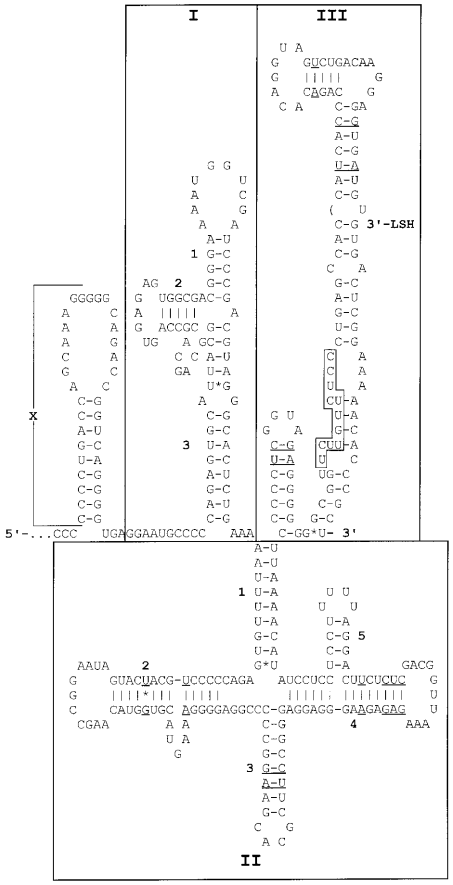


Figure 3. Proposed secondary structure for sequences of the 3′-UTR of the TBE viruses. The sequence used for construction of the secondary structure model is from the Neudoerfl strain (GenBank accession no. U27495), although very similar structures are found in all TBE viruses. Regions and nucleotides of importance are denoted as in Figure 1.

Region II

Unlike region I, where, despite some variation, a structural consensus can be found in all the flaviviruses, region II exhibits differences in the folding patterns which separate flaviviruses into three groups, each with distinct structural features. The first group comprises members of the DEN and JE serological groups. For these viruses, region II forms two symmetrical structures (Figs 1 and 2) each consisting of a long hairpin, with a 6 nt bulge–loop and a stem–loop on the 3′-side of the main hairpin. Interestingly, the duplicated sequences of the lateral bulge– and stem–loop structures (5′-GACUAGAGGUUAGAGGAGACCC-3′) are highly conserved among all DEN and JE serocomplex viruses. These duplicated sequences have been previously described as CS2 and RCS2 (15). In the top loop of hairpin II-2 (TL2) another sequence motif (5′-GAAGCUGUA-3′) is observed, again invariant among all viruses of this group. This motif is also found in the top loop of hairpin II-1 (TL1) of all DEN viruses, while in viruses of the JE serocomplex it was replaced by a non-conserved set of sequences. The stem regions of the main hairpins display a high number of compensatory mutations, demonstrating selection pressure on these structures and strongly suggesting their existence *in vivo*.

The structural pattern of region II constructed for viruses from the TBE serological group (Fig. 3) demonstrates a remarkable difference from that described for the DEN and JE serocomplex viruses, such that no one element of secondary structure or sequence motif is conserved among the mosquito- and tick-borne flaviviruses. Eight sites of compensatory mutation between different strains of TBE and Powassan virus, the most serologically divergent member of the TBE serocomplex (1,2), strongly support our model of folding. Stem II-1 (Fig. 3) is thermodynamically unstable and variable but can be formed in all tick-borne viruses.

Region II of YF viruses can form three stable hairpins (Fig. 4), two of which, II-1 and II-2, contain compensatory mutations. Nucleotides of the loop region of structure II-2 are likely to pair 4 or, in a few cases, 3 nt upstream of this structure, which results in formation of pseudoknot P1. Hairpin II-3 is invariant among all strains of YF. Interestingly, within this structure we have found a sequence motif (CS2) which was also found in the lateral bulge– and stem–loop structures of other mosquito-borne viruses. Despite this similarity, we failed to fold the sequences of region II of YF viruses in any way which would resemble the folding pattern of the other mosquito-borne flaviviruses. Consequently, we conclude that there is a third variant of folding of region II which characterizes YF viruses.

