A Trimerizing GxxxG Motif Is Uniquely Inserted in the Severe Acute Respiratory Syndrome (SARS) Coronavirus Spike Protein Transmembrane Domain†

Eyal Arbely, Zvi Granot, Itamar Kass, Joseph Orly, and Isaiah T. Arkin*

The Alexander Silberman Institute of Life Sciences, Department of Biological Chemistry, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat-Ram, Jerusalem 91904, Israel

Received May 13, 2006; Revised Manuscript Received July 20, 2006

ABSTRACT: In an attempt to understand what distinguishes severe acute respiratory syndrome (SARS) coronavirus (SCoV) from other members of the coronaviridae, we searched for elements that are unique to its proteins and not present in any other family member. We identified an insertion of two glycine residues, forming the GxxxG motif, in the SCoV spike protein transmembrane domain (TMD), which is not found in any other coronavirus. This surprising finding raises an "oligomerization riddle": the GxxxG motif is a known dimerization signal, while the SCoV spike protein is known to be trimeric. Using an in vivo assay, we found that the SCoV spike protein TMD is oligomeric and that this oligomerization is driven by the GxxxG motif. We also found that the GxxxG motif contributes toward the trimerization of the entire spike protein; in that, mutations in the GxxxG motif decrease trimerization of the full-length protein expressed in mammalian cells. Using molecular modeling, we show that the SCoV spike protein TMD adopts a distinct and unique structure as opposed to all other coronaviruses. In this unique structure, the glycine residues of the GxxxG motif are facing each other, enhancing helix−helix interactions by allowing for the close positioning of the helices. This unique orientation of the glycine residues also stabilizes the trimeric bundle during multi-nanosecond molecular dynamics simulation in a hydrated lipid bilayer. To the best of our knowledge, this is the first demonstration that the GxxxG motif can potentiate other oligomeric forms beside a dimer. Finally, according to recent studies, the stabilization of the trimeric bundle is linked to a higher fusion activity of the spike protein, and the possible influence of the GxxxG motif on this feature is discussed.

A previously unidentified member of the coronaviridae has been shown to be the etiologic agent responsible for the recent severe acute respiratory syndrome (SARS)[1] outbreak (*1*). Viral genome sequencing, in combination with protein phylogenetic analyses have shown SARS coronavirus (SCoV) to belong to a new subfamily within the coronaviridae (*2*, *3*).

As enveloped viruses, coronaviridae contain at least three membrane proteins: a small membrane protein (E), an integral membrane protein (M), and the spike protein (S or E2), a class I viral fusion protein (*4*). In an attempt to understand some of the features that distinguish SCoV from other coronaviridae, we performed detailed sequence comparisons between each of the above membrane proteins to those found in other coronaviruses. Interestingly, we have found that the transmembrane domain (TMD) of the SARS spike protein contains an insertion of two glycine residues forming the known GxxxG motif, a unique feature of the SCoV that is the subject of this work.

Coronavirus spike proteins are large multifunctional, homo-oligomeric proteins (*5*−*9*), forming petal-shaped spikes on the virus envelope. It is this morphological feature from which coronaviruses derive their name (*10*). These type-I membrane proteins are responsible for attachment and entry of the virus into the target cell membrane, thereby determining tissue tropism and host specificity (*11*−*13*). The spike protein is also the predominant antigenic determinant of the virus (*10*, *14*).

Structurally, the SCoV spike protein contains 1255 residues and, by analogy to other coronavirus spike proteins, is divided into two functional domains, S1 and S2, comprising the N- and C-terminal halves, respectively. S1 is responsible for binding to cellular receptors, thereby determining host range, while the membrane-anchored S2 is important for viral entry into cells. Unlike other class I fusion proteins, the SCoV S protein is not post-translationally cleaved into the S1 and S2 domains (*8*). In contrast, it was suggested that, after binding to cellular receptors, the SCoV spike protein undergoes proteolysis within endosomes (*15*).

The oligomeric nature of other coronavirus spike proteins was initially shown to be homotrimeric (*5*), although other reports in the literature have shown that the S1 domain forms dimers (*6*). More recent studies on the SCoV spike protein have clearly shown it to be homotrimeric (*8*), but there is

† This work was supported in part by a grant from the Israel Science Foundation (784/01) to I.T.A.

* To whom correspondence should be addressed. E-mail: arkin@cc.huji.ac.il. Telephone/Fax: +972-2-658-4329.

[1] Abbreviations: AMP, ampicillin; CAT, chloramphenicol acetyl transferase; CNS, crystallography and NMR system; DMPC, dimyristoylphosphocholine; MBP, maltose-binding protein; MD, molecular dynamics; PAGE, polyacrylamide gel electrophoresis; rmsd, root-mean-square deviation; SARS, severe acute respiratory syndrome; SCoV, SARS coronavirus; SDS, sodium dodecyl sulfate; TLC, thin-layer chromatography; TMD, transmembrane domain.

10.1021/bi060953v CCC: $33.50
Published on Web 08/31/2006

Table 1: Transmembrane Sequences Used in the TOXCAT System (*24*)[a]

plasmid name	sequence	residues
GPAwt	LIIFGVMAGVIGTIL	75−89
G83I	LIIFGVMAIVIGTIL	75−89
pccKan	no TM	
SARSwt	YIKWPWYVWLGFIAGLIAIVMVTILL	1191−1216
G1201I	YIKWPWYVWLIFIAGLIAIVMVTILL	1191−1216
G1201,1205I	YIKWPWYVWLIFIAILIAIVMVTILL	1191−1216
HCVwt	YVKWPWYVWLLICLAGVAMLVLLFFI	1293−1318
MHVwt	YVKWPWYVWLLIGLAGVAVCVLLFFI	1313−1338
G83I+	YIKWPWYVWLLIIFGVMAIVIGTIL	SARS(1191−1200) + GPA(75−89)
G83I+ran	PYIYLVWWKWLIIFGVMAIVIGTIL	

[a] The indicated sequences were cloned between the ToxR cytoplasmic DNA-binding domains and MBP, to create the ToxR(TM)MBP chimera, used for measuring the relative strength of oligomerization in the *E. coli* inner membrane (*24*).

some evidence that part of the soluble S1 domain forms dimers (*7*). Moreover, it was shown that trimerization is a function of 60 amino acids at the C-terminal region of the protein, encompassing the cytoplasmic tail and the TMD (*8*). Herein, we are able to explicitly delineate the structural elements that potentiate this trimerization reaction.

