Evidence for the Existence of a Pseudoknot Structure at the 3′ Terminus of the Flavivirus Genomic RNA†

Pei-Yong Shi,‡ Margo A. Brinton,*,‡ James M. Veal,§,|| Yi Yi Zhong,§ and W. David Wilson*,§

Departments of Biology and Chemistry, Georgia State University, Atlanta, Georgia 30303

Received October 6, 1995; Revised Manuscript Received January 9, 1996⊗

ABSTRACT: The 3′-terminal nucleotides of the flavivirus genomic RNA form conserved secondary structures that may function as *cis*-acting signals for RNA replication. Here we provide evidence for the existence of a conserved pseudoknot structure at the 3′ terminus of the flavivirus genomic RNA. A truncated version of the West Nile virus (WNV) 3′-terminal RNA sequence was used as the model for these studies. Circular dichroism spectra indicated the presence of a highly structured RNA conformation with a significant amount of A-form helix. Ribonuclease probing not only confirmed the presence of the predicted secondary structure, which consists of a long stem−loop (SL1) and a shorter stem−loop (SL2), but also suggested that base pairing occurs between nucleotides in the loop of SL2 and those in an internal loop strand located on the 5′ side of SL1. Analysis of three mutant RNAs further supported the existence of pseudoknot interactions. UV-melting analysis of the WNV 3′ model RNA showed three transitions with significant hyperchromicity at approximately 46, 62, and 79 °C. UV-melting analysis with either SL1 or SL2 RNA alone suggested that the 62 and 79 °C transitions represent the unfolding of SL2 and SL1, respectively. The 46 °C transition is most likely due to the opening of the proposed tertiary structure. A similar melting curve was obtained for another flavivirus (dengue-3 virus) 3′-terminal RNA, providing further support for the conservation of the structure among flaviviruses. Molecular modeling of the RNA indicated that a pseudoknot structure is a stereochemically and energetically reasonable model for the 3′ terminus of flavivirus genomic RNA.

Flaviviruses annually cause significant human morbidity and mortality. The spectrum of diseases caused by flaviviruses ranges from a mild fever to hepatitis, hemorrhagic disease, and encephalitis (Brinton, 1983). The virion contains a single plus-strand RNA genome of approximately 11 kb in length that encodes a single long open reading frame (Rice et al., 1985). The 5′ and 3′ noncoding regions (NCR) of the genomic RNA are approximately 100 and 500−600 nucleotides in length, respectively. Both the 5′ NCR and the terminal nucleotides of the 3′ NCR can form highly conserved secondary structures (Brinton & Dispoto, 1988; Brinton et al., 1986; Rice et al., 1985). During the flavivirus replication cycle, the plus-strand genomic RNA is first replicated into minus-strand RNA which then serves as the template for the synthesis of more genomic RNA. The conserved 3′-terminal structures as well as some short conserved sequences within the 3′ NCR of the genome RNA may function as *cis*-acting replication signals and interact with viral and, possibly, also cellular proteins during the initiation of the minus-strand RNA synthesis. Deletions in the terminal 3′ NCR sequences are lethal for flavivirus infectious clones (Lai et al., 1992). Three cellular proteins were recently reported to bind specifically to the 3′ termini of the WNV and dengue-3 (Den-3) genomic RNAs (Blackwell & Brinton, 1995).

The 3′-terminal nucleotides of all sequenced flavivirus genomes can be folded into similar 3′-terminal structures consisting of a large stem−loop (SL) followed by a smaller SL (Figure 1A). The existence of the large SL formed by the 3′-terminal 79 nucleotides of the West Nile virus (WNV) 3′ genomic RNA was previously confirmed by RNase probing (Brinton et al., 1986). The mostly conserved 5′-ACACA-3′ sequence is located near the 3′ terminus of the large SL (Figure 1A). The small SL contains a conserved 5′-GANAGA-3′ hexanucleotide loop (Figure 1A, bracketed region). Analysis of the two SL secondary structures revealed that two possible, but mutually exclusive, tertiary structures could be formed by interactions between the two SLs. One of these is a pseudoknot tertiary interaction that could be formed between an unstable region on the 5′ side of the large SL and some nucleotides in the conserved 5′-GANAGA-3′ loop of the small SL (indicated by dotted lines in Figure 1A). In the case of WNV, the 5′-UAGA-3′ (89−86 nt) in the loop of the small SL could potentially base pair with 5′-UCUG-3′ (74−71 nt) located in an unstable region on the 5′ side of the large SL. Similar interactions might occur between 5′-AAGA-3′ (87−84 nt) and 5′-GCUG-3′ (72−69 nt) in the dengue-3 (Den-3) sequence and between 5′-AAGA-3′ (96−93 nt) and 5′-UCUG-3′ (77- 80 nt) in the yellow fever virus (YFV) sequence (Figure 1A). Within these regions, one G/U or one G/A base pair would occur in the WNV and YFV sequences, respectively, while two G/A base pairs would occur in the Den-3 sequence. G/U and G/A base pairs commonly stabilize RNA structures (Heus

† This work was supported by NIH Grants AI-18382 (M.A.B.) and AI-27196 (W.D.W.) and by the Georgia Research Alliance.

* Correspondence can be addressed to either author.

‡ Department of Biology.

§ Department of Chemistry.

|| Current address: Glaxo Research Institute, Research Triangle Park, NC 27709.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1996.

