JOURNAL OF VIROLOGY, Feb. 1989, p. 564–571
0022-538X/89/020564-08$02.00/0

Vol. 63, No. 2

Antigenic Structure of the Flavivirus Envelope Protein E at the Molecular Level, Using Tick-Borne Encephalitis Virus as a Model

CHRISTIAN W. MANDL, FARSHAD GUIRAKHOO, HEIDEMARIE HOLZMANN, FRANZ X. HEINZ,* AND CHRISTIAN KUNZ

Institute of Virology, University of Vienna, Kinderspitalgasse 15, A-1095 Vienna, Austria

Received 6 July 1988/Accepted 18 October 1988

A model of the tick-borne encephalitis virus envelope protein E is presented that contains information on the structural organization of this flavivirus protein and correlates epitopes and antigenic domains to defined sequence elements. It thus reveals details of the structural and functional characteristics of the corresponding protein domains. The localization of three antigenic domains (composed of 16 distinct epitopes) within the primary structure was performed by (i) amino-terminal sequencing of three immunoreactive fragments of protein E and (ii) sequencing the protein E-coding regions of seven antigenic variants of tick-borne encephalitis virus that had been selected in the presence of neutralizing monoclonal antibodies directed against the E protein. Further information about variable and conserved regions was obtained by a comparative computer analysis of flavivirus E protein amino acid sequences. The search for potential T-cell determinants revealed at least one sequence compatible with an amphipathic α-helix which is conserved in all flaviviruses sequenced so far. By combining these data with those on the location of disulfide bridges (T. Nowak and G. Wengler, Virology 156:127–137, 1987) and the structural characteristics of epitopes, such as dependency on conformation or on intact disulfide bridges or both, a model was established that goes beyond the location of epitopes in the primary sequence and reveals features of the folding of the polypeptide chain, including the generation of discontinuous protein domains.

Tick-borne encephalitis (TBE) virus represents one of the major human pathogenic flaviviruses. Of the three structural proteins (core [C], membrane [M], envelope [E]), protein E plays a central role in the biology of flaviviruses (4). It is capable of inducing a protective immune response in vivo. It is responsible for the binding of the virus to the cellular receptor, and it mediates hemagglutination of erythrocytes at acid pH. Furthermore, it may be involved in an intraendosomal acid-catalyzed fusion step. Thus, the structural and functional organization of protein E is of central interest for the understanding of the biology of flaviviruses and the mechanisms of virus-cell interactions.

For TBE virus and other flaviviruses (16), much information concerning topological relationships, functional and structural properties, and serological specificities of epitopes has been gained by monoclonal antibody (MAb) work. These studies led to the establishment of an epitope model of the TBE virus E protein (17; F. Guirakhoo, F. X. Heinz, and C. Kunz, Virology, in press). Based on competitive binding studies, three nonoverlapping antigenic domains (A, B, C) were defined, each composed of several epitopes (A1 to A5, B1 to B5, C1 to C6) that exhibit different functional activities and serological specificities. Furthermore, three MAbs define isolated epitopes (i1 to i3). The epitopes within each domain vary with respect to their serological specificities and functional activities. Domain A not only contains broadly flavivirus-cross-reactive epitopes (A1, A2) but also subtype-specific epitopes (A3, A4). Most of the domain B epitopes are specific for the tick-borne complex of flaviviruses (16), whereas domain C predominantly contains subtype-specific epitopes. Neutralizing activity has been demonstrated for MAbs A3, A4, A5, B1, B2, B4, B5, C1, and i2. Characteristically, the epitopes within each antigenic domain also have similar structural properties (sensitivity to denaturation, low pH, proteolysis, reduction of disulfide bridges). It can thus be assumed that these antigenic domains actually correspond to different structural entities on the protein. This epitope map, however, has so far not been linked to the primary sequence data (24a).

For a precise localization of individual epitopes in the sequence of protein E, we now report comparative sequence analysis data of seven different antigenic variants of TBE virus which were selected in the presence of neutralizing MAbs. In addition, the locations of several immunoreactive fragments of protein E were determined by amino-terminal sequence analysis.

By combining these results with previously determined structural characteristics, including the assignment of disulfide bridges in West Nile (WN) virus (29), we created a structural model of protein E that contains information on the folding of the polypeptide chain into distinct protein domains.

MATERIALS AND METHODS

Preparation of wild-type and variant viruses. The isolation of seven antigenic variants by growing TBE virus (western subtype, strain Neudoerfl) in the presence of neutralizing MAbs has been described elsewhere (H. Holzmann, C. W. Mandl, F. Guirakhoo, F. X. Heinz, and C. Kunz, J. Gen. Virol., in press). The variants were designated VA3a, VA3b, VA4, VA5, Vi2, VB1, and VB4 according to the respective MAbs they were selected with (A3, A4, A5 of domain A; B1 and B4 of domain B; and i2). Both wild-type and variant viruses were grown in primary chicken embryo cells, concentrated by ultracentrifugation, and purified by sucrose density gradient centrifugation as described previously (20).

Genome sequence analysis. Genomic RNA was isolated from purified variant virus preparations after proteinase K digestion by phenol and chloroform extractions as described previously (24a). Sequence data were obtained by a modifi-

* Corresponding author.

cation of the dideoxy method, basically as described by Zimmern and Kaesberg (43), with genomic RNA as a template and synthetic hexadecamer nucleotides as primers. A set of 19 primers was used to sequence completely the 1,500-nucleotide-long protein E-coding region. When direct RNA sequencing yielded ambiguous results, cDNA clones were prepared by the RNase H-DNA polymerase I method of Gubler and Hoffman (13) and cloned into the Bluescript (Stratagene) vector system. cDNA synthesis was primed with the synthetic hexadecamer P1 that is complementary to a sequence at the 3' end of the protein E-coding region. Single-stranded DNA was sequenced by the dideoxy method of Sanger et al. (34) with the same synthetic primers as for RNA sequencing.

