doi:10.1016/j.jmb.2004.06.044 *J. Mol. Biol.* (2004) **341**, 769–779

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

A Highly Unusual Palindromic Transmembrane Helical Hairpin Formed by SARS Coronavirus E Protein

Eyal Arbely[1], Ziad Khattari[2], Guillaume Brotons[2], Mutaz Akkawi[3] Tim Salditt[2] and Isaiah T. Arkin[1]*

[1]*Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University, Givat-Ram Jerusalem 91904, Israel*

[2]*Institut für Röntgenphysik Georg-August-Universität Göttingen Geiststraße 11 D-37073 Göttingen, Germany*

[3]*Faculty of Science and Technology, Al-Quds University, Abu Dis Jerusalem*

**Corresponding author*

The agent responsible for the recent severe acute respiratory syndrome (SARS) outbreak is a previously unidentified coronavirus. While there is a wealth of epidemiological studies, little if any molecular characterization of SARS coronavirus (SCoV) proteins has been carried out. Here we describe the molecular characterization of SCoV E protein, a critical component of the virus responsible for virion envelope morphogenesis. We conclusively show that SCoV E protein contains an unusually short, palindromic transmembrane helical hairpin around a previously unidentified pseudo-center of symmetry, a structural feature which seems to be unique to SCoV. The hairpin deforms lipid bilayers by way of increasing their curvature, providing for the first time a molecular explanation of E protein's pivotal role in viral budding. The molecular understanding of this critical component of SCoV may represent the beginning of a concerted effort aimed at inhibiting its function, and consequently, viral infectivity.

Keywords: membrane proteins; SARS coronavirus; transmembrane helices; viral budding

Introduction

A previously unidentified member of the coronaviridae is the etiologic agent responsible for the recent severe acute respiratory syndrome (SARS) outbreak.[1] Viral genome sequencing, combined with protein phylogenetic analyses, have shown that SARS coronavirus (SCoV) belongs to a new subfamily within the coronaviridae.[2,3] In an effort to better understand the components contributing to the pathogenic mechanism of the virus, we have structurally analyzed the transmembrane domain of SCoV E protein using Fourier Transform Infrared (FTIR) spectroscopy, X-ray scattering, electron microscopy and global searching molecular dynamics simulations.

Functionally, coronavirus E protein's pivotal role in viral morphogenesis has been studied extensively. Co-expression of Mouse Hepatitis Virus (MHV) E and M proteins on their own has been shown to result in the production of virus-like particles.[4,5] This fundamental observation proved that neither the nucleocapsid nor the viral spike are needed for viral budding. Work on other viruses, such as Transmissible Gastroenteritis Virus (TGEV),[6] Bovine coronavirus (BCoV)[6] and Infectious Bronchitis Virus (IBV)[7] has shown this to be a general phenomena of the coronaviridea. The interaction between the cytoplasmic domains of the two proteins is thought to take place in pre-Golgi compartments.[8]

The importance of the role played by E protein in virus budding and morphogenesis is further strengthened by the observation that in a number of coronaviruses, expression of M protein on its own does not produce virus-like particles.[4–7,9] In contrast, expression of IBV or MHV E protein on its own causes the release of vesicles containing E protein, thereby pointing to the importance of

Abbreviations used: SARS, severe acute respiratory syndrome; SCoV, SARS coronavirus; MHV, mouse hepatitis virus; TGEV, transmissible gastroenteritis virus; BCoV, bovine coronavirus; IBV, infectious bronchitis virus; TMD, transmembrane domain; DMPC, dimyristoylphosphocholine; TFA, trifluoroacetic acid; ATR, attenuated total reflection; FTIR, Fourier Transform Infrared.

E-mail address of the corresponding author: arkin@cc.huji.ac.il

```
NH3+ -MYSFVSEETG TLIVNSVLLF LAFVVFLLVT
     LAILTALRLC AYCCNIVNVS LVKPTVYVYS
     RVKNLNSSEG VPDLLV-COO-

AIBV  EE GSFLTALYIIVGF-LALYLLGRALQ----
CEC   DDNGMVISIIFWFLLI-IILILFSIALLNIIK
PTGE  DDNGLVISIIFWFLLI-IILILFSIALLNIIK
PRC   DDNGMVISIIFWFLLI-IILILLSIALLNIIK
HCV1  DDHALVVNVLLWCVVL-IVILLVCITIIK---
EQC   ADTVWYVGQIIFIVAICLLVIIVVVAFLATFK
BCV1  ADTVWYVGQIIFIVAICLLVIIVVVAFLATFK
BCV2  ADTVWYVGQIIFIVAICLLVIIVVVAFLATFK
HECV  ADTVWYVGQIIFIVAICLLVIIVVVAFLATFK
HCV2  ADTVWYVGQIIFIVAICLLVTIVVVAFLATFK
MHV1  TDTVWYVGQIIFIFAVCLMVTIIVVAFLASIK
RSDCV IDTVWYVGQIIFIVAVCLMVTIIVVAFLASIK
MHV2  TDTVWYVGQIIFIVAVCLMVTITVVAFLASIK

SCoV  EETGTLIVNSVLLFLAFVVFLLVTLAILTALR
```

Figure 1. Top panel: sequence of the SCoV E protein (strain TOR2 NP_828854). Sequence of the synthesized peptide is boxed, while gray shading indicates the extent of the hydrophobic stretch of the protein. The iodinated Phe residue is indicated by dark shading. Bottom panel: multiple sequence alignment of the hydrophobic region of E proteins belonging to the four different coronaviridea subgroups. Gray shading indicates the hydrophobic amino acid residues according to the GES scale.[28] Sequence codes are given in the Materials and Methods.

coronavirus E protein in the budding process.[7,9] Interestingly, cells infected with MHV release vesicles containing E protein as well.[9] Finally, mutations in MHV E protein cause marked morphological changes in the resulting viruses.[10]

In the cell, evidence has been accumulating suggesting that IBV E protein localizes to the Golgi apparatus,[7] the site from which coronaviruses are known to invaginate into mature virions. During the expression of E protein the Golgi apparatus changes its morphology dramatically,[11] which could explain in part E protein's ability to induce apoptosis.[12,13]

Structurally, E proteins are conserved within the different coronaviridea groups, yet exhibit little sequence similarity between the groups (Figure 1). In line with its belonging to a new group within the coronaviridea, SCoV E protein differs substantially from other coronavirus E proteins.[3] Whether this has any bearing upon the marked difference in pathogenicity of SCoV *versus* other coronaviruses remains to be seen. Finally, SCoV E protein shares no significant sequence similarity with any other known protein.

It is possible, however, to make the following generalizations regarding the sequence of all coronavirus E proteins: they are small proteins (~75 residues), with an unusually long hydrophobic stretch (25–30 residues) located between hydrophilic N and C termini (~8 and ~40 residues, respectively). SCoV E protein shares the same overall characteristics, although only 20% identical with other coronaviridea E proteins:[3] 74 residues in length with a stretch of 26 hydrophobic amino acid residues (Figure 1).

