JOURNAL OF VIROLOGY, Aug. 1991, p. 4070–4077
0022-538X/91/084070-08$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 65, No. 8

Presence of Poly(A) in a Flavivirus: Significant Differences between the 3′ Noncoding Regions of the Genomic RNAs of Tick-Borne Encephalitis Virus Strains

CHRISTIAN W. MANDL,* CHRISTIAN KUNZ, AND FRANZ X. HEINZ
Institute of Virology, University of Vienna, Kinderspitalgasse 15, A-1095 Vienna, Austria

Received 27 February 1991/Accepted 3 May 1991

A poly(A) tail was identified on the 3′ end of the prototype tick-borne encephalitis (TBE) virus strain Neudoerfl. This is in contrast to the general lack of poly(A) in the genomic RNAs of mosquito-borne flaviviruses analyzed so far. Analysis of several closely related strains of TBE virus, however, revealed the existence of two different types of 3′ noncoding (NC) regions. One type (represented by strain Neudoerfl) is only 114 nucleotides long and carries a 3′-terminal poly(A) structure. This was also found in several TBE virus strains isolated from different geographic regions over a period of almost 30 years. The other type (represented by strain Hypr) is 461 nucleotides long and not polyadenylated. The sequence homology between the two types of TBE virus 3′ NC regions terminates at a specific position 81 nucleotides after the stop codon. The second type of 3′ NC region more closely resembles the common flavivirus pattern, including the potential for the formation of a 3′-terminal hairpin structure. However, it lacks primary sequence elements that are conserved among other flavivirus genomes.

The family *Flaviviridae* consists of about 70 classified members, most of which are arthropod-borne viruses (4, 32). The flaviviruses are subdivided into eight serocomplexes, each of which contains human pathogens. The majority of these pathogens, such as yellow fever (YF) virus, Japanese encephalitis (JE) virus, West Nile (WN) virus, and the four dengue (DEN) virus serotypes, are transmitted by mosquitos. Most of the viruses of serocomplex 1, however, are transmitted by ticks to their vertebrate hosts. The most prominent representative of this subgroup is tick-borne encephalitis (TBE) virus, which is endemic in many parts of Europe and Asia, causing thousands of cases of severe human brain infections every year. Strains of TBE virus have been isolated from both patients and ticks from various geographic regions throughout Europe and Asia over a period of almost 40 years (10, 14). By serology, at least two subtypes of TBE virus have been distinguished, the European and the Far Eastern subtypes (13, 15).

The flavivirus genome consists of a single positive-stranded RNA molecule that is about 10.5 kb long. As with many positive-stranded viral RNAs and cellular mRNAs, the 5′ terminus carries a cap structure (6, 24, 30). However, the 3′ end of various flaviviruses was found to lack a poly(A) structure (1, 24, 29, 31), and this was considered a characteristic feature of the family *Flaviviridae* (1, 5, 24), distinct from most other positive-stranded RNA viruses of plants and animals.

During the past few years, a considerable number of flavivirus genome sequences have been elucidated (summarized in reference 5). These data revealed that the structural proteins (capsid [C], membrane [M] and its precursor protein [prM], and envelope [E]) and the nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) are encoded within a single long open reading frame that makes up more than 90% of the genome. As deduced from nucleotide sequence and protein chemical analyses, the viral proteins are encoded in the order 5′-C-prM-E-NS1 to NS5-3′. The viral polyprotein is cleaved co- and posttranslationally by cellular proteases and at least one viral protease to release the individual proteins. Both sides of the long open reading frame are flanked by short noncoding (NC) regions. The 5′-terminal NC region is approximately 100 nucleotides long and was described to have the potential to form secondary structures that may be conserved among several flaviviruses (2). The 5′ terminus consists of a type 1 cap structure attached to the conserved dinucleotide 5′-AG-3′ (5).

In mosquito-borne flaviviruses, the 3′-terminal NC region has been generally found to be between 400 and 600 nucleotides long. It includes at least two short stretches of high primary sequence conservation (termed CS1 and CS2 [12]) and terminates with the dinucleotide 5′-CU-3′. A portion of CS1 is complementary to another conserved sequence near the 5′ terminus and was proposed to function as a cyclization sequence during replication or encapsidation (12). The approximately 90 most 3′-terminal nucleotides have the potential to form a stable secondary structure (3, 9, 28).

Recently, the complete genome sequences of the European subtype TBE virus strain Neudoerfl (17, 19) and the Far Eastern subtype TBE virus strain Sofyn (23) were determined. Both the 5′ NC region and the long open reading frame closely match the pattern described for mosquito-borne flaviviruses. However, for both TBE virus sequences the 3′ NC region was found to be only 114 nucleotides long and to terminate with the nucleotide C, thereby lacking both the conserved primary (CS1 and CS2) and secondary (3′-terminal stem and loops) structure elements.

In this report, we describe the 3′-terminal NC regions of several strains of TBE virus and provide evidence for the presence of a poly(A) tail. The sequence of another type of 3′ NC region present only in certain strains of TBE virus was elucidated. Our results indicate that strains of TBE virus may contain at least two fundamentally different types of 3′ NC regions, both of which differ significantly from the mosquito-borne pattern.

* Corresponding author.

MATERIALS AND METHODS

Virus strains, growth, and purification. The following strains of European subtype TBE virus were used: prototype strain Neudoerfl, isolated in 1971 from a tick (*Ixodes ricinus*) in Burgenland, Austria (14); strain NE I/12, isolated in 1985 from a tick at the same natural focus as strain Neudoerfl (10); strains ZZ/9 and ZZ/16, both isolated from ticks in the Tyrol, Austria, in 1985 (10); strain A-52, isolated from a tick on Kumlinge Island, Finland, in 1959 (14); strain Hochosterwitz, isolated from a tick in Carinthia, Austria, in 1971 (14); strain Hypr, isolated from human blood in Czechoslovakia in 1953 (14); and strain Scharl, isolated from human brain in Lower Austria in 1956 (14).

