

Secondary structure of the 3'-untranslated region of yellow fever virus: implications for virulence, attenuation and vaccine development

Vitali Proutski,¹ Michael W. Gaunt,² Ernest A. Gould² and Edward C. Holmes¹

¹The Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

²NERC Institute of Virology and Environmental Microbiology, Mansfield Road, Oxford OX1 3SR, UK

A genetic algorithm-based RNA secondary structure prediction was combined with comparative sequence analysis to construct models of folding for the distal 380 nucleotides of the 3'-untranslated region (3'-UTR) of yellow fever virus (YFV). A number of structural elements that are thermodynamically stable, conserved in shape, and confirmed by compensatory mutations were revealed. At the same time structural polymorphisms were observed among strains of YFV. These polymorphisms showed an association with virulence: all wild and pathogenic strains were likely to be folded in a signifi-

cantly different way from vaccine strains with reduced virulence. Structural divergence was also found among vaccine strains, with 17DD, the most virulent in the mouse model, exhibiting an intermediate pattern of folding, combining structural features of both wild and vaccine strains. The observation of a strong association between secondary structure of the 3'-UTR and virulence of YFV may help elucidate the molecular mechanisms of virus attenuation and lead to new strategies of vaccine development directed towards rational modification of secondary structure of the 3'-UTR.

Introduction

Yellow fever virus (YFV) is one of the mosquito-borne representatives of the genus *Flavivirus* (Rice, 1996) and was the first human virus from this group to be isolated. The virus is both viscerotropic and neurotropic and is a major public health problem, causing epidemic and endemic disease in equatorial Africa and South America.

The genetic organization of YFV is typical of flaviviruses: the single-stranded genomic RNA of approximately 11 kb is the only viral messenger RNA and encodes a single open reading frame from which a large protein precursor is translated. Subsequent proteolytic cleavage of this precursor gives rise to ten mature viral proteins. Flanking the coding region are the 5' and 3' untranslated regions (5'- and 3'-UTR), usually of 118 and 565 bases in length, respectively. These regions are believed to play a crucial role in the regulation of virus translation, replication and assembly. Though the precise mechanisms of this regulation are not well understood, some

conserved elements of RNA primary and potential secondary structure have been observed. For example, computer-predicted folding patterns and RNase cleavage experiments 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 sequence (Grange *et al.*, 1985; Brinton *et al.*, 1986; Hahn *et al.*, 1987*a*; Mohan & Padmanabhan, 1991; Mandl *et al.*, 1993), suggesting that this structural element is functionally important. Furthermore, a specific interaction between the 3'-LSH of some flaviviruses and certain host cellular proteins that are thought to be the components of the virus replicase complex has recently been revealed (Blackwell & Brinton, 1995). Although other conserved RNA sequence motifs have been found within the 3'-UTR of YFV (Hahn *et al.*, 1987*a*), little is known about its secondary structure upstream of the 3'-LSH.

Given the importance of YFV as an infectious agent, live-attenuated vaccines have been developed from two distinct wild-type origins: French viscerotropic virus (YF-FVV) and YF-Asibi strain. Studies involving the intranasal inoculation of mice have demonstrated that different wild and vaccine strains of YFV display different virulence (Barrett & Gould, 1986),

Author for correspondence: Vitali Proutski.

Fax +44 1865 310447. e-mail Vitali.Proutski@zoology.oxford.ac.uk

although the molecular basis of this variation and of virus attenuation is still poorly understood. Comparisons of wild strains of YFV with the vaccine variants derived from them have paid particular attention to protein-coding regions of the YFV genome and as a result have identified substitutions located mainly within the envelope (Hahn *et al.*, 1987*b*; Barrett *et al.*, 1990; Jennings *et al.*, 1994), membrane and non-structural 4B (NS4B) (Wang *et al.*, 1995) genes as being the possible cause of attenuation and differences in virulence. The possible role of the untranslated regions of YFV, especially the 3'-UTR, in influencing virulence and attenuation was not considered, although this role has been demonstrated and discussed for the 3'-UTR of other flaviviruses (Men *et al.*, 1996; Wallner *et al.*, 1996; Khromykh & Westaway, 1997).