In relation to other proteins in the virus, the spike protein (mostly the S1 segment) is not as conserved: 24% identity to other coronaviruses, versus an average 39% identity for all SCoV proteins (*2*, *3*). However, it does contain a highly conserved juxtamembranuous sequence element in S2, part of the known PFAM motif: PF01601. It is this sequence element next to which the unique GxxxG motif was identified.

The GxxxG motif is a known transmembrane dimerization signal (*16*−*21*). The gap of three amino acids between the glycine residues in an α-helical structure aligns the two glycines on the same face of the helix. This alignment creates a flat platform, to which the two glycine residues from another helix can associate and form a symmetric homodimer. In this close positioning, the glycine residues can donate their C_α hydrogen to form a hydrogen bond with a carbonyl oxygen atom from an adjacent helix (*22*). Such a C_α−H···O hydrogen bond can stabilize the helix−helix interactions by 0.88 kcal/mol (*23*).

Here, we report the influence of the GxxxG motif upon the oligomerization of the SCoV spike protein TMD and upon the trimerization of the full-length protein. We also present a structural model, demonstrating the effect of this motif on the trimeric structure of the SCoV spike protein TMD, compared to other coronaviruses. As opposed to all other cases reported in the literature, this is the first example of the involvement of the GxxxG motif in a trimerization reaction.

MATERIALS AND METHODS

Plasmid Construction. Molecular cloning was carried out using standard procedures (*36*). The cloning of the plasmid for the ToxR experiments proceeded as follows: the sense oligonucleotide corresponding to the sequence described in Table 1 was synthesized including 5′ *Nhe*I and 3′ *Bam*HI restriction sites. Double-stranded inserts were synthesized using a Taq-polymerase enzyme, after hybridizing a corresponding anti-sense 30-mer to the 3′ extension of the sense oligonucleotide. The resulting purified double-stranded products were digested with *Nhe*I and *Bam*HI and gel-purified. The expression vector pccKAN (a kind gift of D. M. Engelman) (*24*) was digested with *Nhe*I and *Bam*HI and isolated by gel purification. The inserts were ligated in-frame into the digested vector using phage T4 DNA ligase, at 16 °C for about 18 h. The resulting plasmids, encoding the ToxR(TM)maltose-binding protein (MBP) chimera were transformed into *Escherichia coli* DH5α competent cells and plated on Luria−Bertani (LB) agar containing 200 μg/mL ampicillin (AMP). All assays were preformed on transformed *E. coli* NT326 cells (a kind gift of D. M. Engelman) expressing the ToxR(TM)MBP chimera. This strain lacks the endogenous MBP (*mal*E$^-$), resulting in its inability to transport maltose into the cytoplasm.

Maltose Complementation Assay. NT326 cells expressing the ToxR(TM)MBP chimera were cultured on M9 minimal medium plates containing 0.4% maltose as the only carbon source, 1% Bacto agar (Difco, MD), and 100 μg/mL AMP, for 3 days at 37 °C. Transformed cells expressing a chimera not containing a TMD (pcckan) were used as a control. The *mal*E$^-$ NT326 strain is unable to grow on media in which the only carbon source is maltose. Proper orientation of the ToxR(TM)MBP protein in the inner membrane will localize the MBP portion in the periplasm, allowing it to complement the NT326 *mal*E$^-$ phenotype and support growth on maltose as the only carbon source. As seen in Figure 2, all constructs enabled the survival on M9-maltose medium, indicating correct membrane insertion and orientation. As expected, control cells expressing a construct with no TMD (pccKan) were unable to grow on M9-maltose media.

Chloramphenicol Acetyl Transferase (CAT) Assay. In a typical experiment, competent NT326 cells were transformed with the appropriate plasmid and plated on LB agar plates containing 200 μg/mL AMP. Three different colonies were incubated overnight at 37 °C in LB medium containing 200 μg/mL AMP, diluted 100-fold into fresh medium, and regrown at 37 °C to mid-logarithmic phase. A total of 200 μL of cell culture normalized to an OD_{600} of 0.6 was pelleted and resuspended in 500 μL of 100 mM Tris·HCl (pH 8.0). Lysis was achieved by the addition of 20 μL of lysis buffer [100 mM dithiothreitol (DTT), 100 mM ethylenediaminetetraacetic acid (EDTA), and 50 mM Tris·HCl at pH 8.0] and 40 μL of toluene followed by incubation at 30 °C for 30 min. A total of 5 μL of cell lysate was diluted 1:10 with 100 mM Tris·HCl at pH 8.0, and CAT activity was measured with 10 μL of the diluted cell lysate, in the presence of 4 mM/h acetyl CoA and [^{14}C]chloramphenicol (Amersham Biosciences, U.K., 0.05 μCi/assay). Reactions proceeded by incubation at 37 °C for 20 min and were terminated by the

addition of 200 μL of 250 mM Tris·HCl at pH 7.5 and extraction with 250 μL of ethyl acetate. Except for the incubation time at 37 and 30 °C, reaction mixtures were always kept on ice. [^{14}C]chloramphenicol and its acetylated products were separated by thin-layer chromatography (TLC) on silica-gel plates (60 F_{250}, Merck, Germany) using a solvent system composed of chloroform/methanol (95:5, v/v). The amount of radioactivity in the substrate and its acetylated products was quantified using a phosphorimager (Fuji Bio-Imaging analyzer, BAS-100). Results are represented as the percentage of [^{14}C]chloramphenicol converted to the acetylated products.

Expression in Tissue-Culture Cells. The plasmid expressing full-length SCoV spike protein was a kind gift of Prof. G. J. Nabel from the National Institutes of Health (*13*). The sequence is according to accession number AY278741 (strain Urbani), and the protein is expressed under the cytomegalovirus (CMV) promoter/enhancer. The plasmid encoding the double-mutated G1201,1205I S protein was generated using the QuickChange-XL kit (Stratagene) from the plasmid encoding the wild-type (wt) sequence. HEK 293T cells were maintained in Dulbecco's modified Eagle's medium (Beit Haemek) containing penicillin (100 units/mL), streptomycin (0.1 mg/mL), and amphotericin (0.25 mg/mL) and supplemented with 10% fetal bovine serum. The cells were transfected by using Escort V Transfection Reagent (Sigma, St. Louis, MO) according to the instructions of the manufacturer. A total of 40 h post-transfection, the cells were washed twice with ice-cold phosphate-buffered saline (PBS) and lysed with 1× lysis buffer (50 mM Tris·HCl at pH 7.4, 150 mM NaCl, 1 mM EDTA, and 1% Triton X-100) containing complete mini protease inhibitor (Roche, Indianapolis, IN). After a 30 min incubation on ice, cell debris were cleared by centrifugation. The cleared lysate was used for Western blot analysis.