Viruses were grown in primary chicken embryo cells, concentrated by ultracentrifugation, and purified by two cycles of sucrose density gradient centrifugation as described previously (14).

5'-3' ligation of viral RNA and generation of cDNA. We recently described a new protocol that allows rapid generation of cDNA fragments representing the 5'- and 3'-terminal regions of linear RNA molecules (18). Briefly, RNA was prepared from purified virus suspensions as described previously (17). This RNA was decapped by treatment with tobacco acid pyrophosphatase, and then the 5' and 3' termini of the RNA were ligated with RNA ligase. Then double-stranded cDNA was synthesized by using oligonucleotide P33 (5'-AGGATGGCCTTCTTGA-3'), complementary to a sequence on the RNA 150 bases away from the 5' terminus. From this cDNA, a fragment representing the ligated 5'- and 3'-terminal genomic regions was amplified by the polymerase chain reaction (PCR), using the 20-mer primers PHind-1 (5'-CGTAAAAGATTGTCCGGAAC-3') and PHind-2 (5'-CTCCTAAGCTTTCTTTTTCT-3'). These oligonucleotides are complementary (except for single mismatches that create *Hin*dIII restriction sites) to the negative- and positive-strand sequences of TBE virus strain Neudoerfl, 70 and 100 bases away from the 3' and 5' termini, respectively. Amplified DNA was cloned into the single *Hin*dIII site of the phagemid Bluescript (Stratagene), and single-stranded DNA was prepared from individual clones for sequence analysis.

Sequence analysis. Single-stranded DNA was sequenced by the dideoxy method of Sanger et al. (25). In more recent experiments, radioactive labeling was replaced by the use of fluorescent dyes and the reaction products were analyzed on an automated sequencer (ABI model 373A).

Sequencing reactions using RNA as a template were performed by a modification of the dideoxy method as described by Zimmern and Kaesberg (34). Typically, 1 μg of RNA was mixed with 50 ng of one of the $(T)_{15}$-N oligonucleotide primers.

Reverse transcription of poly(A). Template nucleic acid (200 ng or 1 μg) was mixed with 20 ng of a 15-mer oligo(dT) primer, 5 U of human placental RNase inhibitor, 5 μCi of $[^{32}P]$TTP (Amersham), and unlabeled TTP (Boehringer) to a final concentration of 200 μM TTP in 5 μl of the reverse transcription reaction buffer supplied by the manufacturer of the enzyme (Boehringer). After the addition of 5 U of reverse transcriptase, the sample was incubated at 42°C for 15 min. Then the reaction was stopped by the addition of 5 μl of loading buffer. The sample was heated to 90°C for 1 min, and a 4-μl aliquot was loaded onto an 8% polyacrylamide sequencing gel. Electrophoresis was run until unincorporated nucleotides reached the bottom of the gel. Gels were dried and exposed to X-ray films (Hyperfilm β-max; Amersham) for variable lengths of time (several hours to several days).

Computer analyses. Comparisons of different flavivirus sequences, including sequence alignments and homology calculations, were performed by using programs of the Beckman Microgenie software package (version 4.0). Sequences were screened for homologies against the EMBL data bank (release May 1990), using the FASTA algorithm of Pearson and Lipman (22). RNA secondary structure calculations were performed by the method of Williams and Tinoco (33).

RESULTS

Analysis of TBE virus strain Neudoerfl. The 3'-terminal NC region of the genome sequence of TBE virus strain Neudoerfl (19) was recently reported to be considerably shorter than that of mosquito-borne flaviviruses and to lack all the primary and secondary structural elements conserved among these viruses. This prompted us to reinvestigate the genomic termini of this virus. A new method was developed (18) that involves the ligation of the 5' and 3' termini of viral RNA, the generation of double-stranded cDNA, and the PCR amplification of a DNA fragment corresponding to the linked terminal regions of the genome. DNA prepared by this method was cloned, and 35 recombinant bacterial colonies were randomly picked for sequence analysis. The results obtained by this approach are summarized in Fig. 1A. Every clone analyzed contained sequences from both the 3'- and 5'-terminal regions of the genome. Almost half of the clones (16 of 35) contained sequences representing the complete 5'- and 3'-terminal genomic regions as reported previously (17, 19). The other clones exhibited truncated 3'- and/or 5'-terminal sequences (summarized in Fig. 1A). Surprisingly, every clone with a complete 3' end (24 of 35) also contained a poly(A) structure next to this 3' end. This homopolymeric stretch varied in length from 20 to approximately 80 nucleotides. It was not present in any of the clones with truncated 3'-terminal sequences. An example of a sequence pattern obtained from a clone with complete 5' and 3' ends separated by a poly(A) tract is shown in Fig. 1B. In this clone, the homopolymeric structure is interrupted by a single G residue (asterisk in Fig. 1B), which was also found at different positions within the poly(A) tracts of three other clones. These G's were probably introduced by an inaccuracy of reverse transcriptase or *Taq* polymerase. The 5' and 3' truncations (Fig. 1A) and the varying lengths of the poly(A) tract are possibly due to partial breakdown of RNA and polymerase stuttering, respectively.

None of the 35 clones investigated exhibited any flavivirus-specific sequences varying from the previously reported sequence of TBE virus strain Neudoerfl (17, 19). Two clones contained sequences that were unrelated to any known flaviviral or other sequence (EMBL data bank, release May 1990) and did not hybridize to TBE viral RNA (data not shown). Probably, they originated from contaminating DNA or RNA that was picked up during the PCR reaction.