Fragmentation of protein E. The digestion of protein E rosettes (19) with trypsin was performed under conditions that have been described previously (21). Briefly, 100 μg of rosettes prepared as described previously (19) was incubated with 1 μg of trypsin (210 U/mg; Worthington Diagnostics, Freehold, N.J.) at 37°C for either 30 min or 24 h. The chemical degradation of protein E with cyanogen bromide (CNBr) was performed as described previously (21). For the digestion of TBE virus with α-chymotrypsin the following protocol was used. A 10-μg sample of virus was denatured by boiling for 5 min in the presence of 0.5% sodium dodecyl sulfate (SDS). Then 2.5 μg of enzyme (50 U/mg; Serva, Heidelberg, Federal Republic of Germany) was added and allowed to react at 37°C for 80 min.

SDS-polyacrylamide gel electrophoresis and immunoblotting. The fractionation of polypeptides on 15% denaturing polyacrylamide gels was performed as described by Laemmli and Favre (23). For the immunochemical analysis, the fragments were transferred from the polyacrylamide gels to nitrocellulose paper at 40 V overnight, reacted with ^{125}I-labeled MAbs, and visualized by autoradiography essentially as previously described (18).

Amino-terminal sequencing of protein fragments. Amino-terminal sequence analysis of fragments blotted from urea-polyacrylamide gels was performed with a gas-phase sequencer (ORPEGEN, Heidelberg, Federal Republic of Germany).

Computer-assisted sequence analysis. The amino acid sequences of 15 flaviviral E proteins were aligned for maximum homology with the help of the Microgenie Software package (version 4.0; Beckman Instruments, Inc., Fullerton, Calif.). The sequences utilized for this analysis were dengue virus type 1 (DEN-1) (27), DEN-2 strain PUO-218 (12), DEN-2 strain Jamaica (9), DEN-2 strain S1 (15), DEN-4 (42), Saint Louis encephalitis virus (36), Murray Valley encephalitis virus (7), Japanese encephalitis virus strain JaOArS982 (35), Japanese encephalitis virus strain Nakayama (28), WN (38), Kunjin (5), yellow fever virus strain 17D (32), yellow fever virus strain Asibi (14), TBE virus eastern subtype strain Sofyn (31), and TBE virus western subtype strain Neudoerfl (24a). From the alignment of all 15 sequences, the number of different amino acid residues was determined for each position. Gaps were counted as mismatches. This yielded a sequence of numbers ranging from 1 (=100% conservation) to a theoretical maximum value of 15 (=different amino acid residues in each sequence). The variability was plotted as a moving average (window of nine positions) of these values.

For determination of potential amphipathic sequences, the MOMENT plotting program (kindly provided by H. Wolf and S. Modrow, Pettenkofer Institut, Munich, Federal Republic of Germany) was utilized, which is part of the sequence analysis software package of the University of Wisconsin Genetics Computer Group, version 4 (copyright 1986 by John Devereux and Paul Haeberli). This program calculates the hydrophobic moments of an amino acid sequence by the method of Eisenberg (10).

TABLE 1. Antigenic variants of TBE virus

Variant	Selecting MAb	Antigenic domain	Nucleotide change (position)[a]	Amino acid change (position)[a]
VA3a	A3	A	C→U (212)	Ala→Val (71)
VA3b	A3	A	A→G (200)	Asp→Gly (67)
VA4	A4	A	G→A (697)	Gln→Lys (233)
VA5	A5	A	A→G (620)	Glu→Gly (207)
Vi2	i2		A→G (511)	Lys→Glu (171)
VB1	B1	B	U→G (1167)	Ser→Arg (389)
VB4	B4	B	U→C (1150)	Tyr→His (384)

[a] Position numbers are counted from the first nucleotide of the protein E-coding region and the first amino acid of protein E, respectively.

RESULTS

Nucleotide sequence analysis of antigenic variants. Antigenic variants of TBE virus had been generated by growing the virus in chicken embryo cells in the presence of neutralizing MAbs (H. Holzmann et al., in press). The MAbs used for selection were A3, A4, and A5, B1 and B4, and i2. Sequence analysis of the entire protein E-coding region revealed for each variant only single nucleotide exchanges leading to single amino acid changes of the protein E sequences. These results are summarized in Table 1. Variants VA3a and VA3b were both selected by the same MAb (A3) in two independent experiments. They exhibited distinct genotypes, the respective amino acid changes being four residues apart on the primary sequence. On the other hand, two variants independently derived by selection with MAb B4 were found to have identical mutations (data not shown).

Figure 1 indicates the positions of the mutations present in the sequences of the seven antigenic variants on a linear map of protein E. The variants selected by domain A MAbs (A3, A4, A5) had mutations spaced fairly distant from each other

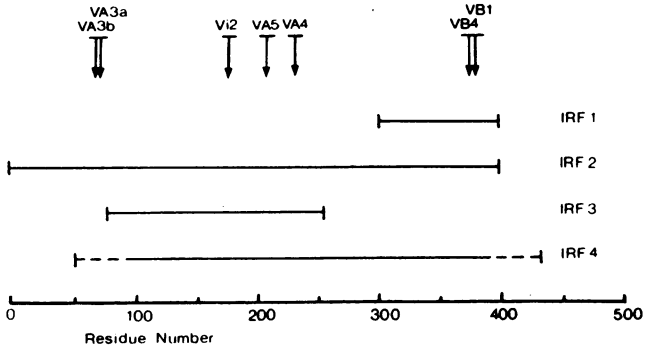

FIG. 1. Location of four IRFs (IRF1 to IRF4) and the mutations present in seven antigenic variants of TBE virus (VA3a, VA3b, Vi2, VA4, VA5, VB1, VB4) on a linear map of protein E. The locations of IRF1, IRF2, and IRF3 were determined by amino-terminal sequencing. The ambiguity of the exact location of IRF4 is indicated by broken lines at the ends of this fragment. The arrows depict the positions of the mutations present in the respective antigenic variants. Data on the variants and IRFs are summarized in Tables 1 to 3.