The length of the hydrophobic segment of SCoV E protein has posed a problem with respect to assigning a topology to the protein (Figure 2(a)). The average length of a transmembrane α-helix is

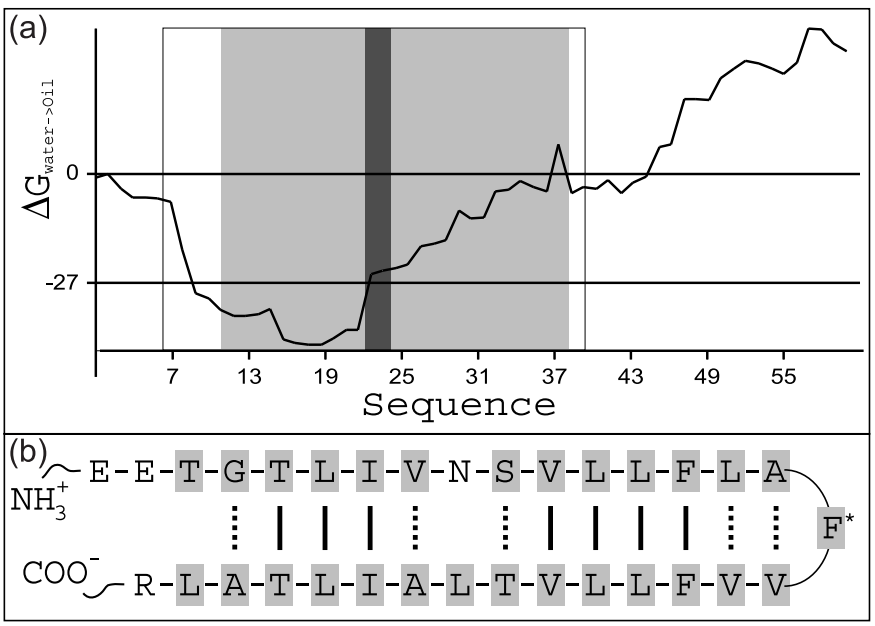

Figure 2. (a) Hydrophobicity analysis of the SCoV E protein according to the GES scale[28] using a window size of 15 and a hydrophobicity cutoff of $\Delta G_{water \rightarrow oil} = -27$ kcal/mol.[29] The region of the synthesized peptide is boxed, while gray shading indicates the extent of the hydrophobic stretch of the protein. The location of the iodinated Phe residue is indicated by dark shading. (b) Inversion diagram of the SCoV E protein hydrophobic region outlining identity and similarity between the two inverted sequence elements, in continuous and broken lines, respectively. Hydrophobic amino acid residues according to the GES scale[28] are indicated by gray shading.

21 residues,[14] far shorter than the hydrophobic stretch of E proteins. In contrast, if a hairpin were to be formed by the long hydrophobic stretch, the resulting helices would be much shorter than the average.

Protease digestion of MHV does not affect E protein's molecular mass, suggesting that no part of the protein protrudes beyond the viral membrane.[11] Antibodies raised against the N terminus of the protein could detect the protein only after treatment of the cells with Triton X-110, but not with digitonin.[7] Both of the above studies concluded that the protein traverses the lipid bilayer once, with its C terminus located in the viral interior.[11,7] More recent results with an epitope tagged MHV E protein are indicative of the protein traversing the lipid bilayer twice, whereby both termini of the protein reside in the virus lumen.[15]

In the current study we present the results of a detailed structural examination of SCoV E protein, in which we determine the topology of the protein and the effects upon the lipid bilayer thereof. The results conclusively show that SCoV E protein has a highly unusual topology, consisting of a very short transmembrane helical hairpin. The hairpin forms an inversion about a previously unidentified pseudo-center of symmetry which seems to be unique to SCoV. Electron microscopy studies indicate that the protein dramatically distorts lipid bilayers, causing tubulation. Taken together, the structure of the protein facilitates a molecular explanation of the effects upon the bilayer curvature underpinning the function of the protein *in vivo* and may represent the beginning of a concerted effort aimed at inhibiting its function and consequently viral infectivity.

Results and Discussion

E protein is decidedly helical

In order to examine the secondary structure of the transmembrane domain (TMD) of SCoV E protein, peptides were made which encompassed the