In the present study, we have built secondary structure models for the last 380 nucleotides of the 3'-UTR of various strains of YFV which have revealed a number of well-conserved structural elements upstream of the 3'-LSH. Some of these elements contain covariant substitutions among different strains, which further implies that they are of functional importance. We also found regions of structural divergence among different strains of YFV which were associated with viral pathogenicity.

Methods

■ **YFV 3'-UTR nucleotide sequences.** A total of 21 complete or partial sequences of the 3'-UTR of YFV, varying in length from 380 to 386 nt, were used in this analysis; 14 sequences were classified as wild strains and 7 as vaccine strains. The names of the sequences used and their GenBank accession numbers are given in Table 1.

Alignment of the sequences was done with the ClustalW program (Thompson *et al.*, 1994) and subsequently adjusted manually. The alignment is available from the authors upon request. Since only the last 380 3' nucleotides of all sequences exhibited relatively high similarity, this region alone was used for construction of models of secondary structure.

■ **YFV 3'-UTR secondary structure prediction.** Prediction of the possible folding of the 3' part of the 3'-UTR of YFV was done with the STAR program (Gulyaev *et al.*, 1995; van Batenburg *et al.*, 1995) and, in particular, the genetic algorithm (GA) available in this package. Unlike the most widely used algorithms for RNA secondary structure prediction, which are based on a search for the minimal free energy state (Zuker, 1989), the GA simulates the natural folding pathway which takes place during RNA synthesis. 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 other pairings are found to be more favourable. The GA also allows the prediction of certain tertiary interactions, including RNA pseudoknots.

Secondary structure prediction was done individually for each sequence in the data set. Results of this folding analysis were then compared manually and by means of a programming module, 'CovarSearch' (available from the authors upon request), which we have developed to reveal possible covariant (compensatory) mutations. This program uses an alignment file as input data and outputs a list of covariant positions within the alignment, accepting both canonical and non-canonical (G-U) pairings. Substitutions are considered to be

Table 1. Names and GenBank accession numbers of the YFV strains used in this analysis

| Name | GenBank acc. no. |
|-----------------------------|------------------|
| Wild strains | |
| Uga48 | U52423 |
| Trinidad79 | U52420 |
| Trinidad54 | U52417 |
| Senegal65 | U52414 |
| Peru81 | U52411 |
| Peru95(153) | U52407 |
| Peru95(149) | U52405 |
| Nigeria46 | U52401 |
| Ecuador79 | U52399 |
| Car77(900) | U52396 |
| Car77(883) | U52393 |
| Brazil35 | U52390 |
| French viscerotropic strain | U21056 |
| YF-Asibi strain* | – |
| Vaccine strains | |
| 17D | X03700 |
| 17D | K02749 |
| 17D-204 | X15062 |
| 17D.204 | X02807 |
| 17D-213 | U17067 |
| 17DD | U17066 |
| French neurotropic strain | U21055 |

* Sequence not available in GenBank. Refer to Hahn *et al.* (1987*b*).

indicative of covariation if they occur in both strands of the possible helix (G-C to G-U substitutions were not counted as covariant).

Results

The computer simulation of RNA folding and the comparative analysis of predicted structures allowed us to construct models of the secondary structure for the 380 terminal nucleotides of the 3'-UTR of YFV. These models revealed a number of elements common to all strains of YFV. In Figs 1, 2 and 3 these conserved elements are shown in bold. Some, like hairpins 1 and 9, are invariant in both the stem and loop regions, while structures 6 and 7 completely preserve their stem parts. Hairpins numbered 2, 3, 4 and 5 preserve their overall shape in all available sequences and retain the characteristic compensatory mutations (covariant base pairs are underlined). This conservation can be interpreted as the result of selection at the level of secondary structure and suggests that these structural elements may have functional importance. Stem number 8 shows some variation between strains but none which disrupts the stem itself and together with hairpin 9 forms the structure corresponding to the 3'-LSH which is typical of all flaviviruses (Grange *et al.*, 1985; Brinton *et al.*, 1986; Hahn *et al.*, 1987*a*; Mohan & Padmanabhan, 1991; Mandl *et al.*, 1993). Although only the distal 380 nucleotides of