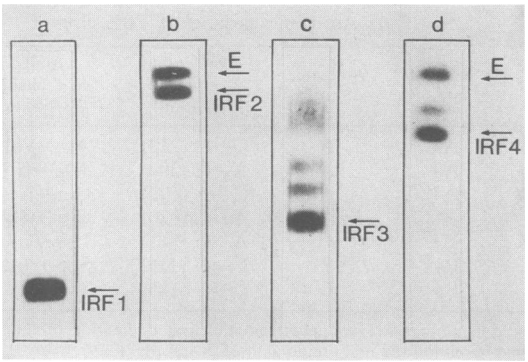

FIG. 2. Immunoblots of fragments of protein E. Protein E fragments were separated on SDS gels, blotted onto nitrocellulose paper, and reacted with ^{125}I-labeled MAbs as described in Materials and Methods. Lane a, 24-h tryptic digest reacted with MAb B1; only IRF1 is detectable. Lane b, 30-min tryptic digest reacted with MAb A1; IRF2 and undigested protein E are visible. Lane c, CNBr fragments reacted with MAb C2; besides IRF3, a number of larger partial degradation products react with the MAb. Lane d, α-Chymotryptic digest reacted with MAb A1; IRF4, undegraded protein E, and another larger degradation product are visible.

on the primary sequence. The MAbs B1 and B4 of domain B, however, selected for mutations only five amino acids apart. It seems reasonable to assume that the mutated amino acid is itself part of the epitope recognized by the selecting MAb, as had been shown by X-ray crystallography to be the case for variants of influenza virus (37). However, the possibility cannot be ruled out that mutations at distant sites cause conformational changes in the respective epitopes.

Immunoblotting of four IRFs of protein E. The reactivity of four different fragments of protein E (designated immunoreactive fragment 1 [IRF1] to IRF4) with MAbs was analyzed by immunoblotting. A representative immunoblot of each fragment is shown in Fig. 2. IRF1 was obtained by digestion of protein E rosettes (19) with trypsin for 24 h. This fragment migrates in SDS gels with an apparent molecular mass of 9 kilodaltons (kDa) and has been characterized previously (41). IRF2 is observed as an intermediary degradation product of tryptic digestion and migrates corresponding to a size of 47 kDa.

IRF3 is a 19-kDa fragment obtained by chemical degradation with cyanogen bromide that has been described previously (18). IRF4 was generated by α-chymotryptic digestion of SDS-denatured virus and migrated corresponding to a molecular size of 38 kDa. Each of the four fragments was tested for its reactivity with the whole panel of MAbs. Results are summarized in Table 2. IRF1 reacted exclusively with all domain B MAbs. IRF2 reacted with all MAbs except A3, A4, A5, i1, and i2. The epitopes recognized by MAbs A3, A4, A5, and i2 are sensitive to denaturation by SDS (Guirakhoo el al., in press), and therefore these MAbs are not expected to react under the conditions of immunoblotting. IRF3 was found to contain most of the epitopes of domain C, and IRF4 reacted with the same MAbs as IRF2.

Localization of IRFs in protein E sequence. The amino termini of the fragments IRF1, IRF2, and IRF3 were determined, permitting us to localize exactly the positions of these fragments on the primary sequence of protein E. Table 3 lists the amino-terminal sequences and the derived sequence position numbers of the amino- and carboxy-terminal ends of these fragments. IRF1 starts at amino acid residue 301. In agreement with the cleavage specificity of trypsin, a Lys residue is found at position 300. Judging from its

TABLE 2. Reactivity of MAbs with fragments of protein E

Fragment	Molecular size (kDa)[a]	Generated by cleavage with:	Reactive MAbs[b]
IRF1	9	Trypsin	B1–B5
IRF2	47	Trypsin	A1, A2, B1–B5, C1–C6, i3
IRF3	19	CNBr	C1, C3–C6
IRF4	38	α-Chymotrypsin	A1, A2, B1–B5, C1–C6, i3

[a] As determined from SDS gels.
[b] The reactivity with MAbs was determined by a Western blot (immunoblot) analysis (see Materials and Methods). MAbs recognizing SDS-sensitive epitopes do not react.

apparent molecular mass of 9 kDa, IRF1 is thought to extend approximately toward residue 384. Closest to this position is Lys-395, which is assumed to most likely form the carboxy terminus of IRF1 (Table 3). Thus, the calculated molecular mass is 10.5 kDa despite the apparent size of 9 kDa derived from SDS gels.

The amino terminus of IRF2 was determined to be identical to that of protein E. From its molecular weight, it is thought to extend to approximately residue 400.

The first residue of IRF3 was determined to be glycine 78. According to the cleavage specificity of CNBr, residue 77 of TBE virus protein E is a methionine. From its molecular weight, it is concluded that IRF3 extends beyond the next Met residue (176) until Met-252, which is assumed to be its carboxy terminus (Table 3).

The amino terminus of IRF4 was not determined. However, some information about its position can be derived by the following considerations. Residue 389 is required for the structural integrity of the epitope of MAb B1 as shown by the mutation present in VB1. MAb B1 reacts with IRF4. Consequently, IRF4 can be assumed to extend at its carboxy terminus at least until residue 389. As deduced from its molecular weight, this fragment is thought to lack at least 50 amino acid residues at the amino terminus of protein E. Figure 1 shows the positions of the IRFs on a linear representation of protein E. The following conclusions concerning the positions of antigenic epitopes and domains can be drawn. At least the first 50 amino-terminal and the 100 most carboxy-terminal amino acid residues are not involved in the epitopes recognized by MAbs A1, A2, B1 to B5, C1 to C6, and i3. The entire domain B is contained between residues 301 and 395, whereas domain C is located somewhere between residues 78 and 252. Thus, the antigenic domains B and C correspond to two distinct protein domains which are built up by nonoverlapping sections of the primary sequence of protein E. Since MAbs A1 and A2 do not react with either IRF1 or IRF3, but react with IRF2 and IRF4,

TABLE 3. Localization of IRFs of protein E

Fragment	Amino-terminal sequence[a]	Amino terminus (position)[b]	Carboxy terminus[c] (position)
IRF1	Gly-Leu-Thr-Tyr-Thr-Met	Gly (301)	Lys (395)
IRF2	Ser-Arg-?-Thr-His-Leu-Glu	Ser (1)	Approx. residue 400
IRF3	Gly-Pro-Ala-Thr-Leu-Ala	Gly (78)	Met (252)

[a] Determined by gas-phase sequencing.
[b] Position numbers are counted from the first amino acid of protein E.
[c] As deduced from molecular weight (see text).