The cell lysate was diluted with sample buffer [60 mM Tris·HCl at pH 6.8, 100 mM DTT, 3% sodium dodecyl sulfate (SDS), 10% glycerol, and 0.01% bromophenol blue] and heated for 5 min in either 80 or 100 °C, as indicated. Cell lysates were resolved by SDS−4% polyacrylamide gel electrophoresis (SDS−4% PAGE) and transferred to a nitrocellulose membrane. The membrane was blocked with blocking buffer (5% bovine serum albumine and 0.2% Tween 20 in PBS) for 1 h in room temperature. The blot was further incubated over night at 4 °C with mouse anti-SARS S protein monoclonal antibody (Zymed, San Francisco, CA) dilluted 1:300 in blocking buffer. After the membrane was washed 4 times with washing buffer (0.2% Tween 20 in PBS), it was incubated with horseradish-peroxidase-conjugated goat anti-mouse antibody (Jackson, Baltimore, PA) for 1 h at room temperature and washed. Detection was performed with ECL reagent (Beit Haemek, Israel) and quantified using Intelligent Dark Box II (LAS-1000, Fujifilm, Japan) and Image Gauge software (Fujifilm, Japan).

Molecular Modeling. A global search, in which the helices were rotated around their axis, was carried out as described in detail elsewhere (*28*). Briefly, all calculations were performed using the crystallography and NMR system (CNS, version 1.1) (*37*) assuming initial symmetrical interaction between the helices in the homo-oligomer. The OPLS parameter set with a united atom topology was used, representing explicitly polarand aromatic hydrogens (*38*). Calculations were carried out using the following sequences, corresponding to the predicted TMD of the spike protein (see Figure 6): SARS strain Urbani (P59594), residues W1194−L1210; bovine coronavirus (BCoV) strain LY-138 (P25192), residues W1306−I1328; murine hepatitis virus (MHV) strain 4 (P22432), residues W1316−I1338; human coronavirus (HCoV1) strain OC43 (P36334), residues W1296−I1318; human coronavirus (HCoV2) strain 229E (P15423), residues W1114−C1136; avian infectious bronchitis virus (AIBV1) strain 6/82 (P05135), residues W1093−I1115; avian infectious bronchitis virus (AIBV2) strain KB8523 (P12650), residues W1092−V1114; porcine respiratory coronavirus (PRCoV) strain 86/137004 (P27655), residues W1165−C1187; porcine epidemic diarrhea virus (PEDV) strain CV777 (NP_598310), residues W1323−C1345; feline infectious peritonitis virus (FIPV) strain 79-1146 (P10033), residues W1392−C1414.

All calculations were carried out in vacuo with the initial coordinates of a canonical α helix (3.6 residues/turn). Symmetric trimers were generated from the various sequences by duplicating and rotating the helix by 360°/3. An initial crossing angle of 25° for left-handed and −25° for right-handed structures was introduced by rotating the long helix axis with respect to the long bundle axis. The symmetrical search was carried out by rotating all of the helices simultaneously between $\phi = 0°$ and 360° in 10° steps. Four trials were carried out for each starting structure, using different initial random atom velocities. This procedure results in a total of 288 different structures (i.e., $36 \times 4 \times 2$). Each structure was subjected to a simulated-annealing and energy-minimization protocol.

The resulting structures were grouped in clusters, defined by having more than 10 similar structures and that the C_α root-mean-square deviation (rmsd) of the coordinates between every two structures within a cluster was not larger than 1.0 Å. For each cluster, an average structure was calculated, energy-minimized, and subjected to a simulated-annealing protocol identical to that used in the systematic search. This average structure was taken as a representative of each cluster.

Hydrated Lipid Simulations. The simulation presented herein was conducted using version 3.2.1 of the GROMACS (*39*) molecular dynamics (MD) simulation package (www.gromacs.org). An extended version of the GROMOS96 force field (*40*), implemented in GROMACS, was used. Dimyristoylphosphocholine (DMPC) force-field parameters were as described elsewhere in detail (*41*). In all simulations, the LINCS algorithm (*42*) was used to constrain bond lengths. A simulation time step of 2 fs was used, and atomic coordinates were saved every 0.5 ps. The simulation was conducted at a constant temperature of 310 K. Solvent, lipid, and protein were coupled separately to a Berendsen temperature bath (*43*), with a coupling constant of $\tau = 0.1$ ps. The pressure was kept constant by coupling the system to a Berendsen pressure bath (*43*) of 1 bar, with a coupling constant of $\tau = 3$ ps. The membrane plane (*xy*) and membrane normal (*z*) directions were separately coupled. A cutoff of 1 nm was used for van der Waals interactions. Electrostatic interactions were computed using the PME algorithm (*44*), with a 1 nm cutoff for the direct space calculation.

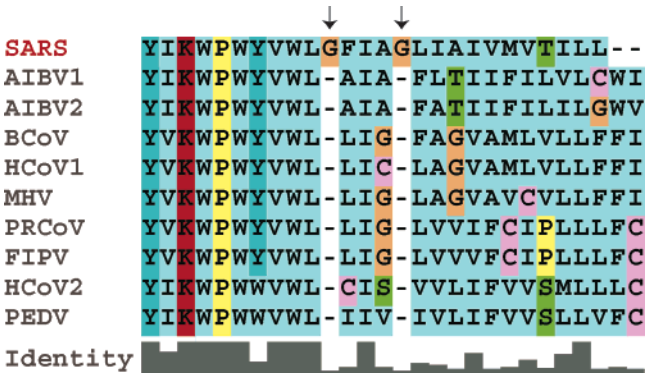

FIGURE 1: Multiple sequence alignment of the putative spike protein TMD of 10 different coronaviruses. The SCoV spike protein TMD corresponds to residues Y1194−L1216. The accession numbers for each of the sequences are given in the Materials and Methods. The coloring is according to the chemical nature of the amino acid. The arrows indicate the insertion of the glycine residues in the SCoV spike protein TMD.