Region III

Region III is the most well-studied distal part of the flaviviral genomic 3′-UTR. A number of authors have reported the results of computer predictions and direct experimental evidence for the existence of a stable and conformationally conserved secondary structure in this region among all flaviviruses and its possible role in viral replication (6,8,9,13–17,23). The models of secondary structure we propose generally comply with these previous findings and show that the secondary structure of the most distal part of the 3′-UTR is not disturbed by pairings with sequences from upstream regions I and II. However, for dengue viruses our model of secondary structure for region III (Figs 1, 5 and 6IV and V) differs from that proposed previously (13,16–18). Specifically, in our model the 3′-LSH common to all flaviviruses is much shorter than has been shown before, although it preserves the upper part of the stem with the characteristic lateral stem–loop structure on top. This discrepancy is explained by the fact that only the truncated part of the DEN 3′-terminal sequence was used previously in determination of the secondary structure (17). The addition of 5′ nucleotides that do not participate in any pairings with upstream sequences results in the formation of several stable stems that prevent the formation of a longer 3′-LSH. For example, the 3′-terminal sequence of DEN3 virus has a free energy of –30.1 kcal/mol according to our model (Fig. 5A), whereas it is significantly higher, –25.6 kcal/mol, for the model previously proposed (Fig. 5B; 13,17,18). This previous model is kinetically as well as energetically unfeasible, since the longer stem can only be formed late in RNA synthesis, which means that energetically more stable structures formed earlier must be disrupted. Recently, experimental evidence for the formation of a pseudoknot structure by the truncated 3′-terminal sequence of DEN3 and other flaviviruses has been published (17). Our calculations also show that for the truncated sequence a structure with this pseudoknot (Fig. 5C) is thermodynamically more stable (–28.0 kcal/mol) and kinetically more likely than the long straight stem (Fig. 5B), although for the longer sequence of the

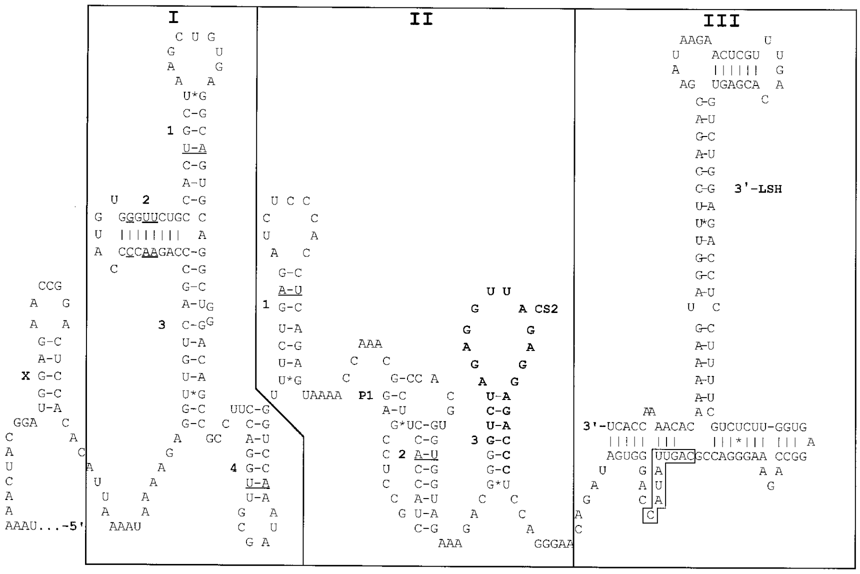


Figure 4. Proposed secondary structure for sequences of YF virus 3′-UTR. The sequence used for construction of the secondary structure model is from strain Trinidad79 (GenBank accession no. U52420). Very similar structures are found in all wild strains of YF. The conservative motif among all mosquito-borne flaviviruses sequence motifs (CS2) is shown in bold. Other regions and nucleotides of importance are denoted as in Figure 1.