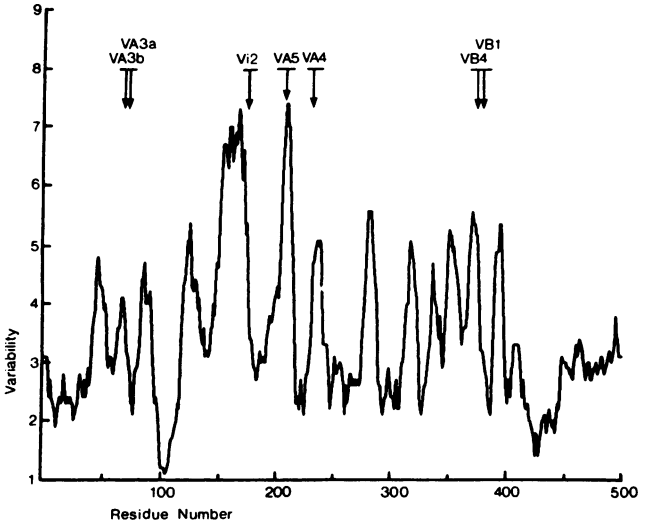

FIG. 3. Variability profile of flavivirus protein E. The amino acid sequences of 15 flaviviral E proteins were aligned, the number of different amino acid residues were calculated for each position, and these numbers were plotted as a moving average of 9. Thus, a value of 1 represents 100% conservation, and 15 would be maximum variability. The arrows depict the positions of the mutations present in the seven antigenic variants of TBE virus.

they most likely involve amino acid residues between positions 50 and 77 or 253 and 300 (Fig. 1).

The positions of the mutations present in antigenic variants of TBE virus are also indicated in Fig. 1 (see above). The mutations of variants VB1 and VB4 map within the range of IRF1, which carries all epitopes of domain B. IRF3, which carries domain C, also contains amino acid residues involved in the epitopes defined by MAbs A4, A5, and i2. The epitope of MAb A3 is located within the range of IRF2 and perhaps IRF4. MAbs A3, A4, A5, and i2 did not react with any of these fragments owing to the SDS sensitivity of the corresponding epitopes (Guirakhoo et al., in press).

Sequence comparisons. The amino acid sequences of the E proteins of 15 different flaviviruses were aligned for maximum homology, and a variability profile of the flavivirus protein E was calculated and plotted as described in Materials and Methods (Fig. 3). Consistent with previous observations (33), it is obvious from Fig. 3 that the most conserved region of protein E is located around residue 100. In fact, 12 of 14 amino acids from positions 98 to 111 are 100% conserved among all flavivirus sequences, and at each of the other two positions two different amino acid residues can be found. Two further regions that are significantly conserved in protein E are located at the amino terminus (residues 1 to 42) and close to the carboxy terminus (residues 410 to 445). A hypervariable region extends from residues 145 to 170, and a second one is found from residues 200 to 215. Also included in Fig. 3 are the positions of the mutations present in the antigenic variants of TBE virus, which are generally located within variable regions.

T-cell determinant. Peptides that are capable of forming an amphipathic α-helix represent potential T-cell antigenic determinants (3). A computer scanning of the primary sequence of TBE virus protein E indicated only one region longer than six amino acids with the characteristics of a potential amphipathic α-helix. This sequence, extending from residues 398 to 413, was further analyzed (Fig. 4). Hydrophobicity values were plotted for each residue according to the scale of Fauchére and Pliška (11), which had been found to perform best for the identification of immunodominant T-cell sites (6). Figure 4 shows that between residues 398 and 413, these hydrophobicity values are well approximated by a sinusoid with a period of 3.6 residues per cycle. Other flavivirus sequences exhibited a similar potential T-cell determinant at exactly the same positions when sequences were aligned for maximum homology (Fig. 4). Data from the analyses of the protein E sequences of the viruses DEN-1 and DEN-4, and Murray Valley encephalitis virus and Japanese encephalitis virus, are not shown in Fig. 4, since these results were very similar to those shown for DEN-2 and WN virus, respectively.

DISCUSSION

The envelope (E) proteins of all flaviviruses sequenced so far show perfect conservation of 12 cysteine residues which form six disulfide bridges. In addition, they exhibit highly homologous hydrophilicity profiles, suggesting that all members of the flavivirus family share a fairly common architecture of their E proteins. The assignment of disulfide bridges for WN virus by Nowak and Wengler (29) was a major breakthrough toward the elucidation of the protein E structure. Based on these analyses, it was proposed that protein E was made up of three major regions (R1 to R3) separated by two intervening sections (L1 and L2), as well as two transmembrane segments at the carboxy terminus. The model of Nowak and Wengler (29) also included the locations of potential antigenic sites as deduced from the hydrophilicity profile of the protein. By using a direct approach, including the analysis of antigenic variants and IRFs, we were able to determine the location of structurally defined antigenic determinants. The compilation of these data with those on the location of disulfide bridges and the structural and functional characteristics of epitopes (Guirakhoo et al., in press) led to a model (Fig. 5) showing a different arrangement of the polypeptide chain into distinct protein domains that also correspond to the antigenic domains A, B, and C.