0006-2960/96/0435-4222$12.00/0

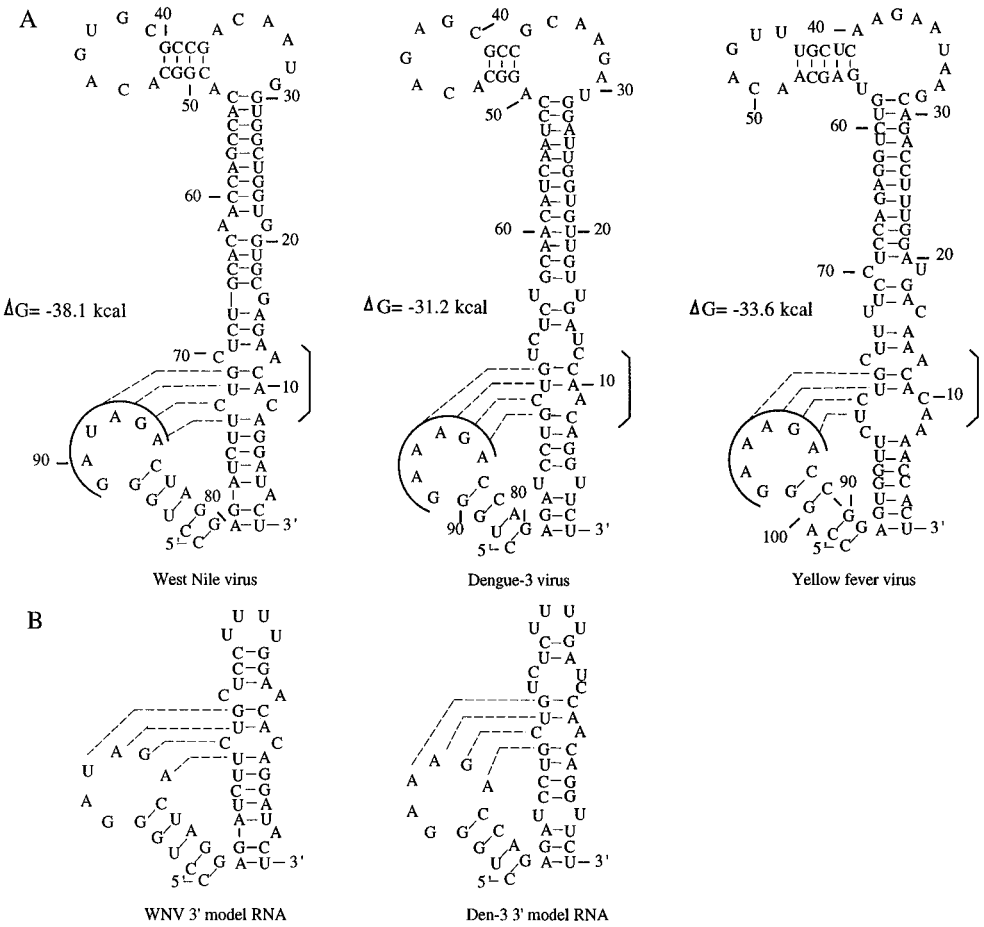

FIGURE 1: Computer-predicted secondary structures formed by the 3′-terminal noncoding region sequences of three flavivirus genomic RNAs. The optimal ΔG is shown next to each of the predicted stem−loop structures. (A) Computer-predicted secondary structures for the 3′-terminal genomic RNAs of West Nile virus (WNV), dengue-3 virus (Den-3), and yellow fever virus (YFV). In spite of the sequence divergence among flaviviruses, similar secondary structures were predicted for the 3′-terminal sequences of each of the genomic RNAs, a large 3′ hairpin and a small 5′ hairpin. Highly conserved sequences are bracketed. (B) WNV and Den-3 3′ model RNAs (see text for description). *In vitro* transcribed 3′ model RNAs contain four extra nucleotides at the 5′ end (5′-GGGC-3′) and one extra G at the 3′ end derived from cloning. The RNA sequence is numbered from the 3′ end.

& Pardi, 1991; Wimberly et al., 1993; Pley et al., 1994). Recently, two G/A base pairs, between 5′-ACG-3′ and 3′-GGA-5′ in the sarcin/ricin loop in rat 28 S rRNA, were reported (Szewczak & Moore, 1995).

An alternative tertiary structure could be formed if the helical segments of the two flavivirus 3′ SLs coaxially stack with their loops oriented in opposite directions. Pseudoknots and coaxial stacking are both commonly involved in RNA tertiary structures (Wyatt & Tinoco, 1993; Turner & Bevilacqua, 1993). Pseudoknots have been found at the 3′ ends of a number of viral RNA genomes (Pleij et al., 1985; Drepher & Hall, 1988; Jacobson et al., 1993) and are frequently found as structural components of RNAs derived from in vitro selection schemes (Pyle & Green, 1995; Lorsch & Szostak, 1994). Coaxial stacking of helices can direct the overall folding of an RNA (Murphy et al., 1994). Analysis of coaxial stacking energetics illustrates that such an interaction can add significantly to the stability of a folded RNA conformation (Walter & Turner, 1994).

As a means of further understanding flavivirus RNA replication, we have analyzed the flavivirus 3′-terminal genomic sequence to determine if it can form tertiary interactions and, if so, whether these interactions result in the formation of a pseudoknot, coaxial stacking, or some other structure. We have used the WNV and Den-3 3′ genomic RNAs as models to analyze possible tertiary interactions. Results from RNase probing, circular dichroism spectra, UV-melting experiments, mutagenesis analysis, and molecular modeling support the existence of a pseudoknot structure at the 3′ terminus of the flavivirus genomic RNA.

MATERIALS AND METHODS

Construction of DNA Templates for in Vitro Transcription. Truncated versions of the WNV and Den-3 3′-terminal genomic RNAs were used as RNA model systems for structural analysis (Figure 1B, see Results for details). To generate the DNA template for *in vitro* transcription of the WNV 3′ model RNA with T7 DNA-dependent RNA polymerase, two complementary oligonucleotides with tails containing the equivalent of cut *Apa*I and *Sac*I restriction enzyme digestion sites positioned appropriately for unidirectional cloning were annealed and ligated to plasmid pCR1000 DNA (Invitrogen Co.) that had been predigested with *Apa*I and *Sac*I. The same strategy was used to construct the plasmid templates for *in vitro* synthesis of three WNV 3′ base-substitution mutant RNAs and the Den-3 3′ model RNA. The plasmid DNA constructs were transfected into *Escherichia coli*, strain DH5α, and after amplification were purified with a QIAGEN plasmid purification kit, and analyzed on agarose gels prior to use.

In Vitro Transcription and Purification of RNA. Purified plasmid DNA was first linearized by digestion with *Sac*I,

phenol−chloroform extracted, and then ethanol precipitated. The transcription reaction mixture consisted of 80 mM HEPES (pH 7.5), 16 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT, 3 mM each NTP, 80 units/mL RNasin inhibitor, 50 ng/μL linearized plasmid DNA template, and 0.8 unit/μL T7 polymerase. The reaction was incubated at 37 °C for 2 h, then another 0.8 unit/μL T7 polymerase was added to the reaction mixture, and incubation was continued at 37 °C for another 2 h. The *in vitro* transcribed RNA was phenol−chloroform extracted, ethanol precipitated, and separated by electrophoresis on an 8% denaturing polyacrylamide gel. The RNA band was excised, and the RNA was eluted from the gel slices by incubation in 500 mM ammonium acetate, 1 mM EDTA, and 0.1% SDS at 37 °C overnight. The eluted RNA was filtered through a 0.45 μm cellulose acetate filter unit (Costar Co.) to remove gel pieces, precipitated in ethanol, and stored in −80 °C.

SL1 or SL2 RNA was directly transcribed from two annealed complementary oligonucleotides which contained the T7 promoter sequence followed by the desired sequence (Milligan et al., 1987). Briefly, the complementary oligonucleotide DNAs were added to 1× transcription buffer (100 mM HEPES, pH 7.2, and 15 mM $MgCl_2$), annealed at 85 °C for 5 min, slowly cooled to room temperature for 10 min, and then allowed to sit on ice for 5 min. The RNA transcription reaction contained 10 mM DTT, 2 mM each NTP (Boehringer Mannheim Biochemicals), 25 unitsmL RNasin inhibitor (Boehringer Mannheim Biochemicals), 20 μg/mL template DNA, and approximately 600 units/mL T7 polymerase. The reaction was incubated at 38 °C for 2 h. An additional 400 units/mL T7 polymerase was then added to the reaction and incubation continued at 38 °C for 1.5 h. The reaction was phenol−chloroform extracted, ethanol precipitated, and gel purified as described above except that a 12% denaturing polyacrylamide gel was used. The purified RNA was precipitated with ethanol and stored at −80 °C. The T7 polymerase was overexpressed in *E. coli* and purified as described by Davanloo et al. (1984).