To rule out the possibility of PCR artifacts, the existence of a homopolymeric A sequence in the genome of TBE virus strain Neudoerfl was confirmed by two alternative approaches. (i) Viral RNA was incubated with reverse transcriptase in the presence of a 15-mer oligo(dT) primer, using $[^{32}P]$TTP as the only substrate (Materials and Methods). Fractionation of the reaction products on a polyacrylamide sequencing gel yielded a ladder of bands on the autoradiograph consistent with the presence of a poly(A) structure in

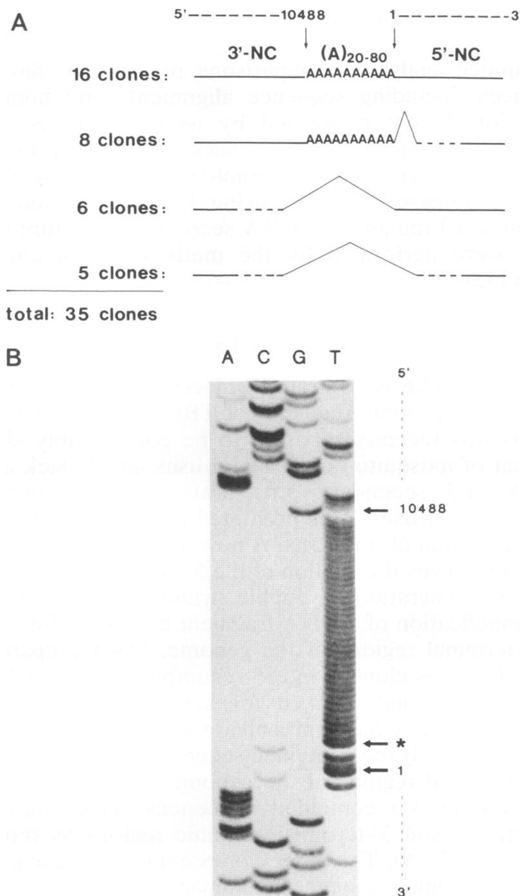

FIG. 1. 5'-3'-Ligated cDNA clones of TBE virus strain Neudoerfl. The first and the last nucleotides of the full-length genome sequence [excluding poly(A)] are shown by arrows and their respective position numbers. (A) Schematic summary of the structures of all clones analyzed. Dashed lines indicate truncations of variable lengths. (B) Section of the sequence pattern obtained from one of the 16 clones with complete 3' and 5' NC regions separated by a homopolymeric A structure. The sequence corresponds to the negative strand. An asterisk indicates a G residue interrupting the poly(A) structure.

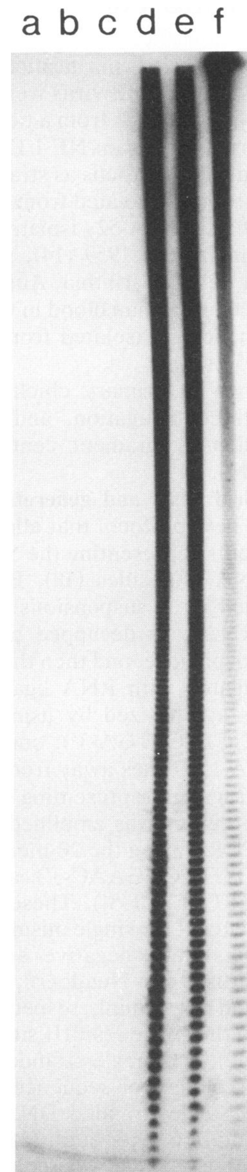

FIG. 2. Oligo(dT)-primed reverse transcription. Template nucleic acid (1 μg) was reverse transcribed by using oligo(dT) as a primer and [^{32}P]TTP as the only substrate. Reaction products were fractionated on an 8% polyacrylamide gel. Exposure time of the autoradiograph was 24 h. Lanes: a, rRNA; b, M13 single-stranded DNA; c, JE virus RNA; d, TBE (strain Neudoerfl) virus RNA; e, SIN virus RNA; f, *neo* mRNA.

the genome of TBE virus strain Neudoerfl (Fig. 2, lane d). The apparent number of A's observed in this experiment is considerably larger than that revealed in the PCR clones described above. This is probably due to stuttering of the reverse transcriptase. Polyadenylated RNAs, such as cellular mRNA and alphavirus genomic RNA, were used as positive controls and gave rise to similar reaction product ladders (lanes e and f). However, incubation of nucleic acids that lack a poly(A) structure, such as rRNA, single-stranded M13 DNA, and JE virus RNA, did not yield detectable amounts of reaction products (Fig. 2, lanes a to c).

(ii) Viral RNA was subjected to dideoxy sequence analysis using $(T)_{15}$-G as a specific primer. G is complementary to the last residue (C) of the sequence of TBE virus strain Neudoerfl (19; Fig. 1B). If this C residue is indeed followed by several A residues, this approach should yield the 3'-terminal sequence. The sequencing pattern obtained (Fig. 3) clearly corresponds to the 3'-terminal NC region of TBE virus strain Neudoerfl, thus confirming the presence of a 3'-terminal poly(A) tail. As a control, Sindbis (SIN) virus RNA, which also terminates in C-poly(A) (26), was likewise subjected to sequence analysis after priming with $(T)_{15}$-G. The sequence pattern shown in Fig. 3 corresponds to the published 3'-terminal sequence of SIN virus (26). As a nonpolyadenylated control, JE virus RNA, which terminates with a U residue (27), was assayed by using the primer $(T)_{15}$-A. Only after very prolonged exposure of the gel could a sequence pattern be visualized (Fig. 3). The derived sequence matches an internal region of the genome of JE virus (positions 6179 to 6163). This may be explained by nonspecific hybridization of the $(T)_{15}$-A oligonucleotide,

