Total Chemical Synthesis of the Integral Membrane Protein Influenza A Virus M2: Role of Its C-Terminal Domain in Tetramer Assembly †

ARTICLE *in* BIOCHEMISTRY · OCTOBER 1999

Impact Factor: 3.02 · DOI: 10.1021/bi990720m · Source: PubMed

Total Chemical Synthesis of the Integral Membrane Protein Influenza A Virus M2: Role of Its C-Terminal Domain in Tetramer Assembly†

Gerd G. Kochendoerfer,*,‡,§ David Salom,§,|| James D. Lear,*,|| Rosemarie Wilk-Orescan,⊥ Stephen B. H. Kent,‡ and William F. DeGrado||

Gryphon Sciences, South San Francisco, California 94080, Department of Biochemistry and Biophysics, Johnson Research Foundation, School of Medicine of the University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059, and DuPont Pharmaceuticals Co., Wilmington, Delaware 19880

Received March 29, 1999; Revised Manuscript Received July 1, 1999

ABSTRACT: The M2 protein from influenza A virus is a 97-residue homotetrameric membrane protein that functions as a proton channel. To determine the features required for the assembly of this protein into its native tetrameric state, the protein was prepared by total synthesis using native chemical ligation of unprotected peptide segments. Circular dichroism spectroscopy of synthetic M2 protein in dodecylphosphocholine (DPC) micelles indicated that approximately 40 residues were in an α-helical secondary structure. The tetramerization of the full-length protein was compared to that of a 25-residue transmembrane (TM) fragment. Analytical ultracentrifugation demonstrated that both the peptide and the full-length protein in DPC micelles existed in a monomer−tetramer equilibrium. Comparison of the association constants for the two sequences showed the free energy of tetramerization of the full-length protein was more favorable by approximately 7 kcal/mol. Partial proteolysis of DPC-solubilized M2 was used as a further probe of the structure of the full-length protein. A 15−20-residue segment C-terminal to the membrane-spanning region was found to be highly resistant to digestion by chymotrypsin and trypsin. This region, which we have modeled as an extension of the TM helices, may help to stabilize the tetrameric assembly.

The molecular mechanism by which ion channels assemble and conduct ions is not fully understood (*1*−*3*). This is largely due to the inadequacy of recombinant DNA-based expression methods in producing large quantities of integral membrane proteins for biophysical and structural studies (*4*). The chemical synthesis of proteins by native chemical ligation of unprotected peptide segments exhibits demonstrated utility for water-soluble proteins (*5*), and is rapidly becoming an effective alternative to producing proteins by standard molecular biology expression systems. However, direct chemical synthesis of membrane proteins has to date been limited to such relatively short membrane peptides as the transmembrane (TM)[1] segments of various ligand-gated channels (*6*, *7*), various small, de novo designed channel-forming peptides (*8*), and phospholamban, a 57-residue peptide involved in Ca^{2+} regulation in the sarcoplasmic reticulum (*9*, *10*). Here, we report the application of the chemical ligation approach to the total chemical synthesis of the translation product of the influenza A virus M2 open reading frame, a 97-residue channel-forming polypeptide that self-assembles to form a tetrameric integral membrane protein.

The M2 protein functions as a highly selective, pH-regulated proton channel and is the target of the anti-influenza drugs rimantadine and amantadine (*11*−*14*). The influenza virus enters cells through internalization into the endocytic pathway, with virus uncoating taking place in acidified endosomal compartments. The M2 ion channel activity permits protons to enter the virion interior, and this acidification weakens the interactions of the matrix protein (M1) with the ribonucleoprotein core (*15*). Considerable experimental evidence (*16*, *17*) indicates that the functional M2 channel is formed from a tetrameric array of parallel, membrane-spanning peptides with their N-termini oriented toward the outside of the virus. A synthetic 25-residue peptide (residues 22−46 of the A/Udorn variant) spanning the hydrophobic region of the protein forms α-helical secondary structure in lipid vesicles (*18*) and produces amantadine-sensitive ion channels when incorporated into planar bilayers (*19*). An atomic resolution structure of the M2 proton channel has yet to be determined, although infrared spectroscopic (*20*) and solid state NMR (*21*) of the TM peptides indicates that the TM helices are oriented with an approximately 30° tilt relative to the bilayer normal axis. Several models for the TM tetramer are in good agreement with the spectroscopic data; one model is based on Fourier analysis of an extensive set of mutations to obtain helix orientations and crossing

† This work was partially supported by NIH Grant GM56423 (W.F.D. and J.D.L.). D.S. was a recipient of a postdoctoral fellowship from the Dirección General de Enseñanza Superior [Ministerio de Educación y Cultura (Spain)].

* Corresponding authors.
‡ Gryphon Sciences.
§ These authors contributed equally to this work.
|| School of Medicine of the University of Pennsylvania.
⊥ DuPont Pharmaceuticals Co.

[1] Abbreviations: Boc, *tert*-butoxycarbonyl; CD, circular dichroism; D_2O, deuterium oxide; DIEA, *N*,*N*-diisopropylethylamine; DNP, 2,4-dinitrophenyl; DPC, dodecylphosphocholine; ES, electrospray; HBTU, 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HOBt, *N*-hydroxybenzotriazole; MALDI, matrix-assisted laser desorption/ionization; M2TM, transmembrane segment of M2 (residues 22−46); MS, mass spectrometry; PAL, tris(alkoxy)benzylamide; PAM, phenylacetomidomethyl; POPC, 1-palmitoyl-2-oleoylphosphatidylcholine; RP, reverse phase; SPPS, solid phase peptide synthesis; TCEP, tris(2-carboxyethyl)phosphine hydrochloride; TM, transmembrane; TFA, trifluoroacetic acid; TFE, trifluoroethanol.

10.1021/bi990720m CCC: $18.00
Published on Web 08/21/1999

angles (*22*), while the others are based on molecular dynamics calculations (*20*, *23*, *24*).

The similarity of two of the models (rmsd = 1.7 Å; *25*) suggests that many of the sequence-specific structural determinants for assembly of the channel lie within this region. However, it is quite possible that residues outside of the membrane also contribute to the channel's structural stability. To experimentally address these issues, significant quantities of highly pure protein are required. Unfortunately, the M2 protein tends to be toxic to cells, and could only be expressed at relatively low levels in yeast (*26*), or from a recombinant baculovirus in *Spodoptera frugiperda* (*Sf* 9) cells (*27*).

We therefore attempted to directly synthesize the full 97-residue M2 polypeptide using native chemical ligation (*4*) of unprotected peptide segments at the Cys^{50} residue found in the native sequence. A single-site mutant (C17S), previously shown to produce M2 tetramers (*28*), was chosen to reduce potential problems with oxidation in biophysical studies. With the synthetic protein, we could determine and quantitatively compare its secondary structure and oligomerization state to those of a synthetic peptide comprising TM residues 22−46.