RNase Probing of RNA. Gel-purified RNA was 5′ dephosphorylated by treatment with calf intestinal alkaline phosphatase according to the instructions of the manufacturer (Boehringer Mannheim Biochemicals) and then ethanol precipitated. The RNA was 5′ end labeled by incubation with [γ-^{32}P]ATP (3000 Ci/mmol; Amersham) and T4 polynucleotide kinase (Promega) at 37 °C for 10 min. The 5′ end labeled RNA was repurified on an 8% denaturing gel and eluted as described above. The labeled RNA was precipitated with ethanol using tRNA as a coprecipitate and stored at −80 °C.

For RNase probing experiments, the ^{32}P end labeled RNA was pelleted, washed with 70% ethanol, dried in a Speedvac, and dissolved in RNA renaturing buffer (20 mM HEPES, pH 7.6, 200 mM NaCl, and 1 mM DTT, containing either 0.2 mM or 10 mM $MgCl_2$). For each experiment, approximately 5×10^4 cpm of RNA was dissolved in 20 μL of renaturing buffer. The RNA was renatured by incubating at 83 °C for 2 min, slowly cooled to room temperature, and placed in an ice-slurry bath. Prior to RNase digestion, 0.15 mg/mL tRNA was added, and the RNA sample was aliquoted at 2 μL per tube. One double-strand-specific RNase, cobra venom nuclease V1 (0.175, 0.35, and 0.70 unit/mL), or one of the two single-strand-specific RNases, RNase A (5×10^{-3}, 1×10^{-2}, and 5×10^{-2} μg/mL) or RNase phyM (0.25, 0.5, and 1 unit/mL), was added to the RNA at several different concentrations. The reactions were incubated in an ice slurry for 20 min or at room temperature for 16 min, and then 2 μL of stop buffer (10 M urea, 10% glycerol, 0.05% xylene cyanol, and 0.05% bromophenol blue) was added to inactivate the nucleases. All digested RNA samples were analyzed on 10% polyacrylamide sequencing gels.

To produce an alkaline hydrolysis ladder, 10^4 cpm of RNA was dissolved in 4 μL of hydrolysis buffer (50 mM $NaHCO_3$/Na_2CO_3, pH 9.2, and 1 μg/μL freshly added tRNA), incubated at 90 °C for 6 min, and immediately placed in an ice-slurry bath. For RNA sequencing reactions, 7×10^3 cpm of RNA was incubated in 4 μL of sequencing buffer (20 mM sodium citrate, pH 5.0, 1 mM EDTA, 7 M urea, 0.025% xylene cyanol, 0.025% bromophenol blue, and 1 μg/μL freshly added tRNA) with either 0.1 unit/mL RNase T1, 0.005 μg/mL RNase A, or 0.5 unit/mL RNase U2 at 50 °C for 15 min. The reactions were terminated by putting them into an ice-slurry bath. The RNases T1 and A were from Boehringer Mannheim and RNases phyM, V1, and U2 were from Pharmacia.

Melting Curves. Absorbance vs temperature melting curves were measured at 260 nm with a heating rate of 0.5 °C min^{-1} on a Cary 4 spectrometer as described previously (Kibler-Herzog et al., 1990). RNA samples were prepared in a buffer consisting of 5 mM Na_2HPO_4, pH 7.0, 0.1 mM EDTA, and 100 mM NaCl. The derivatives of the melting curves were taken with a five-point parabolic fit with binomial weighing at each temperature. Additional thermal melting experiments were also performed with the same buffer in the presence of 0.2 mM Mg^{2+}.

Circular Dichroism. Circular dichroism spectra were measured with a Jasco-J710 spectropolarimeter, equipped with a programmable Neslab temperature controller as previously described (Zou et al., 1990). Spectra were recorded from 350 to 200 nm and averaged over five scans. RNA samples were prepared in the same buffer as described for the thermal melting studies. Cylindrical cuvettes with a path length of 1 cm were used.

Molecular Modeling of the RNA Conformation. RNA molecular models were constructed with the SYBYL software package (version 6.0) from Tripos and were energy minimized by using a modification (Veal & Wilson, 1991; Yao & Wilson, 1992) of the AMBER force field of Kollman and co-workers (Weiner et al., 1984; Weiner & Kollman, 1986). Modifications in the molecular mechanics force field were made because they provide better reproduction of available experimental results with respect to RNA and its complexes. A distance-dependent dielectric constant of the form $\epsilon = 4r_{ij}$, where r_{ij} is the distance between atom pairs, was used to model solvent effects. Nucleic acid partial atomic charges were from Kollman and co-workers (Weiner et al., 1984; Weiner & Kollman, 1986) except for the phosphate group charge which was reduced to a value of −0.2 as previously described (Veal & Wilson, 1991). Both experimental data and theory have shown that nucleic acids do not behave as fully charged molecules due to the association of counterions (Manning, 1978; Record et al., 1978). The starting conformations for all RNA duplexes, including the pseudoknot duplex, were generated from general fiber diffraction data for RNA (Chandrasekaran & Arnott, 1989). Full relaxation of atomic bond stretch, angle

bend, and tortional parameters was allowed. A conjugate gradient method was used for minimizations, and the convergence criterion for a minimized structure was the change in the RMS gradient [0.02 kcal/(mol·Å^2)]. Details of construction of the RNA model systems for molecular mechanics energy minimization are presented in Results.

Thermodynamic Calculation of RNA Secondary Structure. Optimal folding of secondary structures for the model RNAs and mutated versions of these RNAs was calculated through minimization of free energy using the Fold program in the GCG package (Genetics Computer Group, Inc., Version 7, University of Wisconsin; Jaeger et al., 1989; Williams & Tinoco, 1986).

RESULTS

The RNA Model System. In order to simplify the flavivirus 3′ structure and to focus on the tertiary structure of the flavivirus 3′-terminal RNA, we utilized a model RNA sequence that contained the thermodynamically predicted small SL and the bottom portion of the larger SL as depicted in Figure 1B. The deleted portion of the large SL was replaced by a loop of four U's. The predicted 5′ SL was designated SL2, while the 3′ SL was designated SL1 (Figure 1B). In the WNV 3′ model RNA, an extra G/C base pair (at positions 15 and 20, respectively) was added at the top of SL1 to further stabilize the truncated stem. In addition to the nucleotides shown for the model RNA structures in Figure 1B, the *in vitro* synthesized RNAs also contained four extra nucleotides at the 5′ end (5′-GGGC-3′) and one extra G at the 3′ end derived from cloning. Both of the 3′ model RNAs maintained the structural and sequence elements necessary to form the proposed coaxial stacking or pseudoknot conformations as in the native RNAs (Figure 1B). The WNV and Den-3 3′ model RNAs were used in the experiments described below. Comparative analysis of different flavivirus sequences (for example, Figure 1A) supports the prediction of SL1 from thermodynamic data; however, sequence alignment uncertainties make the prediction of SL2 less sure because of the number of sequences currently available. Such comparisons will become more useful when a large number of divergent flavivirus sequences are available.

Circular Dichroism (CD) Spectroscopic Analysis. The conformation of the WNV 3′ model RNA was first analyzed by CD spectroscopy at various temperatures. At low temperatures, the CD spectra indicated the presence of a highly structured RNA conformation with a significant amount of A-form helix (Figure 2). Features characteristic of A-form helical conformations include a positive peak at 262 nm, a small negative signal at approximately 235 nm, and a large negative signal at 209 nm. When the temperature was increased, the CD spectra changed significantly. At temperatures above 90 °C, the CD spectra was typical of an unstructured single-stranded RNA.

