Solution Structure and Structural Dynamics of Envelope Protein Domain III of Mosquito- and Tick-Borne Flaviviruses†

Shaoning Yu,§ Alice Wuu,‡,§ Reneeta Basu,‡,§ Michael R. Holbrook,⊥ Alan D. T. Barrett,*,⊥ and J. Ching Lee*,§

Department of Human Biological Chemistry & Genetics and Department of Pathology and Center for Biodefense and Emerging Infectious Diseases, The University of Texas Medical Branch at Galveston, Galveston, Texas 77555-1055

Received April 6, 2004; Revised Manuscript Received May 14, 2004

ABSTRACT: The mosquito-borne West Nile (WNV) and dengue 2 (DEN2V) viruses and tick-borne Langat (LGTV) and Omsk hemorrhagic fever (OHFV) viruses are arthropod-borne flaviviruses (family Flaviviridae, genus *Flavivirus*). These viruses are quite similar at both the nucleotide and amino acid level, yet they are very divergent in their biological properties and in the diseases they cause. The objective of this study was to examine the putative receptor-binding domains of the flaviviruses, the envelope (E) protein domain III (D3), which assume very similar structures either as part of the whole envelope protein or as individual entities, and to define the biophysical properties that distinguish among these viruses. Circular dichroism and Fourier transform infrared spectroscopy were employed to monitor the solution structure of these proteins. While the spectroscopic results found that the D3 from each of these viruses is composed of either β-sheets or β-turns, which is consistent with X-ray crystal data for tick-borne encephalitis and dengue viruses, these results reveal that recombinant D3s (rED3s) derived from tick-borne flaviviruses (LGT-rED3 and OHF-rED3) were similar to each other, while those from mosquito-borne flaviviruses (WN-rED3 and DEN-rED3) were similar to each other yet distinct from rED3 of the tick-borne viruses. Protein dynamic studies probed by fluorescence quenching and hydrogen/deuterium exchange found that the rED3s are dynamic entities. The tick-borne proteins again exhibit very similar dynamic properties, which are different from the mosquito-borne proteins. The WN-rED3 is significantly less stable than the other three rED3s. Overall, these differences in biophysical properties correlate with biological properties of these viruses that tick-borne flaviviruses are more stable than mosquito-borne flaviviruses.

The flaviviruses are a genus of approximately 70 small single-stranded positive-sense RNA viruses that are transmitted primarily by arthropods. Among the flaviviruses are a number of major human pathogens including the mosquito-borne dengue (DENV), yellow fever (YFV), Japanese encephalitis (JEV), and West Nile (WNV) viruses, as well as tick-borne viruses including the tick-borne encephalitis (TBE) serocomplex of viruses. The viral genome is approximately 11 kb in size encoding a single polyprotein of just over 3400 amino acids. The viral polyprotein is co- and posttranslationally cleaved to generate three structural proteins and seven nonstructural proteins. Among the structural proteins is the major viral surface protein, the envelope (E) protein, which serves as both a receptor-binding protein and fusion protein (*1*). The viral E protein exists as a heterodimer on the surface of the virion along with the second viral surface protein, the membrane (M) protein. X-ray crystal structures of the flavivirus E protein have been solved for both DENV type 2 (DEN2V) and a member of the TBE serocomplex, central European encephalitis (CEEV) virus (*2*, *3*). These structures divide the viral E protein into three distinct domains (I, II, and III). Domains I and II are called the central and dimerization domains, respectively, with the viral fusion peptide at the tip of domain II. Domain III (D3) of the flaviviruses has been shown to function as the putative receptor-binding domain for these viruses (*4*, Holbrook et al.[1]). Crystallographic data has shown that the D3s of CEEV (*2*) and DEN2V (*3*) have a β-barrel configuration and an Ig-like fold. The D3 of mosquito-borne flaviviruses is slightly larger than that of tick-borne flaviviruses because there is a four amino acid loop in the D3 of mosquito-borne viruses. This extra loop contains an RGD integrin-binding motif in some viruses that may be associated with binding to receptors in the mosquito midgut (*4*).

Since the introduction of WNV into North America in 1999, the public awareness of flavivirus infection has increased. WNV is a neurotropic flavivirus closely related to JEV. The virus primarily cycles between birds and mosquitoes with humans being accidental hosts of infection. Since the introduction of WNV into North America more than 13 000 human infections have been documented.

† Supported by NIH Grants GM-45579 (J.C.L.) and T32 AI 07536 (M.R.H.), CDC Grant U90/CCU618754 (A.B.), and Robert A. Welch Foundation Grants H-0013 and H-1238 (J.C.L.).

* To whom correspondence should be addressed. For J.C.L., tel (409) 772-2282, fax (409) 772-4298, e-mail jclee@utmb.edu. For A.B., tel (409) 772-6662, e-mail abarrett@utmb.edu.

§ Department of Human Biological Chemistry & Genetics.

‡ Summer Undergraduate Research Program.

⊥ Department of Pathology and Center for Biodefense and Emerging Infectious Diseases.

[1] Holbrook and Barrett, submitted for publication.

10.1021/bi049324g CCC: $27.50
Published on Web 06/22/2004

```
WN-rD3      KGTTYGVCS- KAFKFLGTPA DTGHGTVVLE LQYTGTDGPC KVPISSVASL NDLTPVGRLV
DEN-rD3     KGMSYSMCT- GKFKVVEEIA ETQHGTIVIR VQYEGDGSPC KIPLEIMDLD NRHV-LGRLI
LGT-rD3     KGLTYTVCDK TKFTWKRAPT DSGHDTVVME VGFSGT-RPC RIPVRAVAHG VPEVNVAMLI
OHF-rD3     KGLTYTMCDK AKFTWKRAPT DSGHDTVVME VAFSGT-KPC RIPVRAVAHG SPDVDVAMLI

WN-rD3      TVNPFVSVAT ANAKVLIELE -PPFGDSYIV VGRGEQQINH HWHKSGS
DEN-rD3     TVNPIVTEKD S--PVNVEAE -PPLGDSYII IGVEPGQLKL NWFKKGS
LGT-rD3     TPNPTMENNG G---GFIEMQ LPP-GDNIIY VG----DLNH QWFQKGS
OHF-rD3     TPNPTIENNG G---GFIEMQ LPP-GDNIIY VG----ELSH QWFQKGS
```

FIGURE 1: Amino acid sequence alignment of WN-rED3, DEN-rED3, OHF-rED3, and LGT-rED3.

The dengue viruses consist of four serologically closely related viruses (DEN1V−DEN4V) that are a major public health problem in tropical and subtropical regions of the world. These mosquito-borne viruses cause an acute febrile illness in the majority of cases, though it can progress to the more severe dengue shock syndrome or dengue hemorrhagic fever, which is potentially fatal. DENV are the only flaviviruses in which humans are a part of the natural cycle of the virus. DEN2V is perhaps the most widely distributed of the DENV.

The major human pathogens among the tick-borne flaviviruses include members of the TBE serocomplex of viruses. Most, such as CEEV, cause an encephalitic disease. Closely related viruses, such as Omsk hemorrhagic fever virus (OHFV), cause a primarily viscerotropic disease rather than neurotropic disease. Langat virus (LGTV) is a naturally attenuated tick-borne flavivirus that does not cause disease in humans.