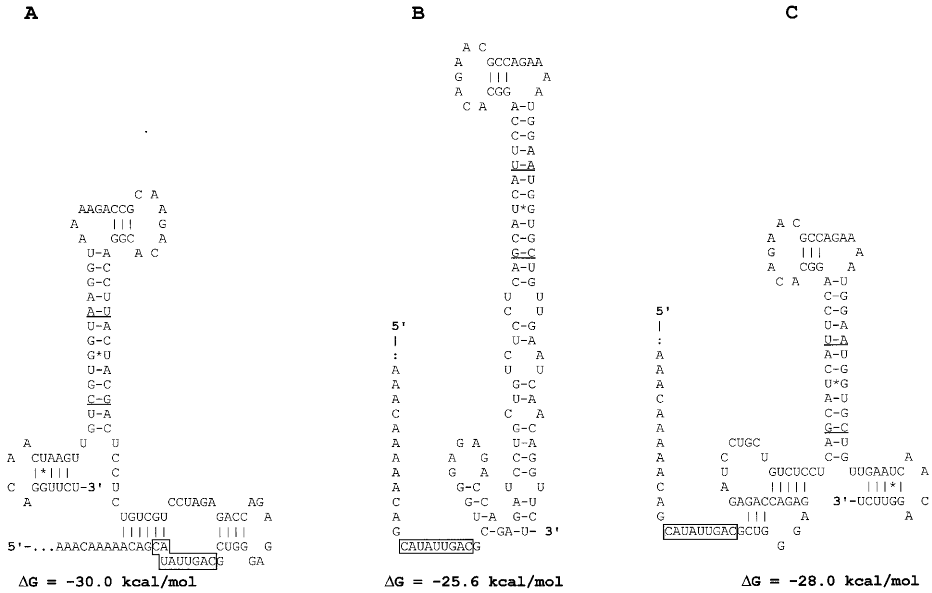


Figure 5. The folding pattern of region III of DEN3 virus (GenBank accession no. M93130). (**A**) Structure according to our prediction, formed by a longer sequence of region III. (**B**) Less stable structure as predicted previously (13,17,18), formed by a truncated sequence of region III. (**C**) More stable structure with a pseudoknot, formed by a truncated sequence of region III. Regions and nucleotides of importance are denoted as in Figure 1.

3′-terminus an even more stable folding is possible (Figs 1 and 5A). Nevertheless, taking into account the folding pattern of region III of other flaviviruses and the fact that the conformation of RNA can be significantly changed by interaction with proteins (25,26), we cannot entirely exclude the existence of a longer stem for DEN viruses as well. This issue can only be resolved by experimentally testing the secondary structures formed by the longer sequences of the DEN virus 3′-UTR.

DISCUSSION

We have used all currently available 3′-UTR sequences of flaviviruses in order to determine the potential folding of the distal part of this region. The models reveal three structurally independent regions within this part of the 3′-UTR (Figs 1–4).

Region I exhibits a high rate of sequence divergence even between closely related viruses, a divergence which has caused

some structural variation (Figs 1–4). However, comparison of the structural motifs (many of which are supported by compensatory mutations) reveals a clear similarity among all flaviviruses. Previously we found an association between the conformation of region I and the virulence of YF virus (Proutski *et al.*, manuscript submitted). This finding and the observation that this region can preserve a similar overall topology in even the most distantly related flaviviruses suggests that these proposed structures do exist *in vivo* and have some functional importance. However, the variants of folding observed in different viruses remain to be interpreted: either the function performed by this region does not require a strict structural conservation or the variants of structure reflect the biological differences between viruses.

The structural properties of region II allowed us to separate the flaviviruses studied into three groups, with dramatic differences between them and remarkable structural similarity within each group. The first group comprises viruses from the JE and DEN serological groups. Despite extensive variation in sequence, these viruses share a general topology for the secondary structure of this region which exposes highly conserved double- and single-stranded motifs: duplicated lateral bulge– and stem–loop structures and the top loop regions of the main hairpins. This structural identity of region II supports a close evolutionary relationship between the DEN and JE serological groups (3,4).