MATERIALS AND METHODS

Peptide Synthesis. Peptide segments for ligation into full-length M2 were synthesized using a custom-modified Applied Biosystems 430A peptide synthesizer following established protocols (*29*). Peptide segments were purified by preparative gradient reverse phase (RP) HPLC on a Rainin dual-pump high-pressure mixing system with 214 nm UV detection using a Vydac C-4 preparative column (10 μm particle size, 2.2 cm × 25 cm), and analytical RP-HPLC was performed on a Vydac C-18 analytical column (5 μm particle size, 0.46 cm × 15 cm), using a Hewlett-Packard model 1100 quaternary pump high-pressure mixing system with 214 and 280 nm UV detection. Electrospray mass spectra (ES-MS) of the peptide products were obtained using a PE-Sciex API-1 quadripole ion spray mass spectrometer. Peptide masses were calculated from all the observed protonation states, and peptide mass spectra were reconstructed using MACSPEC (PE-Sciex, Thornhill, ON). Theoretical masses were calculated using MACPROMASS (T. Lee, City of Hope, Duarte, CA).

M2 segment 1 (amino acid residues 1−49) was synthesized on a thioester generating resin using the in situ neutralization protocol for Boc (*tert*-butoxycarbonyl) chemistry stepwise solid phase peptide synthesis (SPPS) (*29*), using established side chain protection strategies. M2 segment 2 (amino acid residues 50−97) was synthesized analogously on a OCH_2-PAM resin (*29*). The peptides were deprotected and simultaneously cleaved from the resin support using HF/*p*-cresol according to standard Boc chemistry procedures (*29*). The putative membrane spanning M2 segment 1 was purified by preparative RP-HPLC with a linear gradient of 55 to 75% buffer B (6:3:1 2-propanol/acetonitrile/H_2O containing 0.1% TFA) versus 0.1% aqueous TFA over the course of 45 min. The putative cytoplasmic M2 segment 2 was purified by preparative RP-HPLC with a linear gradient of 25 to 45% buffer C (100% acetonitrile containing 0.1% TFA) versus 0.1% aqueous TFA over the course of 45 min. In both cases, fractions containing pure peptide were identified using ES-MS, pooled, and lyophilized for subsequent ligation. The purified peptides were characterized by ES-MS {segment 1 thioester peptide with three protecting groups [one His(DNP) and two Trp(formyl) groups], observed MW of 5993 ± 1 Da and calcd MW of 5993.8 Da (average isotope composition); segment 2 with one protecting group [His(DNP)], observed MW of 5937 ± 1 Da and calcd MW of 5937.4 Da (average isotope composition)}.

The M2TM peptide (residues 22−46, C-terminally amidated) was synthesized on PAL resin using free amino acids (activated in situ by HBTU/HOBt/DIEA). Side chain deprotection and cleavage from the resin were carried out using a mixture of TFA/thioanisole/1,2-ethanedithiol/anisole (90:5:3:2 by volume) at room temperature under nitrogen for 2 h. The resin was filtered off, and the peptide was precipitated with cold diethyl ether. Purification was achieved by RP-HPLC on a preparative C4 column using linear gradients of 55 to 75% buffer B (6:3:1 2-propanol/acetonitrile/H_2O containing 0.1% TFA) versus 0.1% aqueous TFA.

Chemical Protein Synthesis. A 50% excess of the purified unprotected M2(50−97) peptide (segment 2) was added to a solution of the purified unprotected thioester peptide M2-(1−49)$^{\alpha}$COSR (segment 1) (2 mM) in 0.1 M sodium phosphate/6 M guanidinium chloride (pH 7.5) containing 30% trifluoroethanol (TFE, Aldrich) and 1% thiophenol. The ligation mixture was stirred for 20 h at room temperature, and the reaction was monitored by RP-HPLC and ES-MS. The reaction mixture was subsequently treated with an equal volume of a solution of acetonitrile/piperidine/β-mercaptoethanol (1:1:1 by volume) for 30 min to remove the Trp(formyl) protecting group and any residual His(DNP) protecting groups. Reactants and products were separated by preparative RP-HPLC with a linear gradient of 55 to 75% buffer B in water. Fractions containing full-length M2 were identified by ES-MS [amino acid residues 1−97, observed MW of 11 170 ± 1 Da and calcd MW of 11 169.7 Da (average isotope composition)], pooled, and lyophilized.

Circular Dichroism (CD). CD measurements were made in an Aviv 62A DS circular dichroism spectrometer at 25 °C. M2-containing vesicles were prepared as follows. M2 was codissolved with 1-palmitoyl-2-oleoylphosphatidylcholine (POPC, Avanti Polar Lipids) in TFE. The solvent was removed under a nitrogen stream, and the film was kept overnight under high vacuum. Then it was hydrated with 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl, and 0.5 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Pierce), vortexed vigorously, and sonicated using a bath sonicator (Laboratory Supplies, Hicksville, NY) to clarity. M2 (and M2TM)-containing dodecylphosphocholine (DPC, Avanti Polar Lipids) micelles were prepared following the same procedure, but without the sonication step. Peptide concentrations were determined by diluting the samples with an equal volume of 2-propanol and then measuring the UV absorbance. The peptide extinction coefficients were calculated from the values of the Trp and Tyr extinction coefficients in water and 2-propanol (*30*). Calculated values were as follows: for M2, $\epsilon_{280} = 14\,600\ cm^{-1}\ M^{-1}$; and for M2TM, $\epsilon_{280} = 5853\ cm^{-1}\ M^{-1}$. CD spectra were recorded

from 210 to 260 nm in a 0.1 cm quartz cell, at a scan speed of 6 nm/min and then in a 0.01 cm quartz cell from 188 to 210 nm at a scan speed of 2 nm/min. A baseline was recorded and subtracted after each spectrum.