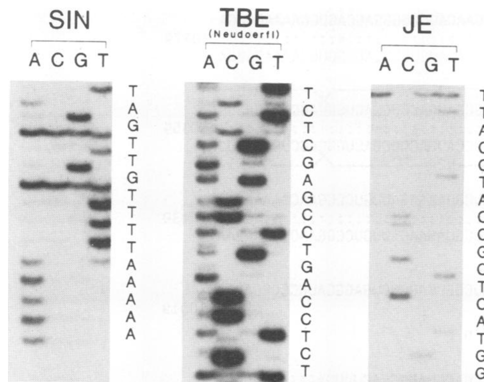

FIG. 3. RNA dideoxy sequencing patterns. A 1-μg sample of each viral RNA was sequenced by using $(T)_{15}$-G (SIN and TBE) or $(T)_{15}$-A (JE) as the primer. Reaction products were analyzed on 12% polyacrylamide gels. The derived nucleotide sequences are indicated next to the sequence patterns. Exposure times of the autoradiographs were 12 h for the SIN and TBE sequences and 4 days for the JE sequence.

possibly to the sequence 6190-TGAAGAAAAGAAAAAC-6205 of the JE virus genome.

Analysis of different TBE virus strains. To determine whether the presence of poly(A) is characteristic of TBE virus in general, we screened several strains of TBE virus that were available in our laboratory. Equal amounts (200 ng) of each viral RNA were subjected to reverse transcription, using oligo(dT) as a primer and [^{32}P]TTP as the only substrate as described above. After fractionation of the reaction products on polyacrylamide gels, five TBE virus strains (NE I/12, Hochosterwitz, A-52, ZZ/9, and ZZ/16) yielded results nearly identical to those obtained from strain Neudoerfl (Fig. 4). In the cases of two strains (Scharl and Hypr), however, no such reaction patterns were observed, even after very long exposure of the autoradiograph. The intensities of the ladders obtained from the five positive strains were either about as strong as (NE I/12 and Hochosterwitz) or weaker than that from strain Neudoerfl.

To verify the presence of a poly(A) tract on the 3′ terminus of the genome of one of the strains with weaker signal intensities, RNA from TBE virus strain A-52 was subjected to dideoxy sequence analysis employing the $(T)_{15}$-G oligonucleotide as a primer. A clear sequence pattern corresponding to the 3′ NC region was obtained (data not shown). In contrast, no reaction pattern was obtained in this assay when RNA from strain Hypr was used. Similarly, the use of $(T)_{15}$-A and $(T)_{15}$-C as primers did not yield a sequence pattern for the RNA of strain Hypr (data not shown).

Sequence analysis of the 3′- and 5′-terminal regions of TBE virus strain Hypr. Since it appeared from the experiments described above that the genome of TBE virus strain Hypr was not polyadenylated, we set out to sequence the terminal regions of this viral RNA. Sequence analysis of short cDNA clones representing randomly spaced regions of the genome had indicated an RNA sequence homology of more than 98% to TBE virus strain Neudoerfl (unpublished observation). This prompted us to use the same primers that had been used for the analysis of strain Neudoerfl for the generation of cDNA. As before, 5′ and 3′ termini of viral RNA were ligated, and cDNA fragments corresponding to the 5′- and 3′-terminal genomic regions were generated, amplified by use of the PCR, and cloned in order to sequence individual clones.

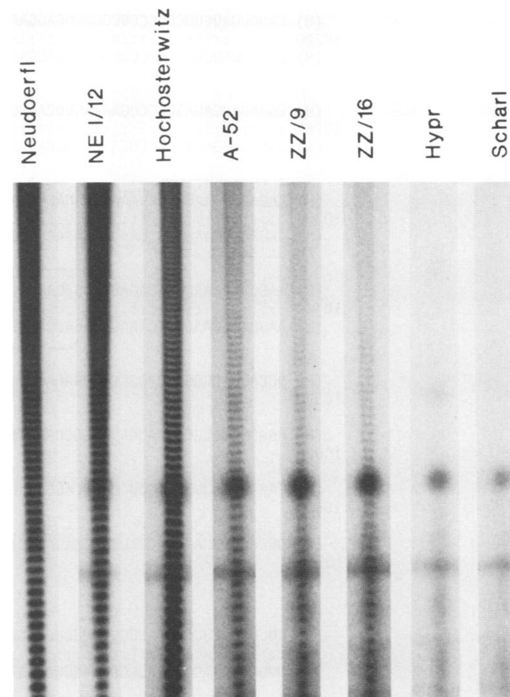

FIG. 4. Oligo(dT)-primed reverse transcription of TBE virus strains. Genomic RNAs (200 ng) of TBE virus strains indicated were transcribed by using oligo(dT) as a primer and TTP as the only substrate. Reaction products were fractionated on 8% polyacrylamide gels. Only a section of the autoradiographs (size range approximately 20 to 100 bases) is shown. Exposure times were 24 h for strains Neudoerfl, NE I/12, and Hochosterwitz and 48 h for strains A-52, ZZ/9, ZZ/16, Scharl, and Hypr.

Sequence analysis of seven clones thus obtained indicated that all of these clones contained sequences that corresponded to the 3′ terminus (up to nucleotide 10458) and the 5′ terminus (starting with nucleotide 1) of TBE virus strain Neudoerfl. (Nucleotide numbers refer to the complete genome sequence of TBE virus strain Neudoerfl, i.e., +20 compared with the numbering in the corresponding figures in references 17 and 19.) Between these homologous regions, five clones contained 377 nucleotides which are not present in the sequence of strain Neudoerfl. In two clones, this additional sequence was truncated. A poly(A) structure was not observed in any of the clones.