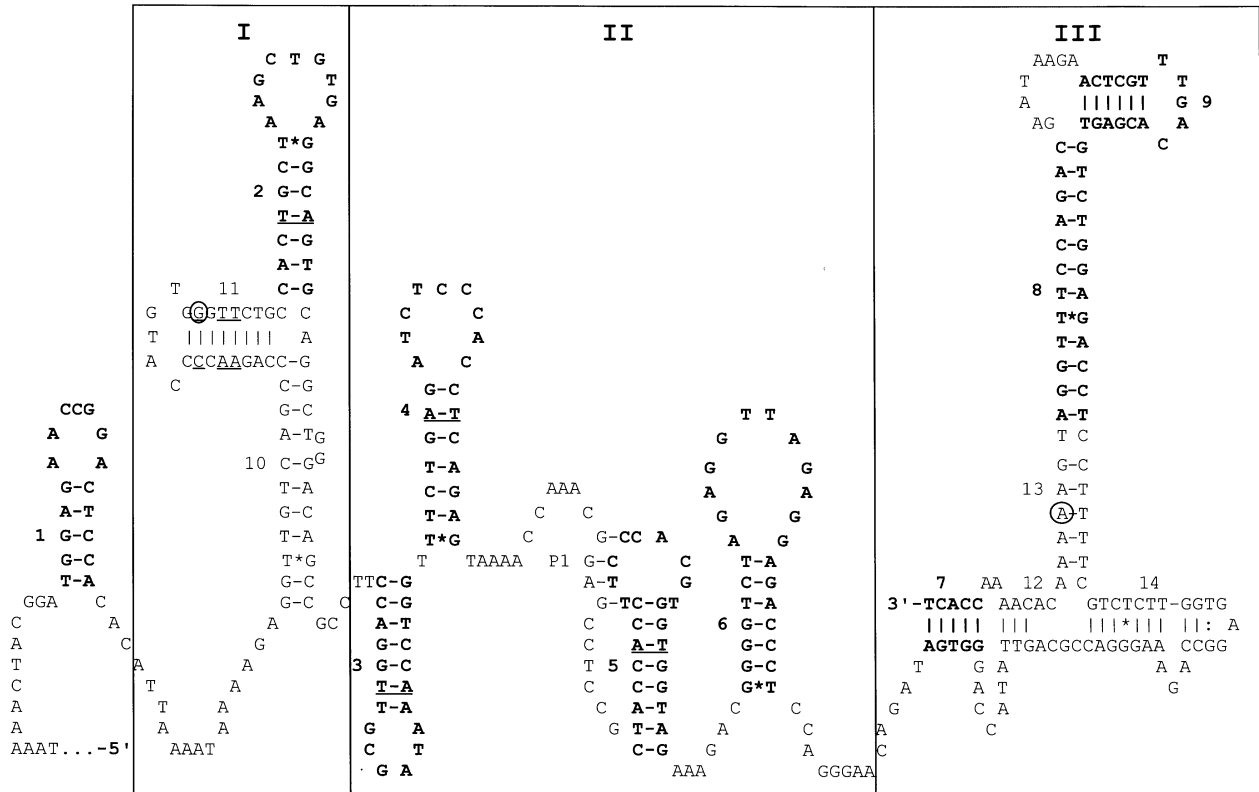


Fig. 1. Proposed secondary structure of the 3'-UTR for wild sequences of YFV. The sequence used for construction of the secondary structure model is from strain Trinidad79, although very similar structures are found in all wild strains. The conserved structures, present in both wild and vaccine strains, are shown in bold and the three distinct structural regions (labelled I to III) are delineated by boxes. Covariant nucleotides are underlined and the variable sites which lead to changes in folding pattern are circled. The nucleotides are numbered from the 3' terminus.