The objective of this study was to determine the biophysical properties of recombinant E protein D3 (rED3)[2] derived from four different flaviviruses and to determine whether these properties were associated with biological characteristics of the viruses. The four viruses described above were chosen for this study because two are transmitted by mosquitoes (WNV and DEN2V) and two by ticks (OHFV and LGTV). Two of these viruses are associated with encephalitic disease (WNV and LGTV) and two with viscerotropic disease (DEN2V and OHFV). The amino acid sequence alignment of these proteins is summarized in Figure 1. Using several different methods, this study found that these proteins exhibit subtle yet distinct differences in secondary and tertiary structures in addition to protein dynamics, which affect protein stability. These data correlate with biological properties that indicate that tick-borne flaviviruses are more tolerant of environmental conditions than are those transmitted by mosquitoes. These properties include aerosol transmission of tick-borne flaviviruses as well as infection via contact with infected tissues or milk from infected goats or sheep. These observations contrast sharply with those about mosquito-borne flaviviruses, which are not thought to survive outside either arthropod or animal hosts.

METHODS AND MATERIALS

Chemicals. Ultrapure guanidine HCl was a product of ICN Biochemical. Blue Dextran 2000, sodium azide, and acrylamide were purchased from Pharmacia Biotech., Kodak, and Boehringer Mannheim, Germany, respectively. All experiments were conducted in TN200 (50 mM Tris, 200 mM NaCl at pH 8.5 and 25 °C) buffer.

Protein Preparation. Recombinant E protein D3 (rED3) incorporates approximately amino acids 300−400 of the E protein (TBE amino acid numbering (*2*)). The ED3 was amplified from the respective viruses using specific primers in RT-PCR as previously described.[3] The PCR product was then cloned into the pMal-c2x vector from a maltose-binding protein (MBP) fusion protein system (New England Biolabs, Beverley, MA). Expression and purification of the fusion protein was performed essentially using the manufacturer's instructions. Briefly, the protein was expressed by induction of a 0.5-OD_{600} culture with 0.4 mM IPTG for 4 h at 37 °C. The cells were pelleted and lysed, and the resulting supernatant was passed over an amylose affinity column (New England Biolabs). The fusion protein was eluted with 10 mM maltose in column buffer, concentrated, and cleaved with factor Xa (Novagen). Following cleavage, the protein was passed over a Superdex75 column. The rED3-containing fractions were collected and concentrated. Proteins were further purified with a Sephacryl HR S-100 column and concentrated. The purity of these proteins was greater than 95% homogeneous based on results from SDS−PAGE. The extinction coefficients for WN-rED3, DEN-rED3, OHF-rED3, and LGT-rED3 are 10 095, 9650, 15 595, and 14 060 M^{-1} cm^{-1} at 280 nm, respectively.

Circular Dichroism Spectroscopy. CD measurements were performed on an AVIV model 60 DS spectropolarimeter. Far-UV CD spectra of proteins in the native state were measured over the range of 190−260 nm by using cells of 0.01 cm path length. Each spectrum was recorded in 0.5 nm wavelength increments. For each sample, three repetitive scans were obtained and averaged. Protein solutions with a concentration of 35−40 μM were used.

FT-IR Spectroscopy. FT-IR spectra were measured with a Bomem MB series Fourier transform infrared spectrometer (Quebec, Canada) equipped with a dTGS detector and purged constantly with dry air. Protein samples (∼18 mg/mL) were warmed to room temperature and loaded in a CaF_2 cell with

[2] Abbreviations: GdnHCl, guanidine hydrochloride; TN200 buffer, 50 mM Tris, 200 mM NaCl at pH 8.5 and 25 °C; PAGE, polyacrylamide gel electrophoresis; CD, circular dichroism; HPLC, high-pressure liquid chromatography; SEC, size-exclusion chromatography; FT-IR, Fourier transform infrared spectroscopy; H/D, hydrogen/deuterium; WN-rED3, West Nile recombinant envelope protein domain 3; DEN-rED3, Dengue 2 virus domain 3; OHF-rED3, Omsk hemorrhagic fever domain 3; LGT-rED3, Langat virus domain 3.

[3] Beasley, D. W., Holbrook, M. R., Travassos da Rosa, A. P., Tesh, R. B., Shope, R. E., and Barrett, A. D. (2004) *J. Clin. Microbiol. 42*, 2759−2765.

a 7.5 μm spacer. For each spectrum, a 256-scan interferogram was collected in single-beam mode with a 4 cm^{-1} resolution at a rate of 3 scans/s. Reference spectra were recorded under identical conditions with only the corresponding buffer in the cell. Protein spectra were obtained using a previously established protocol (*5*). A straight baseline between 2000 and 1750 cm^{-1} was used as the standard to judge the success of water subtraction. Second-derivative spectra were obtained with a seven-point Savitsky−Golay derivative function, baseline-corrected, and area-normalized as described (*5*, *6*). The secondary structure content of WN-rED3 was calculated by curve-fitting analysis of the inverted second-derivative spectrum in the amide I band range of 1600−1700 cm^{-1} (*5*). This band is ascribed to >C═O stretching vibration of the peptide bond (*7*). It was assumed that the fraction of residues composing each secondary structural element is proportional to the relative percent area of the associated vibrational band (*8*).

Analytical Ultracentrifugation. Sedimentation velocity experiments were carried out in a Beckman Optima XL-A analytical ultracentrifuge equipped with absorbance optics and An60Ti rotor at 60 000 rpm. Velocity data were collected at 280 nm at a spacing of 0.002 cm with no averaging in a continuous scan mode and were analyzed using the SetFit Program. The reported weight-average sedimentation coefficient values ($\bar{S}_{20,w}$) were calculated by a weighted integration over the entire range of sedimentation coefficient covered by the $C(s)$ distribution and corrected for the solution density and viscosity. The apparent weight-average molecular weights were obtained by fitting the sedimentation equilibrium data with the following equation:

$$C = E + C_{1,r_0} \exp\left[\frac{(1 - \bar{\nu}\rho)\omega^2}{2RT} M(r^2 - r_0^2)\right] \quad (1)$$

where C is the observed protein concentration in absorbance at radial position r, E is the baseline offset, and C_{1,r_0} is the concentration of monomeric protein. At the meniscus, r_0, ν is the partial specific volume, ρ is the solvent density, ω is the angular velocity, M is the apparent weight-average molecular weight, and R and T are the gas constant and temperature in Kelvin, respectively. The value of $\bar{\nu}$ is 0.742 for studied proteins in TN200 buffer, derived from the amino acid composition of these proteins using the method of Cohn and Edsall (*9*).