Ribonuclease Digestion Analysis. To further define the structure of the WNV 3′ model RNA, ribonuclease probing analysis was carried out. The RNA was *in vitro* transcribed, 5′ end labeled with ^{32}P, and subjected to limited digestion under native conditions with double-strand-specific RNase V1 and single-strand-specific RNase A or RNase phyM in the presence of 10 mM Mg^{2+} as described in Materials and Methods. The digested products were analyzed on a 10% polyacrylamide sequencing gel (Figure 3A). A summary of the RNase probing results is presented in Figure 3B. A similar RNase cleavage pattern was obtained in the presence of 0.2 mM Mg^{2+} (Figure 1 in supporting information). RNase V1 cleavage of the predicted stems of SL1 and SL2

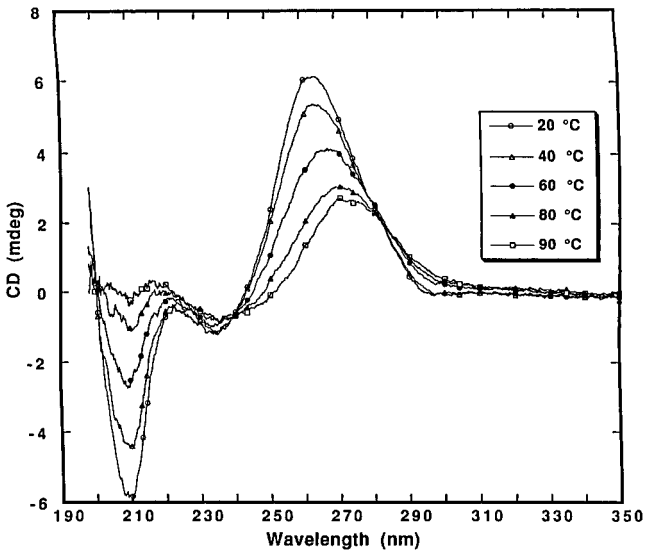

FIGURE 2: Circular dichroism spectra of WNV 3′ model RNA as a function of temperature. At low temperatures, the CD spectra showed a highly structured RNA conformation with a significant amount of A-form helix. The spectra were collected in 5 mM Na_2HPO_4, pH 7.0, 100 mM NaCl, and 0.1 mM EDTA in a 10 mm pathlength cell on a Jasco J-600 spectropolarimeter.

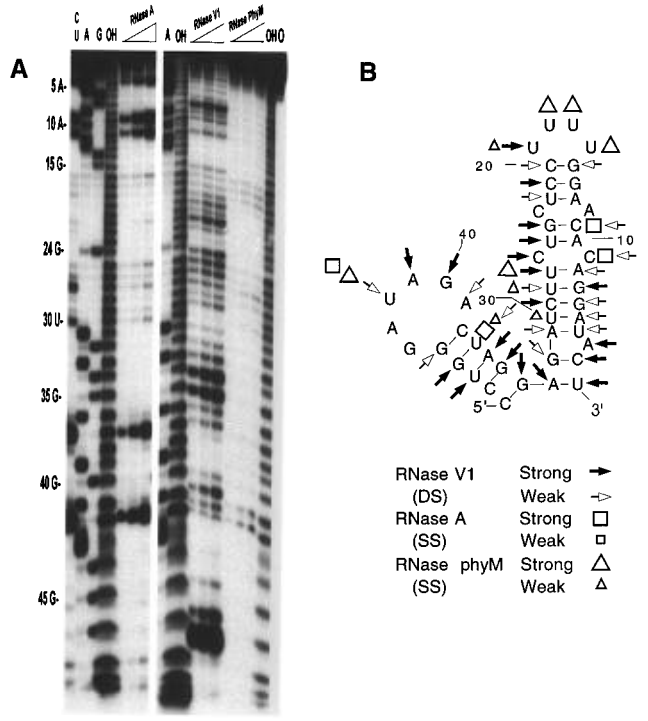

FIGURE 3: RNase probing of the WNV 3′-terminal model RNA. (A) ^{32}P-5′-labeled WNV 3′ model RNA was digested under native conditions with increasing amounts of RNase A (single-strand-specific nuclease), RNase PhyM (single-strand-specific nuclease), or RNase V1 (double-strand-specific nuclease). The cleavage products were separated by a 10% polyacrylamide sequencing gel. RNA sequencing reactions and alkaline hydrolysis ladders were used as markers. Lane C/U: RNase A digested RNA. Lane A: RNase T2 digested RNA. Lane G: RNase T1 digested RNA. Lane OH: alkaline hydrolysis ladder. Lane 0: no nuclease treatment. (B) Summary of the RNase cleavage results. Both strong and weak ribonuclease cleavages are indicated.

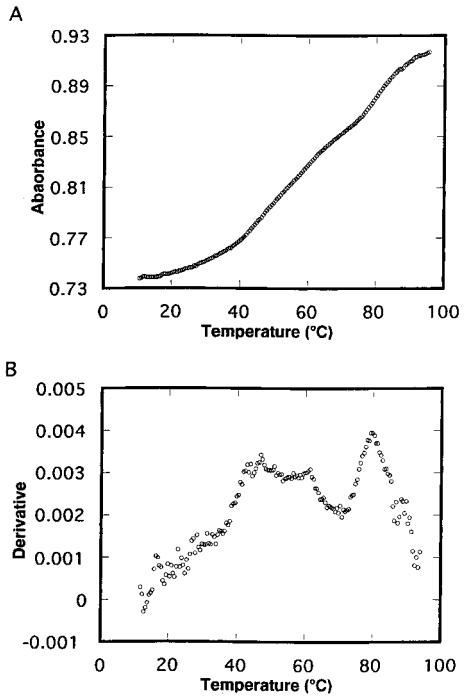

FIGURE 4: Thermal melting curve (A) and its derivative (B) of the WNV 3′ model RNA. Three transitions are easily visible in the derivative plot. The experiment was conducted at 260 nm on a Cary 4 spectrometer in the same buffer as described in Figure 2.

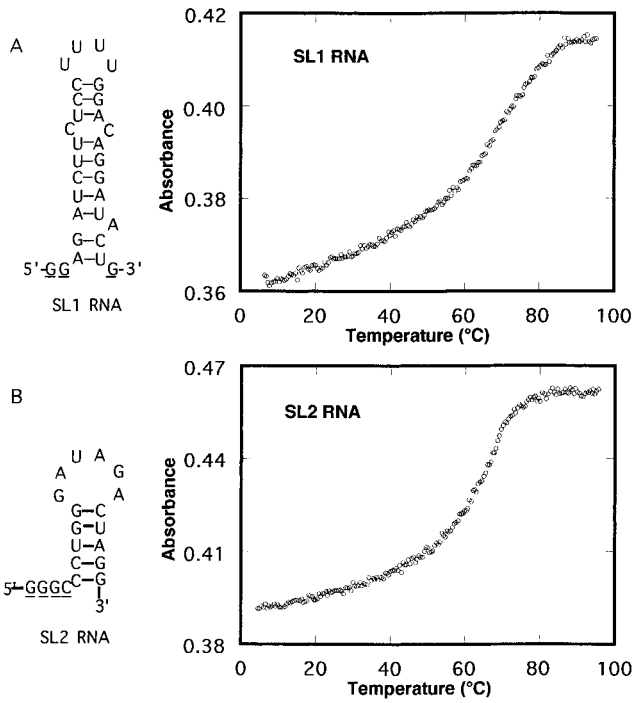

FIGURE 5: Transition assignment of the WNV 3′ model RNA. SL1 and SL2 of the WNV 3′ model RNA were separately synthesized *in vitro* and subjected to UV-melting analysis. (A) SL1 secondary structure and its UV-melting curve. (B) SL2 secondary structure and its UV-melting curve. The underlined nucleotides are nonviral sequences present in the complete WNV 3′ model RNA structure from cloning. The experimental conditions were the same as described in Figure 4.