Figure 5 shows a comparison of the terminal nucleotide sequences of strain Hypr as deduced from the five long PCR clones (complemented with sequence information obtained from random primed cDNA clones) with the homologous 3′- and 5′-terminal sequences of TBE virus strain Neudoerfl. The 3′ NC region of the genome of strain Hypr as deduced from Fig. 5 is 461 nucleotides long and not polyadenylated, whereas that of the strain Neudoerfl genome is only 114 nucleotides long and carries a poly(A) tail.

As can be seen from Fig. 5, the sequences of the two strains are almost identical in the 5′ NC region and in the 3′-terminal region until nucleotide position 10458. In the sequence between positions 1 and 107 there are only four nucleotide differences. Between nucleotides 10200 and 10458 there are only six nucleotide differences. One of these, at

```
                                                         R2                    R1
(H)   CAUGUUAUGUUCCUCCCUCGUUGGGAGGAAAGAAAGAGCAGAAUGGGCCAAGAACAUCUGGGGAGCAGUGGAAAAGGUGA
10200 ::::::::::::::::::x::::::::::x::::::::::::::::::::::::::::::::::::x:::::::::::::   10279
(N)   CAUGUUAUGUUCCUCCCUUGUUGGGAGGAGAGAAAGAGCAGAAUGGGCCAAGAACAUCUGGGGAGCGGUGGAAAAGGUGA

(H)   GGAAGAUGAUAGGGCCUGAAAAGUUCAAGGACUAUCUCUCUUGUAUGGACCGCCAUGACCUGCACUGGGAGCUCAGACUG
10280 :::::::::::::x::::::::::::::::::::::::::x:::::::::::::::::::::::::::::::::::::::   10359
(N)   GGAAGAUGAUAGGUCCUGAAAAGUUCAAGGACUAUCUCUCCUGUAUGGACCGCCAUGACCUGCACUGGGAGCUCAGACUG
                                                         R2                    R1

(H)   GAGAGCUCAAUAAUCUAAACCCAGACUGUGACAGAGCAAAACCCGGAGGGCUCGUAAAAGAUUGUCCGGAACCAAAAGAA
10360 ::::::::::::::::::::::::::::::::::::::::::::::x:::::::::::::::::::::::::::::::::   10439
(N)   GAGAGCUCAAUAAUCUAAACCCAGACUGUGACAGAGCAAAACCCGGAAGGCUCGUAAAAGAUUGUCCGGAACCAAAAGAA

                          *  *    *  * R2  *   **
(H)   AAGCAAGCAACUCACAGAGUGUGAGACCCCCUGACCAGCAAAGGGGCAGAUCGGUCAGGGGUGAGGGAUGCCCCCAGAG
10440 :::::::::::::::::::xxx:::xx:xxx:::xxx:::xxxxxxxxx                                  10519
(N)   AAGCAAGCAACUCACAGAGAUAGAGCUCGGACUGGAGAGCUCUUUAAAC(A)n
                                 *     R1      **  *

(H)   UGCAUUACGGCAGCACGCCAGUGAGAGUGGCGACGGGAAAAUGGUCGAUCCCGACGUAGGGCACUCUGAAAAACUUUGUG
10520                                                                                    10599

(H)   AGACCCCCUGCAUCAUGAUAAGGCCGAACAUGGUGCAUGAAAAGGGAGGCCCCCGGAAGUACGCUUCCGGGAGGAGGGAA
10600                                                                                    10679

(H)   GAGAGAAAUUGGCAGCUCUCUUCAGGAUUUUUCCUCCUCCUAUACAAAAUUCCCCCUCGGUAGAGGGGGGCGGUUCUUG
10680                                                                                    10759

(H)   UUCUCCCUGAGCCACCAUCACCCAGACACAGAUAGUCUGACAAGGAGGUGAUGUGUGACUCGGAAAAACACCCGCUAGAU
10760                                                                             ::::   4
(N)                                                                               AGAU

(H)   UUUCUUGCGCGUGCGUGCGUUUGCUUCGGACAGCAUUAGCAGCGGUUGGUUUGAAAGAAAUAUUCUUUUGUUUUUACCAG
5     ::::::::x:::::x:::::::::::::::::::::::::::::::::::::::::::x::::::::::::::x::::::   84
(N)   UUUCUUGCACGUGCAUGCGUUUGCUUCGGACAGCAUUAGCAGCGGUUGGUUUGAAAGAGAUAUUCUUUUGUUUCUACCAG

(H)   UCGUGAACGUGUUGAGAAAAGA
85    ::::::::::::::::::::::
(N)   UCGUGAACGUGUUGAGAAAAGA
```

FIG. 5. Comparison of the 3′- and 5′-terminal nucleotide sequences of TBE virus strains Hypr (H) and Neudoerfl (N). Nucleotide numbers refer to TBE virus strain Neudoerfl and are continued arbitrarily for the 3′ NC region of the strain Hypr sequence. A box marks the position of the stop codon. The 3′ NC region of the strain Neudoerfl sequence terminates with a poly(A) structure [$(A)_n$]. Colons indicate identical nucleotides, x's mark sequence differences. The positions of the direct repeat sequences R1 and R2 are shown. Sequence differences between the first and the second copies of these repeats are marked by asterisks. A dashed line indicates the part of the strain Hypr sequence that may form a stable secondary structure.

position 10229, which is still within the coding region of protein NS5 (stop codon at position 10375), causes a conservative difference at the amino acid level (Arg in strain Neudoerfl versus Lys in strain Hypr). However, the 30 3′-terminal nucleotides (positions 10459 to 10488) of strain Neudoerfl that are followed by the poly(A) tract do not have a homologous counterpart in the sequence of strain Hypr, but consist in part (nucleotides 10463 to 10488) of a direct repeat of the sequence immediately preceding the stop codon that terminates the long open reading frame (repeat R1 in Fig. 5). Interestingly, the sequence of strain Hypr at the corresponding position is itself a direct repeat (repeat R2 in Fig. 5) of the sequence adjacent to the first copy of the R1 repeat. Both repeats are imperfect (differences between the first and second copies are marked with asterisks in Fig. 5), and their first copies overlap by a single G residue.