The folding pattern of region II of another mosquito-borne flavivirus, YF virus, appears to be different from that observed for the DEN and JE serogroup viruses. However, the sequence motif (CS2) of the duplicated lateral bulge– and stem–loop structures which characterizes other mosquito-borne viruses was also found in a highly conserved hairpin within region II of YF. Conservation of this sequence and the fact that in all cases it is exposed in the conserved structural elements suggest that it has functional significance in the mosquito-borne flaviviruses.

The secondary structure model of region II of tick-borne flaviviruses, which is supported by a number of compensatory mutations, is very different from that proposed for the mosquito-borne flaviviruses. In particular, no single secondary structure or primary structure motif defined in the mosquito-borne flaviviruses has been found in any of the tick-borne viruses. The folding of region II may therefore be an important structural determinant in the tick-borne flaviviruses.

The conservation in shape of region III, with the formation of a long stable hairpin (3′-LSH) was previously shown to be characteristic for all flaviviruses. The models of secondary structure for the longer part of the 3′-UTR we have constructed confirm that nothing from the upstream regions interferes with typical folding of the 3′-terminus. However, our model of the secondary structure of region III of the DEN viruses slightly contradicts the previously reported models and implies that they possess a shorter 3′-LSH than found in other flaviviruses (Fig. 5). This discrepancy remains to be resolved experimentally.

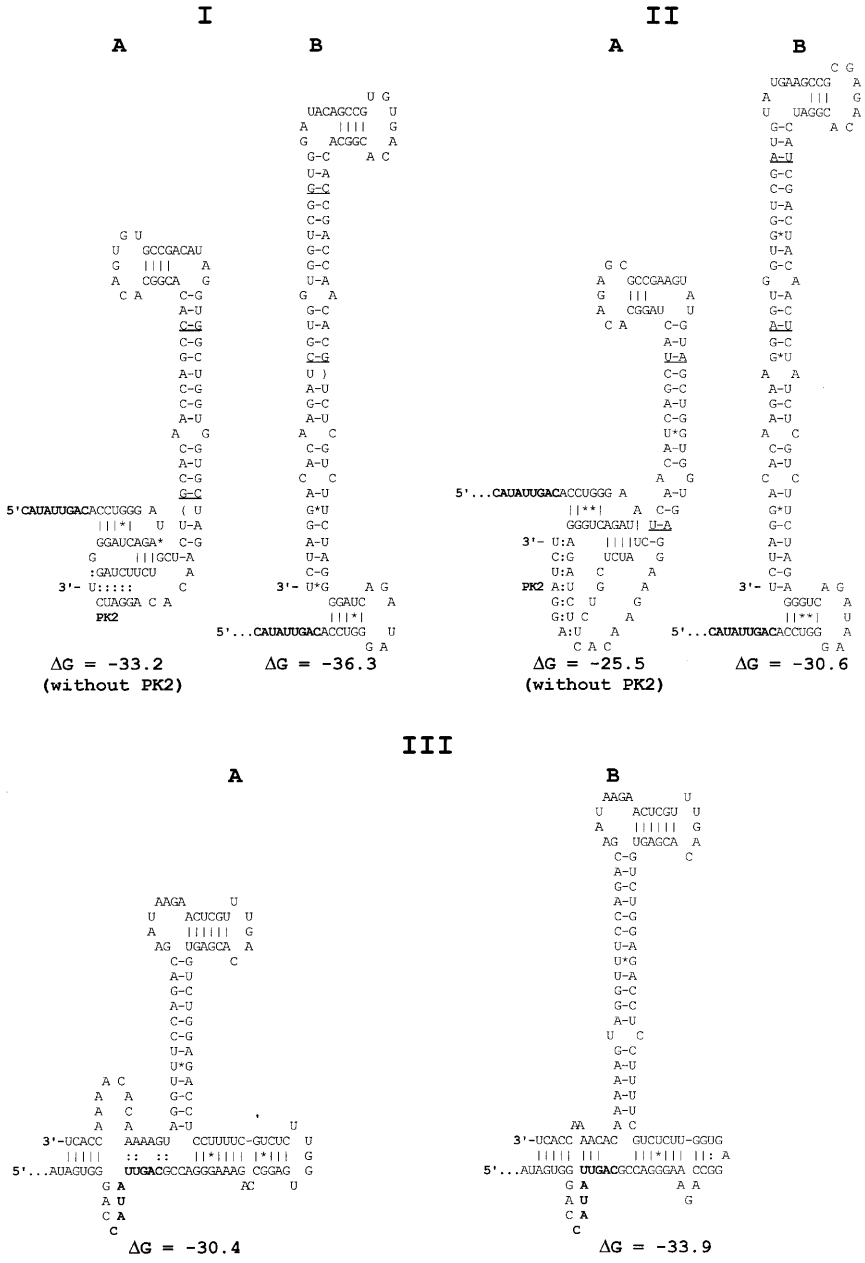
Conformational conservation within the 3′-terminus of the flaviviral genome suggests that the structures have some important functions, which may include initiation and regulation of transcription of the minus strand RNA, regulation of translation, participation in virion assembly and stabilization of the RNA genome, which lacks a poly(A) tail.