The membrane environment in which the proteins exist was mimicked using a lipid bilayer initially made of 128 DMPC lipids (downloaded from Prof. Tieleman D.P. site; http://moose.bio.ucalgary.ca/) embedded in 3655 molecules of SPC water (*45*). To set up the protein-membrane systems, a hole was generated in the bilayer using each protein solvent-accessible surface as a template. This methodology has been described in more detail elsewhere (*46*).

The protein-bilayer-solvent system was energy-minimized followed by equilibration in two stages: (i) 0.5 ns of MD run with positional restraints on the protein and DMPC atoms and (ii) 0.5 ns of MD run with positional restraints on the protein atoms, to improve the packing of lipids around the protein. Thereafter, the equilibrated system was subjected to 7.5 ns of MD runs with hydrogen-containing bonds constrained. The simulation was set up and performed on a Dual 2 GHz G5 workstation (Apple, Inc.). The time for one dual CPU machine was less than 21 h/ns of simulation.

RESULTS AND DISCUSSION

In this work, we try to find whether the GxxxG motif of the SCoV TMD is involved in oligomeric structures other than dimers and what is the structural outcome of this involvement. The contradiction between the classical role of the GxxxG motif as a dimerization signal and the trimeric structure of the SCoV spike protein is more intriguing when considering the fact that the TMD and carboxy terminus of the spike protein play a role in stabilizing the trimeric structure of the protein (*7*−*9*).

SCoV S Contains an Insertion of Two Glycine Residues in Its TMD. Sequence alignments of the coronavirus spike proteins revealed that, unique to the SARS virus, two glycine residues are inserted into the membrane segment of the spike protein. Moreover, the insertion is adjacent to a highly conserved sequence segment, as shown in Figure 1. The exact spacing of three amino acids between the two glycine residues identifies the insertion as the GxxxG motif, a known transmembrane dimerization signal (*16*−*21*).

The identification of an inserted GxxxG motif in the SCoV spike protein TMD poses an immediate "oligomerization riddle": the GxxxG motif is a known dimerization signal (*16*, *17*), while the SCoV spike protein is known to be trimeric (*8*). Other questions that come up from this finding are what, if any, is the role of the SCoV spike protein TMD in the trimerization reaction and what is the significance of the unique insertion of the GxxxG motif on the biology of SCoV? Below, we attempt to experimentally address the first two questions and discuss the third.

SCoV Spike Protein TMD Oligomerization in Bacterial Membranes. The effects of inserting two glycine residues, forming the GxxxG motif, upon the oligomerization of the SCoV spike protein TMD, were investigated experimentally using the ToxR/TOXCAT system (*18*, *24*). In this system, the TMD of interest is fused between the cytoplasmic domain of the *Cholera vibrio* transcriptional activator (ToxR) and the periplasmic domain of the MBP. Oligomerization of the resulting chimera, driven by the TMD, results in transcription of the reporter gene, CAT, the relative concentration of which can be quantitated by a simple CAT assay.

The different transmembrane segments analyzed in this study by the TOXCAT assay are indicated in Table 1. A monomeric mutant of human glycophorin A (G83I) (*24*) was used as a control for basal CAT activity arising even without transmembrane oligomerization. The homo-oligomerization of wild-type SARS (SARSwt) was compared to the singly mutated G1201I and doubly mutated G1201,1205I mutants. Wild-type human coronavirus (HCVwt, strain OC43) and wild-type murine hepatitis virus (MHVwt, strain 4) were chosen as representatives of other coronaviruses. The plasmid G83I+ was constructed from the TMD of glycophorin A containing the mutation G83I (*24*) with the addition of the 10 highly conserved residues (1191−1200) of the SCoV spike protein, to investigate the possible role of that conserved sequence in oligomerization. The same 10 conserved residues were also added in random order to the TMD of G83I, to create the G83I+ran, which served as a control.

Orientation and localization of the various transmembrane domains were confirmed by a maltose complementation assay. Indicated proteins were expressed in the NT326 strain that lacks the endogenous MBP (*mal*E$^-$), resulting in its inability to grow on media in which the only carbon source is maltose. Proper orientation of the ToxR(TM)MBP protein in the inner membrane will localize the MBP portion in the periplasm, allowing it to complement the NT326 *mal*E$^-$ phenotype and support growth when maltose is the only carbon source. As seen in Figure 2, all constructs were allowed for survival on M9-maltose medium, indicating the correct membrane insertion and orientation. As expected, cells expressing a construct with no TMD (pccKan) were unable to grow on M9-maltose media and served as a negative control.

As shown in Figure 3, cells expressing a chimera containing the SARSwt TMD exhibited significantly higher levels of CAT activity than did cells containing the G83I plasmid, which was taken to represent basal CAT activity (see above). The results indicate for the first time that the TMD of the SCoV spike protein is capable of oligomerization.

The extent of SCoV spike protein TMD oligomerization was lower than that found for the TMD of human glycophorin A, a highly dimerizing element that retains oligomerization even in the presence of SDS micelles (*25*). However, the oligomerization observed for the SARS spike protein TMD is similar or larger than that found in other proteins such as the TMD of the ErbB receptor (*26*). Moreover, recent

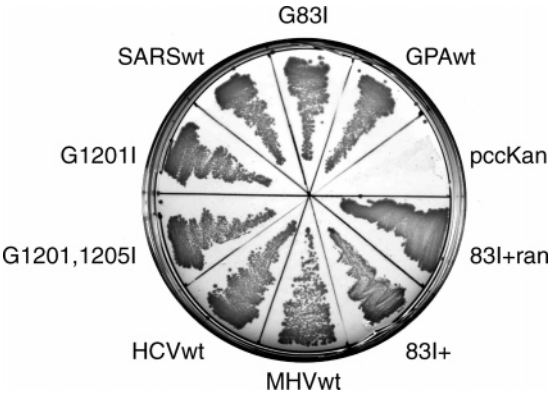

FIGURE 2: *mal*E$^-$ complementation assay to test for proper insertion and orientation of the ToxR(TM)MBP chimeric proteins. NT326 cells expressing the indicated chimera were cultured on M9-maltose plates. Survival of the cells indicates that the chimera is correctly inserted into the membrane, as can be seen for all of the constructs. Cells expressing the pccKan plasmid, which contains no TMD, served as a control.