indicated that these regions are double stranded. Also, strong cleavage by RNase phyM at nucleotides 16−18 and weak cleavage at nucleotide 19 indicate that the thermodynamically predicted loop located at the top of SL1 is single stranded. No cleavage was obtained at nucleotides 10, 12−14, 23, 38, or 43−44. These results indicate that the thermodynamically predicted secondary structure of the two adjacent SLs formed by the WNV 3′ RNA was substantially correct.

Cleavage at several nucleotides was not consistent with a structure consisting of only two SLs. Nucleotides 39−42 in the loop of SL2 were hypersensitive to digestion by the double-strand-specific nuclease, RNase V1, although nucleotide 42 was also cleaved by the single-strand-specific nucleases, RNase A and RNase phyM. In addition, nucleotides 24−27 showed hypersensitivity to RNase V1, although nucleotide 27 was also cleaved by RNase phyM. In the complementary portion (nt 8−11) in SL1, C at position 9 and C at position 11 showed hypersensitivity to RNase A. The A at position 8 and the C's at positions 9 and 11 were weakly cleaved by RNase V1. These results are consistent with the existence of a higher order structure in the flavivirus 3′-terminal sequence, in which nucleotides 40−42 in the loop of SL2 may base pair with nucleotides 24−26 located on the 5′ side of an unstable region of SL2. The U at position 27 might be in an equilibrium in pairing with either A at position 39 or A at position 8.

Thermal Melting Analysis of the WNV 3′ Model RNA. To obtain further evidence about the structure and thermal transitions of the WNV 3′ model RNA, we carried out a series of UV−thermal melting studies, and a reversible, complex UV−visible melting curve with significant hyperchromicity was obtained. As illustrated in Figure 4A, overlapping low temperature transitions were followed by a resolved high temperature transition in the absence of Mg^{2+}. The derivative curve of the same experiment indicated that there were two partially resolved transitions with peaks around 46 and 62 °C and a high temperature transition at 79 °C (Figure 4B). Melting curves of similar shape were obtained with the RNA in the same buffer but with the addition of 0.2 mM Mg^{2+} (Figure 2 in supporting information). When the Mg^{2+} concentration was increased above 0.2 mM, the low temperature transition exhibited a larger differential shift to high temperature than the other transitions and started to merge with the 62 °C transition (data not shown). Similar melting curves were obtained at wavelengths of both 260 and 284 nm (data not shown).

Transition Assignment of the WNV 3′ Model RNA. In order to understand the structure of the flavivirus genomic RNA, it is necessary to assign the three observed melting transitions to particular structural elements of the 3′ model RNA. The two SLs making up the WNV 3′ model structure were synthesized separately and subjected to thermal melting analysis (Figure 5). SL2 contained the same four extra (5′-GGGC-3′) nonviral nucleotides at the 5′ end as does the complete WNV 3′ model RNA. The melting curve of SL2 RNA contained a single transition with a T_m of 62 °C (Figure 5B), which is identical to the midpoint of the central transition (62 °C) observed in the complete WNV 3′ model RNA. The SL1 contained two viral G's at the 5′ end and an extra nonviral G at the 3′ end as does the complete WNV 3′ model RNA. The SL1 RNA showed a single transition with a T_m of 72 °C (Figure 5A), which is similar to the upper transition (79 °C) observed with the complete WNV 3′ model RNA. No transition around 46 °C was obtained from either the SL1 or SL2 RNA. These results suggest that, in the complete WNV 3′ model RNA, the 62 and 79 °C transitions represent the unfolding of SL2 and SL1, respectively, while the 46 °C transition represents the opening of a tertiary

structure (Figure 1B). We assume that the 79 °C upper transition in the complete WNV 3′ model RNA, versus 72 °C in separate SL1 alone, is due to some additional interaction that occurs between the unfolded SL2 strand and SL1.

Comparison of the Thermal Melting Curves of WNV and Den-3 3′ Model RNAs. Since the 3′-terminal secondary structure predicted from thermodynamic data is highly conserved among divergent flaviviruses, we compared the thermal melting curve of the WNV 3′ model RNA with that of the Den-3 3′ model RNA. Similar thermal melting curves were obtained for the Den-3 and WNV 3′ model RNAs with three transitions at 45, 52, and 80 °C. However, the central transition obtained for the Den-3 3′ model RNA was shifted to lower temperature and was less well resolved from the low temperature transition (data not shown) as compared with the WNV 3′ model RNA (Figure 4). The results obtained for the two different viral RNAs further support the transition assignment, since SL2 of the Den-3 3′ model RNA is one base pair shorter than that of WNV 3′ model RNA (Figure 1B). Because of this difference, the central transition for the Den-3 3′ model RNA would be expected to occur at lower temperature than that for the WNV 3′ model RNA. The results obtained are consistent with these expectations. The similar melting curves obtained for the WNV and Den 3′ model RNAs provide further support for the conservation of structure at the 3′ terminus of the flavivirus genomic RNA.

Thermal Melting Analysis of WNV 3′ Model RNA Mutants. RNase probing analysis suggested the possibility of base pairing between the 5′-UAG-3′ (nt 42−40) and 5′-CUG-3′ (nt 26−24) and the U at position 27 might be in an equilibrium in pairing with either A at position 39 or A at position 8 in the WNV 3′ model RNA (Figure 1B). Such an interaction would result in the bulging out of the mostly conserved 5′-ACAC-3′ (nt 12−8). To further test these interactions in the flavivirus genomic RNA, three mutant RNAs were made (Figure 6A). The mutant RNAs were subjected to thermal melting analysis, and the results are shown in Figure 6B. If the proposed structure is correct, a change from a C to a G at position 9 (mutant 1) would favor the pairing of the C at position 26 with the G at position 9 and at the same time would disrupt potential pseudoknot base pairing of the G at position 40 with the C at position 26. The melting curve of mutant 1 RNA was dramatically different from that of the wild-type RNA, and the low temperature tertiary structural transition was not observed. The mutant 2 RNA, which contained two base substitutions (C at position 9 to a G and G at position 40 to a U), showed a melting curve almost identical to that of the mutant 1 RNA, suggesting that the second change (G to U at position 40) had little further effect on the structure of the RNA. Mutant 3 RNA, in which the A at position 8 was substituted by a U, gave a melting curve similar to that of the wild-type RNA. This indicated that the A at position 8 was not as critical for maintaining the tertiary structure of the RNA. These data strongly support the proposed pseudoknot base pairing of 5′-UAG-3′ (nt 42−40) with 5′-CUG-3′ (nt 26−24).