Analytical Ultracentrifugation. Sedimentation equilibrium experiments were performed at 25 °C on peptide solubilized in DPC micelles using a Beckman XL-I analytical ultracentrifuge (*31*). To eliminate the contribution of the DPC to the buoyant molecular mass of the peptide−DPC complex, experiments were carried out at a solvent density that was adjusted to equal that of the DPC (*32*, *33*). Density adjustment was carried out by centrifugation of DPC micelles in buffer [50 mM Tris-HCl (pH 7.5), 0.1 M NaCl, and 0.1 mM TCEP] containing different percentages of deuterium oxide (D_2O). The density match is calculated from a plot of the radial gradient of fringe displacement versus buffer density. Zero gradient for DPC at 25 °C occurs for buffer containing 52.5% D_2O (solvent density $\rho = 1.05932$ as measured in a Paar densitometer). M2 (and M2TM)/DPC samples were prepared by dissolving the desired amount of peptide in TFE in a glass vial and removing the organic solvent under a nitrogen stream to create a film of material on the glass. After the film had been held overnight under high vacuum, 15 mM DPC in density-matched buffer (with 52.5% D_2O) was added, and the sample was vigorously vortexed until it became clear. In the M2TM/DPC samples, the buffer did not contain TCEP. The M2 monomer molecular mass and partial specific volume were calculated using the program Sedinterp (*34*) and corrected for hydrogen−deuterium exchange using averaged H−D-exchanged amino acid residue masses. (That is, all protons associated with amide nitrogens and side chain amine, alcohol, carboxyl, and guanidinium groups were assumed to be 52.5% deuterated.) Values calculated were as follows: for M2, 11250 Da and 0.73147 cm^3/gm; and for M2TM, 2746 Da and 0.787 cm^3/gm. These values were fixed to be constants in fitting the radial concentration profiles to various monomer−*N*-mer equilibrium association models. The samples were centrifuged for times sufficient to achieve equilibrium in three-compartment carbon−epoxy centerpieces using sapphire windows. Data obtained by both UV absorption at 280 nm and interference fringe displacement were analyzed by nonlinear least-squares curve fitting of radial concentration profiles using the Marquardt−Levenberg algorithm implemented in Igor Pro (Wavemetrics, Oswego, OR) with a user-defined function by encoding the following equation describing reversible association in centrifugation:

$$S(r) = \sum_n \frac{n\left[1.2\varepsilon C_o\right]^n \exp\{n\chi M_w\}}{K_n} + baseline$$

$S(r)$ = Signal due to all sedimenting species (n = 1,2, ...n) at radial position r

1.2ε = Path length times the monomer molar extinction coefficient for absorbance data

$3.3 * M_w$ for interference data (assumes dn/dc = 0.187)

C_o = Molar concentration at r_o of monomer of molecular weight M_w

$\chi = \frac{(1-\bar{v}\rho)\omega^2}{2RT}\left(r^2 - r_o^2\right)$; r_o = arbitrary fixed radius reference

K_n = N - mer dissociation constant in units of $M^{(n-1)}$

$\bar{v}$ = Partial specific volume of sedimenting species (cm^3/gm)

ρ = Density of supporting buffer (gms/cm^3)

ω = Angular velocity of rotor (radians/sec)

M_w = Molecular weight of sedimenting species (gms/mole)

R = Ideal Gas constant (8.315×10^7 ergs K^{-1} mol^{-1})

T = Temperature (K)

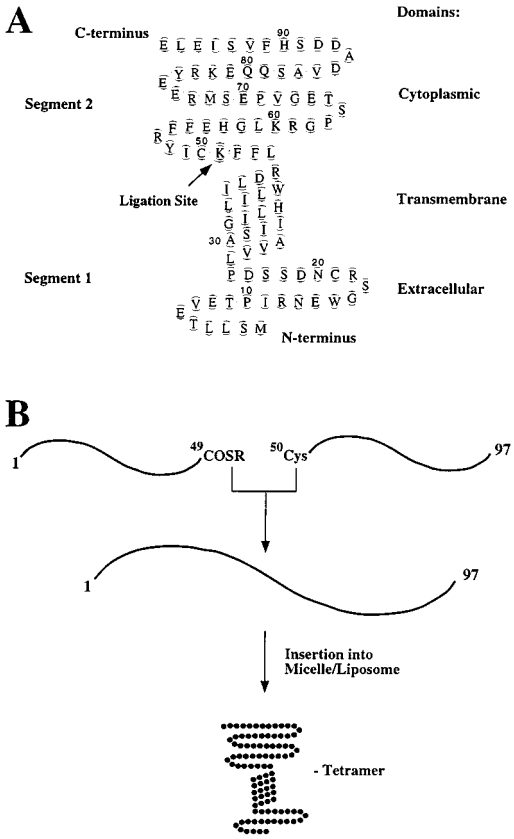

FIGURE 1: (A) Target sequence of the influenza A M2 protein, indicating schematically the predicted protein domains as well as the ligation site at Cys^{50}. (B) Synthetic strategy. The peptide segments M2(1−49)$^{\alpha}$COSR and M2(50−97) are prepared by Boc chemistry stepwise SPPS and then covalently joined to one another by native chemical ligation at a cysteine residue conveniently located near the middle of the native polypeptide sequence. The resulting full-length, 97-residue polypeptide can be inserted into micelles or liposomes for channel assembly, to form the tetrameric integral membrane M2 protein (see the text).

Three data sets were fit simultaneously with equilibrium dissociation constants as global fitting parameters. Baselines and fixed radius signal values for each data set were allowed to vary independently in fitting. Various two-state and three-state equilibria were tried to define the best fits as judged by the magnitude of the squared residuals as well as the random appearance of the residuals distribution.

Partial Proteolysis. Full-length M2 polypeptide was incorporated at a 1:100 molar ratio into 5 mM DPC micelles in 100 mM NaCl and 50 mM Tris-HCl (pH 8.0) as described above for analytical ultracentrifuge samples, and incubated for 2 h at ambient temperature with 0.22 mg/mL trypsin or 0.02 mg/mL chymotrypsin. The partial digest was analyzed by gradient RP-HPLC using 20 to 100% buffer B (6:3:1 2-propanol/acetonitrile/H_2O containing 0.1% TFA) over the course of 40 min at 50 °C on a C-4 column. The resulting major peaks were collected and analyzed by amino acid sequencing and matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS).

RESULTS

Synthetic Strategy. Figure 1 presents the sequence of the 97-residue M2 polypeptide chain, together with a schematic of its predicted topology. The chemical ligation site at residue Cys^{50} of the native sequence is highlighted. The N-terminal

segment M2(1−49) contains many hydrophobic residues and is predicted to constitute the intracellular and membrane-spanning domains of M2. The C-terminal segment M2(50−97) contains fewer hydrophobic residues and is predicted to constitute the cytoplasmic domain of M2.

The major anticipated obstacle in the chemical synthesis of membrane proteins is the difficulty in making, handling, and purifying peptides containing large regions of highly hydrophobic amino acid residues and the difficulty in solubilizing the hydrophobic segments under conditions that are suitable for the native chemical ligation reaction (*4*). Our highly optimized in situ neutralization protocols for Boc chemistry stepwise SPPS (*29*) is efficient for the synthesis of 49-residue hydrophobic membrane-spanning segment 1 of M2, typically yielding material with a crude purity of ∼67%, which could be purified to homogeneity with an overall yield of 18%. Preheating the peptide sample to 60 °C and addition of 30% TFE to the standard ligation buffer (see above) dissolves the peptide segments efficiently, and keeps them in solution during the reaction without interfering with the reaction chemistries.

