Stable secondary structures at the 3′-end of the genome of yellow fever virus (17 *D* vaccine strain)

Thierry Grange+, Michèle Bouloy and Marc Girard

Unité de Virologie Moléculaire, UA CNRS 545, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

Received 7 June 1985

The sequence of the 3′-terminal 565 nucleotides of yellow fever virus has been determined from a cloned cDNA. Several structures are detectable: three tandemly repeated sequences, a series of inverted repeats and a stable secondary structure involving the 3′-terminal 120 nucleotides.

Yellow fever *Flavivirus* *Sequence analysis* *3′-terminal structure*

1. INTRODUCTION

Yellow fever virus (YF); a mosquito borne virus, belonging to the Flaviviridae family, is able to replicate in both arthropod and vertebrate cells. Its genome is a single-stranded RNA molecule of positive polarity (review [1,2]), approx. 11 000 bases long, which lacks a 3′-poly(A) sequence [3]. Cloned cDNA copies of the YF RNA could be powerful tools to study the regulatory mechanisms involved in the expression of the viral genome and to understand the basis of YF virulence.

Here we report the isolation and the nucleotide sequence of a cDNA copy corresponding to the 3′-end of the viral RNA of the vaccinal strain 17D.204 [4].

2. MATERIALS AND METHODS

2.1. *Purification of the viral RNA*

YF strain 17 D 204 (236th passage) was grown in SW13 cells. The virus and its RNA were purified as described [4].

+Present address: Institut Jacques Monod (CNRS), Unité INSERM 257, Université Paris 7, Tour 43,2 Place Jussieu, 75251 Paris Cedex 05, France

2.2. *Synthesis and cloning of cDNA*

When not quoted we have used current procedures reviewed by Maniatis et al. [5].

The RNA preparation was polyadenylated as described [6] following denaturation with 10 mM methylmercury hydroxide (CH_3HgOH). The synthesis of cDNA was performed using AMV reverse transcriptase in the presence of oligo(dT). After second strand synthesis and nuclease S1 treatment, double-stranded cDNA was tailed with oligo(dC) and hybridized to oligo(dG) tailed pBR 322, cut at the *Pst*I site. The mixture was used to transform calcium chloride treated *E. coli* HB101 [7] cells. The in situ hybridization of the library was performed with YF RNA, which was further purified on low melting temperature agarose gel and fragmented during extraction from the gel slice by partial alkaline hydrolysis. The RNA fragments were 5′-labeled with polynucleotide kinase and [γ-^{32}P]ATP. Nucleotide sequencing was performed according to Maxam and Gilbert [8].

2.3. *Ribonuclease T2 mapping*

YF RNA was 3′-labeled with [^{32}P]pCp and RNA ligase [9] after denaturation with CH_3HgOH. The labeled RNA was hybridized with the purified cDNA insert and after ethanol precipitation, the

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

hybrids were subjected to ribonuclease T2 digestion. The products were analysed on CH_3HgOH agarose gels.

3. RESULTS AND DISCUSSION

3.1. *Isolation of YF cDNA clones*

YF RNA lacks poly(A) tail and the only sequence data available when we initiated this work was the identity of the 3′-terminal nucleotide of the viral genome: a uracil [3]. Thus we chose to add, in vitro, a poly(A) tail at the 3′-end of the RNA to prime the synthesis of the first cDNA strand with oligo(dT).

Previous experiments pointed out the low accessibility of this 3′-end to enzymatic modification, probably because of the secondary structure of the molecule [3]. Thus, the intact RNA molecule was denatured with methylmercury hydroxide prior to the polyadenylation reaction. The denaturing agent was complexed just before addition to the reaction mixture. In spite of this treatment, YF RNA was not polyadenylated as efficiently as minor amounts of contaminating RNA molecules (fig.1).