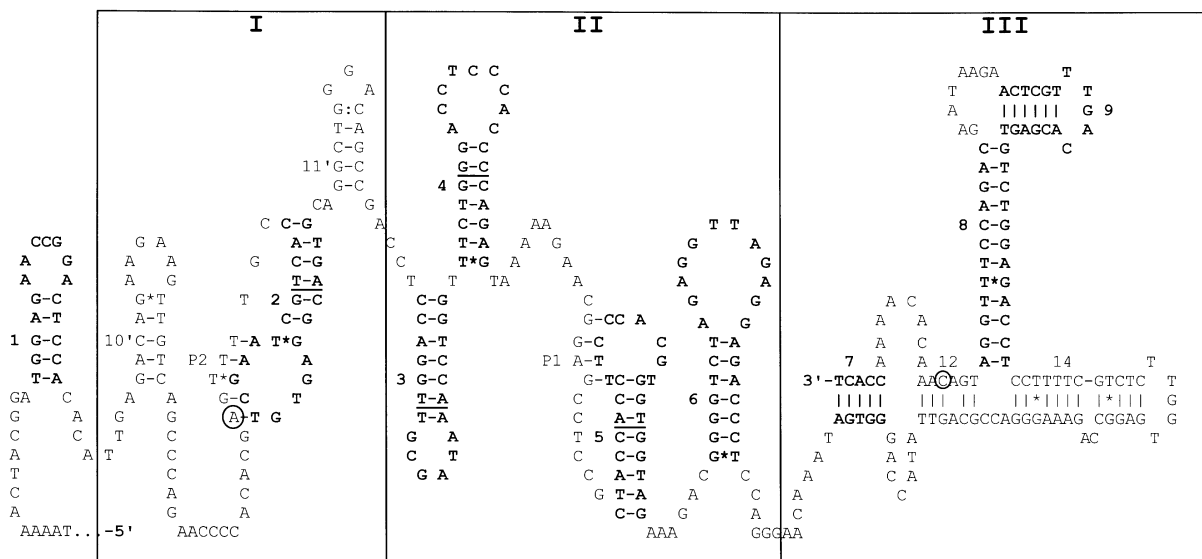


Fig. 2. Proposed secondary structure of the 3'-UTR for vaccine sequences of YFV. The sequence used for construction of the secondary structure model is from strain 17D.204. This folding pattern is characteristic of all vaccine strains except 17DD (see Fig. 3). Regions and nucleotides of importance are denoted as in Fig. 1.

stamp of selection pressure: one covariant substitution. The second branching hairpin, numbered 11, is also likely to be functionally important because there are three covariations within its stem (covariant nucleotides underlined). In contrast, all vaccine strains have a G to A point mutation (circled) within this stem. Furthermore, in all vaccine strains with the exception of 17DD, this mutation does not have a compensatory counterpart on the opposite side of the stem (although two wild strains also have the G to A substitution they have a compensatory C to U change which means they retain the wild-type folding). As a result, the folding in the vaccine strains occurs in such a way that the stems 10 and 11 are disrupted and two new structures, 10' and 11', are formed (Fig. 2). Hairpin 2 remains untouched, but instead of being exposed on the top of a 'wild-type' structure is hidden within the strong pseudoknot, P2. The 17DD vaccine strain, on the other hand, has the compensatory C to U mutation along with the G to A change and consequently preserves the 'wild-type' folding pattern of region I. Therefore, 17DD exhibits a mixed pattern of folding, where region III is folded in the 'vaccine-type' way and region I in the 'wild-type' manner (Fig. 3). Interestingly, 17DD was shown to be the most neurovirulent of the vaccine strains in a mouse model (Barrett & Gould, 1986) implying that an 'intermediate' pattern of folding may be in some way associated with an intermediate level of virulence.

