Analysis of Structural Properties which Possibly Are Characteristic for the 3'-Terminal Sequence of the Genome RNA of Flaviviruses

By G. WENGLER* AND E. CASTLE
Institut für Virologie, Justus-Liebig-Universität Giessen, D-6300 Giessen, F.R.G.

(Accepted 26 February 1986)

SUMMARY

Recently we have shown that an open reading frame comprising 10290 nucleotides is present on the infectious, single-stranded genome RNA of the West Nile flavivirus. We have now isolated cloned cDNA representing the 3'-terminal untranslated region of this molecule. The sequence of this region which comprises 571 nucleotides is given in this report. Recently, the nucleotide sequence of the genome RNA of the yellow fever flavivirus has been described. A comparative analysis of the 3'-terminal untranslated nucleotide sequences present in each genome suggests that in flaviviruses this region probably has the following properties. It has a heteropolymeric sequence at the 3' terminus. It contains one or more oligonucleotide sequences that are repeated. An extensive stem and loop structure can be folded from the nucleotide sequences present at the 3' terminus. The stem of this structure contains a conserved region introducing a defined mismatch into the stem. The loop of this structure probably contains short conserved oligonucleotide sequences in analogous positions. In both viruses the oligonucleotide CAUAUUGAC (~)AGAC lies closely in front of the sequence which can be folded into the stem of the 3'-terminal stem and loop structure and the oligonucleotide CUAGAGGUUAGAGGAGCC is strictly conserved between both viruses. The analyses indicate that the 3'-terminal untranslated region of the genome of flaviviruses probably has rather unique characteristics of primary and secondary structure. Possible implications of these findings are discussed.

The flavivirus genus comprises a number of important pathogens, e.g. the viruses causing yellow fever, Japanese encephalitis, dengue, St. Louis encephalitis and tick-borne encephalitis (for review, see Shope, 1980). The genome of flaviviruses consists of a molecule of infectious, single-stranded RNA containing about 11000 nucleotides and the available evidence indicates that this molecule functions as mRNA for the synthesis of all structural and non-structural viral proteins (Stollar et al., 1967; Naeve & Trent, 1978; Wengler et al., 1978; for review see Westaway, 1980). Recently, we have shown that a long open reading frame (ORF) comprising 10290 nucleotides is present on the genome RNA of the West Nile (WN) flavivirus (Castle et al., 1985, 1986; Wengler et al., 1985) and similar analyses have been done using the genome of the yellow fever (YF) flavivirus (Rice et al., 1985). The 3'-terminal nucleotide sequence of the YF virus genome has also been determined by Grange et al. (1985). Since the cDNA used in our experiments was synthesized using a mixture of random hexanucleotides derived from calf thymus DNA as primer, the total 3'-terminal sequence of the genome RNA was not present in these cDNA molecules. We have now prepared cDNA containing the 3'-terminal nucleotide sequence of the WN virus genome RNA and determined its primary structure. A comparison of this nucleotide sequence with the corresponding sequence derived from YF virus (Rice et al., 1985; Grange et al., 1985) indicates that the 3'-untranslated region of the genome of flaviviruses has characteristic structural properties which are described in this report.
Viral genome 42S RNA was isolated from purified WN virus by phenol extraction followed by sucrose density gradient centrifugation as described previously (Castle et al., 1985). The 3' terminus of the WN virus genome does not contain a poly(A) sequence and is not readily accessible for enzymic modification in the intact molecule (Wengler & Wengler, 1981). In order to polyadenylate the 3'-terminal nucleotide sequence the following procedure was used. Twenty μg genome RNA was solubilized in 40 μl buffer containing 50 mM-NaHCO₃, 0.5 mM-disodium EDTA, pH 9.1 and incubated at 90 °C for 1 min followed by ethanol precipitation at −70 °C for 10 min in order to fragment the RNA and thereby expose the 3' terminus for enzymic modification. The fragmented RNA was redissolved in 32 μl H₂O, and 10 μl buffer containing 250 mM-Tris-HCl pH 7.9, 50 mM-MgCl₂, 1250 mM-NaCl, 0.5 mg/ml autoclaved gelatine was added. After addition of MnCl₂ and ATP to final concentrations of 2.5 mM and 0.5 mM, respectively, 10 units poly(A) polymerase was added at 37 °C and the sample which had a final volume of 50 μl was incubated at 37 °C for 10 min. The RNA was recovered by phenol extraction and precipitated with ethanol at −70 °C.