A pilot study was carried out using 10 mg of pure segment 1 and 15 mg of pure segment 2. Analytical data showing the progress of the ligation reaction in forming full-length M2 polypeptide are depicted in Figure 2. After overnight reaction, the magnitude of the reactant peak at 21.7 min is significantly decreased and a new peak dominates at a retention time of 20.0 min. ES-MS analysis (see inset) correlates this new peak with the desired full-length reaction product [M2(1−97)]. Integration and comparison of the HPLC peaks of reactant and product after overnight reaction detected at 214 nm show that the reaction is ∼65% complete. Longer periods of reaction did not increase the yield under these conditions. Preparative RP-HPLC of the reaction mix yielded 5 mg of pure 97-residue M2 polypeptide.

Circular Dichroism. The CD spectra of full-length M2 incorporated into POPC vesicles (Figure 3A) exhibit evidence of significant α-helical structure as indicated by the strong positive band at 190 nm and negative bands at 208 and 222 nm. A qualitatively similar spectrum was reported (*18*) for the M2TM peptide in DOPC vesicles. A more quantitative comparison of M2 and M2TM secondary structures is afforded in the spectra recorded under the same conditions in the less scattering DPC micelle solutions (Figure 3B). Spectral shape analysis using the LINCOMB program with the "Brahms" data set (*35*) (not shown) indicates that as many as 45 residues could be helical in M2 versus 16 in M2TM. Other methods (*36*) also predicted a greater number of helical residues in the full-length protein relative to the M2TM peptide. Thus, CD spectroscopy strongly suggests that M2 has one or more helices in addition to its TM segment.

Analytical Ultracentrifugation. The M2 protein is known to be tetrameric in vivo. We therefore examined the oligomerization state and stability of the full-length M2 polypeptide and the M2TM peptide by analytical ultracentrifugation, using an approach similar to that described recently for designed membrane-interactive peptides (*37*) and the TM helix of the dimeric protein, glycophorin (*32*). Both peptides were incorporated into DPC micelles, at a DPC concentration (15 mM) considerably higher than its critical micelle concentration (approximately 1 mM). Following the method of Reynolds and Tanford (*33*), the density of the

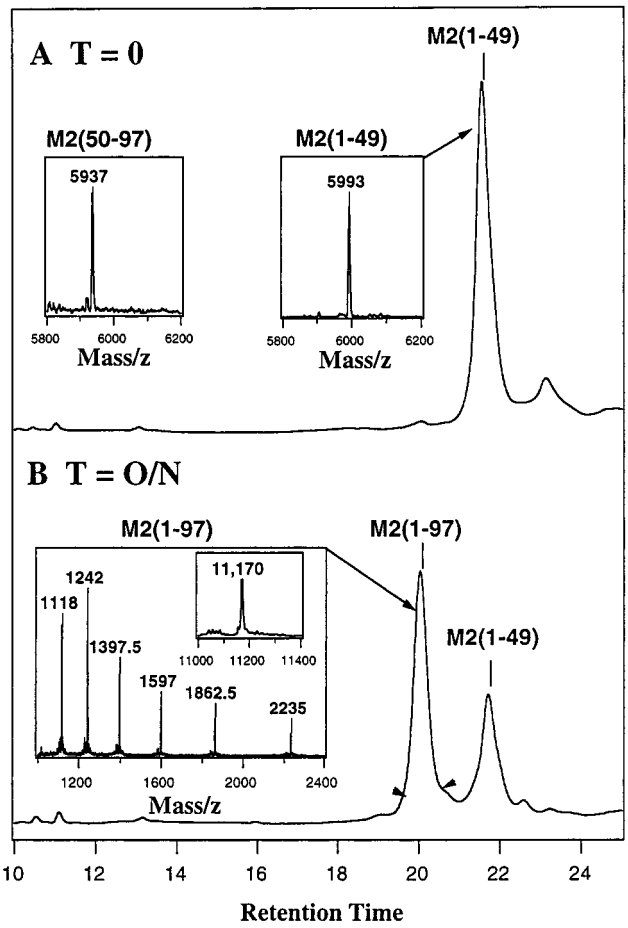

FIGURE 2: Progress of the ligation reaction. Analytical RP-HPLC traces (detected at 280 nm) monitoring the ligation reaction. Aliquots of the reaction mixture were withdrawn, treated with 20% β-mercaptoethanol for 20 min to hydrolyze remaining thioesters, acidified with formic acid, and treated briefly with a 3-fold excess of TCEP, prior to injection. Analytical RP-HPLC was performed with a linear gradient of 45 to 100% buffer B (6:3:1 2-propanol/acetonitrile/H_2O containing 0.1% TFA) versus 0.1% aqueous TFA over the course of 24 min. Panel A depicts the RP-HPLC trace after injection of an aliquot of the reaction mixture at time zero. Only one major peak with a retention time of 21.7 min is observed. The mass of the eluting peptide is identified by ES-MS analysis and correlated to the segment 1 thioester peptide [M2(1−49)]. (The more hydrophilic segment 2 is not retained at the high concentration of organic during loading and elutes in the void volume under these conditions.) The insets in panel A present ES-MS reconstructs of the starting materials demonstrating their identity and giving some indication of purity. Panel B presents the equivalent RP-HPLC trace after overnight reaction. The inset shows ES-MS spectra of the purified ligation product M2(1−97), which eluted at 20.0 min.

solvent was adjusted to that of the detergent. Under these conditions, the DPC molecules make a negligible contribution to the overall sedimentation of the peptide/DPC mixed micelles, and the buoyant molecular mass of the peptide may be directly determined. Care was taken to avoid adventitious oligomerization of the protein, which may occur if the concentration of DPC micelles is not large relative to the protein concentration (*33*). Thus, DPC/peptide ratios were maintained at a value considerably greater than the micelle number of DPC (determined to be 65 by analytical ultracentrifugation, data not shown). To test for a small amount of adventitious oligomerization, the ultracentrifugation was carried out at peptide to DPC ratios ranging from approximately 1:125 to 1:700. If adventitious association were