Secondary structure and sequence comparison. The newly obtained sequence of TBE virus strain Hypr was subjected to secondary structure prediction. This revealed that the 87 3′-terminal nucleotides (depicted by a dashed line in Fig. 5) may potentially fold into a stem-loop structure as drawn in Fig. 6. A ΔG value of −43.9 kcal (−183.7 kJ) was calculated for this RNA structure. The secondary structure exhibits a long stem and two short loops (each 7 nucleotides long), which are separated by a short base-paired region. While this secondary structure is reminiscent of 3′-terminal secondary structures described for mosquito-borne flaviviruses, primary sequence comparisons revealed no significant se-

FIG. 6. 3′-Terminal secondary structure of TBE virus strain Hypr. The structure was calculated by the method of Williams and Tinoco (33). Asterisks depict nucleotides that are exposed in the side loop and are conserved among several flavivirus sequences.

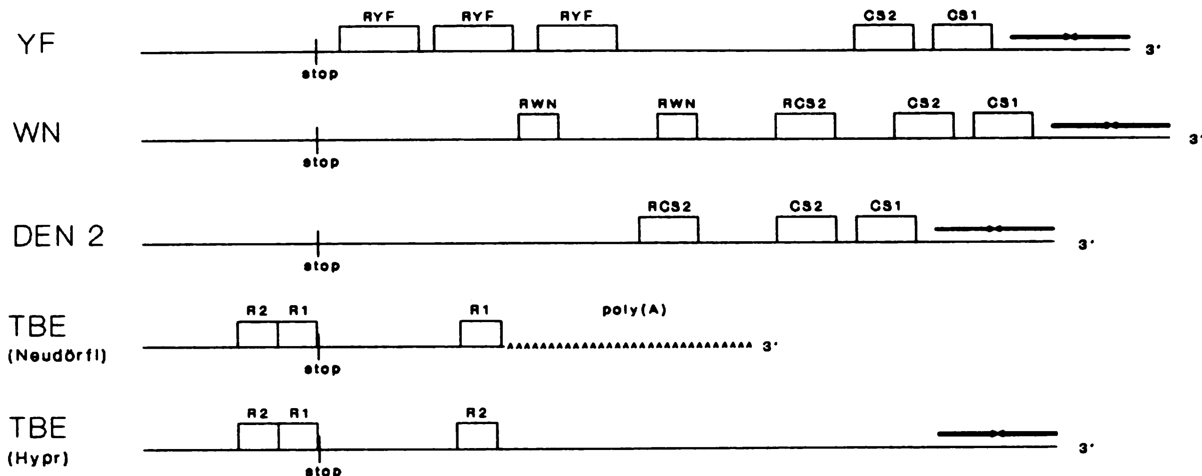

FIG. 7. Schematic comparison of the 3′ NC regions of three mosquito-borne (YF, WN, and DEN 2) flaviviruses as described by Hahn et al. (12) and two strains of TBE virus. Repeat sequences and conserved sequence elements are indicated by boxes. RYF and RWN refer to direct repeats present only in the respective sequences. CS1 and CS2 are sequence boxes conserved among mosquito-borne flaviviruses. RCS2 are inverted repeats of the CS2 sequence. R1 and R2 are TBE virus-specific direct repeats. The sequences of the two TBE virus strains diverge at the position of the second copies of these repeats. A pair of arrows pointing toward each other emphasize potential 3′-terminal secondary structures. The sequences are aligned with respect to the positions of the first stop codons.

quence conservation between the 3′ NC regions of TBE virus strain Hypr and mosquito-borne flaviviruses.

DISCUSSION

The presence of a poly(A) tail in the genome of TBE virus strain Neudoerfl is an unexpected finding. So far the absence of a poly(A) structure and the presence of a characteristic 3′-terminal secondary structure was thought to be a general characteristic of the flavivirus family (1, 5, 24, 31). The 3′ NC region of both the European subtype strain Neudoerfl (19) and the Far Eastern subtype strain Sofyn (23) had recently been found to be much shorter than that of mosquito-borne flaviviruses. However, since in both cases 3′-terminal clones had been prepared only after in vitro polyadenylation, a natural poly(A) tail had previously not been identified. We have now shown that the 3′-terminal NC region of TBE virus strain Neudoerfl is only 114 nucleotides long and carries a poly(A) tail. In addition, our data indicate that polyadenylation of TBE virus is apparently strain specific. Poly(A) tracts were detected for several strains of TBE virus that had been isolated from different geographic regions over a period of more than 20 years. On the other hand, we identified two TBE virus strains which lack a poly(A) structure. Sequence analysis of one of these nonpolyadenylated strains (strain Hypr) yielded a 3′ NC region that is approximately three times longer than that of strain Neudoerfl. Within its 3′-terminal 87 nucleotides, it exhibits a potential secondary structure (Fig. 6) that resembles those predicted for mosquito-borne flaviviruses. In particular, the size, the calculated ΔG value, and the occurrence of two characteristic loops, one of which exposes the sequence 5′-CACAG-3′ (asterisks in Fig. 6), are reminiscent of the corresponding structures postulated for other flaviviruses (3, 9, 28). Moreover, the derived 3′ NC region of strain Hypr terminates with the dinucleotide CU, as do all other flavivirus genomes determined so far (with the exception of TBE virus strains Neudoerfl and Sofyn).