It is well known that *in vivo* flaviviruses demonstrate a remarkable disproportion in the synthesis of plus and minus strand RNAs: 10–100 times more plus strand RNA than minus strand is produced (27,28). The biological explanation for this is the double function of the genomic plus strand RNA: it is used as a template both for transcription of the minus strand and translation of the polyprotein precursor, while the minus strand is only transcribed into the new plus strand. However, nothing is known about the mechanisms that regulate initiation of transcription and translation and control the switch between them.

The GA which we used for prediction of the secondary structure simulates the natural pathway of RNA folding and allows the user to follow this pathway during elongation of the RNA chain. From this we observed that during ‘synthesis’ of the 3′-terminus of genomic RNA of all flaviviruses the growing RNA chain can form similar *intermediate* metastable structures that are disrupted by the algorithm in the later stages of ‘synthesis’ in order to form the more stable ‘final’ structure. In the intermediate folding the 3′-LSH was usually much shorter than that of the ‘final’ structure (Fig. 6). Previously we found an association between the predicted folding structure of the 3′-terminus and the virulence of YF virus: all vaccine strains have a much shorter 3′-LSH stem than all wild (virulent) strains (Proutski *et al.*, manuscript submitted). This finding was supported by the experimental evidence that interaction of the 3′-LSH with proteins that are believed to be components of the virus replication complex was sensitive to the length of the 3′-LSH (18). Thus, we can assume that the shorter 3′-LSH which is formed as part of an intermediate structure has a lower capacity to interact with the components of the replication complex and to initiate transcription. The existence of the intermediate structure may then delay formation of the ‘final’ structure, which is capable of initiating transcription, and therefore provide a sufficient time window to start the process of translation. Indeed, in the life cycle of flavivirus, transcription cannot be started until the virus-specific RNA polymerase is translated and processed. One possible problem with this hypothesis is that the real rate of RNA chain synthesis is ~20 bases/s (29) and individual stem formation takes milliseconds (30), so that the intermediate structure must be disrupted and replaced by the ‘final’ one extremely quickly. However, it has been shown that the lifetime of the metastable structure is long and sufficient to perform its function (23).

The calculated energies of the intermediate metastable and the ‘final’ structures are very close in all flaviviruses. This implies that *in vivo* the population of genomic RNAs may represent a dynamic equilibrium of alternative conformations of the 3′-termini (Fig. 6). We speculate that one of these conformations, with a longer 3′-LSH, is able to bind the proteins of the polymerase complex and thereby initiate transcription, while another, with a truncated stem of the 3′-UTR, is unable or, at least, less able to do this. Some, so far unknown, mechanisms may shift this equilibrium in one direction or another and, as a result, favour transcription or translation.