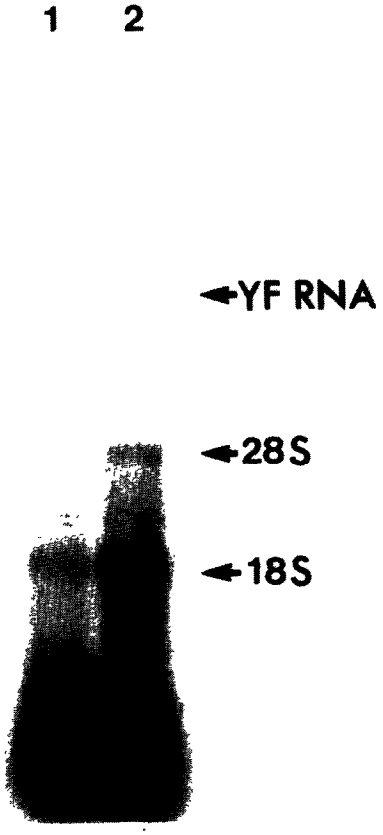

Fig.1. Analysis of the products of polyadenylation of YF RNA on CH_3HgOH agarose (1%) gel. The polyadenylation reaction was performed as described [6] with the following modification. After denaturation, 5 μg YF RNA were incubated for 15 min at 37°C in a reaction mixture of 50 μl containing 10 U *E. coli* poly(A) polymerase (BRL) and 10 μCi [γ^{32}P]ATP. Slot 1: the denaturing agent (CH_3HgOH) was complexed with β-mercaptoethanol just before (less than 1 s) the reaction mixture was added. Slot 2: the reaction mixture was added about 15 s after addition of the complexing agent. The same amount of trichloroacetic acid precipitable radioactivity (5000 cpm) has been loaded on each slot.

This polyadenylated RNA was used as a template for cDNA synthesis using oligo(dT) as a primer. A cDNA library of 1200 bacterial clones was obtained. To identify YF specific clones, in situ hybridization of the library was performed with 5′-labeled fragmented YF RNA. Three clones gave a positive signal with this probe. The largest cDNA containing plasmid, pYF1, was shown to hybridize with YF RNA on a Northern blot and so was further studied.

3.2. *Characterization of a cDNA clone corresponding to the 3′-end of the viral RNA*

The cDNA insert of the plasmid pYF1 was sequenced according to the procedure of Maxam and

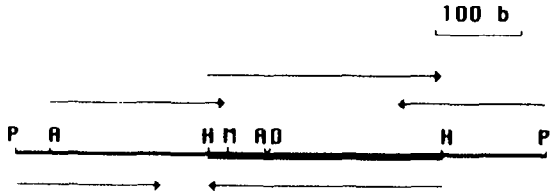

Fig.2. Restriction mapping of the cloned cDNA and strategy for nucleotide sequence determination. The cDNA is oriented in the 5′ to 3′ direction relative to the vRNA. Only the restriction cleavage sites used for DNA sequencing are shown. A, *Alu*I; H, *Hpa*II; P, *Pst*I; M, *Mbo*II; D, *Dde*I. Arrows indicate length and direction of DNA sequences analysed. Arrows above and below the restriction map indicate, respectively, that viral and complementary strands were sequenced. The thick part of the map indicates the cDNA fragments used for primer extension analysis.