Discussion

In this study we used a genetic algorithm-based RNA secondary structure prediction, combined with comparative sequence analysis, to determine the potential folding of the distal 380 nucleotides of the 3'-UTR of YFV. The very 3' terminus of this region was folded in the way typical of all flaviviruses, forming the 3'-LSH structure. A number of structural elements have been revealed upstream of the 3'-LSH. The conservation of these structures among different strains of YFV, along with the presence of compensatory mutations, implies that they are functionally important and formed *in vivo*.

Two regions of the 3'-UTR were likely to be folded in alternative ways for different strains. In one (region III on Figs 1 and 2), a point mutation in the distal part of the 3'-UTR was found to characterize vaccine strains so that their 3'-LSH, although preserving the general shape, had a much shorter stem than that found in the wild strains. Another region (region I on Figs 1 and 2), which also appears to show structural divergence between all wild and the majority of vaccine strains, has not been described previously. Normally, for the wild strains, this region is formed by two stable and conserved structures, with four compensatory mutations within the region hinting at its functional importance. A further computer prediction indicates that the corresponding part of the 3'-UTR of other flaviviruses is likely to be folded in a similar way (Proutski *et al.*, 1997). Most vaccine strains of YFV have a point mutation which completely changes the structural

pattern of this region. 17DD is the only vaccine strain which demonstrates the 'vaccine-type' folding of region III but preserves the 'wild-type' folding of region I (Fig. 3). This strain was found to have higher virulence for mice than the other vaccine strains (Barrett & Gould, 1986). Thus, we observe a strong association between the pathogenicity of various strains of YFV and the structural features of the 3'-UTR of these strains. This is particularly compelling since the vaccine strains were independently derived from two different parental wild YFV strains, Asibi strain and French viscerotropic strain. Nevertheless, all vaccine strains share common nucleotide substitutions within the 3'-UTR that lead to similar conformational changes within this region.

There is some experimental evidence for functional activity of the 3'-LSH of flaviviruses, including a specific interaction with cellular proteins. This interaction may be involved in regulation of the rate of viral RNA transcription initiation and/or may play a regulatory role during viral RNA synthesis (Blackwell & Brinton, 1995). Although these functions may be realized by both conserved primary sequence motifs and secondary structure, the fact that the shape of the 3'-LSH is conserved whilst the nucleotide sequences of the stems are not suggests that it is the secondary structure which is under functional constraint. It is possible that the shortened variant of the 3'-LSH which we observe in all vaccine strains of YFV is not capable of functioning in the same way as the respective structure of the wild strains. This idea is supported by experiments on RNA-protein interactions in the 3'-LSH of West Nile encephalitis virus and dengue virus type 3 where RNA-protein binding was sensitive to the length of the 3'-LSH (Blackwell & Brinton, 1995).

There is no direct evidence for the function of the second dimorphic region (region I). However, the virulence-structure association that we observe and especially the discovery that the 'wild-type' folding of this region, even in combination with the 'vaccine-type' of region III, was found in viruses with relatively high virulence, strongly suggests that this region and its structural organization is somehow important in the development of virus pathogenicity.

There is one more interesting implication of the structural polymorphism we have found for the 3'-UTR of YFV. In the literature there are data which suggest that not all virions in a population possess the same level or type of virulence (Barrett & Gould, 1986; Liprandi, 1981; Liprandi & Walder, 1983). Usually, this is explained by the fact that RNA viruses consist of a heterogeneous and continually evolving population of virions, the 'quasispecies', so that individual virions which differ in sequence may also differ in virulence. However, if the virulence of YFV is associated with certain structural variants of the 3'-UTR, and the energetic difference between these structures is relatively small, then it is possible that genetically identical virions may take on alternative structures which can co-exist in a population, perhaps with different phenotypic properties such as virulence. For example, the 3'-UTR of wild

strains of YFV can be folded in the 'vaccine' way with little difference in energy: the free energy of the 'wild type' folding of region III for the Trinidad79 strain is -142 kJ/mol, while the 'vaccine type' folding of the same sequence will result in a free energy of -127.3 kJ/mol. Although the energetic amplitude between these two states (14.7 kJ/mol) is sufficient to favour one pattern of folding over another, it is small enough that in different conditions, such as those caused by changes in ion concentration or interactions with proteins, the structural equilibrium (i.e. the proportion of structures) can be shifted in one direction or another.