Pseudoknot Molecular Modeling. Molecular modeling of the WNV 3′ model RNA was carried out to determine (1) whether, after formation of the predicted secondary structure, the pseudoknot indicated in Figure 1B is a stereochemically and an energetically reasonable possibility and (2) what general conformational features are exhibited by the energy-minimized tertiary structure. For the modeling studies the 4U loop and the adjacent C/G base pair of SL1 as depicted in Figure 1B were not used as shown in Figure 7. Helical sections H1, H2, H3, and H1′ (of the proposed pseudoknot) were first constructed in A-form conformations and linked together in the sequence described below (Figure 7). Helical segments H1 and H2 were linked to each other; the 3′ U of the H1 helix is the 3′ terminus of the viral RNA model, while the 5′ C in the H2 helix leads to the remainder of the viral RNA genome sequence. The H3 helix was connected to the H2 helix as shown in Figure 7. The remaining connection between the H2 and H3 helices was accomplished by connecting them through an A-G unpaired dinucleotide. At this point, internal rotations of the molecule with the SYBYL software allowed the remaining end of the H3 helix to be connected to the H1 helix. The single-stranded 5′-ACACA-3′ unit was connected to the A nucleotide of H1′, and with additional internal rotations of the helical and single-stranded segments, the remaining connections of the H3 helix to the H1′ helix and of the 5′-ACACA-3′ strand to the H1 helix were made (Figure 7).

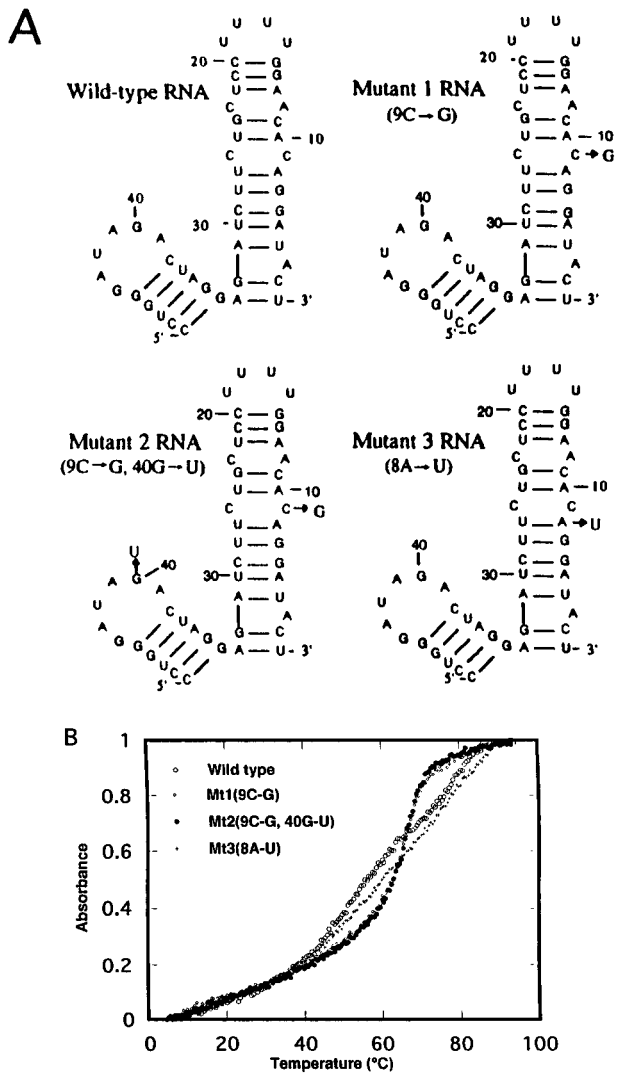

FIGURE 6: Thermal melting transition analysis of the WNV 3′ model RNA and three mutant variants. (A) Computer-predicted secondary structures of the wild-type WNV 3′ model RNA and three mutant RNAs. Substituted nucleotides are indicated with an arrow: mutant 1 (C9→G), mutant 2 (C9→G and G40→U), and mutant 3 (A8→U). (B) Thermal melting curves of wild-type WNV 3′ model RNA and of each of the mutant RNAs.

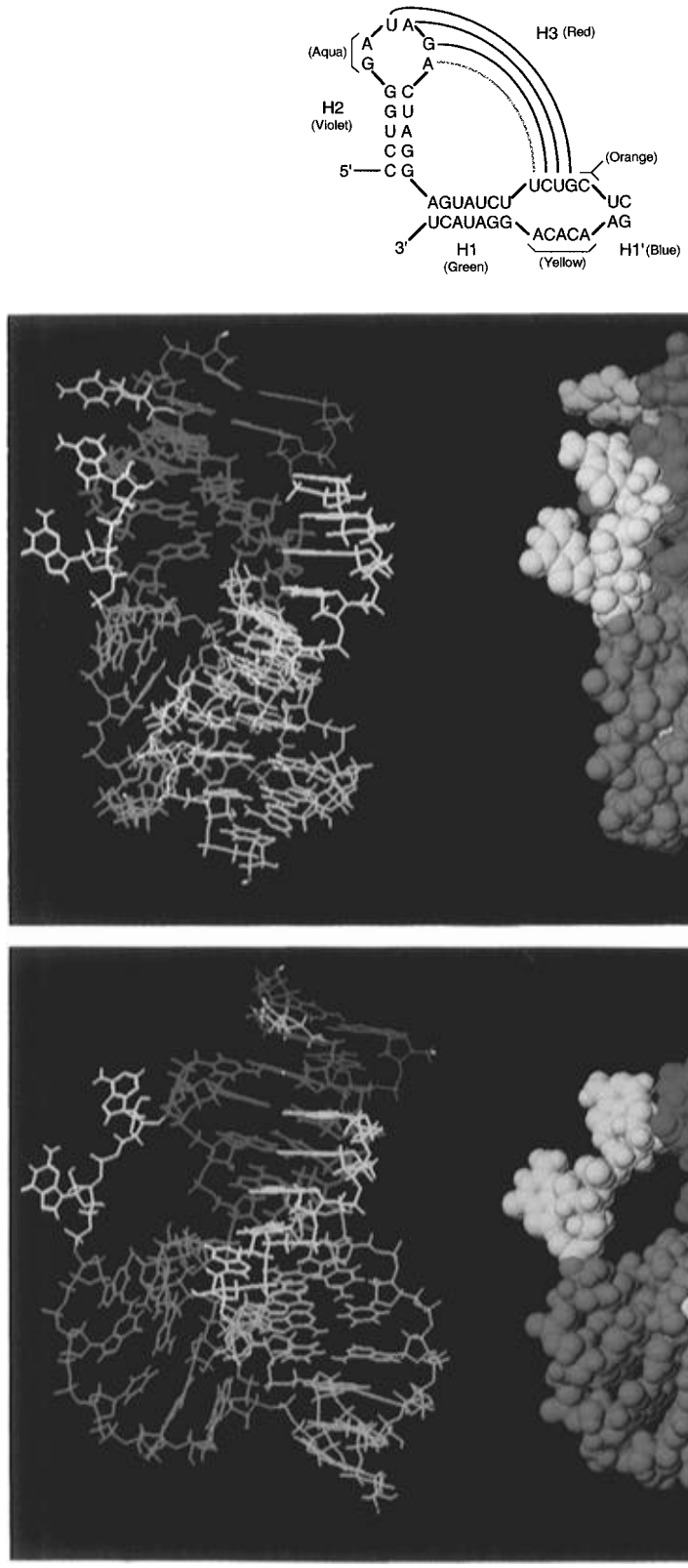

FIGURE 7: Pseudoknot model. A schematic diagram of the sequence of the WNV 3′ model RNA with labeled helical regions (top panel) and two views of the energy-minimized model (middle and bottom panels) are shown. The color scheme is as indicated in the top panel. The view of the structure shown in the middle panel looks directly into the cleft formed by the GA and ACACA strands. The view shown in the bottom panel is rotated approximately 30° clockwise relative to that shown in the middle panel. The light line in the top panel indicates that the U in the internal loop of SL2 could potentially pair with either the A in the loop of SL2 or the A in the other strand of the internal loop of the SL1.