Domain A is shown as a discontinuous structural entity composed of distant regions of the primary sequence, specifically amino acids 50 to 125 (containing three disulfide bridges) and 200 to 250. This arrangement is based on a number of considerations. The antigenic domain A is defined by five different MAbs (A1 to A5). Of these, MAbs A3 to A5 neutralize TBE virus and were used to select antigenic variants. The mutations present in these variants mapped on distant regions of the primary sequence. However, there is strong evidence that the epitopes recognized by these MAbs are located on the same structural entity, since they share the same structural properties (sensitivity to SDS and acid pH) and since A3 and A4 as well as A4 and A5 mutually block each other's binding (Guirakhoo et al., in press). The following considerations provide evidence that the epitopes recognized by MAbs A1 and A2 involve amino acid residues between positions 50 and 77. A1 and A2 are located on IRF2 and IRF4, but not on IRF1 and IRF3 (compare Fig. 1 and Table 2), indicating that they involve any of either amino acids 50 to 77 or 252 to 300. MAbs A2 and A3 mutually block each other's binding, and the mutation present in variants VA3a and VA3b maps to positions 67 and 71. The epitopes recognized by A1 and A2 are not destroyed by SDS treatment but are sensitive to reduction by 2-mercaptoethanol, suggesting that these epitopes are part of a protein domain stabilized by disulfide bridges. Such a disulfide-stabilized

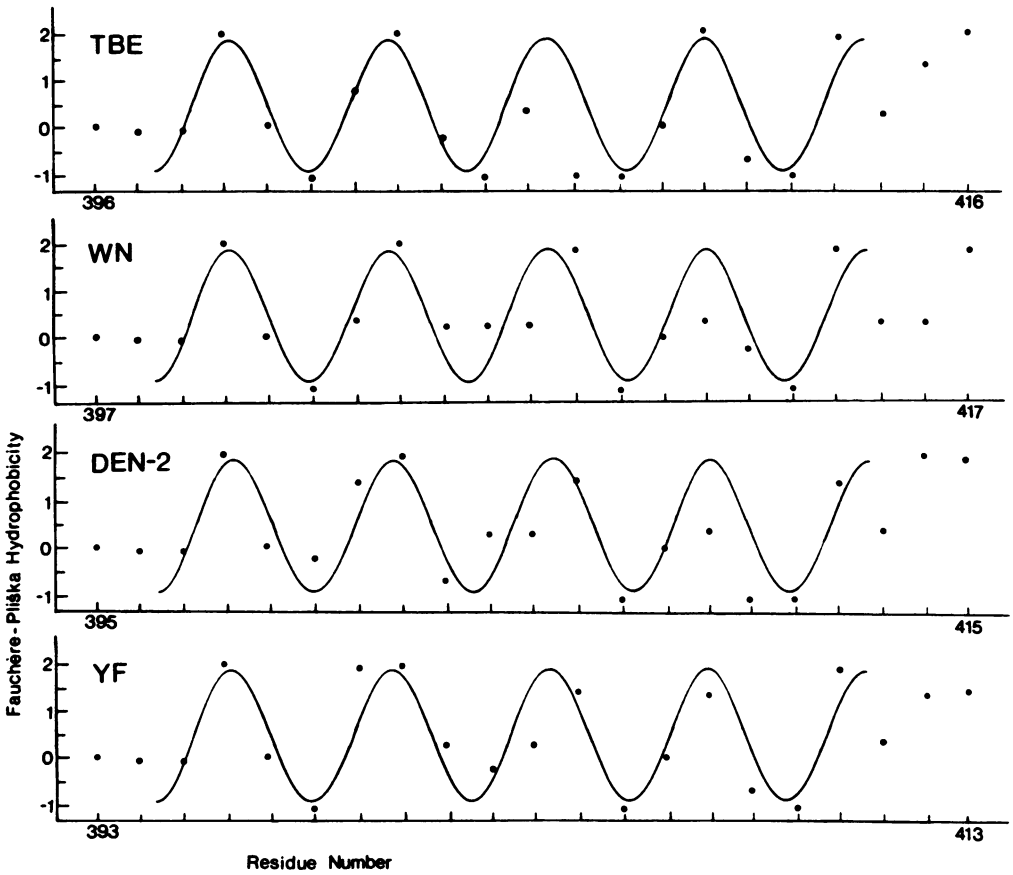

FIG. 4. Flavivirus protein E sequences that potentially form amphipathic α-helices. The amino acid sequences of four flaviviruses (belonging to four different serological subgroups) are aligned for maximum homology. The hydrophobicity value of each amino acid residue between the indicated borders is plotted according to the scale of Fauchère and Pliška (11). A sinusoid with a period of 3.6 residues per cycle approximates the hydrophobicity plots, indicating the potential of these regions to form an amphipathic α-helical structure.

region of the protein is located between residues 60 and 121, which includes as many as three disulfide bridges.

This disulfide-specified region also contains the most conserved sequence (residues 98 to 111) of protein E (indicated by triangles in Fig. 5), suggesting a crucially important function of this sequence element. It represents a moderately hydrophobic sequence surrounded by a hydrophilic and variable sequence environment. This is reminiscent of the receptor-binding area of influenza virus that was shown by X-ray crystallography to be a hydrophobic pocket surrounded by variable surface-exposed sequence elements which are recognized by neutralizing antibodies (40). One may therefore hypothesize that the conserved peptide 98 to 111 is part of the flavivirus receptor-binding site.

Domain B is shown as a distinct protein region located between residues 301 and 395. Its borders are defined by the end of the tryptic fragment IRF1 (indicated by tryp in Fig. 5). This protease-resistant fragment contains all epitopes of domain B, suggesting that no other parts of the polypeptide chain are involved in this structural domain. Domain B contains one disulfide bridge and apparently represents an independently folding protein domain with a strong tendency for renaturation (41). As shown previously, its structural integrity can be destroyed by opening the disulfide bridge, but it fully regains its antigenic reactivity when renaturation of the disulfide bridge is allowed to occur (41). Domain B seems to be strongly folded and stabilized by its single disulfide bond and probably additional intramolecular forces as suggested by the presence of several potential tryptic cleavage sites (small arrows in Fig. 5) which are apparently inaccessible in the native form of this domain. The involvement of the carboxy-terminal part of IRF1 in the antigenic domain B is demonstrated by the mutations present in the antigenic variants VB1 and VB4.