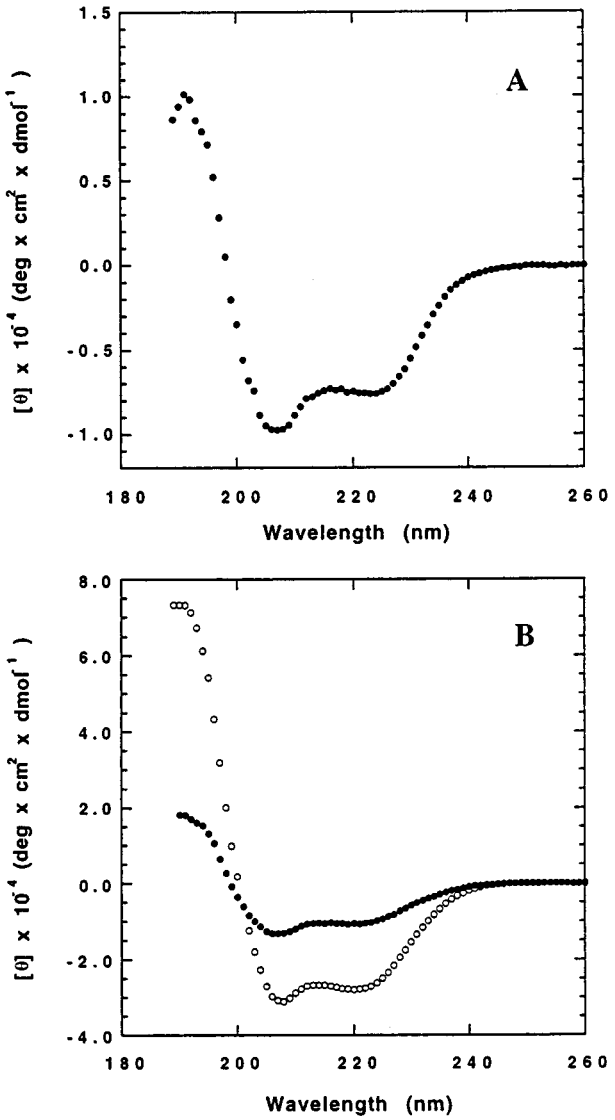

FIGURE 3: (A) CD spectrum of M2 in POPC vesicles. [M2] = 27 μM, and [POPC] = 2 mM. (B) CD spectra of M2 (●) and M2TM (residues 22−46) (○) in DPC micelles. [M2] = 135 μM, and [DPC] = 10 μM. [M2TM] = 191 μM, and [DPC] = 20 μM. The buffer in all samples was composed of 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 0.5 mM TCEP. In the case of M2-containing vesicles, the spectrum from 210 to 260 nm had to be scaled slightly to coincide at 210 nm with the second spectrum (188−210 nm) because the extent of absorption and light scattering in the 1 mm cell was higher than in the 0.1 mm cell.

a major problem, a greater apparent association constant would be observed at the highest peptide:DPC ratio.

Equilibrium radial concentration profiles at three different peptide concentrations for M2TM are well described by a cooperative monomer−tetramer equilibrium (Table 1), as demonstrated by a global nonlinear least-squares fit of the data to this association scheme. Alternatively, monomer−dimer−tetramer or monomer−trimer−tetramer equilibria may be used to fit the data, although there is no significant improvement in the quality of the fit. Thus, the association is reasonably cooperative, and is well described by a monomer−tetramer equilibrium.

At the highest peptide concentrations, there is a small positive deviation in the residuals (data not shown), suggesting the presence of higher-order aggregates. A small improvement in the fit may be realized by inclusion of these

Table 1: Analysis of Analytical Ultracentrifugation Using Different Association Schemes[a]

model		M2			M2TM		
n_1	n_2	χ^2/N	pK_1	pK_2	χ^2/N	pK_1	pK_2
2	4	2.3	6.7	20.4	1.2	6.7	17.6
3		28.7	10.57		2.5	9.06	
3	4	1.9	9.3	20.1	1.5	12.7	18.1
3	6	1.9	11.8	28.2	1.0	8.95	21.43
4		3.7	15.6		1.2	11.64	
4	6	1.6	15.8	25.4	0.9	11.62	18.8
4	8	1.6	16.6	36.8	0.9	11.69	26.5

[a] The model represents the aggregation equilibrium as a monomer−n_1-mer to −n_2-mer equilibrium. χ^2/N is the weighted sum of squared residuals divided by the number of degrees of freedom in the curve fitting. Goodness of fits are indicated by how close this number is to unity (*54*). pK_i is negative $\log_{10}$ of the corresponding n_i-mer dissociation constant in molar units. For comparison of free energies of association, these values can be converted to $\Delta G°$ values per monomer (1 M standard state) by multiplying the pK_i values by $0.6 \times 2.303/n_i$, where n_i is the number of monomers in the corresponding equilibrium.

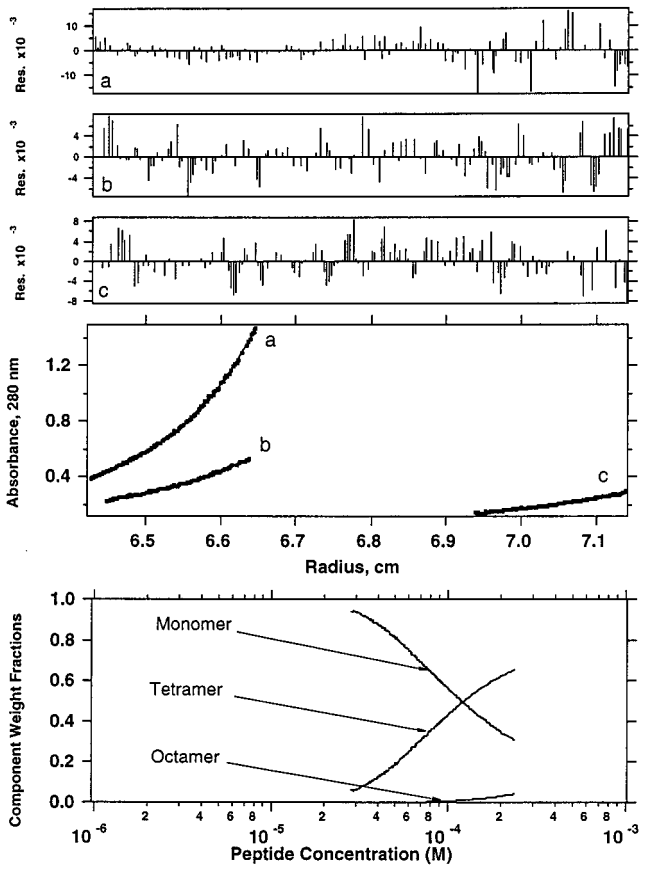

FIGURE 4: Sedimentation equilibrium of M2TM in DPC micelles at 40K rpm. Equilibrium A_{280}−radius profiles for three different cell compartments, all containing DPC concentrations of 15 mM and M2TM at peptide to DPC molar ratios of (a) 1:126, (b) 1:284, and (c) 1:554. The lines are the best (global) fit to a monomer−tetramer−octamer equilibrium model. The residuals of the fit are shown in the three upper panels. The lower panel shows the calculated relative contributions (*y*-axis) of the different species as a function of the total peptide concentration (*x*-axis) over the range observed in the combined experimental data.

in a reversible equilibrium with the monomeric and tetrameric states. Figure 4 illustrates the fitting of the M2TM data to a monomer−tetramer−octamer scheme. As can be seen in the lower panel, the octamer contribution is less than 5% of the total signal at all peptide concentrations. Together, these data show that the TM peptide exists in a cooperative