Finally, the possibility of alternative yet energetically similar patterns of folding for the 3' terminus of the virulent YFV strains raises the question of whether there is some kind of trigger mechanism which can favour one type of folding over another and so alter the function of this region. This hypothesis is supported by the observation that the distal part of the 3'-UTR of other flaviviruses can also take on alternative folding patterns which are close in terms of energetic stability (Proutski *et al.*, 1997). In contrast, the point mutation in region III of the vaccine strains of YFV strongly favours the 'vaccine-type' folding of this region and, possibly, impedes this switching ability.

In conclusion, we have found an association between the predicted folding structure of the 3'-UTR and the virulence of YFV, an observation that needs to be examined experimentally. Our conclusions concerning the role of the 3'-UTR do not diminish the significance of substitutions within the coding region of the YFV genome reported by other groups (Hahn *et al.*, 1987*b*; Barrett *et al.*, 1990; Jennings *et al.*, 1994; dos Santos *et al.*, 1995; Wang *et al.*, 1995) as it is likely that the virulence or avirulence of YFV is the result of the cohesive activity of many factors. However, the association we observe between the folding structure of the 3'-UTR and virulence may potentially lead to a new strategy for vaccine development aimed at the non-coding regions of YFV, and particularly the 3'-UTR. Indeed, the rational modification of structural elements of the 3'-UTR may result in the development of a strain of YFV with high immunogenicity (since the coding regions, and hence the viral proteins, remain untouched) and reduced virulence, which would be an ideal candidate for vaccines. Considering the common structural features shared by the flaviviruses, it is possible that this approach may be of use in the development of vaccines against other viruses of this family.

This work was supported by The Wellcome Trust and The Royal Society.