Antigenic domain C is shown as a single highly variable loop which does not contain any disulfide bridges, but carries the single carbohydrate side chain of the TBE virus E protein (solid diamond in Fig. 5), since IRF1, containing the second potential N-glycosylation site, was shown to be not glycosylated (41). The location of domain C, which is contained within the CNBr fragment IRF3 (borders depicted by CNBr in Fig. 5), was deduced as follows. The mutation present in variants VA4, VA5, and Vi2 defines areas of fragment IRF3 that do not belong to domain C, since the epitopes A4, A5, and i2 do not overlap with any domain C epitope as demonstrated by competitive binding assays. Second, reduction and carboxymethylation do not affect the structural integrity of domain C epitopes, indicating that no disulfide bond is involved in the integrity of this structural entity. Third, domain C is resistant to SDS denaturation, suggesting that it is not composed of discontinuous regions of the protein like those of domain A. However, this denaturation resistance is lost when the carbohydrate side chain is removed (Guirakhoo et al., in press), suggesting that

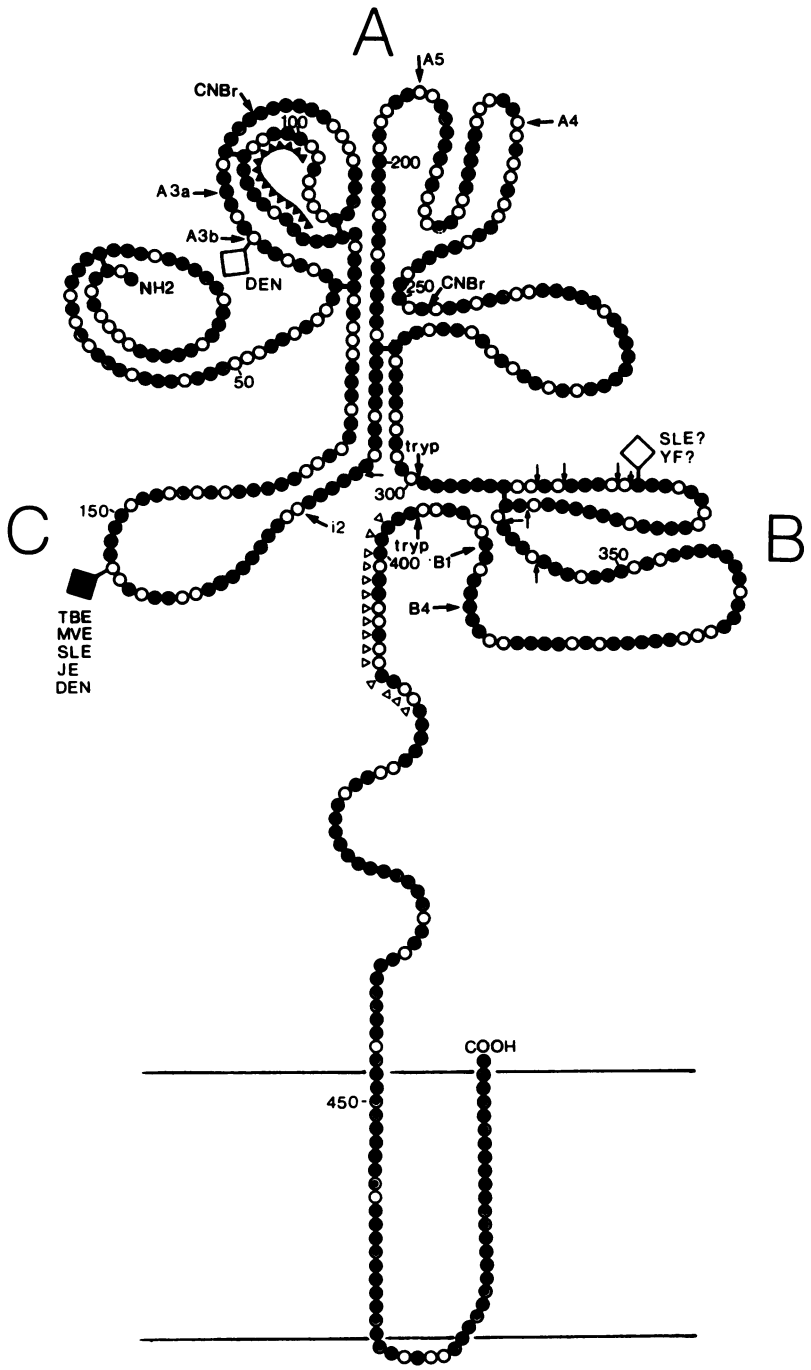

FIG. 5. Model of the TBE virus protein E. Open circles represent hydrophilic amino acid residues (Arg, Lys, Asn, Asp, Gln, Glu, His), dotted circles show intermediate amino acid residues (Pro, Tyr, Ser, Trp, Thr, Gly), and solid circles show hydrophobic amino acid residues (Ile, Val, Leu, Phe, Cys, Met, Ala). Amino acids were classified by the scale of Kyte and Doolittle (22). Position numbers are shown every 50 amino acids. Cysteine residues forming disulfide bridges are connected by solid lines. Arrows depict cleavage sites that liberate IRF1 (tryp) and IRF3 (CNBr), respectively. Small arrows indicate potential cleavage sites within these fragments that are not utilized. Two solid lines stand for the lipid membrane that is spanned by two transmembrane regions of protein E. The polypeptide chain is folded to indicate the antigenic domains A, B, and C, which are designated by large capital letters. Arrows together with the names of neutralizing MAbs depict the locations of the mutations identified in the respective antigenic variants of TBE virus by sequence analysis. A line of solid triangles indicates the almost perfectly conserved sequence within domain A. A line of open triangles marks the region of a potential T-cell determinant. A solid diamond represents the carbohydrate side chain of TBE virus. The Murray Valley (MVE), St. Louis (SLE) and Japanese (JE) encephalitis viruses and DEN viruses have potential N-glycosylation sites at the homologous position. Yellow fever (YF) and St. Louis encephalitis viruses have such a site within domain B, DEN viruses within domain A. The homologous positions of TBE virus are shown by open diamonds.

glycosylation stabilizes domain C, although domain C epitopes do not directly depend on the presence of the sugar residues. Last, MAbs directed against domain C epitopes exhibit very little cross-reactivity (MAbs C3 to C6 are even subtype specific), suggesting the involvement of hypervariable amino acid residues, such as those between positions 145 and 170 (Fig. 3).