Despite these similarities, the sequence of strain Hypr differs significantly in other respects from the other flavivirus sequences. Figure 7 schematically compares the structural elements of three mosquito-borne flaviviruses (YF, WN, and DEN 2 viruses) as derived from the original sequence data by Hahn et al. (12) with those of TBE virus strains Neudoerfl and Hypr. The strict conservation of the sequence elements CS1 and CS2 among mosquito-borne flaviviruses suggested a functional role, possibly for the replication of the genome. These do not have a homologous counterpart in the TBE virus sequences. In addition, Hahn et al. (12) identified a conserved 5′-proximal sequence element within the C protein coding region, which is complementary in part to the CS1 sequence and was hypothesized to function as a cyclization sequence. However, there is no evidence for a conserved cyclization sequence in TBE virus genomes sequenced so far.

The occurrence of two significantly different types of 3′-terminal NC regions among strains of a single virus type is remarkable. Although considerable strain-specific variations within untranslated regions of the genomes of RNA viruses have been reported (8), the differences among the TBE virus strains exceed those observed for other viruses and include the absence or presence of a poly(A) tail at the 3′ terminus. It is unclear how these two different structures could evolve within a single virus type among strains that have diverged less than 2% on the nucleotide level in other genomic regions. The sudden interruption of homology at a certain nucleotide position certainly suggests some kind of recombination event. Recombination has recently been observed for a number of RNA viruses and may involve both cellular or viral RNA (7, 20). In the case of TBE virus, it is striking that each of the sequences following the point of divergence in the genomes is a direct repeat of a sequence at the carboxy terminus of the NS5 protein coding region (compare Fig. 5 and 7). The fact that neither of these repeats is perfect, however, suggests that if they arose by recombination, this event did not occur very recently in evolution. In addition, our results demonstrate that there are TBE virus strains isolated at different times and from various geographic regions which exhibit the poly(A) genotype. These strains

also differ from each other in their passage histories. These findings seem to argue against the possibility that the poly(A) tract was acquired recently under artificial tissue culture growth conditions. One might also speculate that polyadenylated genomes are noninfectious abortive products that are formed in variable abundance. However, we could not obtain evidence for a difference of specific infectivities when standardized virus preparations of TBE virus strain Hypr and strain Neudoerfl were compared with respect to growth in primary chicken embryo cells (unpublished observation). It seems noteworthy that the identified polyadenylated strains were all isolated from ticks, whereas the nonpolyadenylated strains originated from human patients. Because of the low number of strains analyzed to date, it is too early to judge the significance of this finding.

It is an interesting question whether specific biological properties of TBE virus strains can be correlated with their respective types of 3′-terminal NC regions. It has been suggested (23a) that the more "mosquito-borne-like" 3′ NC region present in strain Hypr would enable this strain to grow in mosquito cells. However, both strain Neudoerfl and strain Hypr failed to grow to detectable levels in the C6/36 mosquito cell line even after introduction of viral RNA into the cells by the lipofectin transfection method (unpublished observation). This rules out the possibility of the 3′ NC region being the only determinant of a tick- versus mosquito-specific host range of TBE virus. Nevertheless, differences in biological properties, such as virulence in mice, have been observed for the TBE virus strains Neudoerfl and Hypr (unpublished observation), but these may be caused by sequence differences within other regions of the genome rather than the 3′ NC region. For instance, attenuation of TBE virus can be achieved by a single point mutation within the protein E coding sequence (16). On the other hand, point mutations present in attenuated strains of JE virus (21) and YF virus (11) include nucleotide changes in the NC regions of these viruses. Thus, it remains to be determined which biological functions are specifically affected by elements of the 3′-terminal NC regions.

ACKNOWLEDGMENTS

We gratefully acknowledge the expert technical assistance of Melby Wilfinger, and we thank Steve L. Allison for critical reading of the manuscript.