Fluorescence Acrylamide Quenching. Fluorescence intensity measurements were carried out in 1-cm quartz cuvette at 23 °C using a Perkin-Elmer LS50B spectrofluorometer. The excitation and emission wavelengths were 295 and 340 nm, respectively. The temperature of the cuvette holder was regulated by a circulating water bath at 23 ± 0.1 °C. Protein concentrations were 5 μM. The reaction mixtures were titrated with 5.0 M acrylamide in TN200 at room temperature. Four cuvettes containing different solutions were measured one by one (*10*). A, B, and C cuvettes contained 5 μM protein in TN200 buffer. Cuvette D contained TN200 buffer but without protein. During the experiment, no buffer or acrylamide was titrated into A, the value of which was to check for instrument stability. Into B and C were titrated the same volume of buffer and acrylamide, respectively. Aliquots of acrylamide were titrated to cuvette D. As a consequence of all the effects on the observed fluorescence intensity, the actual obtained fluorescence intensity fraction can be described as

$$F = \left(\frac{F_C}{F_B}\right) - F_D \quad (2)$$

where F_B, F_C, and F_D are the observed signal fractions of protein fluorescence in cuvettes B, C, and D, respectively. To increase the precision, slit widths (both ϵ_x and ϵ_m) were adjusted to make the initial values of F_B and F_C close to 990. For every point, the fluorescence intensity was read at least 10 times to make sure that the result is not located at an extreme deviation region. Quenching data were plotted using the Stern−Volmer equation (*11*),

$$F_0/F = (1 + K_{sv}[Q])(1 + V[Q]) \quad (3)$$

where F_0/F is the fractional decrease in fluorescence due to the quencher ([Q]), and K_{sv} and V are collisional and static quenching constants, respectively.

Size-Exclusion Chromatography. The experiments were carried out using a Phenomenex Biosep SEC-S2000 HPLC gel filtration column, 2000 mm × 7.8 mm, with 25 °C water bath. Protein in different concentrations of GdnHCl was prepared and incubated overnight at 4 °C. Protein samples (~30 μM) were warmed to room temperature before using. The steel-jacketed column was operated with mechanical injection within a fully automated BioCad SPRINT HPLC system. High concentrations of GdnHCl could affect the gel matrix and perturb the SEC chromatogram protein (*12*). The elution volumes of protein sample, Blue Dextran, and sodium azide at each GdnHCl concentration were obtained, and the partition coefficient (K_d) was calculated with eq 4 (*13*):

$$K_d = (V_{protein} - V_{BlueDextran})(V_{azide} - V_{BlueDextran}) \quad (4)$$

where $V_{protein}$, $V_{BlueDextran}$, and V_{azide} represent the elution volumes of protein sample, Blue Dextran, and azide, respectively. Before each experiment, the column was calibrated using proteins with Stokes radii taken from the literature (*14*).

H/D Exchange Measurement. One hundred microliters of protein in TN200 buffer was lyophilized for 3 h at room temperature. Samples for exchange experiments were prepared by dissolving 100 μL of lyophilized protein or buffer solution in 100 μL of D_2O. The reconstituted sample was injected into a CaF_2 window cell with a path length of 50 μm. One minute after the addition of D_2O, single-beam spectra were recorded using kinetic scanning mode. FT-IR spectra were recorded at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15, 20, 30, 40, 50, 60, 90, 120, 150, and 180 min in D_2O. Eight scans were collected for each time interval from 1 to 10 min, while 64 scans were collected for each time interval from 11 to 90 min, and 128 scans were collected for each longer time interval. To compare the FT-IR spectra in H_2O and D_2O, we normalized the amide I band in H_2O to the amide I band in D_2O at 1 min. The spectrum collected after 24 h exchange was used as the spectrum of the fully deuterated protein.

Calculation of Amide Proton Exchange Rate. We monitored the H/D exchange of protein by following apparent intensity changes of the amide II band, located around 1550 cm^{-1}, which is attributed to a combination of N−H in-plane bending and C−N stretching vibrations in the peptide bond,

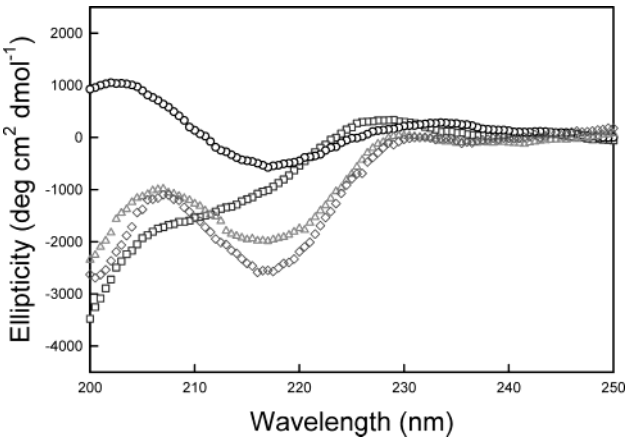

FIGURE 2: CD spectra of the WN-rED3 (○), DEN-rED3 (□), OHF-rED3 (△), and LGT-rED3 (◇) in TN200 buffer. Protein concentration is ~35 μM.

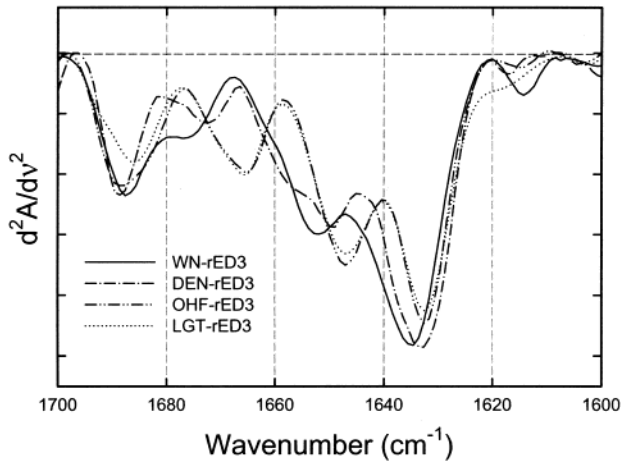

FIGURE 3: FT-IR secondary-derivative spectra of WN-rED3 (—), DEN-rED3 (— · —), OHF-rED3 (— ··), and LGT-rED3 (···) in H_2O.

because this band does not adversely interfere with absorption bands of H_2O, HOD, or D_2O (*7*). As N−H in protein is exchanged into N−D in D_2O, the absorption peak of the N−D bending vibration at about 1450 cm^{-1} is strengthened, while the N−H absorption peak decreases. The fraction of unexchanged amide proton, *F*, was calculated at various time intervals using eq 5,

$$F = (A_{II} - A_{II\infty})/(A_I\omega) \quad (5)$$

where A_I and A_{II} are the absorbance maxima of the amide I and II bands, respectively. $A_{II\infty}$ is the amide II absorbance maximum of fully deuterated rED3, and ω is the ratio of A_{II0}/A_{I0}, A_{II0} and A_{I0} being the respective absorbance maxima for the amide II and amide I bands of rED3 in H_2O (*15*, *16*).

The exchange kinetic parameters were fitted from eq 6:

$$F = A_1 e^{-k_1 t} + A_2 e^{-k_2 t} + C \quad (6)$$

where *F* is amide proton fraction at given time *t* and k_1 and k_2 are the intermediate and slow exchange rates, respectively. A_1, A_2, and *C* are the constants. F_0 is the remaining amide proton fraction at 1 min.

RESULTS

Protein Structure

Circular Dichroism Spectroscopy. The secondary structure of rED3 proteins was monitored by circular dichroism (CD) spectroscopy, as shown in Figure 2. There is a small positive ellipticity at 232 nm, and a strong negative ellipticity at 218 nm, indicating the presence of a significant amount of β-strand structure. The spectra of tick-borne OHF-rED3 and LGT-rED3 are very similar and show a negative ellipticity around 200 nm, whereas mosquito-borne WN-rED3 shows much weaker ellipticity intensity and a positive ellipticity at 205 nm, indicating clearly the presence of β-turn structures. Mosquito-borne DEN-rED3 shows a structure between those of WN-rED3 and OHF-rED3 (or LGT-rED3).