To test this alternative folding hypothesis we constructed secondary structure models for the 3′-termini of the complementary minus strand RNAs of DEN and YF viruses (not shown). We found that this region can also form a long stable hairpin structure which may be recognized by the components of the replication complex and so initiate transcription from the minus strand, but we found no alternative folding which would affect the length of this hairpin. Thus, the alternative folding of the 3′-terminus of flaviviral genomic RNA, which is not observed in the 3′-terminus of minus strand RNA, may work as a switching mechanism between translation and replication. Indeed, in infected cells minus strand RNAs are only found within the replication



complex, whereas the protein-free plus chains are efficiently produced for encapsidation into the virion (13).

One more piece of evidence supporting the functional importance of the alternative folding comes from the tick-borne flaviviruses. In the 3′-terminus of these viruses a conserved 11 nt domain (5′-UCUUGUUCUCC-3′) has been identified which is complementary to a sequence near the 5′-end of the viral genome (8). These sequence motifs are considered functionally analogous to the putative cyclization sequences of mosquito-borne flaviviruses (15). Interestingly, in the ‘final’ structure of the 3′-terminus of TBE this sequence occurs in the lower part of the 3′-LSH stem (Fig. 6VIB). However, in the alternative model of folding, with the truncated 3′-LSH, this sequence can be exposed in the loop region of another hairpin (hairpin Y in Fig. 6VIA) and so is more accessible for complementary interaction with the respective 5′-end sequence. According to our hypothesis, folding with the shorter 3′-LSH is transcriptionally less efficient, so that the complementary interaction between the loop region in the 3′-terminus and the corresponding sequence in the 5′-terminus of genomic RNA, which leads to cyclization of the flaviviral genome, may not be involved in replication, but instead may be the first step in virion packaging. Interestingly, in all models of the secondary structure we have constructed for region III of mosquito-borne flaviviruses the cyclization sequence is either exposed in the single-stranded region or (in YF virus) participates in a very unstable pairing. Hence, it can also be accessed by the 5′-end complement and cause genome cyclization.

In conclusion, the secondary structure models of the flaviviral 3′-UTR we have constructed demonstrate similarities between the viruses and, at the same time, highlight structural differences