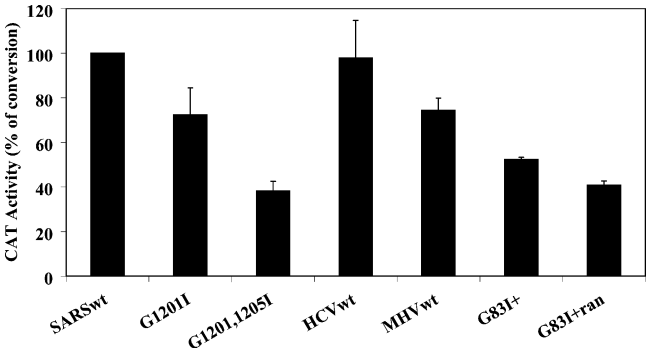

FIGURE 3: Quantitative CAT assay of transmembrane homo-oligomerization employing the TOXCAT system (*24*). CAT assays were performed on normalized quantities of lysates. Bars represents the mean CAT activity ± standard deviation of at least three independent measurements, after subtraction of the basal CAT activity of the G83I mutant and relative to SARSwt.

studies have shown that CAT activity, taken to represent oligomerization in the TOXCAT assay (*24*), is dependent upon the length of the transmembrane segment (*27*). Therefore, comparisons should only be made between transmembrane sequences of equal length. For this reason, the level of wild-type SCoV TMD oligomerization was set as 100% for all future comparisons.

Importance of the GxxxG Motif. To ascertain the importance of the GxxxG motif in the oligomerization of the SCoV spike protein TMD, the CAT activity of cells containing the G1201I chimera was measured. The results were indicative of a 28% loss of oligomerization relative to SARSwt, because of the mutation of a single glycine in the motif. Furthermore, mutation of the second glycine (G1201I + G1205I) reduced the apparent oligomerization by a further 33% (total reduction of 61%). Thus, it can be concluded that not only is the SCoV spike protein TMD oligomeric but that the uniquely inserted GxxxG motif plays a pivotal role in the oligomerization.

Lack of Influence of the Juxtamembranous Part of the PF01601 Motif upon Oligomerization. The influence of the highly conserved juxtamembranous sequence element (part of the PFam PF01601 motif) upon oligomerization was analyzed as well, using a chimera in which it was added to the G83I monomeric mutant of glycophorin A (representing the oligomerization basal level). The oligomerization of this

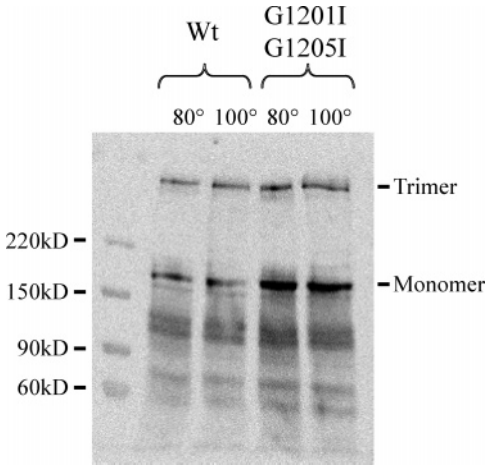

FIGURE 4: Western blot SDS−PAGE of the SCoV wild-type protein and G1201,1205I mutant expressed in HEK293T cells. Anti-SCoV spike protein antibodies were used to detect the presence of the protein. Samples were heated to either 80 or 100 °C prior to electrophoresis as indicted. The sizes of the molecular-weight markers as well as the likely positions of the spike protein monomer and trimer are indicated.

construct was similar to that of the G1201,1205I mutant and smaller than SARSwt TMD by 46%. Most importantly, no significant difference was seen upon the addition of a randomized sequence of the aforementioned sequence element. These results indicate that the difference in the oligomerization state between the chimeric proteins and the G83I mutant is length-dependent and not sequence-dependent. Thus, the highly conserved juxtamembranous sequence element does not seem to contribute significantly toward oligomerization of the TMD of the SCoV spike protein.

Oligomerization of the TMD of the Human Coronavirus Spike Protein. Interestingly, despite the fact that it does not contain a GxxxG motif, the TMD of the human coronavirus (HCVwt) spike protein also seems to be oligomeric to an extent similar to wild-type SCoV TMD (see lane 4 in Figure 3). The murine hepatitis virus (MHVwt) spike protein on the other hand is about 70% oligomeric relative to SARS (lane 5 in Figure 3). It may be possible to gain insight into this second contribution toward oligomerization in coronaviruses by noticing that only three amino acids differ between HCoV1 and MHV (Figure 1). One obvious possibility is that C1305 in HCoV participates in an interhelical disulfide bond, not found in MHV. However, it is clear that even without the GxxxG motif, transmembrane segments of other coronavirus spike proteins are able to oligomerize, albeit differently from the SCoV spike protein.

Importance of the GxxxG Motif in Trimerization of the SCoV Spike Protein Expressed by Mammalian Cells. As stated above, all data in the literature regarding the GxxxG motif pertain to its involvement in dimerization processes. However, full-length SCoV spike protein is known to be trimeric (*8*, *9*). Because the ToxR/TOXCAT system (*18*, *24*) cannot distinguish between the different oligomeric forms, we have expressed the full-length SCoV spike protein in tissue-culture cells to reconcile this conundrum. In addition, such experiments will also enable us to gauge the importance of the TMD in the context of the entire spike protein.

Toward this end, we have transiently expressed SCoV wild-type and mutant proteins in HEK 293T cells. As shown in Figure 4, SDS−PAGE of the full-length SCoV spike

protein shows that the protein migrates in the gel as two species with a molecular weight of 510 and 170 kD. These species correspond to the molecular weights of the trimeric and monomeric forms of the protein, respectively. Trimerization persists even when heating the samples in 3% SDS sample buffer to 80 or 100 °C prior to electrophoresis. Densitometer quantification indicates that the wild-type protein is 31−35% more trimeric than the mutant in which both glycines were replaced by isoleucine.

Thus, not only does the GxxxG motif contribute toward oligomerization of the isolated TMD of the protein, it also promotes the trimerization of the protein in the context of the entire protein. This analysis allows us to unequivocally state that the GxxxG motif in the SCoV spike protein potentiates trimerization. To the best of our knowledge, this is the first experimental proof that a GxxxG motif participates in trimerization. Moreover, the GxxxG motif in SCoV represents the first transmembrane motif that is important for trimerization.