There is no observable difference in the CD spectra between native and refolded protein after an exposure to 4 M GdnHCl, indicating that the unfolding reaction is reversible for all these rED3 proteins (data not shown). Hence, the thermodynamic parameters derived from this study are not compromised by complications due to detectable irreversible processes.

FT-IR Spectroscopy. The secondary structure of rED3 proteins was also monitored by FT-IR, which has an advantage over CD in that the spectral properties of the secondary structural components can be resolved clearly in FT-IR, thus, enabling a more accurate interpretation of the data. The second-derivative FT-IR amide I spectra of rED3 in TN200 buffer are shown in Figure 3. It is most interesting to note that, similar to the CD spectra, the FT-IR spectra of the tick-bone OHF-rED3 and LGT-rED3 are essentially identical whereas the mosquito-borne rED3s exhibit different spectral properties, for example, the peaks of both β-strands and random coil shift to a higher wavenumber. Since the vibrational energy of the amide C═O stretching is inversely related to the strength of hydrogen bonding, these differences imply a difference in the overall strength in hydrogen bonding within the secondary structure elements, reflecting a different microenvironment of secondary structure components in these proteins. These results indicate that the solution structures of the tick-borne proteins are very similar, if not identical, but the mosquito-borne proteins have structural features that are distinguishable from the tick-borne proteins and each other. These data are supported by the moderate variability in the primary amino acid sequence between the mosquito- and tick-borne flaviviruses and the high degree of similarity among the tick-borne flavivirus proteins (Figure 1).

The second derivative spectral component can be assigned to the secondary structure component (*6*, *7*). Quantitative analysis of secondary structure of these proteins by curve fitting revealed that the proteins are mainly composed of β-strands, consistent with crystallographic (*2*, *3*) and CD data (Figure 2). The secondary structure content of these proteins is listed in Table 1. The total percent of β-strands for WN-rED3, DEN-rED3, OHF-rED3, and LGT-rED3 is 60%, 57%, 56%, and 56%, respectively. Although the total percent of β-strands is essentially identical among all the rED3 studied, the microenvironments surrounding these strands are different between the mosquito- and tick-borne groups of proteins. For example, the peaks at frequency 1693−1686 and 1642−1628 are both assigned to β-strands. In the mosquito-borne species, the percentage of β-strand associated with each of

Table 1: Secondary Structure Content of Envelope Protein Domain 3 Determined by FT-IR Spectra

frequency (cm^{-1})	percent (%) WN-rED3	DEN-rED3	OHF-rED3	LGT-rED3	assignment
1693−1686	13.9	10.4	18.5	17.7	β-strand[a]
1678−1671	9.3	9.5	7.2	7.0	β-turn
1663−1659	2.5	4.4	5.3	4.9	3_{10} helix
1645−1650	26.2	25.3	29.3	28.9	random coil
1642−1628	45.7	46.6	37.7	38.1	β-strand
1620−1616	2.4	3.8	2.0	3.4	side chain

[a] The total β strand component is 60%, 57%, 56%, and 56% for WN-rD3, DEN-rD3, OHF-rD3, and LGT-rD3, respectively.

Table 2: Summary of Sedimentation Data

protein	$\bar{S}_{20,w}$	MW (kDa) in GdnHCl 0 M	1.0 M	2.0 M	3.0 M
WN-rED3	1.32 ± 0.02	11.1 ± 1.4	11.9 ± 1.9	10.9 ± 1.2	11.2 ± 1.4
DEN-rED3	1.33 ± 0.02	10.9 ± 2.3	10.3 ± 2.8	11.7 ± 1.1	11.1 ± 3.5
OHF-rED3	1.34 ± 0.02	11.0 ± 1.2	9.6 ± 0.9	9.6 ± 0.9	9.4 ± 0.9
LGT-rED3	1.33 ± 0.02	10.9 ± 1.2	9.3 ± 0.5	8.8 ± 1.8	8.8 ± 1.1

these peaks is 10−14% and 46−47%, respectively. However, the corresponding values for the tick-borne proteins are 18−19% and 38%, respectively.

Hydrodynamic Properties. The hydrodynamic properties of these proteins were monitored by both SEC and sedimentation velocity analysis. The SEC data of the native proteins show K_d values of 0.545 ± 0.006, 0.559 ± 0.005, 0.572 ± 0.001, and 0.575 ± 0.008 for WN-, DEN-, OHF-, and LGT-rED3, respectively. These K_d values correspond to 20.3 ± 0.1, 20.0 ± 0.1, 19.7 ± 0.0, and 19.7 ± 0.2 in Stokes radius, R_s (Å), for WN-, DEN-, OHF-, and LGT-rED3, respectively. These results indicate that the hydrodynamic shape factor of these proteins is very similar. These proteins were also subjected to sedimentation velocity analysis. The weight-average sedimentation coefficient ($\bar{S}_{20,w}$) was determined by fitting the sedimentation velocity data to a single species with the SedFit Program. The results show that the values for $\bar{S}_{20,w}$ were 1.32, 1.33, 1.34, and 1.33 for WN-rED3, DEN-rED3, OHF-rED3, and LGT-rED3, respectively, as summarized in Table 2. There are no observable differences in $\bar{S}_{20,w}$ for tick-borne and mosquito-borne rED3, indicating that these proteins are of similar hydrodynamic shape. These data are consistent with crystallographic data or amino acid sequence comparison (Figure 1), which indicates a high degree of similarity among these proteins.

Protein Dynamics

Solvent Accessibility of Trp Residues (Fluorescence Acrylamide Quenching). Mosquito-borne WN-rED3 and DEN-rED3 have one tryptophan at residue 95, which is close to the C-terminus. Tick-borne OHF-rED3 and LGT-rED3 have two tryptophans, which are located at residues 17 and 95. The accessibility of tryptophan residues was quantitatively assessed using fluorescence collisional quenching by acrylamide. Results of these measurements are shown in Figure 4. The data were fitted to eq 3, and the parameters are summarized in Table 3.

In the native state, the tryptophan of mosquito-borne WN-rED3 and DEN-rED3 was more accessible than that of tick-borne OHF-rED3 and LGT-rED3, as indicated by a steeper slope in Stern−Volmer plots and the higher values for the collisional quenching constant, K_{sv}, and static quenching constant, V, as listed in Table 3. Since there is one tryptophan located close to the C-terminus in all rED3s examined, these results imply that at least this region of the tick-borne OHF-rED3 and LGT-rED3 is less accessible to solvent. The interpretation of these results is consistent with the observation that the λ_{max} of emission is 332 and 340 nm for OHF-rED3 and WN-rED3, respectively. Conversely, in the presence of 2 M GdnHCl, in which the proteins are in a partially unfolding state, the tryptophan of tick-borne OHF-rED3, LGT-rED3, and mosquito-borne DEN-rED3 is more accessible than that in the native state. All of the quenching constants are larger than those in native state. Since the difference between the tick-borne rED3 and mosquito-borne rED3 is the presence of an extra tryptophan near the N-terminus (position 17) of tick-borne rED3, this result implies that the tryptophan located at amino acid residue 17 in the tick-borne rED3 is buried in the native state but is made solvent accessible upon unfolding. These data are supported by the crystal structures for the intact E protein of both DEN2V and CEEV (*2*, *3*).