References

- Barrett, A. D. T. & Gould, E. A. (1986). Comparison of neurovirulence of different strains of yellow fever virus in mice. *Journal of General Virology* **67**, 631–637.
- Barrett, A. D., Monath, T. P., Cropp, C. B., Adkins, J. A., Ledger, T. N., Gould, E. A., Schlesinger, J. J., Kinney, R. M. & Trent D. W. (1990). Attenuation of wild-type yellow fever virus by passage in HeLa cells. *Journal of General Virology* **71**, 2301–2306.
- Blackwell, J. L. & Brinton, M. A. (1995). BHK cell proteins that bind to the 3' stem-loop structure of the West Nile virus genome RNA. *Journal of Virology* **69**, 5650–5658.
- Brinton, M. A., Fernandez, A. V. & Dispoto, J. H. (1986). The 3'-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology* **153**, 113–121.
- dos Santos, C. N., Post, P. R., Carvalho, R., Ferreira, I. I., Rice, C. M. & Galler, R. (1995). Complete nucleotide sequence of yellow fever virus vaccine strains 17DD and 17D-213. *Virus Research* **35**, 35–41.
- Grange, T., Bouloy, M. & Girard, M. (1985). Stable secondary structures at the 3'-end of the genome of yellow fever virus (17 D vaccine strain). *FEBS Letters* **188**, 159–163.
- Gulyaev, A. P., van Batenburg, F. H. D. & Pleij, C. W. A. (1995). The computer simulation of RNA folding pathways using a genetic algorithm. *Journal of Molecular Biology* **250**, 37–51.
- Hahn, C. S., Hahn, Y. S., Rice, C. M., Lee, E., Dalgarno, L., Strauss, E. G. & Strauss, J. H. (1987*a*). Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. *Journal of Molecular Biology* **198**, 33–41.
- Hahn, C. S., Dalrymple, J. M., Strauss, J. H. & Rice, C. M. (1987*b*). Comparison of the virulent Asibi strain of yellow fever with the 17D vaccine strain derived from it. *Proceedings of the National Academy of Sciences, USA* **84**, 2019–2023.
- Jennings, A. D., Gibson, C. A., Miller, B. R., Mathews, J. H., Mitchell, C. J., Roehrig, J. T., Wood, D. J., Taffs, F., Sil, B. K., Whitby, S. N., Whitby, J. E., Monath, T. P., Minor, P. D., Sanders, P. G. & Barrett, A. D. T. (1994). Analysis of a yellow fever virus isolated from a fatal case of vaccine-associated human encephalitis. *Journal of Infectious Diseases* **169**, 512–518.
- Khromykh, A. A. & Westaway, E. G. (1997). Subgenomic replicons of the flavivirus Kunjin: construction and applications. *Journal of Virology* **71**, 1497–1505.
- Liprandi, F. (1981). Isolation of plaque variants differing in virulence from the 17D strain of yellow fever virus. *Journal of General Virology* **56**, 363–370.
- Liprandi, F. & Walder, R. (1983). Replication of virulent and attenuated strains of yellow fever virus in human monocytes and macrophage-like cells (U937). *Archives of Virology* **76**, 51–61.
- Mandl, C. W., Holzmann, H., Kunz, K. & Heinz, F. X. (1993). Complete genomic sequence of Powassan virus: evaluation of genetic elements in tick-borne versus mosquito-borne flaviviruses. *Virology* **194**, 173–184.
- Men, R., Bray, M., Clark, D., Chanock, R. M. & Lai, C.-J. (1996). Dengue type 4 virus mutants containing deletions in the 3' noncoding region of the RNA genome: analysis of growth restriction in cell culture and altered viremia pattern and immunogenicity in Rhesus monkeys. *Journal of Virology* **70**, 3930–3937.
- Mohan, P. M. & Padmanabhan, R. (1991). Detection of stable secondary structure at the 3' terminus of dengue virus type 2 RNA. *Gene* **108**, 837–838.
- Proutski, V., Gould, E. A. & Holmes, E. C. (1997). Secondary structure of the 3'-untranslated region of flaviviruses: similarities and differences. *Nucleic Acids Research* **25**, 1194–1202.
- Rice, C. M. (1996). *Flaviviridae: the viruses and their replication*. In *Fields Virology*, 3rd edn, pp. 931–959. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.

- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.
- van Batenburg, F. H. D., Gulyaev, A. P. & Pleij, C. W. A. (1995).** An APL-programmed genetic algorithm for the prediction of RNA secondary structure. *Journal of Theoretical Biology* **174**, 269–280.
- Wallner, G., Mandl, C. W., Kunz, C. & Heinz, F. X. (1995).** The flavivirus 3'-noncoding region: extensive size heterogeneity independent of evolutionary relationships among strains of tick-borne encephalitis virus. *Virology* **213**, 169–178.
- Wallner, G., Mandl, C. W., Ecker, M., Holzman, H., Stiasny, K., Kunz, C. & Heinz, F. X. (1996).** Characterization and complete genome sequences of high- and low-virulence variants of tick-borne encephalitis virus. *Journal of General Virology* **77**, 1035–1042.
- Wang, E., Ryman, K. D., Jennings, A. D., Wood, D. J., Taffs, F., Minor, P. D., Sanders, P. G. & Barrett, A. D. (1995).** Comparison of the genomes of the wild-type French viscerotropic strain of yellow fever virus with its vaccine derivative French neurotropic vaccine. *Journal of General Virology* **76**, 2749–2755.
- Zuker, M. (1989).** On finding all suboptimal foldings of an RNA molecule. *Science* **244**, 48–52.

Received 21 November 1996; Accepted 28 February 1997