Molecular Modeling of the SARS Spike Protein TMD. The ability of the GxxxG motif to promote dimerization has been well-established in the literature (for example, see refs *16*−*18* and *20*). Specifically, it was postulated (*22*) and recently shown experimentally (*23*) that the close apposition of transmembrane helices, facilitated by the small size of the glycine, promotes the formation of a C_α−H···O hydrogen bond.

To better understand the implications of the insertion of a GxxxG motif upon the structure of the TMD of SCoV spike protein and what a trimeric GxxxG structure might look like, we turned to molecular modeling. Candidate models for the transmembrane helical bundles were derived using global-searching MD simulations (*28*). This procedure attempts to derive a model for a transmembrane helical bundle by exploring different starting configurations using MD. Specifically, 72 different starting configurations are used, whereby each structure was distinguished by the rotational angle of the helices (in increments of 10°) and by right- or left-handed crossing angles made between the helices (±25°). The process is repeated 4 times, leading to a total number of 36 × 2 × 4 = 288 structures sampled. When global-searching MD simulations are taken together, they yield several candidate structures for the corresponding transmembrane helical bundle.

The process of global search MD was conducted for transmembrane segments of spike proteins from 10 different coronaviruses (Figure 1). This enabled us to locate structures that are found in all of the sequences (*29*) and those that are unique to one sequence. Close inspection of Figure 5 reveals a structure in SCoV that stands out from all other structures in any of the other viral variants. It appears at a rotational pitch angle of 113° and a crossing angle of −18° (see dotted circle). However, aside from its unique rotational and crossing angles, what distinguishes this structure from all of the rest is the close positioning of the helices shown by the color scale.

Examination of this unique structure reveals that the GxxxG motif is located in the bundle center, as shown in Figure 6A. Thus, it may be possible to state that the insertion of the GxxxG motif changes the structure of the TMD of the SCoV spike protein in comparison to the structure adopted by spike proteins from all other coronaviruses.

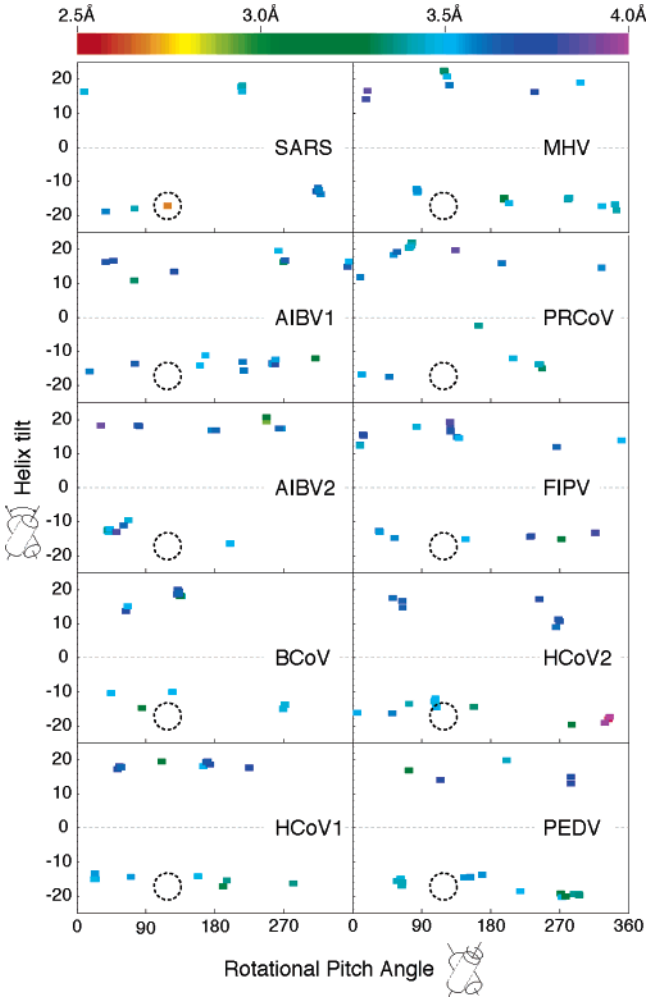

FIGURE 5: Results of the global MD searches (*28*) for 10 different coronavirus spike proteins. TMD trimers plotted as a function of their orientation. The color coding indicates the closest approach between three C_αs in the different helices of the trimer, according to the color scale given at the top of the plot. The dotted gray circle indicates the unique SARS structure, not present in any of the other viruses. Note that it is also the structure in which the helices are the closest among all of the variants. The sequences of the different coronavirus spike protein TMDs are given in Figure 1.

To refine the results obtained in the in vacuo modeling, we further analyzed the unique SARS TMD structure using multi-nanosecond MD simulations of the complex embedded in a fully hydrated phospholipid bilayer (Figure 6C). The 7.5 ns simulation revealed that the structure is stable, leading further credence to the validity of this model, in which the GxxxG motif has been shown, for the first time, to promote trimerization.

Possible Implications upon Viral Function. Considering the fact that the spike protein is responsible for attachment, entry, and antigenicity of the virus, the consequences of inserting an oligomerization domain must be discussed. Because of the health hazard that the SCoV poses, it is difficult, if not impossible, to determine directly the implications of inserting the GxxxG oligomerization motif upon SCoV virulence. However, considering the results presented here and in other recent publications, we would like to conduct a theoretical discussion regarding the possible influence of the GxxxG motif upon viral virulence.

Oligomerization is key to the function of the spike protein (*9*, *10*, *30*); therefore, any modification in its nature, such as