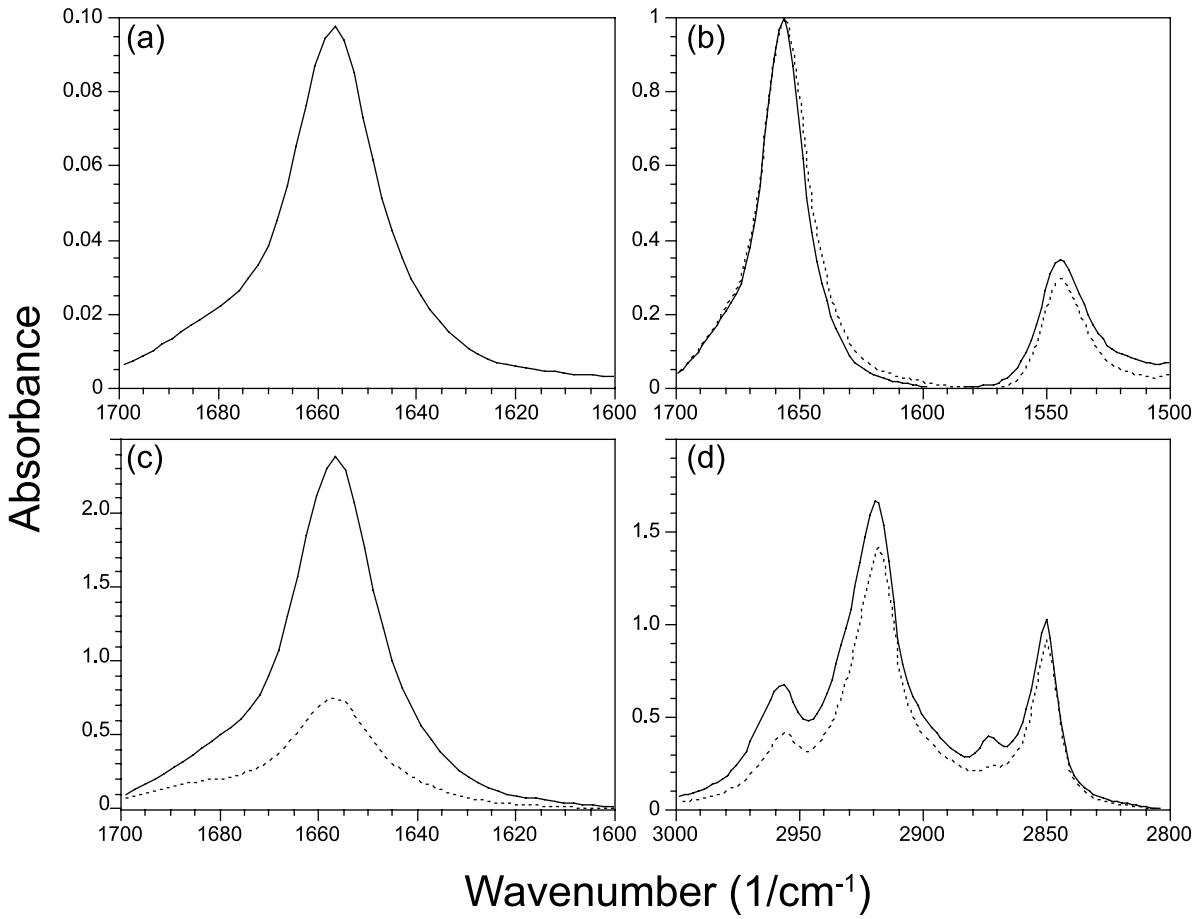

Figure 3. (a) Transmission FTIR spectrum of the SCoV E protein TMD reconstituted in DMPC lipid bilayers. (b) ATR-FTIR spectra of SCoV E protein TMD reconstituted in lipid bilayers, focusing on the amide I and amide II regions. Spectra were obtained after flushing air saturated with H_2O (continuous line) or D_2O (broken line) over the sample. Both spectra were normalized so as to adjust the absorption of the amide I mode to 1.0 OD. (c) ATR-FTIR spectra of SCoV E protein TMD reconstituted in lipid bilayers, focusing on the amide I region. Spectra were obtained with parallel (continuous line) or perpendicular (broken line) polarized light. (d) ATR-FTIR spectra as in c but focusing on the lipid CH_2 and CH_3 stretching modes.

entire hydrophobic region of the protein (Glu7-Arg38), as depicted in Figure 1. The peptide was purified and reconstituted into dimyristoylphosphocholine (DMPC) lipid vesicles, to be used for FTIR spectroscopy. The transmission FTIR spectrum of the resulting proteo-liposomes focusing on the amide I vibrational mode (1600–1700 cm^{-1}) is presented in Figure 3(a). Since the amide I vibrational mode arises mostly from the peptidic C = O stretch, its frequency is well correlated with the secondary structure of the protein.[16] Peaks resonating at 1655 cm^{-1}, 1630 cm^{-1} and 1645 cm^{-1}, correspond to α-helical, β-strand and random coil peptide segments, respectively.[16] The amide I mode of SCoV E protein TMD is centered at 1657 cm^{-1} with a peak-width at half height of 23 cm^{-1}, both indicative of the very high helical content of the protein, and the absence of any other secondary structure component.[16]

E protein contains 26 residues embedded in the lipid bilayer

FTIR spectroscopy can be used to delineate the extent of membrane incorporation of transmembrane proteins. This is achieved by observing the reduction in any vibrational mode containing significant contributions from the amide proton (e.g II mode:peptidic N–H deformation) upon exchanging the solvent from H_2O to D_2O. The reason is that the lipid bilayer protects the peptide from exchange, whereas amide groups exposed to the solvent undergo H^+/D^+ exchange. Any vibrational mode containing the "new" N–D group will resonate elsewhere and H^+/D^+ exchange can be quantitated by measuring the reduction in the amide II mode directly.[16]

As shown in Figure 3(b), upon flushing the membrane with air saturated with D_2O for several hours, the reduction in the amide II peak, centered at 1545 cm^{-1}, is minimal. Since the amide I peak at 1657 cm^{-1} is not expected to undergo any intensity changes upon exchange from H_2O to D_2O, it is possible to calculate the extent of exchange by normalizing both spectra (in H_2O and D_2O) on the amide I peak. The resulting exchange rate of 19 ± 2% is indicative of only six out of 32 residues in the peptide undergoing H^+/D^+ exchange. In other words, 26 residues of SCoV E protein are protected from exchange as a result of their being embedded in the lipid bilayer. Note that this value corresponds exactly to the length of the hydrophobic stretch of the protein.

The transmembrane helical segment of E protein is virtually perpendicular to the lipid bilayer

After establishing the helicity of the peptide and determining that most of it is embedded in the lipid bilayer, experiments were undertaken to determine its orientation using polarized FTIR spectroscopy of oriented membranes. Attenuated total internal reflection (ATR-FTIR) spectra obtained using both parallel and perpendicular polarized light are depicted in Figure 3(c). The dichroism (ratio of absorption of parallel and perpendicular polarized light) for the amide I peak (centered at 1657 cm^{-1}) is 3.76 ± 0.52. From this dichroism it is possible to calculate an exceptionally high order parameter for the peptide of $S = 0.82 \pm 0.17$, a value nearing the theoretical limit.[17] When taking into account sample disorder, it is possible to convert the maximal value of the measured order parameter to a maximal tilt angle. Thus, the tilt angle of SCoV E protein helical elements from the membrane normal must be less than or equal to 6.3°.

SCoV transmembrane domain is a helical hairpin

So far three lines of evidence were obtained characterizing the transmembrane domain of SCoV E protein: (i) it is highly helical, (ii) it is composed of 26 amino acid residues and (iii) the helical elements are oriented normal to the membrane plane. Thus, there are two ways in which to describe the structure of the protein: either as a single long transmembrane helix, or as two short helices forming a transmembrane helical hairpin. Both models however, incur a hydrophobicity mismatch between the protein and the bilayer. In the case of a single helix, the hydrophobic stretch is too long relative to the bilayer thickness.† On the other hand, the helical hairpin would have to be comprised of two very short helices in order to enable the 26 residues to traverse the membrane twice. Since the smallest possible loop is of three residues, each helix cannot contain more than 11–12 hydrophobic residues. This value is much smaller than the average of 21 residues per transmembrane helix.[14]

In order to conclusively determine the topology of SCoV E protein we iodinated Phe23, which is located at the center of the hydrophobic stretch of the protein (see Figure 1). FTIR spectroscopy has shown that the iodinated and unlabeled peptides are indistinguishable in any of the experiments described above (data not shown). Electron density profiles obtained using X-ray scattering should be able to pinpoint the position of the labeled electron-dense iodine. As shown in Figure 4, the electron density of lipid vesicles containing iodinated and unlabeled SCoV E protein transmembrane domain exhibit normal density profiles.[18] However, upon subtracting the densities form one another, clear density resulting from the iodine label is identified, located 16.5 Å from the bilayer core, close to the lipid head-group region. Thus,

† Note that a single long hydrophobic helix cannot compensate for its elongated hydrophobicity by titling, since we have shown that the helix has a maximum tilt angle with respect to the bilayer normal of less than 6.3°.