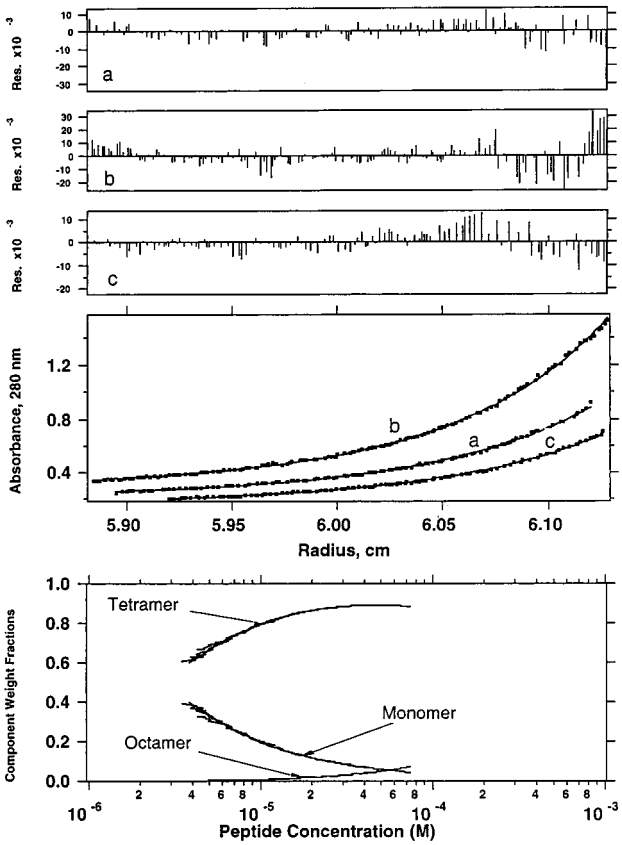

FIGURE 5: Sedimentation equilibrium of M2 in DPC micelles at 20K rpm. The data are treated as described in the legend of Figure 4. The DPC concentration in all cases was 15 mM, and M2 to DPC molar ratios were (a) 1:358, (b) 1:505, and (c) 1:692.

monomer−tetramer equilibrium, with only a small amount (<5%) of a higher-order aggregate.

The sedimentation curves for the full-length M2 polypeptide (Figure 5) indicate that this protein forms more stable tetramers than the corresponding TM peptide. The data are again best described by a monomer−tetramer equilibrium with a small amount (<10%) of higher-order aggregates being formed at the highest protein concentrations. Of the equilibria that were examined, the data are best fit by a monomer−tetramer−hexamer or monomer−tetramer−octamer equilibrium (Table 1). A comparison of the best fit model (monomer−tetramer−octamer) dissociation constants for tetramerization of M2 versus M2TM indicates that the tetramer formed by the full-length peptide is approximately 7 ± 1 kcal/mol more stable than the corresponding TM peptide. Thus, regions outside of the TM helix must contribute to the stability of the tetramer.

It is interesting to note that, although the full-length protein was examined at higher DPC:peptide ratios than M2TM, it nevertheless exhibited a greater tendency to form high-order aggregates. These data indicate that the formation of higher-order aggregates is not due to adventitious oligomerization, which would be favored at lower DPC:peptide ratios by a nonspecific mechanism.

Partial Proteolysis. The M2 polypeptide in DPC micelles was treated with trypsin and chymotrypsin to identify regions that are structured, and hence resistant to degradation by these proteases. Both proteases very rapidly cleaved the M2 protein into smaller fragments, which resisted further proteolysis. Trypsinolysis gave rise to two products that were separated by RP-HPLC. Partial N-terminal sequence analysis and MALDI-MS indicated that the two peptides spanned residues 19−60 and 19−61 of the sequence of M2. The fragment amino acid sequences are presented in Figure 6. Residue 19 is approximately five residues N-terminal from the putative TM helix, and residue 60 is approximately 17 residues from this membrane-spanning helix. Interestingly, a number of potential trypsin cleavage sites (R45, K49, and R53) near the C-terminus of this fragment and outside of the TM-spanning helix are protected from proteolysis.

A chymotryptic fragment (see Figure 6) was similarly isolated and found to span residues G16−Y76. The N-terminal clip again occurs near the N-terminus of the TM helix, while the C-terminal cleavage at Y76 confirms the existence of a structured region following the membrane-spanning segment. Potential chymotryptic sites (F47, F48, Y52, F54, and F55) within this region are strongly protected from proteolysis in the folded form of the protein. Thus, it is likely that the region immediately C-terminal to the TM helix is

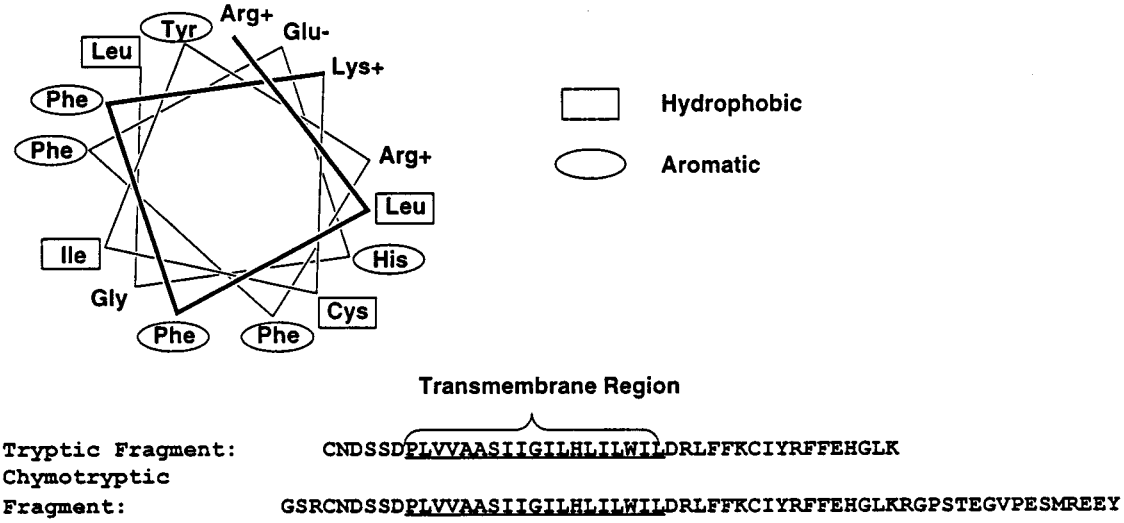

FIGURE 6: Amino acid sequences and helical wheel analysis of proteolytic fragments of the M2 protein. A helical wheel representation is shown of the peptide segment (residues 45−59) immediately C-terminal to the TM helix. Below are shown sequences of a tryptic fragment spanning residues 19−60 and of a chymotryptic fragment spanning residues 16−76.

structured, and possibly contributes to the stability of the tetramer.