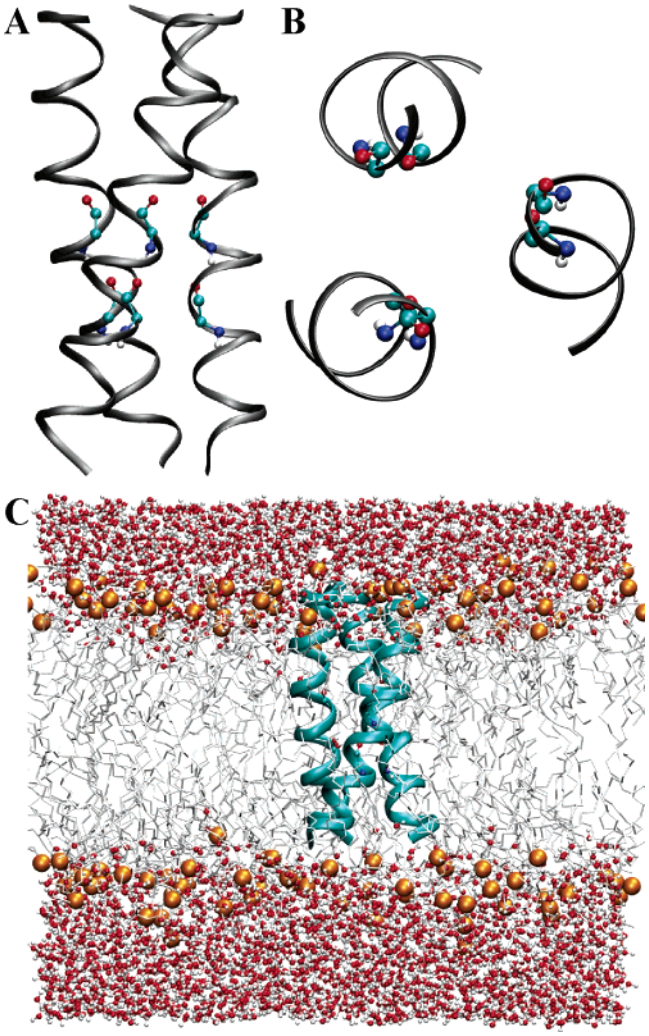

FIGURE 6: Structural model obtained from MD simulations for the SARS spike protein TMD. (A and B) Structures obtained from in vacuo modeling in side and top view, respectively. The glycine residues forming the GxxxG motifs are shown in a ball-and-stick representation. (C) Hydrated lipid bilayer simulation system of the SCoV spike protein TMD. The protein is shown in a ribbon representation in blue, with the two glycine residues in a ball-and-stick representation. The lipid molecules are shown in bond format, with hydrocarbon tails in gray and polar atoms, N, O, and P, in blue, red, and purple, respectively. The solvent water molecules are represented in a ball-and-stick representation. This figure was generated by VMD (*47*) and rendered in PovRay (*48*).

a change in the trimerization surface (i.e., contact regions) or in oligomerization stability, is likely to cause considerable changes in the function of the spike protein. In a recent publication, Corver and co-workers tested the role of the SCoV spike protein TMD during entry to target cells (*9*). The infectivity of the SCoV spike protein contains different TMDs was tested using SCoV spike protein pseudotyped viruses. Also, the cell−cell fusion activity of these proteins was tested using a luciferase-based cell−cell fusion assay.

When the SCoV TMD and cytoplasmic tail where replaced by those originating from MHV, fusion activity was reduced by 40−50%. When the same domains where replaced by those originating from the vesicular stomatitis virus G (VSV-G) protein (also contains the GxxxG motif), infectivity was less than 5% of the wild-type sequence and cell−cell fusion was just 10−30% of the wild-type sequence. In addition, the replacement of only the cytoplasmic domain with the VSV-G cytoplasmic domain had a mild effect, clearly indicating that the TMD is crucial for infectivity and fusion activity. Most importantly, the trimeric structure of the chimeric protein contains the TMD, and the cytoplasmic domain of VSV-G was less stable than all of the other constructs. This result strengthens the assumption that trimerization is linked with higher fusion activity.

Although there is no direct proof for the influence of the trimeric structure stability upon fusion activity, this option was already considered (*30*). Moreover, we found a good correlation between the extent of oligomerization that we measured using the CAT assay and the fusion activity measured by Broer et al. (*9*). In our tests, the TMD of MHV was 70% oligomeric relative to the SCoV wild-type sequence, compared to 50−60% fusion activity measured by Corver and co-workers (*9*).

It should be noted that the VSV-G TMD contains the GxxxG motif and that mutations of the glycines in the GxxxG motif of the VSV spike protein abolished viral infectivity and virulence (*31*).

When these results are taken together, the influence of the oligomerization of the TMD of viral spike proteins upon viral virulence begins to be appreciated. Therefore, our finding that the SCoV spike protein contains an inserted active transmembrane trimerization motif may bear impact upon viral virulence, but this should be further investigated.

Oligomeric Versatility of the GxxxG Motif. As stated above, all data pertaining to the GxxxG sequence signature has thus far implicated the motif in dimeric interactions (for example, see refs *16*, *26*, *27*, and *32*−*35*). Surprisingly, in the current study, we show that not only is a GxxxG motif present in a trimeric transmembrane helical bundle but that it is one of the driving forces of the trimerization reaction. To the best of our knowledge, this is the first demonstration of a sequence motif that drives transmembrane helical trimerization. Finally, it remains to be seen why some GxxxG motifs result in dimerization, while others contribute to trimeric assemblies.

ACKNOWLEDGMENT

The authors wish to thank Prof. D. M. Engelman for the kind gift of the TOXCAT system and Prof. G. J. Nabel for the kind gift of the plasmid containing the SCoV spike protein.

REFERENCES

1. Ksiazek, T. G. et al. (2003) A novel coronavirus associated with severe acute respiratory syndrome, *N. Engl. J. Med. 348*, 1953−1966.
2. Marra, M. A. et al. (2003) The genome sequence of the SARS-associated coronavirus, *Science 300*, 1399−1404.
3. Rota, P. A. et al. (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome, *Science 300*, 1394−1399.
4. Bosch, B. J., van der Zee, R., de Haan, C. A., and Rottier, P. J. (2003) The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex, *J. Virol. 77*, 8801−8811.
5. Delmas, B., and Laude, H. (1990) Assembly of coronavirus spike protein into trimers and its role in epitope expression, *J. Virol. 64*, 5367−5375.
6. Lewicki, D. N., and Gallagher, T. M. (2002) Quaternary structure of coronavirus spikes in complex with carcinoembryonic antigen-related cell adhesion molecule cellular receptors, *J. Biol. Chem. 277*, 19727−19734.