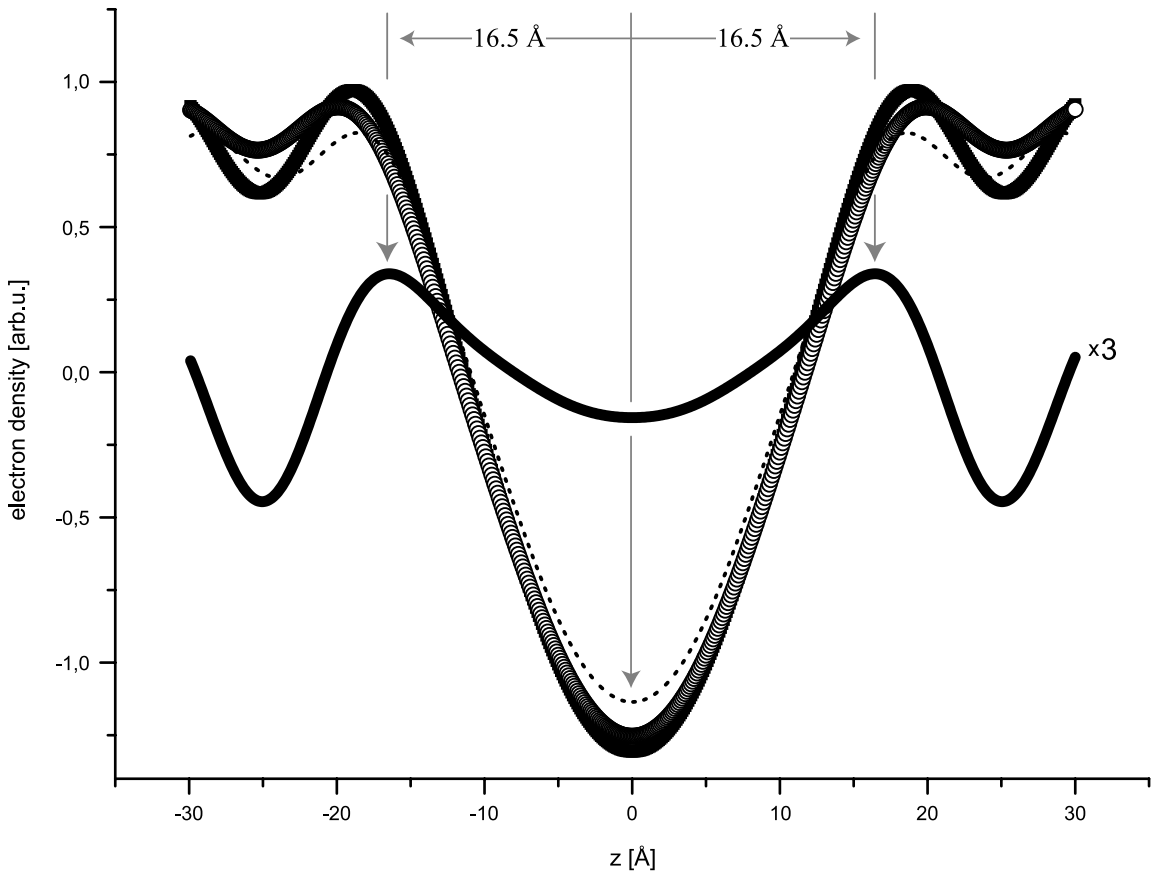

Figure 4. Electron density profile (on arbitrary scale around the mean density) of the lipid bilayers obtained at molar protein-to-lipid (DMPC/SCoV E protein) ratios of $P/L = 0.1$, for iodinated (filled symbols) and non-iodinated (open symbols) protein, as obtained from Fourier synthesis of measured integrated Bragg peak intensities. The difference between the two curves indicates the position of the one iodinated phenylalanine per protein located adjacent to the head-group region, close to the hydrophilic/hydrophobic interface. For comparison, the electron density profile of the Influenza A M2 H^+ channel transmembrane domain, a tetrameric helical bundle,[30] is depicted in broken line.

one can clearly state that the labeled Phe23 is located adjacent to the head-group region, despite the fact that it is in the middle of the hydrophobic segment of the protein. Taken together, our data prove conclusively that the transmembrane domain of SCoV E protein forms a hairpin with unusually short hydrophobic helices.

SCoV E protein contains an unusual palindromic symmetry

After establishing that SCoV E protein forms a helical hairpin, close inspection of its sequence reveals an unusual feature. As shown in Figure 2(b), inverting the protein about itself at Phe23 indicates that the protein is virtually symmetrical with respect to the inversion point, thereby forming a palindrome. The only deviation from palindromic symmetry is located at Asn15, the significance of which is discussed below.

In order to estimate the statistical significance of forming a palindrome within the helical hairpin framework, a Monte–Carlo simulation was performed. The sequence of each helix forming the hairpin was randomized, followed by assessing the extent of palindromic symmetry. The results indicate that the degree of inverted identity between the two helices of SCoV E protein (thereby forming a palindrome) is a rare event, observed at a frequency of ca 10^{-5}. This would explain why to our knowledge palindromic symmetry in helical hairpins has not been observed before.

E protein dramatically effects lipid bilayer structure and morphology

The SCoV E protein contains the shortest transmembrane hydrophobic helical hairpin known to date. Its 26 hydrophobic amino acid residues not only traverse the lipid bilayer as two helices, but must form a connecting turn between the helices. An immediate question then arises: What affect does the unusual structure of SCoV E protein transmembrane domain have upon the lipid bilayer? In order to answer the above question we undertook two lines of experiments, aimed at (i) examining the molecular structure of the lipid in detail, alongside (ii) establishing the effects upon the membrane morphology as a whole.

The order of the lipid bilayer acyl chains was examined using polarized FTIR spectroscopy of oriented multilayered DMPC vesicles containing