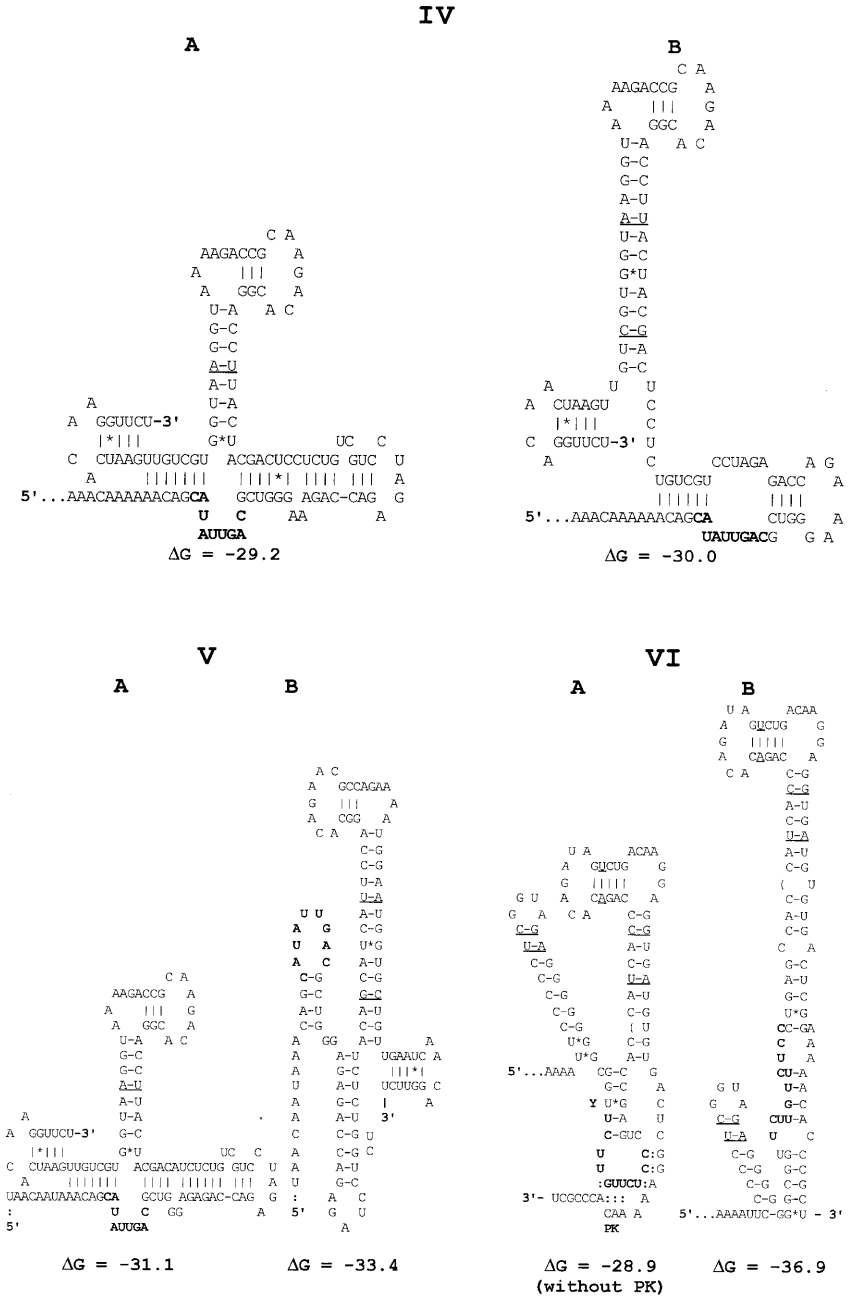


Figure 6. Alternative structures which can be formed by the distal part of the 3′-UTR of various flaviviruses. I, WN virus (GenBank accession no. M12294); II, JE virus, strain JaOArS982 (GenBank accession no. M18370); III, YF virus, Trinidad79 strain (GenBank accession no. U52420); IV, DEN2 virus, Jamaica strain (GenBank accession no. M20558); V, DEN1 virus (GenBank accession no. M87512); VI, TBE virus, Neudoerfl strain (GenBank accession no. U27495). (**A**) Structure with the shortened hairpin; (**B**) structure with a long straight hairpin. – or | denote canonical pairing, * non-canonical pairing and : possible pairings which were not predicted by the program. In some cases (IA, IIA and VIA) it is currently impossible to assess the energy value of the structures, because there are no available data for double pseudoknots. However, the energy was estimated (in kcal/mol) for the structures without taking into account the contribution of these pseudoknots and it is presumed that the real energy is lower. Regions and nucleotides of importance are denoted as in Figure 1.

which may reflect their underlying biological differences. The elements of secondary structure were found to contain a high number of compensatory mutations, which support our models of folding and suggest that these elements are under strong functional constraint. These functions, which may provide an insight into the delicate mechanisms of viral life, remain to be determined experimentally. Furthermore, the functional importance of these structures makes them potentially attractive targets for both antiviral drug design and vaccine development: the rational modification of these structures or the functionally important sequences which are exposed by them could produce an avirulent but immunogenic virus and chemical agents that specifically interact with the structures of the flaviviral 3′-UTR and interfere with their functions may be used against these viruses.

ACKNOWLEDGEMENTS

This work was funded by research grants from The Wellcome Trust and The Royal Society.