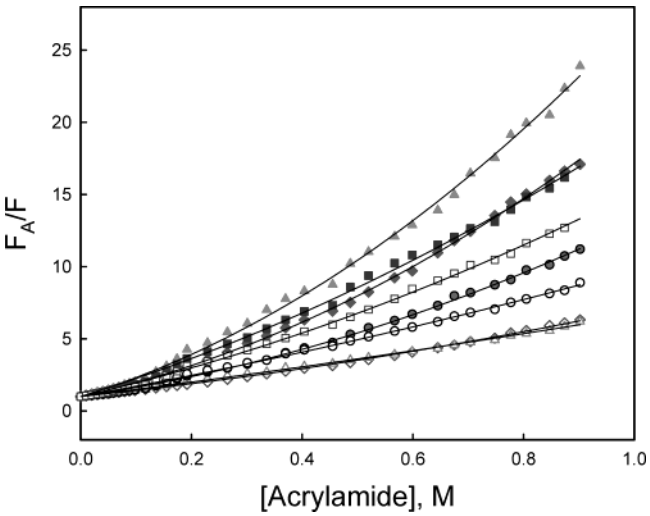

FIGURE 4: Acrylamide quenching of tryptophan fluorescence of proteins in TN200 buffer at 23 °C. Open symbols represent proteins in native state, whereas closed symbols represent proteins in partially unfolding state (DEN-rED3 (□), OHF-rED3 (△), and LGT-rED3 (◇) in 2 M GdnHCl; WN-rED3 (○) in 1 M GdnHCl). The solid curves represent the result of the best fit.

Table 3: Summary of Fluorescence Data

protein	native state K_{sv} (M^{-1})	V (M^{-1})	partially unfolded K_{sv} (M^{-1})	V (M^{-1})
WN-rED3	6.7 ± 0.2	0.27 ± 0.04	4.1 ± 0.3	1.6 ± 0.2
DEN-rED3	8.2 ± 0.3	0.64 ± 0.06	8.3 ± 0.5	1.5 ± 0.1
OHF-rED3	4.7 ± 0.2	0.16 ± 0.04	9.3 ± 0.6	1.5 ± 0.1
LGT-rED3	4.2 ± 0.1	0.22 ± 0.04	7.1 ± 0.3	1.5 ± 0.1

H/D Exchange Measurement. To explore and compare protein dynamics of mosquito-borne and tick-borne rED3, H/D exchange of these proteins was monitored by FT-IR. While the spectra in H_2O exhibited amide I and II band maxima at 1652 and 1552 cm^{-1}, respectively, the spectra of the proteins in D_2O showed a time-dependent isotopic shift of the amide II band from 1548 to 1452 cm^{-1}. This effect is indicative of NH to ND exchange of peptide backbone

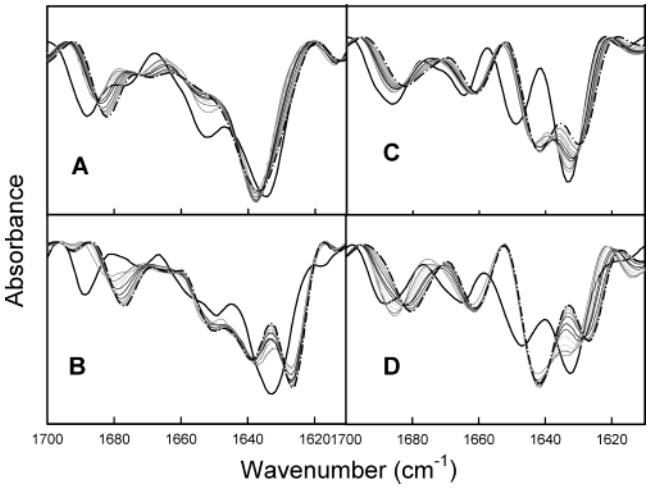

FIGURE 5: WN-rED3 (A), DEN-rED3 (B), OHF-rED3 (C), and LGT-rED3 (D) secondary-derivative amide I spectra change induced by H/D exchange. The spectra of proteins in H_2O and in D_2O after 24 h of exchange, drawn in solid and dashed lines, respectively, were included for comparison.

groups, which causes a downshift of approximately 100 cm^{-1} in the vibrational frequency of the amide II band (*18*).

Figure 5 shows the overlay of the representative second-derivative spectra recorded at 1, 2, 5, 10, 20, 30, 60, 90, 120, 150, 180, and 240 min in D_2O with spectra of the proteins in H_2O and fully deuterated in D_2O plotted as reference. The most interesting and revealing result is the similarity between the patterns of exchange in the tick-borne proteins, as shown in Figure 5C,D. The shifts in the peaks around 1700−1660 and 1640−1620 cm^{-1} as a function of time are very similar. In contrast, the time-dependent pattern of shift in peaks in the WN-rED3, as shown in Figure 5A, is quite different, whereas the pattern for DEN-rED3 is almost a composite of the three, as shown in Figure 5B. These results are consistent with all the spectroscopic data indicating that there are subtle differences in the structures of these proteins. Since the vibrational energy of the amide I is from C═O stretching, the effect of H/D exchange on amide I second-derivative spectra is indirect. Specifically, the H/D exchange affects the second-derivative spectra via changing the strength of hydrogen bonding. All β-turns shift to lower wavenumbers. For tick-borne OHF-rED3 and LGT-rED3, the peak at 1632 cm^{-1} for the β-strands band was upshifted by 4−5 cm^{-1} to 1627 cm^{-1} and was redistributed into two different energy bands (1640 and 1628 cm^{-1} peak). The peak of the β-strand of the mosquito-borne DEN-rED3 was upshifted by 7 cm^{-1} to 1627 cm^{-1}. Unlike the DEN-rED3 and tick-borne envelope proteins, the major 1635 cm^{-1} β-strand band of mosquito-borne WN-rED3 was downshifted to 1637 cm^{-1}, and there was no redistribution into different energy bands.

The overall H/D exchange rates were estimated by plotting the fraction of unexchanged amide protons as a function of time, shown in Figure 6. Essentially, all amide protons in proteins can be divided into three classes (*19*−*21*): (1) fast-exchange protons, which are most likely located on the surface of protein or in regions that are easily solvent-accessible, (2) amide protons with intermediate rates located in flexible buried regions, and (3) the slow-exchange fraction located in the core region of the protein. The fraction of the unexchanged amide protons at the first exchange time point

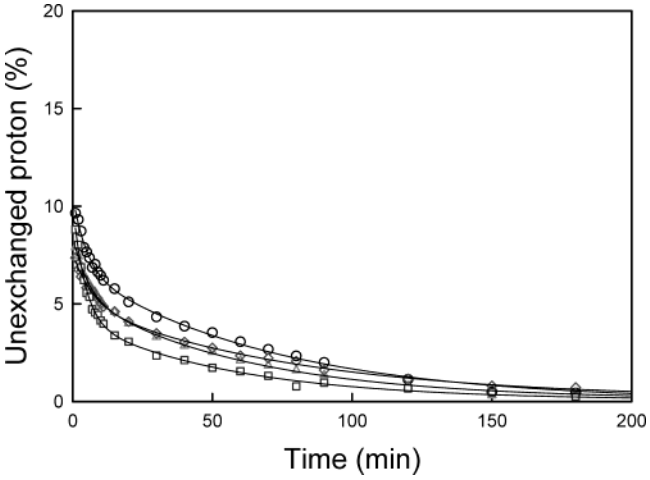

FIGURE 6: Amide proton exchange rates of WN-rED3 (○), DEN-rED3 (□), OHF-rED3 (△), and LGT-rED3 (◇). Fraction of unexchanged amide protons as a function of exposure time in D_2O. The lines represent best fits to the data points using a two-exponential function as described in eq 6.