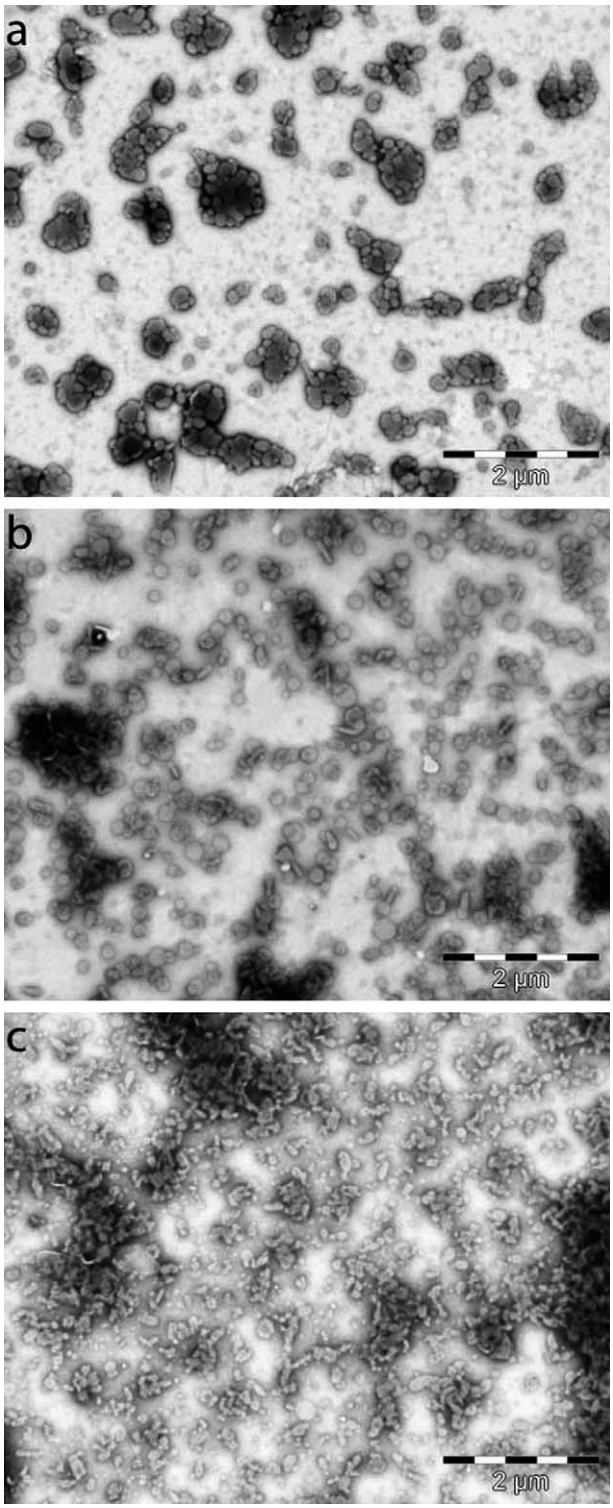

Figure 5. Negative staining electron micro-graphs of lipid DMPC vesicles (a), DMPC vesicles containing MHC class II-associated invariant chain transmembrane domain[19] (b) and DMPC vesicles containing SCoV E protein transmembrane domain (c). All vesicles were extruded though a 200 nm filter.

SCoV E protein. Figure 3(d) presents the ATR-FTIR spectrum of the lipid CH_2 and CH_3 stretching modes of the aforementioned vesicles. The measured dichroism for the lipid CH_2 asymmetric stretching mode (centered at 2920 cm^{-1}) is 1.30 ± 0.13, yielding an order parameter of $S = 0.51 \pm 0.11$. Interestingly, the order parameter of the protein ($S = 0.82 \pm 0.17$) is much higher than that of its ordering environment, the lipid bilayer. The reduced order of the lipid acyl chains is an inherent property of the system caused by SCoV E protein and cannot be attributed to incomplete sample deposition. The reason is that the exceptionally high order parameter obtained for the protein would not be possible under such circumstances. Moreover, the mosaicity of the membranes was measured to be below 0.01°, as determined by X-ray reflectivity rocking scans (not shown). Thus, incorporation of SCoV E protein transmembrane domain in the lipid vesicle deforms the acyl chains of the lipid bilayer.

The global change in the structure and morphology of the lipid bilayer upon incorporating SCoV E protein transmembrane domain was examined using negative staining electron microscopy. While vesicles without protein (Figure 5(a)) exhibited normal globular structure, SCoV E protein-containing vesicles were markedly different (Figure 5(c)), exhibiting extensive tubulation and deformation. Vesicles containing a control transmembrane protein, the MHC class II-associated invariant chain transmembrane domain,[19] were similar to vesicles without protein, and did not exhibit any tubulation (Figure 5(b)).

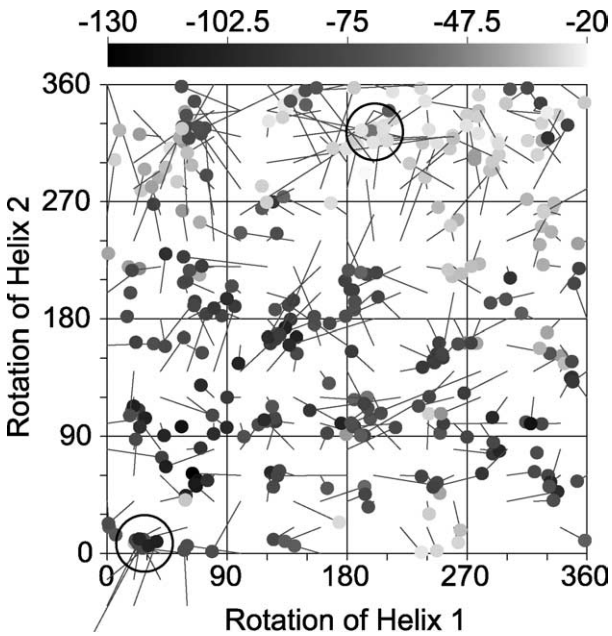

Figure 6. Results of the global searching molecular dynamics simulations of the two helices forming the hairpin of SCoV E protein. Each structure is positioned according to its final rotational angle of helix I and helix II, and colored according to energy of the bundle as indicated by the color legend (values in kcal/mol). The gray lines represent a vector connecting the starting position for each structure and the final one. The two single clusters found are encircled in the Figure.

In summery, FTIR spectroscopy has shown that the incorporation of SCoV E protein in the lipid bilayer results in a molecular disordering of the lipid acyl chains. This leads to the tubulation of the vesicles as we have shown using electron microscopy. Thus, the molecular defects in the lipid molecule are translated into a distortion of the entire bilayer structure.

A structural model of SCoV E protein transmembrane domain

The evidence obtained so far points conclusively to the fact that the SCoV E protein transmembrane domain forms an unusually short transmembrane helical hairpin: (i) The secondary structure of the transmembrane domain is highly helical. (ii) More than 80% of the protein is embedded in the lipid bilayer. (iii) The helical element of the protein is tilted from the membrane normal by no more than 6.3°. (iv) Phe23 residing in the middle of the hydrophobic stretch of the protein is located in the lipid head-group region. (v) The protein contains a center of symmetry upon which it can be inverted.

What might the hairpin structure of SCoV E protein look like? The only deviation from hairpin symmetry is at Asn15 (Figure 2(b)). The groups of Engelman & DeGrado have recently shown that asparagine residues in transmembrane helices promote strong homo-dimerization due to H-bonding.[20,21] In this instance there is only a single asparagine in the two helices and as such the hydrogen bonds that it might form are to a backbone amide group, thereby driving the two helices to close apposition. Alternatively, an asparagine residue from a different protein might promote dimerization, forming a pseudo four-helix bundle. Another interaction that might take place between the helices is through a salt bridge between Glu8 and Arg38. The presence of opposing charged resides at both sites is highly conserved in the coronaviridea (Figure 1), thereby substantiating the possibility of an interaction between them.

In order to derive a model for the protein based on its deduced topology and helix tilt angles, global searching molecular dynamics simulations were employed.[22] Multiple starting structures were generated in which both helices of SCoV E protein (Glu7-Leu21 and Val25-Arg38) formed a hetero-dimer. The sizes of the helices reflected the high helical content of the protein. Each structure was distinguished by the rotational angle of the two helices with respect to one another, in increments of 20°, leading to a total of $18 \times 18 = 324$ structures. The tilt angle between the helices was set to 0°, according to the high degree of orientation found experimentally. The stability of each structure was then determined by a molecular dynamics and energy minimization trajectory. The resulting structures were then compared in order to find regions of the configuration space to which structures from different starting configurations converged.