DISCUSSION

While the features defining the three-dimensional structures and thermodynamic stability of water-soluble proteins are relatively well understood, the understanding of membrane proteins remains in its infancy. This discrepancy lies in the difficulties associated with the expression, structure determination, and thermodynamic characterization of membrane proteins. It is often very difficult to express large quantities of properly folded membrane proteins for biophysical studies. This problem has recently been partially circumvented by screening amino acid sequence databases for bacterial homologues of the protein of interest (*3*, *38*). In this way, it has been possible to identify homologous sequences that are more stable and easier to express in bacteria. However, a bacterial homologue may not be available or its properties may differ from the original target. Here, we describe an attractive alternative strategy, which involves the chemical synthesis of the desired target using the strategy of amide-forming ligation at cysteine residues (*4*, *5*). Thus, whereas the expression of M2 in *Escherichia coli* produces very low yields due to rapid cell lysis (*39*), the work presented here demonstrates that it is now possible to chemically synthesize 10−50 mg of the pure protein within 1 or 2 weeks.

A second problem associated with the study of the thermodynamics of membrane proteins is that their unfolding transitions tend to be thermodynamically irreversible. In fact, the thermodynamic stability under reversible folding conditions of only one monomolecular protein has been described (*40*). Therefore, there has been considerable interest in the study of membrane proteins that contain a single TM helix capable of self-assembling into radially symmetrical helical bundles (*2*, *41*). The study of such simple systems should advance the understanding of membrane protein folding in much the same way that the study of coiled coils has advanced the understanding of water-soluble proteins (*42*−*46*). In particular, the TM helices required for dimerization of glycophorin (*41*, *47*−*49*) and pentamerization of phospholamban have been extensively studied (*10*, *50*−*52*). The oligomerization surfaces of glycophorin and phospholamban have been defined by site-directed mutagenesis. Also, a fusion protein containing the TM region of glycophorin fused to staphylococcal nuclease has been shown (*32*) to dimerize in micelles, although a third "thermodynamically incompetent" tetramer component was required for data fitting, so the meaning of the derived dissociation constants is unclear. Interestingly, the dissociation constants, expressed in units of the aqueous solution concentration of the protein, were independent of micelle concentration, indicating that the protein monomer and dimer are always fully protected by micelle-forming molecules. The M2 protein provides a third, functionally interesting system, which is shown here to reversibly self-assemble in micelles. Thus, this protein provides an ideal system for studying the structural determinants of the thermodynamics of assembly of membrane proteins.

Because the folding of M2 is reversible, it was possible to quantitatively explore the differences in the thermodynamic assembly of the full-length protein versus a peptide representing its TM helix. While both constructs form tetramers, the full-length protein tetramerizes with a more favorable free energy ($\Delta\Delta G^\circ = -7 \pm 1$ kcal/mol). Thus, although the TM region contains the primary determinants for assembly into a four-helix bundle, other regions in the chain appear to help stabilize this assembly. CD spectroscopy also showed that the full-length protein contained a larger number of residues in a helical conformation than could be ascribed to the TM helix, and a region immediately C-terminal to the TM helix was further shown to be highly protected from cleavage by proteases. Examination of the protected sequence on a helical wheel shows that this segment could form a highly amphiphilic structure, when in a helical conformation. This region is particularly rich in aromatic and positively charged side chains, which segregate on opposite sides of a helix. Thus, it is likely that this segment forms a helical extension of the TM helix.

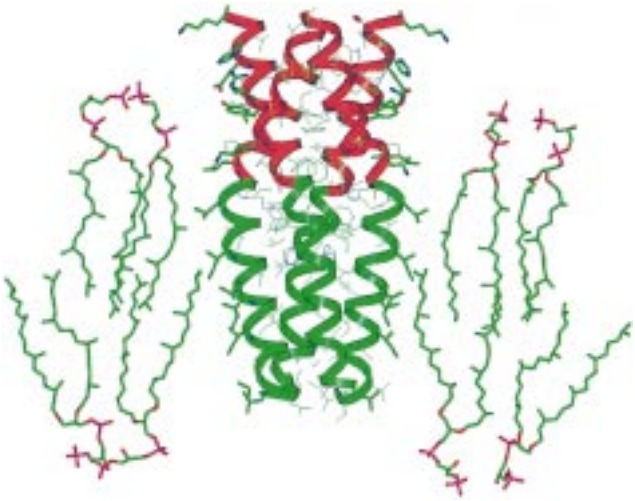

FIGURE 7: Model of the TM peptide of M2 (green ribbon) with the water-soluble helical extension (residues 45−60, red ribbon). The coordinates for the model of the TM region were those described in ref *22*. The C-terminal extension was model built, and then transformed to an α-helical conformation. The entire structure was then minimized without constraints using the method of steepest descents followed by the conjugate gradient method for 10 000 cycles with the CV force field as implemented in DISCOVER. The computation was conducted without added solvent or lipid. A second model (not shown), obtained with slightly different coordinates for the TM helix, showed a larger hole down the middle. More experimental data will be required to provide a high-resolution model of the protein. The phospholipids shown in the figure are included to provide a comparison of the protein structure with that of a typical phospholipid bilayer.

The predicted structure of the TM region of the M2 protein has the appearance of a frustum (a truncated cone). That is, the helices flare slightly from a point of closest approach near the N-terminus of the bundle, which faces the outside of the virus. If the helical conformation is propagated beyond the TM segment into residues 44−60, the apolar residues in the helical extension are well-positioned to engage in numerous interhelical hydrophobic interactions. Further, the polar residues lie on the outside of the structure, where they may form stabilizing interactions with the water, the lipid headgroups, and possibly also the M1 protein (*53*). Computer modeling identified two possible orientations of the helical extension, relative to the TM helix. In the first (not shown), the helical extensions continue to follow the path of the TM

helices. The resulting structure has a large hole down its center, similar to the helical vestibule observed in the crystal structure of the MscL homologue from *Mycobacterium tuberculosis* of a stretch-activated channel (*38*). Alternatively, if a slight bend occurs near the junction of the TM region, the helical extensions form a well-packed helical bundle with no appreciable central cavity (Figure 7). However, protons may be shuttled into the TM channel via polar residues at the helix−helix interface, near the end of the TM region.

In an environment more closely representative of the assembled virus, other possibilities can be envisioned. For example, each amphiphilic helix of the extension could individually make a tight, 90° turn, and then bind hydrophobically to the inner leaflet of the bilayer membrane, creating a "rosette" structure. Clearly, a more detailed model of the M2 protein will have to await spectroscopic or crystallographic characterization of the protein. The availability of procedures for the synthesis of the full-length M2 protein and long fragments of this structure will facilitate these structural studies as well as experiments aimed at functional reconstitution of ion channel activity.

ACKNOWLEDGMENT

G.G.K. thanks Jay Levy for expert peptide assembly and Christie L. Hunter and Paolo Botti for many helpful discussions. Protein sequencing was carried out by the Protein Chemistry Laboratory of the Medical School of the University of Pennsylvania supported by core grants of the Diabetes and Cancer Centers (DK-19525 and CA-16520).