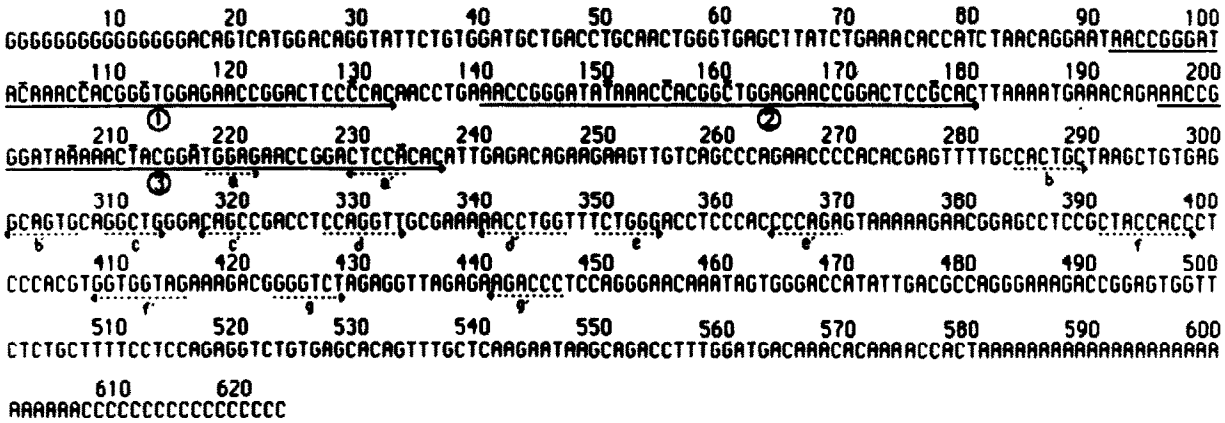

Fig.3. Nucleotide sequence of the cloned cDNA strand corresponding to the viral RNA. The tracts of dG and dC at the 5′- and 3′-end, respectively, were generated during the cloning procedure. The tract of dA preceding the dC tract was generated during the polyadenylation of the viral RNA. The stretches of nucleotides (1–3) underlined by plain arrows correspond to the 3 direct repeats. The nucleotides marked with a dot are not identical in all these repeats. The stretches of nucleotides (a,a′–g,g′) underlined by dotted arrows correspond to the inverted repeats. Only the inverted repeats longer than 4 nucleotides, spaced by less than 15 nucleotides and localized upstream from the region corresponding to the 3′-terminal 120 nucleotides of the viral RNA are figured.

Gilbert [8] following the strategy shown in fig.2. The resulting nucleotide sequence is shown in fig.3.

Due to the strategy employed for cDNA synthesis, the DNA strand corresponding to the positive (or sense) viral RNA should carry a 3′-poly(dA) tract. The sequence of the pYF1 insert indicates the presence of such a dA tract on one DNA strand, thus allowing a putative identification of the positive strand. To test this assignation, the 2 strands of a 266 bp long restriction fragment generated by *Hpa*II cleavage of the cloned cDNA (fig.2) were separated, sequenced and used as primers for reverse transcription of the viral RNA. As expected, only 1 of the 2 strands could be used as a primer, namely the previously assigned negative strand. More than 10 different cDNA products of discrete size could be detected as the result of the primer extension reaction (fig.4). Premature termination of cDNA synthesis at defined sites of a RNA molecule has been described [10] and it seems likely that this is the case here as the amount of products decreases with increasing size. As it is not possible to ascertain that the largest primer extension product detectable is a cDNA molecule which has been elongated to the 5′-end of the viral RNA, these data do not allow the localization of pYF1 cDNA on the viral genome. However, 2 features suggest that this cDNA corresponds to the 3′-end of YF RNA: (i) the poly(dA) tract at the 3′-end of the positive viral DNA strand, which should have been generated during the polyadenylation step of the viral RNA; (ii) the presence of a thymidine residue just preceding this poly(dA) tract which is in agreement with the fact that a uracil is the 3′ ultimate nucleotide of YF RNA [3].

We confirmed the localization of the pYF1 insert by its ability to protect 3′-labeled YF RNA from ribonuclease T2 treatment. Such protection shows also that pYF1 cDNA has retained the overall structure of the 3′-end of YF RNA.

3.3. *Characteristics op the nucleotide sequence of the 3′-end of the YF RNA*

In the 3 possible reading frames, there are at least 6 translational stops and at least one in the 5′ 125 nucleotides of the insert. Although it is not yet possible to localize the end of the coding region of the genome the sequence data indicate that the last 440 nucleotides of the viral RNA are noncoding.

The nucleotide sequence reported here reveals some interesting structural features of the 3′-end of YF RNA. A stretch of 41 nucleotides (under-

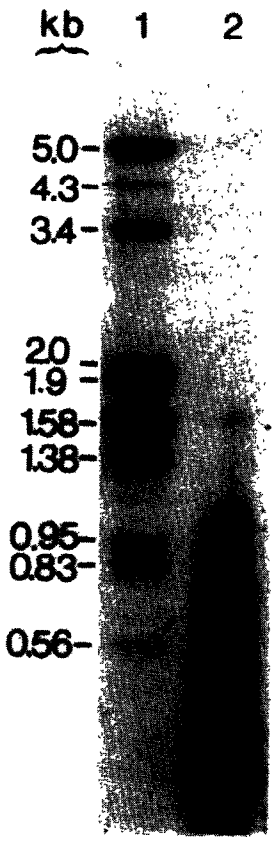

Fig.4. Analysis of the products of reverse transcription of YF RNA from a cloned primer on alkaline agarose (1%) gel. Slot 1: size marker, 3′-labeled *Eco*RI + *Hin*dIII cleaved DNA. Slot 2: primer extension product synthesized as follows: the 2 strands of a 266 bp *Hpa*II restriction fragment (fig.2) were separated. After 5′-labeling with [γ-^{32}P]ATP the strand complementary to the viral RNA was mixed with the viral RNA before denaturation. The reverse transcriptase (Anglian Biotech) directed synthesis of cDNA was then performed as described [5] without addition of oligo(dT) and labeled dNTP.