Three MAbs (i1 to i3) recognize epitopes that do not overlap with any of the three antigenic domains A, B, or C. Epitope i2 is defined by the mutation present in variant Vi2. In the protein E model, i2 is placed between the three protein domains A, B, and C. This intermediate location of i2 is compatible with the findings that (i) i2 shares a number of structural characteristics with domain A epitopes, such as sensitivity to SDS and acid pH, and (ii) binding of MAb i2 is unidirectionally blocked by MAbs B4, B5, and C1.

The first 50 amino acids could not be assigned to any of the three antigenic domains, but preliminary evidence from protein expression studies (W. Tuma et al., manuscript in preparation) locates epitope i1 within this amino-terminal region.

The model also reflects the finding that most of the B-cell determinants described, especially those of domains A and B, are conformation dependent and probably even discontinuous. This is consistent with the view derived from X-ray crystallographic studies that the majority of B-cell determinants depend on the tertiary folding of the polypeptide chain (1, 2).

In a recent study, Wengler et al. (39) identified two surface-exposed regions of WN virus protein E by their accessibility to proteolytic cleavage which, in good agreement with our model, correspond to the carboxy-terminal end of domain B and a section of domain C, respectively. By sequence analysis of neutralization escape mutants, Lobigs et al. (24) localized a neutralization epitope on the yellow fever virus protein E sequence. The mutations present in their antigenic variants mapped at amino acid position 71 or 72, which perfectly corresponds to the mutated position of the TBE virus variant VA3a. By a different approach involving bacterial expression, Mason et al. (25, 26) identified two antigenic domains of protein E. One of these domains is contained between residues 280 and 414 and was found to depend structurally on a disulfide bond (P. Mason, personal communication). This result agrees very well with the description of domain B in the TBE virus protein E model. The other domain maps between residues 76 and 93, corresponding to what we identified to be part of domain A.

Besides these B-cell-specified epitopes, a potential amphypathic α-helix which might serve as a T-cell determinant was identified by primary sequence analysis and is depicted in the model in Fig. 5 by a line of open triangles. It is located adjacent to the carboxy terminus of fragment IRF1 and does not overlap with any of the three B-cell antigenic domains.

The model further indicates the intramembranous parts of the protein, i.e., the hydrophobic anchor of protein E and the signal sequence of protein NS1, which are interrupted by a short hydrophilic protein section. The first transmembrane segment (residues 448 to 469) includes a very hydrophilic amino acid (Lys, Arg, or Gln) at a position conserved among all flaviviruses (number 457 in the TBE protein E sequence), causing controversial predictions of the actual length of this anchor sequence (5, 7, 9, 12, 27, 28, 32, 36, 38, 42). If the hydrophilic residue was not placed within the lipid envelope, the transmembrane segment would be only 12 amino acids long, although such elements usually consist of 20 to 25 hydrophobic amino acids which, folded as an α-helix, are required to span the membrane and provide a stable association (8). If it is a part of the transmembrane region, the hydrophilic residue might be involved in some kind of protein aggregation and thus be protected from the hydrophobic environment. Alternatively, the amino acid might be modified by the addition of a fatty acid side chain (30).

The carboxy terminus of WN virus protein E was determined by Nowak et al. (T. Nowak, P. Farber, G. Wengler, and G. Wengler, Virology, in press) to be the residue preceding the cleavage site that liberates the amino terminus of protein NS1. This cleavage is probably catalyzed by a signalase in the lumen of the endoplasmic reticulum. Therefore, the carboxy terminus of protein E is shown to be exposed at the outer surface of the viral membrane.

Owing to the expected overall structural conservation, the model presented here should be useful as a basis for the further analysis of the antigenic structure and functional activity of flavivirus protein E.

ACKNOWLEDGMENTS

We thank Hans Wolf and Susanne Modrow for allowing us to use their computer facilities at the Pettenkofer Institut (Munich, Federal Republic of Germany) and for providing programs for the prediction of antigenic determinants. We further acknowledge Melby Wilfinger, Angela Dohnal, and Walter Holzer for their expert technical assistance, as well as Susanne Pfauser for typing the manuscript.

LITERATURE CITED

1. **Barlow, D. J., M. S. Edwards, and J. M. Thornton.** 1986. Continuous and dicontinuous protein antigenic determinants. Nature (London) **322**:747–748.
2. **Berzofsky, J. A.** 1985. Intrinsic and extrinsic factors in protein antigenic structure. Science **229**:932–940.
3. **Berzofsky, J. A., K. B. Cease, J. L. Cornette, J. L. Spouge, H. Margalit, I. J. Berkower, M. F. Good, L. H. Miller, and C. DeLisi.** 1987. Protein antigenic structures recognized by T-cells: potential applications to vaccine design. Immunol. Rev. **98**: 9–52.
4. **Brinton, A. B.** 1986. Replication of flaviviruses, p. 327–374. *In* S. Schlesinger and M. J. Schlesinger (ed.), The Togaviridae and Flaviviridae. Plenum Publishing Corp., New York.
5. **Coia, G., M. D. Parker, G. Speight, M. E. Byrne, and E. G. Westaway.** 1988. Nucleotide and complete amino acid sequences of Kunjin virus: definitive gene order and characteristics of the virus-specified proteins. J. Gen. Virol. **96**:1–21.
6. **Cornette, J. L., K. B. Cease, H. Margalit, J. L. Spouge, J. A. Berzofsky, and C. DeLisi.** 1987. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. J. Mol. Biol. **195**:659–685.
7. **Dalgarno, L., D. W. Trent, J. H. Strauss, and C. M. Rice.** 1986. Partial nucleotide sequence of the Murray Valley encephalitis virus genome. J. Mol. Biol. **187**:309–323.
8. **Davis, N. G., and P. Model.** 1985. An artificial anchor domain: hydrophobicity suffices to stop transfer. Cell **41**:607–614.
9. **Deube, V., R. M. Kinney, and D. W. Trent.** 1986. Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. Virology **155**:365–377.
10. **Eisenberg, D., R. M. Weiss, and T. C. Terwillinger.** 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. Proc. Natl. Acad. Sci. USA **81**:140–144.
11. **Fauchére, J. L., and V. Pliška.** 1983. Hydrophobic parameters pi of amino-acid side chains from the partitioning of *N*-acetyl-amino-acid amides. Eur. J. Med. Chem. **18**:369–375.
12. **Gruenberg, A., W. S. Woo, A. Biedrycka, and P. J. Wright.** 1988. Partial nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue virus type 2, New Guinea C and PUO-218 strains. J. Gen. Virol. **69**:1391–1398.
13. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.