REFERENCES

1 Porterfield,J.S. (1980) In Shlesinger,R.W. (ed.), *The Togaviruses.* Academic Press, New York, NY, pp. 13–46.
2 Calisher,C.H., Karabatsos,N., Dalrymple,J.M., Shope,R.E., Porterfield,J., Westaway,E.G. and Brandt,W.E. (1989) *J. Gen. Virol.*, **70**, 37–43.
3 Marin,M.S., de Zanotto,P.M.A., Gritsun,T.S. and Gould,E.A. (1995) *Virology*, **206**, 1133–1139.
4 Zanotto,P.M. de A., Gould,E.A., Gao,G.E., Harvey,P.H. and Holmes,E.C. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 548–553.
5 Rice,C.M. (1996) In Fields,B.N. *et al.* (eds), *Fields Virology*, 3rd Edn. Lippincott-Raven Publishers, Philadelphia, PA, pp. 931–959.
6 Rice,C.M., Lenches,E.M., Eddy,S.R., Shin,S.J., Sheets,R.L. and Strauss,J.H. (1985) *Science*, **229**, 726–735.
7 Chambers,T.J., Hahn,C.S., Galler,R. and Rice,C.M. (1990) *Annu. Rev. Microbiol.*, **44**, 649–688.
8 Mandl,C.W., Holzmann,H., Kunz,K. and Heinz,F.X. (1993) *Virology*, **194**, 173–184.
9 Wallner,G., Mandl,C.W., Kunz,C. and Heinz,F.X. (1995) *Virology*, **213**, 169–178.
10 Poidinger,M., Hall,R.A. and Mackenzie,J.S. (1996) *Virology*, **218**, 417–421.
11 Mandl,C.W., Kunz,C. and Heinz,F.X. (1991) *J. Virol.*, **65**, 4070–4077.
12 Coia,G., Parker,M.D., Speight,G., Byrne,M.E. and Westaway,E.G. (1988) *J. Gen. Virol.*, **69**, 1–21.
13 Brinton,M.A., Fernandez,A.V. and Dispoto,J.H. (1986) *Virology*, **153**, 113–121.
14 Grange,T., Bouloy,M. and Girard,M. (1985) *FEBS Lett.*, **188**, 159–163.
15 Hahn,C.S., Hahn,Y.S., Rice,C.M., Lee,E., Dalgarno,L., Strauss,E.G. and Strauss,J.H. (1987) *J. Mol. Biol.*, **198**, 33–41.
16 Mohan,P.M. and Padmanabhan,R. (1991) *Gene*, **108**, 837–838.
17 Shi,P.Y., Brinton,M.A., Veal,J.M., Zhong,Y.Y. and Wilson,W.D. (1996) *Biochemistry*, **35**, 4222–4230.
18 Blackwell,J.L. and Brinton,M.A. (1995) *J. Virol.*, **69**, 5650–5658.
19 Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
20 Gultyaev,A.P., van Batenburg,F.H.D. and Pleij,C.W.A. (1995) *J. Mol. Biol.*, **250**, 37–51.
21 van Batenburg,F.H.D., Gultyaev,A.P. and Pleij,C.W.A. (1995) *J. Theor. Biol.*, **174**, 269–280.
22 Zuker,M. (1989) *Science*, **244**, 48–52.
23 Gultyaev,A.P., van Batenburg,F.H.D. and Pleij,C.W.A. (1995) *Nucleic Acids Res.*, **23**, 3718–3725.
24 Wengler,G. and Castle,E. (1986) *J. Gen. Virol.*, **67**, 1183–1188.
25 Puglisi,J.D., Tan,R., Calnan,B.J., Frankel,A.D. and Williamson,J.R. (1992) *Science*, **257**, 76–80.
26 Battiste,J.L., Tan,R., Frankel,A.D. and Williamson,J.R. (1994) *Biochemistry*, **33**, 2741–2747.
27 Stollar,V., Shlesinger,R.W. and Stevens,T.M. (1967) *Virology*, **33**, 650–658.
28 Cleaves,G.R., Ryan,T.E., Schlesinger,R.W. (1981) *Virology*, **111**, 73–83.
29 Matsuzaki,H., KassavetisG.A. and Geiduschek,E.P. (1994) *J. Mol. Biol.*, **235**, 1173–1192.
30 Tinoco,I., Puglisi,J.D. and Wyatt,J.R. (1990) *Nucleic Acids Mol. Biol.*, **4**, 205–226.