Twenty μg fragmented, polyadenylated WN genome RNA and 10 μg oligo(dT)₁₂₋₁₈ were incubated with 25 units reverse transcriptase at 37 °C for 60 min in buffer containing 50 mM-Tris–HCl pH 8.3, 10 mM-MgCl₂, 100 mM-KCl, 1 mM-dTTP, 1 mM-dATP, 1 mM-dGTP, 0.5 mM-dCTP 10 mM-dithiothreitol and 15 μCi [α-³²P]dCTP. At the end of incubation the template RNA was hydrolysed by alkali treatment and the first-strand cDNA was recovered by phenol extraction followed by chromatography on Sephadex G-100 as described by Maniatis et al. (1982). The second-strand cDNA was synthesized using the endogenous priming activity of the first cDNA strand by the Klenow fragment of Escherichia coli DNA polymerase I exactly as described by Maniatis et al. (1982). Double-stranded cDNA was treated with S1 nuclease, tagged by addition of oligo(dC) and cloned using pBR322 containing oligo(dG) tails added to the PstI site exactly as described by Castle et al. (1985). Transfected E. coli 803 colonies were stored in microtitre plates.

Filter paper replicas of bacterial colonies present in microtitre plates prepared as described by Gergen et al. (1979) were lysed by alkali, and colony hybridization using the oligonucleotide 5'-³²P-pGGAUCCUGUGUUOH which is complementary to the T-terminal sequence of WN genome RNA (Wengler & Wengler, 1981) was performed as described by Dalbadie-McFarland et al. (1982) and Singer-Sam et al. (1983).

Computer analyses of the primary and secondary structure of the individual RNA molecules and of the structural features that are conserved between the viruses were performed using the sequence analysis programs developed by Queen & Korn (1984).

A poly(A) sequence was added onto the 3' terminus of the WN genome RNA, double-stranded cDNA was synthesized using oligo(dT) as primer and the cDNA was cloned in pBR322 as described above. An oligonucleotide complementary to the 3' terminus of the WN genome RNA was used as a probe to detect colonies containing recombinant plasmid molecules containing the 3'-terminal sequence of the WN genome. Plasmid DNA was prepared from appropriate colonies and the cDNA insert was recovered by treating this DNA with PstI followed by agarose gel electrophoresis. These analyses allowed us to conclude that the cDNA present in clone 37/H5 comprised approximately 1100 nucleotides and probably contained the 3'-terminal nucleotide sequence of the WN genome RNA. The primary structure of this cDNA insert was determined by subcloning restriction fragments derived from this molecule into vectors M13 mp8, mp9, mp10 and mp11 and determination of the primary structures of these fragments using the chain termination method (Sanger et al., 1977). Besides the terminal region of the long ORF of the WN genome (Castle et al., 1986) the total nucleotide sequence of the untranslated 3' terminus of the WN genome RNA which comprises 571 nucleotides is present in this molecule. This sequence is shown in Fig. 1 (a).

The results obtained from computer analyses for the presence of repeated nucleotide sequences and for inverted repeated sequences are also indicated. Three different nucleotide sequences which are underlined by different symbols in Fig. 1 are each repeated once rather faithfully and an extensive stem and loop structure can be formed from inverted repeat sequences present at the 3’ terminus of the molecule.
Fig. 1. Nucleotide sequence of the 3'-terminal untranslated region of the genomes of (a) WN and (b) YF viruses. The plasmid DNA present in colony 37/H5 was isolated, the cDNA insert present in this plasmid was recovered and its primary structure was determined. (a) The nucleotide sequence following the UAA stop codon which terminates the long ORF of the WN virus genome RNA (Castle et al., 1986) is shown. The homopolymeric sequences which were introduced during the cloning procedure are not included in the sequence. The data are in accordance with our earlier finding that the 3' terminus of the 42S WN genome RNA has the structure AACACAGGAUCU...

(b) The 3'-terminal nucleotide sequence which follows the UGA codon terminating the long ORF on the genome RNA of YF virus (Rice et al., 1985) is shown. The three oligonucleotide sequences which, with small imperfections, are repeated in the WN virus-derived sequence are underlined by open circles, closed circles and arrowheads, respectively (a). The oligonucleotide which is repeated with slight imperfections three times in the YF virus-derived sequence is underlined by triangles (b). Inverted repeat sequences, which by hybridizing with each other would give rise to a stem and loop structure, are underlined by wavy lines. Long oligonucleotide sequences which are conserved between both viruses are included in boxes. The box made up of broken lines contains a nucleotide sequence which is imperfectly conserved between the two viruses; the three positions in which different nucleotides are present are indicated by asterisks. The closed box contains a sequence which is repeated exactly in the WN virus-derived sequence and which is also present without alterations in the YF virus-derived sequence.
Recently the primary structure of the genome RNA of YF virus has been described by Rice et al. (1985) and the 3'-terminal sequence of this RNA was independently determined by Grange et al. (1985). The nucleotide sequence of the 3' untranslated region of this molecule is shown in Fig. 1 (b). The results of computer analyses for the presence of repeated sequences and of inverted repeats are also indicated. In accordance with the observations made by Rice et al. (1985) and by Grange et al. (1985) it can be seen that a single oligonucleotide sequence is repeated rather exactly three times in this region and that inverted repeat sequences are present.