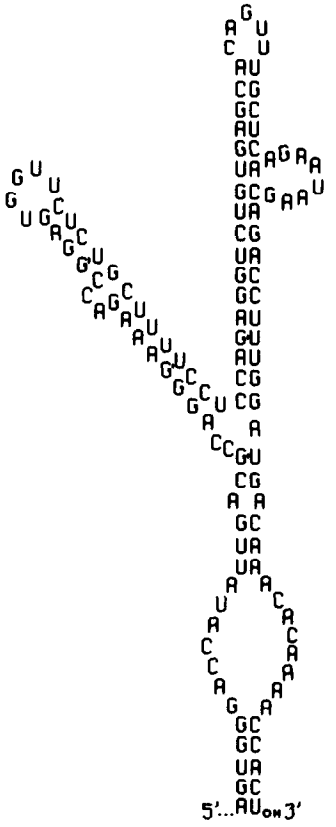

Fig.5. Representation of the most stable ($\Delta G_0 = -52$ kcal/mol) secondary structure adoptable by the 3′-terminal 120 nucleotides of the YF RNA as predicted by computer program analysis [11].

lined by plain arrows in fig.3) is repeated 3 times in this part of the genome with only 4 variable positions in the 3 repeats. Following these direct repeats there is a series of inverted repeats (underlined by dotted arrows in fig.3) which might be involved in a secondary structure of the stem loop type.

In the 3′-terminal 120 nucleotides there are long stretches of complementary sequences. Computer program analysis [11] allows the prediction of many slightly different stable secondary structures ($\Delta G_0 \leqslant -40$ kcal/mol). The most stable one ($\Delta G_0 = -52$ kcal/mol), for which two thirds of these nucleotides are base paired, has been represented (fig.5).

All these stable conformations have a feature in common: the 5 terminal nucleotides at the 3′-end form part of a base paired structure. Such secondary structures might therefore account for the low reactivity of the 3′-end observed during our attempts to modify it enzymatically. One might speculate that the 3′-end of pYF RNA is also a poor substrate for cytoplasmic exonucleases. Such a secondary structure could therefore have an effect on RNA stability similar to the 3′-poly(A) tail of eukaryotic mRNA [12].

Stable secondary structures are also likely to exist at multiple sites elsewhere in the genome and could be responsible for the preferential stops observed during reverse transcription of the viral RNA. More nucleotide sequence data are necessary to test this hypothesis, and isolation of cDNA clones covering the entire genome is currently under way. They should help in determining the function of the different region of the viral genome and in analyzing the biological significance of these secondary structures.

ACKNOWLEDGEMENTS

We thank Pierre Vialat for preparing virus and its RNA. We are grateful to Dr Vincent Deubel and Dr Jean Pierre Digoutte for stimulating discussions. T.G. was supported by DRET. Biohazards associated with the experiments described in this publication have been examined previously by the French National Control Committee.

REFERENCES

[1] Russel, P.K., Brandt, W.E. and Dalrymple, J.M. (1980) in: The Togaviruses (Schlesinger, R.W. ed.) pp. 503–529, Academic Press, New York.

[2] Westaway, E.G. (1980) in: The Togaviruses (Schlesinger, R.W. ed.) pp. 531–581, Academic Press, New York.

[3] Deubel, V., Crouzet, J., Benichou, D., Digoutte, J.P., Bouloy, M. and Girard, M. (1983) Ann. Virol. (Inst. Pasteur) 134, 581–588.

[4] Monath, T.P., Kinney, R.M., Schlesinger, J.J., Brandriss, R.W. and Bres, P. (1983) J. Gen. Virol. 64, 627–637.

[5] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor, NY.

[6] Cohen, J., Mason, B.B. and Estes, M.K. (1983) in: Double Stranded RNA Viruses (Compans, R.W. and Bishop, D.H.L. eds) pp. 65–72, Elsevier, Amsterdam, New York.

[7] Boyer, H.W. and Roulland-Dussois, D. (1969) J. Mol. Biol. 41, 459–472.

[8] Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.

[9] England, T.E. and Uhlenbeck, O.C. (1978) Nature 275, 560–561.

[10] Hagenbüchle, O., Bovey, R. and Young, R.A. (1980) Cell 21, 179–187.

[11] Papanicolaou, C., Goyu, M. and Ninio, J. (1984) Nucleic Acids. Res. 12, 31–44.

[12] Nevins, J.R. (1983) Annu. Rev. Biochem. 52, 441–466.