REFERENCES

1. Montal, M. (1996) *Curr. Opin. Struct. Biol. 6*, 499−510.
2. Dieckmann, G. R., and DeGrado, W. F. (1997) *Curr. Opin. Struct. Biol. 7*, 486−494.
3. Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1998) *Science 280*, 69−77.
4. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) *Science 266*, 776−779.
5. Wilken, J., and Kent, S. B. (1998) *Curr. Opin. Biotechnol. 9*, 412−426.
6. Oblattmontal, M., Yariazaki, M., Nelson, R., and Montal, M. (1995) *Protein Sci. 4*, 1490−1497.
7. Montal, M. O., Iwamoto, T., Tomich, J. M., and Montal, M. (1993) *FEBS Lett. 320*, 261−266.
8. Lear, J. D., Wasserman, Z. R., and DeGrado, W. F. (1994) in *Membrane Protein Structure* (White, S. H., Ed.) pp 335−354, Oxford University Press, New York.
9. Arkin, I. T., Adams, P. D., Brunger, A. T., Smith, S. O., and Engelman, D. M. (1997) *Annu. Rev. Biophys. Biomol. Struct. 26*, 157−179.
10. Simmerman, H. K., and Jones, L. R. (1998) *Physiol. Rev. 78*, 921−947.
11. Lamb, R. A., Holsinger, L. J., and Pinto, L. H. (1994) in *Receptor-Mediated Virus Entry into Cells* (Wimmer, E., Ed.) pp 303−321, Cold Spring Harbor Press, Cold Spring Harbor, NY.
12. Hay, A. J. (1992) *Sem. Virol. 3*, 21−30.
13. Chizhmakov, I. V., Geraghty, F. M., Ogden, D. C., Hayhurst, A., Antoniou, M., and Hay, A. J. (1996) *J. Physiol. 494*, 329−336.
14. Shimbo, K., Brassard, D. L., Lamb, R. A., and Pinto, L. H. (1996) *Biophys. J. 70*, 1336−1346.
15. Helenius, A. (1992) *Cell 69*, 577−578.
16. Sakaguchi, T., Tu, Q., Pinto, L. H., and Lamb, R. A. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5000−5005.
17. Panayotov, P. P., and Schlesinger, R. W. (1992) *Virology 186*, 352−355.
18. Duff, K. C., Kelly, S. M., Price, N. C., and Bradshaw, J. P. (1992) *FEBS Lett. 311*, 256−258.
19. Duff, K. C., and Ashley, R. H. (1992) *Virology 190*, 485−489.
20. Kukol, A., Adams, P. D., Rice, L. M., Brunger, A. T., and Arkin, I. T. (1999) *J. Mol. Biol. 286*, 951−962.
21. Kovacs, F. A., and Cross, T. A. (1997) *Biophys. J. 73*, 2511−2517.
22. Pinto, L. H., Dieckmann, G. R., Gandhi, C. S., Shaughnessy, M. A., Papworth, C. G., Braman, J., Lear, J. D., Lamb, R. A., and DeGrado, W. F. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 11301−11306.
23. Sansom, M. S. P., Kerr, I. D., Smith, G. R., and Son, H. S. (1997) *Virology 233*, 163−173.
24. Zhong, Q., Husslein, T., Moore, P. B., Newns, D. M., Pattnaik, P., and Klein, M. L. (1998) *FEBS Lett 434*, 265−271.
25. Forrest, L. R., DeGrado, W. F., Dieckmann, G. R., and Sansom, M. S. (1998) *Folding Des. 3*, 443−448.
26. Kurtz, S., Luo, G., Hahnenberger, K. M., Brooks, C., Gecha, O., Ingalls, K., Numata, K., and Krystal, M. (1995) *Antimicrob. Agents Chemother. 39*, 2204−2209.
27. Schroeder, C., Ford, C. M., Wharton, S. A., and Hay, A. J. (1994) *J. Gen. Virol. 75*, 3477−3484.
28. Holsinger, L. J., and Lamb, R. A. (1991) *Virology 183*, 32−43.
29. Schnölzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S. B. H. (1992) *Int. J. Pept. Protein Res. 40*, 180−193.
30. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) *Protein Sci. 4*, 2411−2423.
31. Furst, A. (1997) *Eur. Biophys. J. 35*, 307−310.
32. Fleming, K. G., Ackerman, A. L., and Engelman, D. M. (1997) *J. Mol. Biol. 272*, 266−275.
33. Tanford, C., and Reynolds, J. A. (1976) *Biochim. Biophys. Acta 457*, 133−170.
34. Laue, T., Shaw, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C., Eds.) pp 90−125, The Royal Society of Chemistry, Cambridge, U.K.
35. Brahms, S., and Brahms, J. (1980) *J. Mol. Biol. 138*, 149−178.
36. Chen, Y.-H., Yang, J. T., and Martinez, H. M. (1972) *Biochemistry 11*, 4120−4131.
37. Javadpour, M. M., and Barkley, M. D. (1997) *Biochemistry 36*, 9540−9549.
38. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T., and Rees, D. C. (1998) *Science 282*, 2220−2226.
39. Guinea, R., and Carrasco, L. (1994) *FEBS Lett. 343*, 242−246.
40. Lau, F. W., and Bowie, J. U. (1997) *Biochemistry 36*, 5884−5892.
41. Lemmon, M. A., and Engelman, D. M. (1994) *Q. Rev. Biophys. 27*, 157−218.
42. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. (1995) *Science 270*, 935−941.
43. Cohen, C., and Parry, D. A. D. (1990) *Proteins 7*, 1−15.
44. Gibney, B. R., Rabanal, F., and Dutton, P. L. (1997) *Curr. Opin. Chem. Biol. 1*, 537−542.
45. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science 262*, 1401−1407.
46. Hodges, R. S. (1996) *Biochem. Cell. Biol. 74*, 133−154.
47. Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J., and Engelman, D. M. (1992) *Biochemistry 31*, 12719−12725.
48. MacKenzie, K. R., Prestegard, J. H., and Engelman, D. M. (1997) *Science 276*, 131−133.
49. Adams, P. D., Engelman, D. M., and Brünger, A. T. (1996) *Proteins 26*, 257−261.
50. Arkin, I. T., Rothman, M., Ludlam, C. F. C., Aimoto, S., Engelman, D. M., Rothschild, K. J., and Smith, S. O. (1995) *J. Mol. Biol. 248*, 824−834.

51. Herzyk, P., and Hubbard, R. E. (1998) *Biophys. J. 74*, 1203−1214.
52. Simmerman, H. K., Kobayashi, Y. M., Autry, J. M., and Jones, L. R. (1996) *J. Biol. Chem. 271*, 5941−5946.
53. Hay, A. J., Wolstenholme, A. J., Skehel, J. J., and Smith, M. H. (1985) *EMBO J. 4*, 3021−3024.
54. Straume, M., and Johnson, M. L. (1992) in *Numerical Computer Methods* (Brand, L., and Johnson, M. L., Eds.) pp 87−105, Academic Press, San Diego, CA.

BI990720M