REFERENCES

1. **Brinton, M. A.** 1986. Replication of flaviviruses, p. 327–374. *In* S. Schlesinger and M. J. Schlesinger (ed.), The Togaviridae and Flaviviridae. Plenum Publishing Corp., New York.
2. **Brinton, M. A., and J. H. Dispoto.** 1988. Sequence and secondary structure analysis of the 5′-terminal region of flavivirus genome RNA. Virology **162:**290–299.
3. **Brinton, M. A., A. V. Fernandez, and J. H. Dispoto.** 1986. The 3′-nucleotides of flavivirus genomic RNA form a conserved secondary structure. Virology **153:**113–121.
4. **Calisher, C. H., N. Karabatsos, J. M. Dalrymple, R. E. Shope, J. S. Porterfield, E. G. Westaway, and W. E. Brandt.** 1989. Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. J. Gen. Virol. **70:**37–43.
5. **Chambers, T. J., C. S. Hahn, R. Galler, and C. M. Rice.** 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. **44:**649–688.
6. **Cleaves, G. R., and D. T. Dubin.** 1979. Methylation status of intracellular dengue type 2 40 S RNA. Virology **96:**159–165.
7. **Dominguez, G., C.-Y. Wang, and T. K. Frey.** 1990. Sequence of the genome RNA of rubella virus: evidence for genentic rearrangement during togavirus evolution. Virology **177:**225–238.
8. **Faragher, S. G., and L. Dalgarno.** 1986. Regions of conservation and divergence in the 3′ untranslated sequences of genomic RNA from Ross River virus isolates. J. Mol. Biol. **190:**141–148.
9. **Grange, T., M. Bouloy, and M. Girard.** 1985. Stable secondary structure at the 3′ end of the genome of yellow fever virus (17D vaccine strain). FEBS Lett. **188:**159–163.
10. **Guirakhoo, F., A. C. Radda, F. X. Heinz, and C. Kunz.** 1987. Evidence for antigenic stability of tick-borne encephalitis virus by the analysis of natural isolates. J. Gen. Virol. **68:**859–864.
11. **Hahn, C. S., J. M. Dalrymple, J. H. Strauss, and C. M. Rice.** 1987. Comparison of the virulent Asibi strain of yellow fever virus with the 17D vaccine strain derived from it. Proc. Natl. Acad. Sci. USA **84:**2019–2023.
12. **Hahn, C. S., Y. S. Hahn, C. M. Rice, E. Lee, L. Dalgarno, E. G. Strauss, and J. H. Strauss.** 1987. Conserved elements in the 3′ untranslated region of flavivirus RNAs and potential cyclization sequences. J. Mol. Biol. **198:**33–41.
13. **Heinz, F. X.** 1986. Epitope mapping of flavivirus glycoproteins. Adv. Virus Res. **31:**103–186.
14. **Heinz, F. X., and C. Kunz.** 1981. Homogeneity of the structural glycoprotein from European isolates of tick-borne encephalitis virus: comparison with other flaviviruses. J. Gen. Virol. **57:**263–274.
15. **Heinz, F. X., C. W. Mandl, F. Guirakhoo, H. Holzmann, W. Tuma, and C. Kunz.** 1990. The envelope protein E of tick-borne encephalitis virus and other flaviviruses: structure, functions and evolutionary relationships. Arch. Virol. Suppl. **1:**125–135.
16. **Holzmann, H., F. X. Heinz, C. W. Mandl, F. Guirakhoo, and C. Kunz.** 1990. A single amino acid substitution in envelope protein E of tick-borne encephalitis virus leads to attenuation in the mouse model. J. Virol. **64:**5156–5159.
17. **Mandl, C. W., F. X. Heinz, and C. Kunz.** 1988. Sequence of the structural proteins of tick-borne encephalitis virus (Western subtype) and comparative analysis with other flaviviruses. Virology **166:**197–205.
18. **Mandl, C. W., F. X. Heinz, E. Puchhammer-Stöckl, and C. Kunz.** 1991. Sequencing the termini of capped viral RNA by 5′-3′ ligation and PCR. BioTechniques **10:**485–486.
19. **Mandl, C. W., F. X. Heinz, E. Stöckl, and C. Kunz.** 1989. Genome sequence of tick-borne encephalitis virus (Western subtype) and comparative analysis of nonstructural proteins with other flaviviruses. Virology **173:**291–301.
20. **Meyers, G., T. Rümenapf, and H.-J. Thiel.** 1989. Ubiquitin in a togavirus. Nature (London) **341:**491.
21. **Nitayaphan, S., J. A. Grant, G.-J. J. Chang, and D. W. Trent.** 1990. Nucleotide sequence of the viruslent SA-14 strain of Japanese encephalitis virus and its attenuated vaccine derivative, SA-14-14-2. Virology **177:**541–552.
22. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
23. **Pletnev, A. G., V. F. Yamshchikov, and V. M. Blinov.** 1990. Nucleotide sequence of the genome and complete amino acid sequence of the polyprotein of tick-borne encephalitis virus. Virology **174:**250–263.

23a. **Rice, C.** Personal communication.

24. **Rice, C. M., E. G. Strauss, and J. H. Strauss.** 1986. Structure of the flavivirus genome, p. 279–320. *In* S. Schlesinger and M. J. Schlesinger (ed.), The Togaviridae and Flaviviridae. Plenum Publishing Corp., New York.
25. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
26. **Strauss, E. G., C. M. Rice, and J. H. Strauss.** 1984. Complete nucleotide sequence of the genomic RNA of Sindbis virus. Virology **133:**92–110.
27. **Takegami, T., M. Washizu, and K. Yasui.** 1986. Nucleotide sequence at the 3′ end of Japanese encephalitis virus genomic RNA. Virology **152:**483–486.
28. **Wengler, G., and E. Castle.** 1986. Analysis of structural properties which possibly are characteristics for the 3′-terminal sequences of the genome RNA of flaviviruses. J. Gen. Virol. **67:**1183–1188.

29. **Wengler, G., and G. Wengler.** 1981. Terminal sequences of the genome and replicative-form RNA of the flavivirus West Nile virus: absence of poly(A) and possible role in RNA replication. Virology **113**:544–555.
30. **Wengler, G., G. Wengler, and H. J. Gross.** 1978. Studies on virus-specific nucleic acids synthesized in vertebrate and mosquito cells infected with flaviviruses. Virology **89**:423–437.
31. **Westaway, E. G.** 1987. Flavivirus replication strategy. Adv. Virus Res. **33**:45–90.
32. **Westaway, E. G., M. A. Brinton, S. Y. Gaidamovitch, M. C. Horzinek, A. Igarashi, L. Kaariainen, D. K. Lvov, J. S. Porterfield, P. K. Russel, and D. W. Trent.** 1985. Flaviviridae. Intervirology **24**:183–192.
33. **Williams, A. L., and J. Tinoco, Jr.** 1986. A dynamic programming algorithm for finding alternative RNA secondary structures. Nucleic Acids Res. **14**:299–315.
34. **Zimmern, D., and P. Kaesberg.** 1978. 3′-Terminal nucleotide sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **75**:4257–4261.