The results obtained from a computer search for the presence of nucleotide sequences that are common to the 3'-terminal untranslated sequence of the WN and YF virus genomes are also indicated in Fig. 1. These nucleotides are included in boxes.

The stem and loop structure which according to these computer analyses possibly is present at the 3' terminus of the viral genome, and the conserved nucleotide sequence preceding this structure are shown in Fig. 2 for both viruses. It can be seen that a similar 3'-terminal stem and loop structure is derived by these analyses for the two viruses. Whereas the computer analyses performed by Rice et al. (1985) have led to a similar prediction, those of Grange et al. (1985) have led to a somewhat different secondary structure prediction of the 3' terminus of YF RNA. It can be seen from the data shown in Fig. 2 that the 3' terminus of the YF genome is complexed to an AGUGG sequence and that this sequence is also present between the two large oligonucleotides which are conserved between WN and YF viruses as indicated by dots in Fig. 2 (b). In the secondary structure model of Grange et al. the sequence indicated by these dots is complexed to the 3' terminus of the genome. Inspection of Fig. 2 (a) shows that the WN virus-derived nucleotide sequence present between the oligonucleotides conserved between WN and YF virus does contain a nucleotide sequence complementary to the WN virus 3' terminus.

Fig. 2. Representation of the possible stem and loop structure and of the conserved oligonucleotide sequences present on the genomes of (a) WN and (b) YF viruses. The oligonucleotides underlined by wavy lines in Fig. 1 are aligned with each other, leading to the stem and loop structure shown here. The long nucleotide sequences which are conserved between both viruses are indicated as in Fig. 1 by boxes. The three positions within these conserved sequences in which different nucleotides are present in each virus are indicated by asterisks. Short nucleotide sequences which are strictly conserved in corresponding positions of the stem and loop structures of both viruses are also marked.
The analyses presented in Fig. 1 and 2 indicate that the 3'-terminal untranslated nucleotide sequence of the genome RNA of flaviviruses probably has the following properties. (i) In accordance with earlier findings (Wengler & Wengler, 1981) it contains no poly(A) sequence but has a heteropolymeric sequence at the 3' terminus. (ii) It contains one or more oligonucleotide sequences which are repeated; the nucleotide sequences present in these repeats are not necessarily conserved between different viruses. (iii) An extensive stem and loop structure can be folded from the nucleotide sequence present close to the 3' terminus of the molecule. (iv) The stem of this structure probably contains a conserved sequence region introducing a defined mismatch into the stem. (v) The loop of this structure probably contains short conserved oligonucleotide sequences in analogous positions. (vi) In both viruses the oligonucleotide sequence CAUAAUGAC(\(\delta\))CC(\(\delta\))GGGA(\(\delta\))AGAC lies closely in front of the sequence which can be folded into the stem of the 3'-terminal stem and loop structure. (vii) The oligonucleotide sequence CUAGAGGUUAGAGGAGACCC is strictly conserved between both viruses; it is located in an analogous position preceding the conserved oligonucleotide mentioned above in the viruses and in addition it is repeated in the WN virus-derived nucleotide sequence.

The analyses presented in Fig. 1 and 2 immediately give an explanation for our finding (data not shown) that an oligonucleotide complementary to the 3' terminus of the WN genome RNA was inactive as a primer for the synthesis of first-strand cDNA. Conserved nucleotide sequences present in the genomes of flaviviruses might be useful for the identification and characterization of flavivirus-infected cells by nucleic acid hybridization experiments. The oligonucleotides described above might be of value in such analyses. Further experiments are necessary to clarify the functional roles of the different conserved elements of primary and secondary structure identified above.

We thank Dr R. Rott for encouragement and support and U. Leidner for excellent technical assistance. We thank Drs Rice, Lench, Eddy, Se Jung Shin, Sheets and Strauss for sending us a preprint of their manuscript on the nucleotide sequence of yellow fever virus. This study was supported in part by the Sonderforschungsbereich 47 (Virologie) and by the Fonds der Chemischen Industrie.

REFERENCES


Short communication


(Received 19 November 1985)