The molecular model was constructed without any unusual bond lengths or angles and with the helical segments in A-form conformations. The pseudoknot helix dramatically constrains the conformational space available to the RNA model. The model was then subjected to molecular mechanics energy minimization as described in Materials and Methods. In the initial stages of energy minimization the terminal base pairs of the helical segments were constrained

to remain base paired, and these constraints were removed as an energy minimum was approached. A low energy, compact, stereochemically reasonable RNA conformation was obtained from the minimization (Figure 7), and very similar final structures were obtained for all models that were built with the pseudoknot helix. This result clearly illustrates that the pseudoknot conformation is a reasonable proposal for the 3′ region of flavivirus RNA.

DISCUSSION

The 3′-terminal nucleotides of the flavivirus genomic RNA can form two adjacent SLs that may interact to form a tertiary structure (Figure 1). These predicted structures are highly conserved among divergent flaviviruses, suggesting that they are functionally important. The experiments described here were designed to obtain evidence to support the existence of a tertiary structure and to test two hypothetical tertiary structures, (1) a pseudoknot formed by base pairing between nucleotides in the loop of SL2 and on the 5′ side of an unstable region of SL1 (Figure 1B) or (2) a coaxially stacked long A-form helix of SL1 and SL2 with loops at each end and a nick in the middle region on one side of the stem.

The CD spectral results and the observed UV hypochromicity during melting transitions confirmed that the flavivirus 3′ model RNA contains extensive regions of stacked base pairs in an A-form helical conformation. Although two SLs were predicted by secondary structural folding of the 3′ model RNA, three transitions were observed in UV-melting experiments. Additional transitions are frequently observed during the thermal denaturation of RNA that contains hairpins organized into a tertiary structure (Cole et al., 1972; Crothers et al., 1974; Jaeger et al., 1990; Banerjee et al., 1993; Chang & Tinoco, 1994; Laing & Draper, 1994; Gluick & Draper, 1994). The low temperature transition often represents the unfolding of the tertiary structure, while the high temperature transitions represent the unfolding of individual hairpin helices. In agreement with these reports, separate preparation and UV-melting analysis of each of the two SLs gave single melting transitions for each with transition midpoints that correlated with the two high temperature transitions observed for the complete 3′ model RNA. These results strongly suggest that the 3′ model RNA contains a tertiary structural element with a 46 °C melting transition**.** This low temperature transition could be due to loss of coaxial stacking of the two hairpins, perhaps with additional unfolding of terminal base pairs in the hairpins to give the observed hypochromicity, or it could be due to disruption of the proposed pseudoknot (Figure 1B).

To distinguish between the two tertiary-structure motifs, ribonuclease probing experiments were carried out on the WNV 3′ model RNA. Extensive cleavage by double-strand-specific RNase V1 of the thermodynamically predicted stem regions of the two SLs confirmed the existence of the predicted secondary structures (Figure 3). However, cleavage at several nucleotides was not consistent with a structure consisting of only two SLs. Strong RNase V1 cleavages at nt 42−39 (5′-UAGA-3′) and nt 27−24 (5′-UCUG-3′) as well as strong RNase A cleavage of the C at position 11 were not predicted by the secondary structure (Figure 3). A possible explanation for these observations is that there is higher order folding of the RNA, involving base pairing between the 5′-UAG-3′ (nt 42−30) and 5′-CUG-3′(nt 26−24). The U at position 27 might be in an equilibrium pairing with either the A at position 39 or the A at position 8 (Figure 7, top panel). This explanation was further supported by mutagenesis analysis. The UV-melting curves of two RNAs with mutations that disrupted the proposed pseudoknot base pairings did not contain the low temperature tertiary structure transition, while RNA with a mutation outside the pseudoknot base pair region showed a UV-melting curve similar to that of wild-type RNA (Figure 6). The proposed pseudoknot model is the best fit for all of the available experimental results. In addition, phylogenetic comparison suggests that similar pseudoknot base pairs can be formed by nucleotides in the 3′-terminal regions of all sequenced flavivirus genomes and suggests a functional relevance for this structure.

Similar RNase digestion patterns, as shown in Figure 3 and Figure 1 in supporting information, were obtained over a broad range of Mg^{2+} concentrations, indicating that the tertiary structure is stable from low to high Mg^{2+} concentrations. Melting studies at increasing Mg^{2+} concentrations indicated that the low temperature transition, which we have assigned to the disruption of tertiary structure, is stabilized to a greater extent than the hairpin to strand transitions of SL1 and SL2. Additional investigations of the effects of metal ions on the flavivirus 3′ RNA structure are in progress.

Even with the strong experimental support for the pseudoknot, we were concerned that the pseudoknot, as indicated in Figure 1B, might not be stereochemically feasible within the context of the highly constrained hairpins of the 3′ model RNA. To evaluate the structural possibilities for the pseudoknot, the structure shown in Figure 7 was built and energy miniminized. The unstable internal loop of SL1 (Figure 7) was unstacked, and the 5′-UCUG-3′ sequence was base paired with 3′-AGAU-5′ from the loop of SL2 to form the pseudoknot (H3 in red in Figure 7). The structure was further stabilized by stacking the H3 helix with the upper base pairs of SL1 (H1′ in blue in Figure 7). The remainder of the helical stem of SL1 (H1 in green in Figure 7) and the stem of SL2 (H2 in violet in Figure 7) are approximately perpendicular in the energy-minimized structure. The three major helical segments are linked by the GA from the loop of SL2 and the internal loop (5′-ACACA-3′) of SL1 (as shown in Figure 7). It is clear from this energy-minimized structure that the pseudoknot as suggested by the phylogenetic comparison, and by the experiments described above, is a stereochemically reasonable conformation.

Although the pseudoknot structure shown in Figure 7 is only a model, it has several interesting and potentially important features. The helical segments H1 and H2 form a solid base for the structure while the helical segment contributed by H3 and H1′ forms the back and top of the structure (Figure 7). The sides of the structure are composed of the mostly conserved residues 5′-ACACA-3′ from the internal loop of SL1 and GA from the loop of SL2. These single-stranded regions are displayed on the same side and in the same region of the structure in a manner that would allow recognition by a single protein. The two strands enclose a cleft that could form additional contact sites for a bound protein and that could also be a strong metal binding site. The loop in SL2 is more conserved than expected for base pairing alone, and it may also be involved in protein binding. Another interesting feature of the model is that the U at position 27 could potentially pair with either the A at position 8 of the internal loop of SL1 or the A at position

39 of the loop of SL2. The A at position 8 stacks onto H1 while the A at position 39 stacks onto H3. In the model, U at position 27 forms a distant base pair with the A at position 8, but it appears that it could rotate to interact with the A at position 39. We plan to investigate this in more detail by dynamics calculations and by preparation of additional mutant RNAs.