14. **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.
15. **Hahn, C. S., R. Galler, T. Hunkapiller, J. M. Dalrymple, J. H. Strauss, and E. Strauss.** 1988. Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. Virology **161**:167–180.
16. **Heinz, F. X.** 1986. Epitope mapping of flavivirus glycoproteins. Adv. Virus Res. **31**:103–186.
17. **Heinz, F. X., R. Berger, W. Tuma, and C. Kunz.** 1983. A topological and functional model of epitopes on the structural glycoprotein of tick-borne encephalitis virus defined by monoclonal antibodies. Virology **126**:525–537.
18. **Heinz, F. X., R. Berger, W. Tuma, and C. Kunz.** 1983. Location of immunodominant antigenic determinants on fragments of the tick-borne encephalitis virus glycoprotein: evidence for two different mechanisms by which antibodies mediate neutralization and hemagglutination inhibition. Virology **130**:485–501.
19. **Heinz, F. X., and C. Kunz.** 1980. Formation of polymeric glycoprotein complexes from a flavivirus: tick-borne encephalitis virus. J. Gen. Virol. **49**:125–132.
20. **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.
21. **Heinz, F. X., W. Tuma, F. Guirakhoo, R. Berger, and C. Kunz.** 1984. Immunogenicity of tick-borne encephalitis virus glycoprotein fragments: epitope-specific analysis of the antibody response. Virology **65**:1921–1929.
22. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
23. **Laemmli, U. K., and M. Favre.** 1973. Maturation of the head of bacteriophage T4. I. DNA packing events. J. Mol. Biol. **80**: 575–599.
24. **Lobigs, M., L. Dalgarno, J. J. Schlesinger, and R. C. Weir.** 1987. Location of a neutralization determinant in the E protein of yellow fever virus (17D vaccine strain). Virology **161**:474–478.

24a. **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.

25. **Mason, P. W., T. L. Mason, and M. J. Fournier.** 1988. Current approaches for the development of dengue vaccines and related aspects of the molecular biology of flaviviruses. J. Infect. Dis. **157**:1105–1111.
26. **Mason, P. W., P. C. McAda, J. M. Dalrymple, M. J. Fournier, and T. L. Mason.** 1987. Expression of Japanese encephalitis virus antigens in Escherichia coli. Virology **158**:361–372.
27. **Mason, P. W., P. C. McAda, T. L. Mason, and M. J. Fournier.** 1987. Sequence of the dengue-1 virus genome in the region encoding the three structural proteins and the major nonstructural protein NS1. Virology **161**:262–267.
28. **McAda, P. C., P. W. Mason, C. S. Schmaljohn, J. M. Dalrymple, T. L. Mason, and M. J. Fournier.** 1987. Partial nucleotide sequence of the Japanese encephalitis virus genome. Virology **158**:348–360.
29. **Nowak, T., and G. Wengler.** 1987. Analysis of disulfides present in the membrane proteins of the West Nile flavivirus. Virology **156**:127–137.
30. **Pillai, S., and D. Baltimore.** 1987. Myristoylation and the post-translational acquisition of hydrophobicity by the membrane immunoglobulin heavy-chain polypeptide in B lymphocytes. Proc. Natl. Acad. Sci. USA **84**:7654–7658.
31. **Pletnev, A. G., V. F. Yamshchikow, and V. M. Blinov.** 1986. Tick-borne encephalitis virus genome. FEBS Lett. **200**:317–321.
32. **Rice, C. M., E. M. Lenches, S. R. Eddy, S. J. Shin, R. L. Sheets, and J. H. Strauss.** 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science **229**:726–733.
33. **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.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Sumiyoshi, H., K. Morita, C. Mori, I. Fuke, T. Shiba, Y. Sakaki, and A. Igarashi.** 1986. Sequence of 3000 nucleotides at the 5′ end of Japanese encephalitis virus RNA. Gene **48**:195–201.
36. **Trent, D. W., R. M. Kinney, B. J. B. Johnson, A. V. Vorndam, J. A. Grant, V. Deubel, C. M. Rice, and C. Hahn.** 1987. Partial nucleotide sequence of St. Louis encephalitis virus RNA: structural proteins, NS1, ns2a, and ns2b. Virology **156**:293–304.
37. **Webster, R. G., G. M. Air, D. W. Metzger, P. M. Colman, J. N. Varghese, A. T. Baker, and A. G. Laver.** 1987. Antigenic structure and variation in an influenza virus N9 neuraminidase. J. Virol. **61**:2910–2916.
38. **Wengler, G., E. Castle, U. Leidner, T. Nowak, and G. Wengler.** 1985. Sequence analysis of the membrane protein V3 of the flavivirus West Nile virus and of its gene. Virology **147**:264–274.
39. **Wengler, G., G. Wengler, T. Nowak, and K. Wahn.** 1987. Analysis of the influence of proteolytic cleavage on the structural organization of the surface of the West Nile flavivirus leads to the isolation of a protease-resistant E protein oligomer from the viral surface. Virology **160**:210–219.
40. **Wiley, D. C., and J. J. Skehel.** 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu. Rev. Biochem. **56**:365–394.
41. **Winkler, G., F. X. Heinz, and C. Kunz.** 1987. Characterization of a disulfide bridge-stabilized antigenic domain of tick-borne encephalitis virus structural glycoprotein. J. Gen. Virol. **68**: 2239–2244.
42. **Zhao, B., E. Mackow, A. Buckler-White, L. Markoff, R. M. Cannock, C.-J. Lai, and Y. Makino.** 1986. Cloning full-length dengue type 4 viral DNA sequences: analysis of genes coding for structural proteins. Virology **155**:77–88.
43. **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.