Table 4: Fitted Exchange Parameters for rED3 Proteins

parameters[a]	WN-rED3	DEN-rED3	OHF-rED3	LGT-rED3
A_1	3.7 ± 0.2	6.0 ± 0.2	3.0 ± 0.2	2.6 ± 0.2
A_2	6.8 ± 0.2	4.1 ± 0.2	5.4 ± 0.2	4.9 ± 0.2
k_1 (min^{-1})	0.21 ± 0.03	0.24 ± 0.02	0.19 ± 0.02	0.15 ± 0.02
k_2 (min^{-1})	0.01 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.01 ± 0.00
C	0.09 ± 0.13	0.09 ± 0.08	0.10 ± 0.07	0. 13 ± 0.10
F_0	0.096	0.088	0.079	0.074

[a] Parameters were derived from fitting the exchange data in Figure 10 to a two-exponential model as described by eq 6; k_1 and k_2 are the intermediate and slow exchange rate, respectively; A_1, A_2 and C are the constants, while F_0 is the remaining amide proton fraction after 1 min of D_2O exposure.

(1 min) for all studied proteins is less than 10%, suggesting that the majority of the amide protons exchanged so rapidly that their exchange was completed within the time interval of the acquisition of the first time point. This indicates that all of these proteins are very dynamic. Therefore, only the intermediate and slow-exchange protons can be practically monitored semiquantitatively over the time range employed in this study. A two-exponential decay model (eq 6) was used to describe the exchange reaction of the remaining amide protons within the experimental time frame and the resolved parameters are summarized in Table 4. Due to the complexity of the overall H/D exchange reaction and technical limitations, no attempt was made to quantitatively associate these parameters with any actual physical properties or compare the dynamics of these proteins.

Protein Stability

Protein stability studies can provide information on the folding stability of protein molecules. Both CD and fluorescence were employed initially to monitor structural changes as a function of denaturant concentrations. However, due to the weak absorbance of the CD spectra the data were not very precise. An example is shown for the unfolding data of OHF-rED3 in Figure 7. Attempts were made to monitor the unfolding process by tryptophan fluorescence. Only small changes in fluorescence intensity were observed during the unfolding process in WN-rED3, as shown in

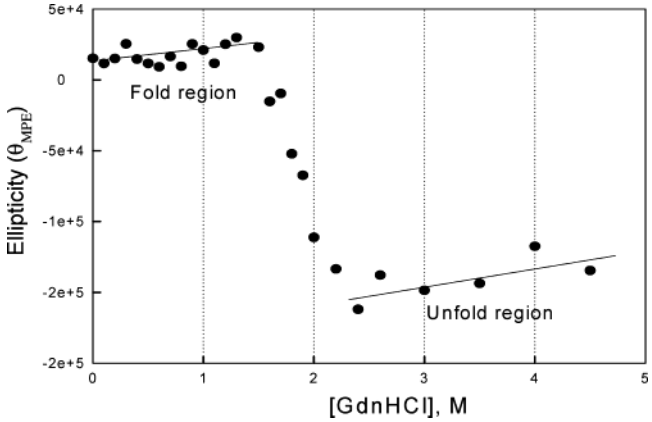

FIGURE 7: Denaturation of OHF-rED3 monitored by change of ellipticity at 222 nm.

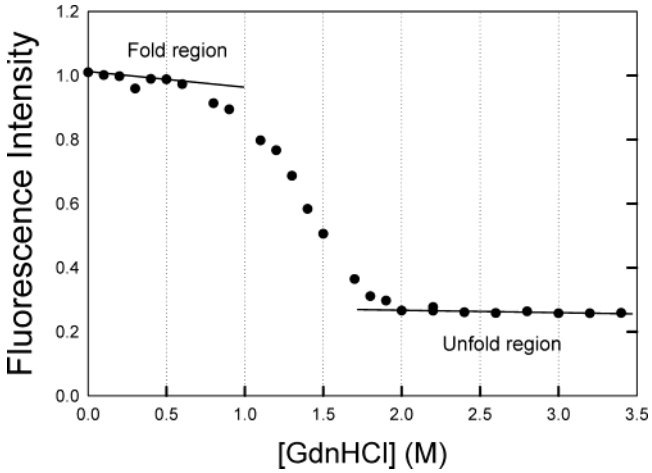

FIGURE 8: Denaturation of WN-rED3 monitored by fluorescence intensity. The excitation and emission wavelengths are 280 and 340 nm, respectively.

Figure 8, while no signal changes in other proteins. For these limitations, these spectroscopic approaches could not be employed to monitor the unfolding process of all these domain 3 proteins. Instead, size-exclusion chromatography was employed to monitor the denaturation of rED3. Data analysis methods and procedures are the same as previously published (*12*). The protein concentration was about 35 μM. The SEC column was calibrated with proteins of known Stokes radii, the values of which are cited in the literature (*14*). The partition coefficient, K_d, is derived from measurements of the elution volumes of the chromatographic peaks from the SEC column. From these measurements, a relation of $1/K_d$ to protein Stokes radii was established (data not shown) with a linear correlation between $1/K_d$ and Stokes radii covering the range of dimensions for the rED3.

Typical elution profiles of OHF-rED3 as a function of GdnHCl concentration are shown in Figure 9. The chromatographic peaks are narrower at low GdnHCl concentrations, become wider at intermediate concentrations, and eventually narrow again at higher GdnHCl concentrations. The change in the width of the elution profile is an indication of the change in the distribution of species, for example, native and unfolded domain 3 protein. The unfolding of rED3 proteins was monitored as a function of GdnHCl concentration, and results are shown in Figure 10A. In all cases, there is a gradual change in $1/K_d$ at low GdnHCl concentrations. This is followed by the transition region in which significant

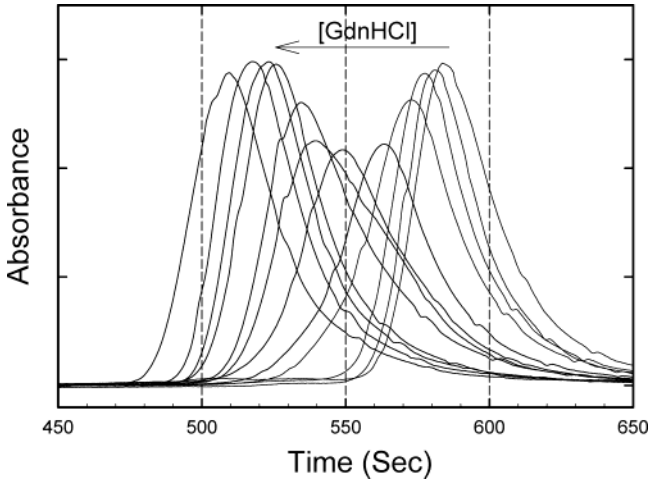

FIGURE 9: Elution profiles of OHF-rED3 in the presence of different GdnHCl concentrations: 0.0 (the right-most peak), 0.4, 0.8, 1.2, 1.6, 1.8, 2.0, 2.2, 2.6, 3.5, 4.0, and 4.5 M GdnHCl (the left-most peak).