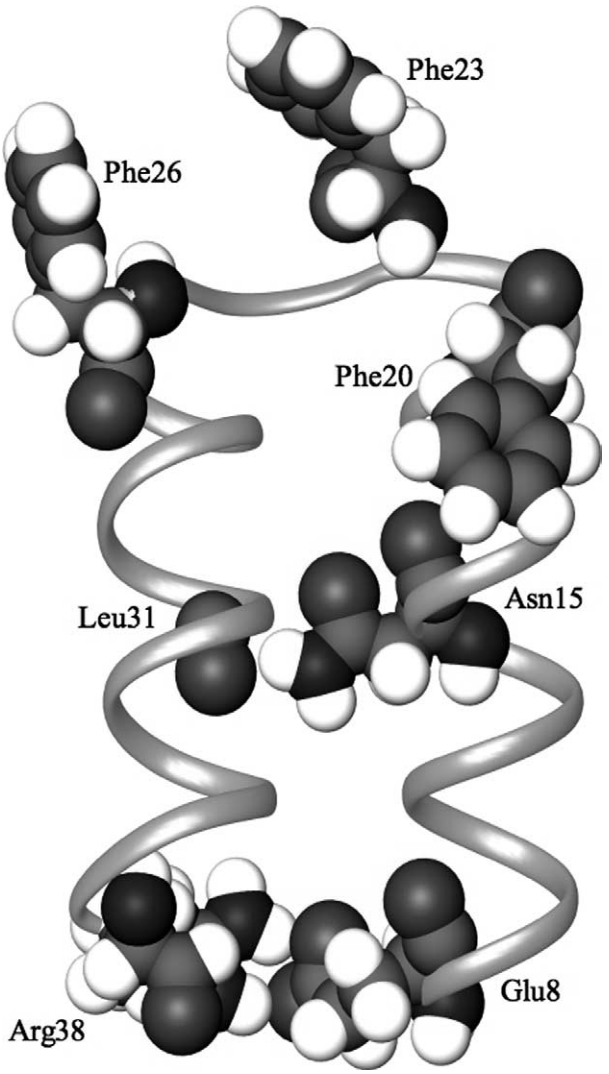

Figure 7. Ribbon diagram of the structural model derived for the transmembrane domain of SCoV E protein. The side chains of Glu8, Asn15, Phe20, Phe 23, Phe26 and Arg38 as well as the carbonyl group of Leu31 are shown in CPK representation. Figure drawn with the program Ribbons.[31]

The results of the global searching molecular dynamics simulations are shown in Figure 6, depicting two clusters to which structures have converged. The energy of the structures and the two clusters show a clear preference to lower rotational angles of the two helices (most likely due to the formation of a salt bridge, as discussed below). Thus the cluster found at rotational angles of 50° and 10° (helix I and II, respectively) was taken to represent a model for the interactions between the two helices (see Figure 6).

After the packing between the two helices forming the hairpin was obtained, the loop connecting the helices formed by amino acid residues Ala22-Val24 was constructed. The entire hairpin structure containing the loop was subsequently subjected to

energy minimization and molecular dynamics. The resulting model is shown in Figure (7), and exhibits three prominent features: (i) a salt bridge between Glu8 and Arg38, (ii) an inter-helical H-bond between the amide side-chain of Asn15 and Leu31 peptide carbonyl and (iii) three phenylalanine residues located at the tip of the hairpin pointing outwards forming an aromatic belt.

Finally, based on the evidence that IBV E protein is palmitoylated,[23] it is possible to speculate that SCoV E protein might be as well. In contrast, experiments attempting to label TGEV E protein with palmitic acid have failed.[24] If present in SCoV E protein, palmitoylation will take place on one or more of the three juxtamembranous cysteine residues. It is difficult to estimate the effects, if any of palmitoylation on the SCoV E protein. One might speculate that the intercalation of the palmitic acid in the cytoplasmic membrane leaflet might further affect the local membrane structure.

Relationship to *in vivo* studies

In any *in vitro* study the relevance of the results to those found under native conditions is obviously a concern. However, the following facts negate any suspicions of artifactuality:

Comparison to the full length protein. The protein under study encompasses all the hydrophobic domain of the SCoV E protein. The only thing known about the endogenous protein is that it traverses the lipid bilayer twice.[15] That is exactly what was found herein. Thus it is highly unlikely that the full length protein and its membrane domain both traverse the lipid bilayer twice, but do so in different ways.

Structural similarity to that found in native conditions. The endogenous function of SCoV E protein is in viral budding by way of causing membrane tubulation.[4–7,9] Moreover, it is the only protein in the virus that is required for this process.[7,9] Thus, the finding that the membranous segment of SCoV E protein (analyzed *in vitro*) causes membrane tubulation in an indistinguishable manner to the wild-type protein, *in vivo*, is proof that the membranous segment is necessary and sufficient for SCoV E protein role in membrane tubulation.

Finally, the membranous segment of SCoV E protein was not studied in detergent micelles (as is customary), but rather in hydrated lipid bilayers, a system that is closest to the native conditions as possible. It was shown that the protein has a defined helical secondary structure and that it is highly embedded in the lipid bilayer. It is difficult to think of an example in which a small membrane protein was reconstituted in lipid vesicles, shown that it was helical as well as highly embedded in the lipid bilayer but is nevertheless in a non-native conformation.

Functional implications of SCoV E protein structure

With a structural model of SCoV E protein at hand, is it possible to rationalize the effects of the protein upon the lipid bilayer? The hydrophobic stretch of both hairpin helices is only 11 residues long, which is much shorter than the average of 21 residues.[14] At the apex of the hairpin three amino acid residues have their amide carbonyl and amine groups unpaired, surrounded by an aromatic belt. At the other end of the hairpin a cluster of charged residues is located. The distance between these two polar elements is much shorter than the hydrophobic thickness of the lipid bilayer. The polarity of both elements is not equal however, whereby the charged residues at the opening of the hairpin predominate. Thus, one can speculate that the hairpin opening is situated at the lipid head-group region of one leaflet. Therefore, the hairpin apex is situated in the hydrophobic region of the bilayer, with its aromatic belt (known for its affinity to the polar head group) and unpaired amide groups poised to attract the lipid head-group region. It is this attraction that exerts a deformation force on the bilayer, causing an increase in curvature.

In comparison to other coronaviridea E proteins, only the length of the hydrophobic segment lined by opposing charged residues at the either end, is conserved (Figure 1). Thus, E proteins from different coronavirus subfamilies do not exhibit any sequence similarity to those from other subfamilies. It would seem therefore, that members of the coronaviridea have "found" at least four independent sequence motifs to form a functional E protein. The effects of the unique features of SCoV E protein—the aromatic belt, the inter-helical H-bond of Asn15 and the inversion symmetry on the overall virus life cycle and resulting pathogenicity, remain to be seen.