Three cellular proteins have recently been found to bind specifically to the 3′ terminus of flavivirus genomic RNA (Blackwell & Brinton, 1995). The binding site of one of the identified host proteins (p84) resides within the 3′ model RNA region. It is well documented that the local conformation of an RNA can be changed upon binding to protein (Wikman et al., 1982; Rould et al., 1989; Matsumoto et al., 1990; Andreazzoli & Gerbi, 1991; Puglisi et al., 1992; Battiste et al., 1994). We hypothesize that the pseudoknot structure of the flavivirus 3′-terminal RNA may play an important role in the recognition of the 3′ RNA by host and viral proteins during the initiation of viral RNA replication.

ACKNOWLEDGMENT

We thank a reviewer for suggestions on sequence comparison for RNA structure prediction.

SUPPORTING INFORMATION AVAILABLE

Figure 1 showing RNase probing of the WNV 3′-terminal model RNA in the presence of 0.2 mM Mg^{2+} and Figure 2 showing the thermal melting curve (A) and its derivative (B) of the WNV 3′ model RNA in the presence of 0.2 mM Mg^{2+} (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Andreazzoli, M, & Gerbi, S. A. (1991) *EMBO J. 10*, 767−777.
Banerjee, A. R., Jaeger, J. A., & Turner, D. H. (1993) *Biochemistry 32*, 153−163.
Battiste, J. L., Tan, R., Frankel, A. D., & Williamson J. R. (1994) *Biochemistry 33*, 2741−2747.
Blackwell, J. L., & Brinton, M. A. (1995) *J. Virol. 69*, 5650−5658.
Brinton, M. A. (1983) *J. Virol. 46*, 860−870.
Brinton, M. A. (1986) in *Togaviridae and Flaviviridae, the viruses* (Schlesinger, S., & Schlesinger, M., Eds.) pp 329−376, Plenum, New York, NY.
Brinton, M. A., & Dispoto, J. H. (1987) *Virology 162*, 290−299.
Brinton, M. A., Fernandez, A. V., & Dispoto, J. H. (1986) *Virology 153*, 113−121.
Chambers, T. J., Hahn, C. S., Galler, R., & Rice, C. M. (1990) *Annu. Rev. Microbiol. 44*, 649−688.
Chandrasekaran, R., & Arnott, S. (1989) in *New Series* (Madelung, O., Ed.) Vol. VII/1b, pp 31−170, Landolt-Bornstein.
Chang, K.-Y., & Tinoco, I., Jr. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8705−8709.
Cole, P. E., Yang, S. K., & Crothers, D. M. (1972) *Biochemistry 11*, 4358−4368.
Crothers, D. M., Cole, P. E., Wilbers, C. W., & Shulman, R. G. (1974) *J. Mol. Biol. 87*, 63−88.
Drepher, T. W., & Hall, T. C. (1988) *J. Mol. Biol. 201*, 31−40.
Gluick, T. C., & Draper, D. E. (1994) *J. Mol. Biol. 241*, 246−262.
Heus, H. A., & Pardi, A. (1991) *Science 253*, 191−194.
Jacobson, S. J., Konings, D. A. M., & Sarnow, P. (1993) *J. Virol. 67*, 2961−2971.
Jaeger, J. A., Turner, D. H., & Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706−7710.
Jaeger, J. A., Zuker, M., & Turner, D. H. (1990) *Biochemistry 29*, 10147−10158.
Kibler-Herzog, L., Kell, B., Zon, G., Shinozuka, K., Mizan, S., & Wilson, W. D. (1990). *Nucleic Acids Res.18*, 3545−3555.
Lai, C.-J., Men, R., Pethel, M., & Bray, M. (1992) in *Vaccines 92* (FBrown, F., Chanock, R. M., Ginsberg, H. S., & Lerner, R. A., Eds.) pp 265−270, Cold Spring Harbor Press, Cold Spring Harbor, NY.
Laing, L. G., & Draper, D. E. (1994) *J. Mol. Biol. 237*, 560−576.
Lorsch, J. R., & Szostak, J. W. (1994) *Biochemistry 33*, 973−982.
Manning, G. S. (1978) *Q. Rev. Biophys. 11*, 179−246.
Matsumoto, T., Nishikawa, K., Hori, H., Ohta, T., Miura, K. I., & Watanabe, K. (1990) *J. Biochem. 107*, 331−338.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
Murphy, F. L., Wang, Y.-H., Griffith, J. D., & Cech, T. R. (1994) *Science 265*, 1709−1712.
Pleij, C. W. A., Rietveld, K., & Bosch L. (1985) *Nucleic Acids Res. 13*, 1717−1731.
Pley, H. W., Flaherty, K. M, & McKay D. B. (1994) *Nature 372*, 68−74.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson J. R. (1992) *Science 257*, 76−80.
Pyle, A. M., & Green J. B. (1995) *Curr. Opin. Struct. Biol. 5*, 303−310.
Record, M. T., Anderson, C. F., & Lohman, T. M. (1978) *Q. Rev. Biophys. 11*, 103−178.
Rice, C. M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheets, R. L., & Strauss, J. H. (1985) *Science 229*, 726−735.
Rould, M. A., Perona, J. J., Söll, D., & Steitz, T. A. (1989) *Science 246*, 1135−1142.
Szewczak, A. A., & Moore, P. B. (1995) *J. Mol. Biol. 247*, 81−98.
Turner, D. H., & Bevilacqua, P. C. (1993) in *the RNA world* (Gesteland, R. F., & Atkins, J. F., Eds.) pp 447−464, Cold Spring Harbor Press, Cold Spring Harbor, NY.
Veal, J. M., & Wilson, W. D. (1991) *J. Biomol. Struct. Dyn. 8*, 1119−1145.
Walter, A. E., & Turner, D. H. (1994) *Biochemistry 33*, 12715−12719.
Weiner, S. J., & Kollman, P. A. (1986). *J. Comput. Chem. 7*, 230−252.
Weiner, S. J., Kollman, P. A., Case D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765−784.
Wikman, F. P., Siboska G. E., Petersen, H. U., & Clark, B. C. (1982) *EMBO J.* 1, 1095−1100.
Williams, A. L., & Tinoco, I., Jr. (1986) *Nucleic Acids Res. 14*, 299−315.
Wimberly, B., Varani, G., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 1078−1087.
Wyatt, K. R., & Tinoco, I., Jr. (1993) in *the RNA world* (Gesteland, R. F., & Atkins, J. F., Eds.). pp 465−496, Cold Spring Harbor Press, Cold Spring Harbor, NY.
Yao, S., & Wilson, W. D. (1992) *J. Biomol. Struct. Dyn. 10*, 363−387.
Zou, E. T., Tanious, F. A., Wilson, W. D., Zon, G., Tan, G., & Wartell, R. M. (1990) *Biochemistry 29*, 4446−4456.

BI952398V