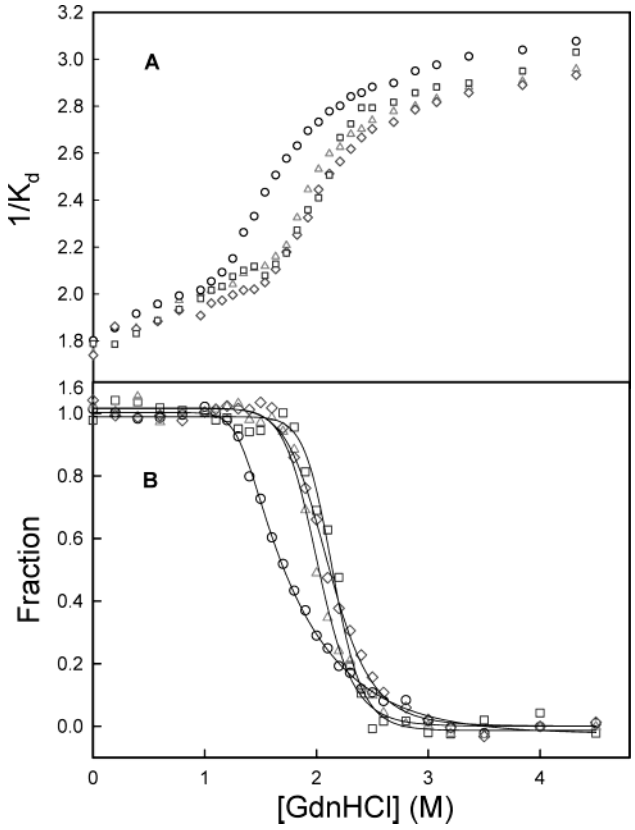

FIGURE 10: GdnHCl-induced equilibrium denaturation of WN-rED3 (○), DEN-rED3 (■), OHF-rED3 (△), and LGT-rED3 (◆) in TN200 buffer monitored by size-exclusion chromatography (SEC).

changes in $1/K_d$ were observed. In the posttransition region, there is an increase in $1/K_d$ without reaching a plateau value. The appearance of nonzero slopes in both the pre- and posttransitional regions indicates that these proteins were continuously expanding in these conditions while unfolded cooperatively in the transition region. The data were further analyzed and expressed as fraction of folded structure as a function of GdnHCl concentration as shown in Figure 10B. The solid lines are the nonlinear least-squares fits of the data to a two-state model using the previously published method (*17*). $\Delta G°$ of the unfolding reaction was evaluated by the

Table 5: Unfolding Data

	protein			
	WN-rED3	DEN-rED3	OHF-rED3	LGT-rED3
$\Delta G°$ (kcal/mol)	3.4 ± 0.4;[a] 3.7[b]	7.8 ± 0.4	7.0 ± 0.4;[a] 7.2[c]	6.6 ± 0.3

[a] Results derived from SEC data. [b] Results derived from fluorescence data. [c] Results derived from CD data.

linear extrapolation method. The $\Delta G°$'s for WN-rED3, DEN-rED3, OHF-rED3, and LGT-rED3 are 3.4 ± 0.4, 7.8 ± 0.4, 7.0 ± 0.4, and 6.6 ± 0.3 kcal/mol, respectively, as summarized in Table 5. WN-rED3 is significantly less stable than DEN-rED3 and tick-borne OHF-rED3 or LGT-rED3 proteins. These values of $\Delta G°$ are reflective of the stability of the monomeric protein since the molecular weights of these proteins are consistent with that of a monomeric domain 3 in the presence of various concentrations of GdnHCl as monitored by sedimentation equilibrium, as summarized in Table 2.

DISCUSSION

One of the intriguing issues of flaviviruses is the similarity of sequences of gene products and the very significant differences in the specific vectors that transmit these viruses and the diseases related to them. Similarity in amino acid sequence is expected to lead to structural similarity of gene products. Thus, it is not surprising that the structures of the envelope protein are essentially identical (*2*, *3*). In addition to X-ray crystallography structures for the DEN2V and CEEV whole E protein extramembrane region, the structure of the domain 3 of the JEV envelope protein has been determined recently by NMR (*22*). The structure of the domain 3 of JEV is very similar to the structure of the domain 3 of DEN2V, even though the latter structure was determined as part of the whole envelope protein. The root-mean-square deviation of backbone atoms involved in β-sheets, which are the core of the protein, is 1.65 Å. Thus, one might conclude that the core of the domain 3 is very similar as either an isolated domain or part of the whole molecule. The observations and conclusions derived from these solution properties most likely reflect the behavior of these domains.

All of these domain 3 protein structures consist of a barrel of at least six β-sheets and β-turns connecting the sheets. The points of contact of a neutralizing monoclonal antibody have been mapped to these turns in JEV. While these high-resolution structures provide important information to investigate some of their biological properties, there is still limited information on the biophysical properties of these viral proteins and their correlations to biological function. The important result of this study is that while the composition of the secondary structures of these rED3 proteins is essentially identical and the hydrodynamic property is identical, there are distinct differences in the solution properties, so much so that the WN-rED3 is $\sim$3−4 kcal/mol less stable than the other three rED3s examined. The stability data correlate with biological properties that indicate that tick-borne flaviviruses are more tolerant of environmental conditions than are those transmitted by mosquitoes. These properties include relative resistance to acidic pH and biological properties of aerosol transmission of tick-borne flaviviruses and infection via contact with infected tissues or milk from infected goats or sheep (*23*). This contrasts sharply with mosquito-borne flaviviruses that do not tolerate acidic pH and are not thought to survive outside either arthropod or animal hosts.

Although WN-rED3, DEN-rED3, OHF-rED3, and LGT-rED3 are mainly composed of β-strands ($\sim$60%, Table 1), the solution microenvironments of these strands are different, depending on the virus from which the proteins were isolated. The CD spectra of the tick-borne OHF-rED3 and LGT-rED3 are essentially identical, with the exception of the magnitude of ellipticity at $\sim$218 nm. The same conclusion is supported by the FT-IR data (Figure 3) in that the structures of OHF-rED3 and LGT-rED3 are essentially identical, although a small but significant difference can be observed for the peak at $\sim$1690 cm^{-1}, which is assigned to β-strand. This is consistent with the two proteins only differing by 11 of 95 amino acids. The spectroscopic data of the tick-borne proteins are easily distinguishable from that of WN-rED3 (Figures 2 and 3), while the spectrum for DEN-rED3 has the appearance of a cross between the spectra of WN-rED3 and the tick-borne proteins. Thus, all spectroscopic data are consistent and show that the microenvironments surrounding these β-strands are different, depending on the source of these proteins.