Materials and Methods

Viral strains

Abbreviations for viral strains listed in Figure 1, are as follows, with the accession numbers in parentheses: AIBV—Avian infectious bronchitis virus Beaudette US (CAC39303); CEC—Canine enteric coronavirus strain INSAVC-1 (P36696); PTGE—Porcine transmissible gastroenteritis coronavirus strain FS772/70 (P2265); PRC—Porcine respiratory coronavirus strain RM4 (P24415); HCV1—Human coronavirus strain 229E (P19741); Equine coronavirus strain NC99 (AAQ67208); BCV1—Bovine coronavirus strain Mebus (AAA42914); BCV2—Bovine coronavirus strain F15 (P15775); HECV—Human enteric coronavirus strain 4408 (AAQ67200); HCV2—Human coronavirus strain OC43 (AAR01017); MHV1—Murine hepatitis virus strain A59 (NP_068673); RSDCV—Rat sialodacryoadenitis coronavirus strain SDAV-681 (AAF97741); MHV2—Murine hepatitis virus strain S (P29076); SCoV—SARS coronavirus strain TOR2 (NP_828854).

Protein synthesis, purification and reconstitution

Peptides encompassing the predicted transmembrane domain of SARS coronavirus E protein (Figure 1, residues

7 to 38) were synthesized by standard solid-phase *N*-(9-fluorenyl) methoxycarbonyl chemistry. The peptides were cleaved from the resin with trifluoroacetic acid (TFA, Aldrich) and lyophilized. Two different synthetic peptides were made: an unlabeled peptide and one containing a *para*-iodo phenylalanine at position 23 of the sequence, as shown in Figure 1. Peptide purity was confirmed by mass spectrometry. The lyophilized peptide was dissolved in TFA (final concentration ~5 mg/ml), and immediately injected on a Jupiter 5μ C4 300 Å column (Phenomenex, Chesire, UK), equilibrated with 80% H_2O, 8% (v/v) acetonitrile (J.T. Baker and 12% (v/v) 2-propanol (J.T. Baker Peptide elution was achieved with linear gradient to a final solvent composition of 40% (v/v) acetonitrile, and 60% (v/v) 2-propanol. All solvents contained 0.1% (v/v) TFA. After over night lyophilization of pooled fractions in the presence of 10 mM HCl (to remove traces of TFA adducts), ca 1 mg of the dried peptide was dissolved in a solution of 10 mg dimyristoylphosphocholine (DMPC, Avanti) in 900 μl of 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, Merck). The solvent was evaporated over night under reduced pressure, by means of a rotary evaporator. One millilitre of H_2O was added to the dried product and the solution was mixed for 20 minutes at 30 °C.

FTIR spectroscopy

Data were recorded on a Nicolet Magna-560 infrared spectrometer (Nicolet Instrument Corporation, USA) purged with dry air and equipped with an MCTA detector, cooled with liquid nitrogen. A total of 1000 interferograms were collected at a resolution of 4 cm^{-1}. Attenuated total reflection (ATR) spectra were measured with a 25 reflections ATR accessory from Grasbey Specac (Kent, UK) and a wire grid polarizer (0.25 μM, Graseby Specac). Three hundred microlitres of sample (ca 1 mg/ml of peptide and 10 mg/ml of lipid) were deposited onto a Ge trapezoidal internal reflection element (50 × 2 × 10 mm) flowed by removal of bulk solvent. For the purpose of solvent exchange, air was followed over the sample that was bubbled through ^2H_2O or H_2O. Transmission FTIR spectra were collected by depositing 50 μl of sample (ca 1 mg/ml of peptide and 10 mg/ml of lipid) on a CaF_2 window.

The dichroic ratios of the Amide I band were calculated by integrating between 1670 cm^{-1} and 1645 cm^{-1}. All integrations were performed by using a straight baseline that contains points immediately before and after the band. Reported standard deviations values represent a minimum of three data sets measured.

Order parameter calculation

Order parameters for the transmembrane helices were calculated as described.[17] In brief, the order parameter S is defined as follows:

$$S = \frac{3\langle\cos^2\theta\rangle - 1}{2} \quad (1)$$

whereby θ represents the angle between the helix director and the z axis which is coincident with the membrane normal. The order parameter can be determined experimentally from the ATR dichroic ratio R^{ATR} by the following equation:

$$S = \frac{E_x^2 - R^{ATR}E_y^2 + E_z^2}{E_x^2 - R^{ATR}E_y^2 - 2E_z^2}\left(\frac{3\cos^2\alpha - 1}{2}\right)^{-1}$$

whereby the electric field components of the evanescent wave for a Ge internal reflection element, $E_x = 1.398$, $E_y = 1.516$ and $E_z = 1.625$ are given by Harrick,[25] and α is the angle between the helix director and the transition dipole moment of the vibrational mode: 39° for the amide I mode and 27° for the amide A mode.[26] Lipid order parameters are obtained from the lipid methylene symmetric (2852 cm^{-1}) and asymmetric (2924 cm^{-1}) stretching modes using the same equation by setting $\alpha = 90°$. These equations are based on the reasonable assumption that the thickness of the deposited film (>20 μm) is much larger than the penetration depth (ca 1 μm) of the evanescent wave.[25]

X-ray scattering measurements

For X-ray reflectivity measurements, samples of varied molar peptide-to lipid ratio ($P/L = 0, 0.002, 0.01, 0.05, 0.1, 0.13$) were deposited on silicon surfaces by spreading from HFIP/chloroform solvent (0.1 ml of 20 mg/ml solution on wafers cut to 15 × 25 mm^2). The orientational distribution (mosaicity) of the highly oriented membranes was measured to be below 0.01°, as determined by a rocking scan. High resolution reflectivity scans up to vertical momentum transfer $q_z \simeq 1$ $Å^{-1}$ were taken at the D4 bending magnet beamline of the DORIS storage ring at HASYLAB/DESY, at a photon energy of 11 keV, set by a Si(111) monochromator. The curves were measured with a fast scintillation counter (Cyberstar, Oxford instruments), using motorized collimating slits on the incident and exit arms. The curves were corrected for ring current, sample illumination, and diffuse background (offset-scan).

General aspects of reflectivity experiment and analysis are discussed in Ref. 18 The samples were kept in an environmental chamber at controlled relative humidity of 98% and a temperature of $T = 45$ °C to assure that the bilayers were in the fluid L_α state. The electron density profile was obtained by the Fourier synthesis method from the integrated peak intensities, using a Lorentz correction factor $1/q_z$, and phases (+, −, −, −, −). The integrated intensity of the first Bragg peak was normalized to 1 for the iodinated SCoV E protein and for the Influenza A M2 curve, while the non-iodinated SCoV E protein curve was properly scaled to the iodinated curve corresponding to the respective scattering signals. Full q_z-range fits to determine the density profiles on an absolute scale and analysis of all samples at different R.H.% is in progress (Z.K. *et al.*, to be published).

Electron microscopy

Liposomes containing the transmembrane domain of the SCoV E protein (residues 7–38) and the MHC class II-associated invariant chain (residues 29–60)[19] where prepared as described. Liposomes composed of pure DMPC were prepared in an identical manner, with no addition of peptide. Unilamellar vesicles where obtained by two cycles of freeze-thaw sonication (40 seconds) using an inverted probe sonicator (Vibra cell model vc130, Sonics and Materials Inc., CT). The liposomes where passed twice through 200 nm polycarbonate filter using an Avanti Mini-extruder (Avanti Polar Lipids Inc., AL).