The similarity and differences in spectral properties are extended to the protein dynamics as revealed by the H/D exchange data monitored by FT-IR. Focusing on the exchange pattern around 1640−1620 cm^{-1}, there is significant similarity between the data for the two tick-borne proteins (Figure 5C,D), which are quite distinguishable from that of WN-rED3 (Figure 5A). Interestingly, the data for DEN-rED3 are similar to the tick-borne proteins in this region but more resemble that of WN-rED3 around 1680 cm^{-1}. Thus, a consistent picture emerges indicating the presence of subtle differences in the structural dynamics among these proteins.

The FT-IR data cannot provide specific information to correlate the dynamic aspects with the specific secondary structural element. However, the fluorescence quenching data can reveal the accessibility of the tryptophan residues, and thus indirectly, the dynamic nature of the microenvironments surrounding the tryptophan residues. All rED3s investigated in this study contain a tryptophan residue located in β-sheet 6, while only the tick-borne OHF- and LGT-rED3 contain a second tryptophan residue located in β-sheet 1, which is within the structural region constrained by the only disulfide bond. Based on the acrylamide quenching data, neither tryptophan residue in the tick-borne domains is accessible to solvent in their native states. The values of K_{SV} and V imply that either both tryptophan residues are more inaccessible than those in β-sheet 6 of DEN- and WN-rED3 or the one in β-sheet 1 is more buried and not very accessible to solvent. Thus, these data may reflect on the differential dynamic nature of different parts of the molecule, a conclusion that is consistent with the other spectroscopic data.

Evidently these subtle but detectable differences in structure can be translated into differences in structural stability. Such information cannot be gleaned from high-resolution data alone. Solution and high-resolution data complement each other to reveal a complete understanding in the design of this structural domain in carrying out specific functions. Additional information on the protein dynamics is needed to identify the location of these dynamic structural elements

and their functional roles with the ultimate goal of designing small molecules that bind to these regions in an analogous manner reported recently in design of a specific inhibitor against thymidylate synthase (*24*). Results of this study serve as the foundation for further studies with the whole protein to investigate the interaction between various domains in the envelope proteins from various sources.

ACKNOWLEDGMENT

We thank Wayne Bolen for the use of the BioCad SPRINT gel filtration system.

REFERENCES

1. Heinz, F. X., and Allison, S. L. (2001) The machinery for flavivirus fusion with host cell membranes, *Curr. Opin. Microbiol. 4*, 450−455.
2. Rey, F. A., Heinz, F. X., Mandl, C. W., Kunz, C., and Harrison, S. C. (1995) The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution, *Nature 375*, 291−298.
3. Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein, *Proc. Natl. Acad. Sci. U.S.A. 100*, 6986−6991.
4. Hung, J. J., Hsieh, M. T., Young, M. J., Kao, C. L., King, C. C., and Chang, W. (2004) An external loop region of domain III of dengue virus type 2 envelope protein is involved in serotype-specific binding to mosquito but not mammalian cells, *J. Virol. 78*, 378−388.
5. Dong, A., Malecki, J. M., Lee, L., Carpenter, J. F., and Lee, J. C. (2002) Ligand-induced conformational and structural dynamics changes in *Escherichia coli* cyclic AMP, *Biochemistry 41*, 6660−6667.
6. Dong, A., and Caughey, W. S. (1994) Infrared methods for study of hemoglobin reactions and structures, *Methods Enzymol. 232*, 139−175.
7. Susi, H., and Byler, D. M. (1986) Resolution-enhanced Fourier transform infrared spectroscopy of enzymes, *Methods Enzymol. 130*, 290−311.
8. Dong, A., Huang, P., and Caughey, W. S. (1990) Protein secondary structures in water from second-derivative amide I infrared spectra, *Biochemistry 29*, 3303−3308.
9. Cohn, E. J., and Edsall, J. T. (1943) *Protein, Amino Acids and Peptides*, p 372, Van Nostrand-Reinhold, Princeton, NJ.
10. Lin, S.-H., and Lee, J. C. (2002) Communications between the high-affinity cyclic nucleotide binding sites in *E. coli* cyclic AMP receptor protein: effect of single site mutations, *Biochemistry 41*, 11857−11867.
11. Chen, R., and Lee, J. C. (2003) Functional roles of Loops 3 and 4 in the cyclic nucleotide binding domain of cyclic AMP receptor protein from *Escherichia coli*, *J. Biol. Chem. 278*, 13235−13243.
12. Baskakov, I. V., and Bolen, D. W. (1998) Monitoring the sizes of denatured ensembles of staphylococcal nuclease proteins: implications regarding *m* values, intermediates, and thermodynamics, *Biochemistry 37*, 18010−18017.
13. Yang, M., Ferreon, A. C. M., and Bolen, D. W. (2000) Structural thermodynamics of a random coil protein in guanidine hydrochloride, *Protein: Struct., Funct., Genet.* (Suppl. 4), 44−49.
14. Corbett, R. J. T., and Roche, R. S. (1984) Use of high-speed size-exclusion chromatography for the study of protein folding and stability, *Biochemistry 23*, 1888−1894.
15. Barksdale, A. D., and Rosenberg, A. (1982) Acquisition and interpretation of hydrogen exchange data from peptides, polymers, and proteins, *Methods Biochem. Anal. 28*, 1−113.
16. Yu, S., Lee, L. L.-Y., and Lee, J. C. (2003) Effects of metabolites on the structural dynamics of rabbit muscle pyruvate kinase, *Biophys. Chem. 103*, 1−11.
17. Cheng, X., Gonzalez, M. L., and Lee, J. C. (1993) Energetics of intersubunit and intrasubunit interactions of *Escherichia coli* adenosine cyclic 3′,5′-Phosphate receptor protein, *Biochemistry 32*, 8130−8139.
18. Blout, E. R., de Lozé, C., and Asadourian, A. (1961) The deuterium exchange of water-soluble polypeptides and proteins as measured by infrared spectroscopy, *J. Am. Chem. Soc. 83*, 1895−1900.
19. Kim, K. S., Fuchs, J. A., and Woodward, C. K. (1993) Hydrogen exchange identifies native-state motional domains important in protein folding, *Biochemistry 32*, 9600−9608.
20. de Jongh, H. H., Goormaghtigh, E., and Ruysschaert, J. M. (1995) Tertiary stability of native and methionine-80 modified cytochrome *c* detected by proton-deuterium exchange using online Fourier transform infrared spectroscopy, *Biochemistry 34*, 172−179.
21. Li, J., Cheng, X., and Lee, J. C. (2002) Structure and dynamics of the modular halves of *Escherichia coli* cyclic AMP receptor protein, *Biochemistry 41*, 14771−14778.
22. Wu, K. P., Wu, C. W., Tsao, Y. P., Kuo, T. W., Lou, Y. C., Lin, C. W., Wu, S. C., and Cheng, J. W. (2003) Structural basis of a flavivirus recognized by its neutralizing antibody: solution structure of the domain III of the Japanese encephalitis virus envelope protein, *J. Biol. Chem. 278*, 46007−46013.
23. Burke, T. S., and Monath, T. P. (2001) Flaviviruses, in *Field's Virology* (Knipe, D. M., and Howley, P. M., Eds.) pp 1043−1125, Lippincott Williams & Wilkens, Philadelphia.
24. Ferrari, S., Costi, P. M., and Wade, R. C. (2003) Inhibitor specificity via protein dynamics: insights from the design of antibacterial agents targeted against thymidylate synthase, *Chem. Biol. 10*, 1183−1193.

BI049324G