Negative staining for transmission electron microscopy was done by mixing the liposome suspension at 1 : 1 ratio with 2% (v/v) phosphotungstic acid (Electron Microscopy Sciences, Hatfield, PA) and incubating for

one minute at room temperature. The mixed sample was absorbed onto carbon-coated grids that had been rendered hydrophilic by glow discharge in air. Excess solution was removed by filter paper absorbance. After drying, the samples were imaged with a TECNAI 12 electron microscope (Philips) operating at 100 kV.

Molecular modeling

Molecular modeling of the transmembrane domain of SCoV E protein was undertaken using the CHI suite of macros[22] for CNS 1.1.[27] Briefly, bundles of helices were generated in which the angle of the helices about their director was rotated independently in increments of 20°. The crossing angle of the helices was set close to that found experimentally: 0°. Each structure was subjected to the following simulated annealing and energy minimization protocol as follows: 600 steps of Powell energy minimization followed by torsion angle molecular dynamics at 600 K for 5 ps and at 300 K for 10 ps. Finally, an additional 250 steps of Powell energy minimization were undertaken. Throughout the simulations, harmonic restraints were employed to maintain helical geometry and distance restraints were employed to maintain the helix centers within 10.5 Å one another.[22]

The resulting 324 structures were compared, whereby clusters were defined as those that contained at least seven structures and a C^{α} RMSD <1.0 Å. The clusters were than averaged and subjected to the same molecular dynamics/energy minimization protocol as above.

Acknowledgements

This research was supported in part by a grant from the Israel Science Foundation (784/01) to ITA, a grant from the Deutsche Forschungsgemeinschaft grant to ITA, MA and TS and by a grant from Niedersachsen to ITA. ITA wishes to thank Professors A. Panet, I. Ohad and Dr M. Kosloff for helpful discussions.

References

1. Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S. *et al.* (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953–1966.
2. Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S. *et al.* (2003). The Genome sequence of the SARS-associated coronavirus. *Science*, **300**, 1399–1404.
3. Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P. *et al.* (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, **300**, 1394–1399.
4. Bos, E. C., Luytjes, W., van der Meulen, H. V., Koerten, H. K. & Spaan, W. J. (1996). The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus. *Virology*, **218**, 52–60.
5. Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J. & Rottier, P. J. (1996). Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* **15**, 2020–2028.
6. Baudoux, P., Carrat, C., Besnardeau, L., Charley, B. & Laude, H. (1998). Coronavirus pseudoparticles formed with recombinant M and E proteins induce alpha interferon synthesis by leukocytes. *J. Virol.* **72**, 8636–8643.
7. Corse, E. & Machamer, C. E. (2000). Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. *J. Virol.* **74**, 4319–4326.
8. Lim, K. P. & Liu, D. X. (2001). The missing link in coronavirus assembly. Retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-Golgi compartments and physical interaction between the envelope and membrane proteins. *J. Biol. Chem.* **276**, 17515–17523.
9. Maeda, J., Maeda, A. & Makino, S. (1999). Release of coronavirus E protein in membrane vesicles from virus-infected cells and E protein-expressing cells. *Virology*, **263**, 265–272.
10. Fischer, F., Stegen, C. F., Masters, P. S. & Samsonoff, W. A. (1998). Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. *J. Virol.* **72**, 7885–7894.
11. Raamsman, M. J., Locker, J. K., de Hooge, A., de Vries, A. A., Griffiths, G., Vennema, H. & Rottier, P. J. (2000). Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E. *J. Virol.* **74**, 2333–2342.
12. An, S., Chen, C. J., Yu, X., Leibowitz, J. L. & Makino, S. (1999). Induction of apoptosis in murine coronavirus-infected cultured cells and demonstration of E protein as an apoptosis inducer. *J. Virol.* **73**, 7853–7859.
13. Chen, C. J., An, S. & Makino, S. (2001). Induction of apoptosis in murine coronavirus-infected 17Cl-1 cells. *Advan. Expt. Med. Biol.* **494**, 615–620.
14. Arkin, I. T. & Brunger, A. T. (1998). Statistical analysis of predicted transmembrane alpha-helices. *Biochim. Biophys. Acta*, **1429**, 113–128.
15. Maeda, J., Repass, J. F., Maeda, A. & Makino, S. (2001). Membrane topology of coronavirus E protein. *Virology*, **281**, 163–169.
16. Braiman, M. S. & Rothschild, K. J. (1988). Fourier transform infrared techniques for probing membrane protein structure. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 541–570.
17. Arkin, I. T., MacKenzie, K. R. & Brunger, A. T. (1997). Site-directed dichroism as a method for obtaining rotational and orientational constraints for oriented polymers. *J. Am. Chem. Soc.* **119**, 8973–8980.
18. Salditt, T., Li, C., Spaar, A. & Mennicke, U. (2002). X-ray reflectivity of solid supported, multilamellar membranes. *Eur. Phys. J.* E**7**, 105–116.
19. Kukol, A., Torres, J. & Arkin, I. T. (2002). A structure for the trimeric MHC class II-associated invariant chain transmembrane domain. *J. Mol. Biol.* **320**, 1109–1117.
20. Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T. & Engelman, D. M. (2000). Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nature Struct. Biol.* **7**, 154–160.
21. Choma, C., Gratkowski, H., Lear, J. D. & DeGrado, W. F. (2000). Asparagine-mediated self-association of a model transmembrane helix. *Nature Struct. Biol.* **7**, 161–166.
22. Adams, P. D., Arkin, I. T., Engelman, D. M. &

Brunger, A. T. (1995). Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nature Struct. Biol.* **2**, 154–162.
23. Corse, E. & Machamer, C. E. (2002). The cytoplasmic tail of infectious bronchitis virus E protein directs Golgi targeting. *J. Virol.* **76**, 1273–1284.
24. Godet, M., L'Haridon, R., Vautherot, J. F. & Laude, H. (1992). TGEV coronavirus ORF4 encodes a membrane protein that is incorporated into virions. *Virology*, **188**, 666–675.
25. Harrick, N. (1967). *Internal Reflection Spectroscopy*, 1st edit., Interscience, New York.
26. Tsuboi, M. (1962). Infrared dichroism and molecular conformation of a-form poly-*g*-benzyl-L-glutamate. *J. Polym. Sci.* **59**, 139–153.
27. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P. & Grosse-Kunstleve, R. W. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D Biol. Crystallog.* **54**, 905–921.
28. Engelman, D. M., Steitz, T. A. & Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321–353.
29. Stevens, T. J. & Arkin, I. T. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins*, **39**, 417–420.
30. Kovacs, F. A. & Cross, T. A. (1997). Transmembrane four-helix bundle of influenza A M2 protein channel: structural implications from helix tilt and orientation. *Biophys. J.* **73**, 2511–2517.
31. Carson, M. & Bugg, C. E. (1986). Algorithm for ribbon models of proteins. *J. Mol. Graph.* **4**, 121–122.

Edited by G. von Heijne

(Received 23 April 2004; received in revised form 8 June 2004; accepted 8 June 2004)