7. Xiao, X., Feng, Y., Chakraborti, S., and Dimitrov, D. S. (2004) Oligomerization of the SARS-CoV S glycoprotein: Dimerization of the N-terminus and trimerization of the ectodomain, *Biochem. Biophys. Res. Commun. 322*, 93−99.
8. Song, H. C. et al. (2004) Synthesis and characterization of a native, oligomeric form of recombinant severe acute respiratory syndrome coronavirus spike glycoprotein, *J. Virol. 78*, 10328−10335.
9. Broer, R., Boson, B., Spaan, W., Cosset, F. L., and Corver, J. (2006) Important role for the transmembrane domain of severe acute respiratory syndrome coronavirus spike protein during entry, *J. Virol. 80*, 1302−1310.
10. Gallagher, T. M., and Buchmeier, M. J. (2001) Coronavirus spike proteins in viral entry and pathogenesis, *Virology 279*, 371−374.
11. Casais, R., Dove, B., Cavanagh, D., and Britton, P. (2003) Recombinant avian infectious bronchitis virus expressing a heterologous spike gene demonstrates that the spike protein is a determinant of cell tropism, *J. Virol. 77*, 9084−9089.
12. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., and Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, *Nature 426*, 450−454.
13. Yang, Z. Y., Huang, Y., Ganesh, L., Leung, K., Kong, W. P., Schwartz, O., Subbarao, K., and Nabel, G. J. (2004) pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN, *J. Virol. 78*, 5642−5650.
14. Zhang, H., Wang, G., Li, J., Nie, Y., Shi, X., Lian, G., Wang, W., Yin, X., Zhao, Y., Qu, X., Ding, M., and Deng, H. (2004) Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies, *J. Virol. 78*, 6938−6945.
15. Simmons, G., Gosalia, D. N., Rennekamp, A. J., Reeves, J. D., Diamond, S. L., and Bates, P. (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry, *Proc. Natl. Acad. Sci. U.S.A. 102*, 11876−11881.
16. Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J., and Engelman, D. M. (1992) Sequence specificity in the dimerization of transmembrane α-helices, *Biochemistry 31*, 12719−12725.
17. Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brunger, A. T., and Engelman, D. M. (1994) A dimerization motif for transmembrane α-helices, *Nat. Struct. Biol. 1*, 157−163.
18. Langosch, D., Brosig, B., Kolmar, H., and Fritz, H. J. (1996) Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator, *J. Mol. Biol. 263*, 525−530.
19. Arkin, I. T., and Brunger, A. T. (1998) Statistical analysis of predicted transmembrane α-helices, *Biochim. Biophys. Acta 1429*, 113−128.
20. Russ, W. P., and Engelman, D. M. (2000) The GxxxG motif: A framework for transmembrane helix−helix association, *J. Mol. Biol. 296*, 911−919.
21. Senes, A., Gerstein, M., and Engelman, D. M. (2000) Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with β-branched residues at neighboring positions, *J. Mol. Biol. 296*, 921−936.
22. Senes, A., Ubarretxena-Belandia, I., and Engelman, D. M. (2001) The C_α−H···O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions, *Proc. Natl. Acad. Sci. U.S.A. 98*, 9056−9061.
23. Arbely, E., and Arkin, I. T. (2004) Experimental measurement of the strength of a C_α−H···O bond in a lipid bilayer, *J. Am. Chem. Soc. 126*, 5362−5363.
24. Russ, W. P., and Engelman, D. M. (1999) TOXCAT: A measure of transmembrane helix association in a biological membrane, *Proc. Natl. Acad. Sci. U.S.A. 96*, 863−868.
25. Bormann, B. J., Knowles, W. J., and Marchesi, V. T. (1989) Synthetic peptides mimic the assembly of transmembrane glycoproteins, *J. Biol. Chem. 264*, 4033−4037.
26. Mendrola, J. M., Berger, M. B., King, M. C., and Lemmon, M. A. (2002) The single transmembrane domains of ErbB receptors self-associate in cell membranes, *J. Biol. Chem. 277*, 4704−4712.
27. Li, R., Gorelik, R., Nanda, V., Law, P. B., Lear, J. D., DeGrado, W. F., and Bennett, J. S. (2004) Dimerization of the transmembrane domain of integrin α_{IIb} subunit in cell membranes, *J. Biol. Chem. 279*, 26666−26673.
28. Adams, P. D., Arkin, I. T., Engelman, D. M., and Brunger, A. T. (1995) Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban, *Nat. Struct. Biol. 2*, 154−162.
29. Briggs, J. A., Torres, J., and Arkin, I. T. (2001) A new method to model membrane protein structure based on silent amino acid substitutions, *Proteins 44*, 370−375.
30. Luo, Z., Matthews, A. M., and Weiss, S. R. (1999) Amino acid substitutions within the leucine zipper domain of the murine coronavirus spike protein cause defects in oligomerization and the ability to induce cell-to-cell fusion, *J. Virol. 73*, 8152−8159.
31. Cleverley, D. Z., and Lenard, J. (1998) The transmembrane domain in viral fusion: Essential role for a conserved glycine residue in vesicular stomatitis virus G protein, *Proc. Natl. Acad. Sci. U.S.A. 95*, 3425−3430.
32. Bustos, D. M., and Velours, J. (2005) The modification of the conserved GXXXG motif of the membrane-spanning segment of subunit G destabilizes the supramolecular species of yeast ATP synthase, *J. Biol. Chem.*, in press.
33. Gerber, D., Sal-Man, N., and Shai, Y. (2004) Two motifs within a transmembrane domain, one for homodimerization and the other for heterodimerization, *J. Biol. Chem. 279*, 21177−21182.
34. Sulistijo, E. S., Jaszewski, T. M., and MacKenzie, K. R. (2003) Sequence-specific dimerization of the transmembrane domain of the "BH3-only" protein BNIP3 in membranes and detergent, *J. Biol. Chem. 278*, 51950−51956.
35. Arselin, G., Giraud, M. F., Dautant, A., Vaillier, J., Brethes, D., Coulary-Salin, B., Schaeffer, J., and Velours, J. (2003) The GxxxG motif of the transmembrane domain of subunit e is involved in the dimerization/oligomerization of the yeast ATP synthase complex in the mitochondrial membrane, *Eur. J. Biochem. 270*, 1875−1884.
36. Sambrook, J., and Russell, D. W., Eds. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
37. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54* (part 5), 905−921.
38. Jorgensen, W. L., and Tirado-Rives, J. (1988) The OPLS potential function for proteins, energy minimization for crystals of cyclic peptides and crambin, *J. Amer. Chem. Soc. 110*, 1657−1666.
39. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis, *J. Mol. Model. 7*, 306−317.
40. van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hunenberger, P. H., Kruger, P., Mark, A. E., Scott, W. R. P., and Tironi, I. G. (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag AG an der ETH Zuerich, Zuerich, Germany.
41. Berger, O., Edholm, O., and Jahnig, F. (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature, *Biophys. J. 72*, 2002−2013.
42. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS. A linear constraint solver for molecular simulations, *J. Comput. Chem. 18*, 1463−1472.
43. Berendsen, H. J. C., Postma, J. P. M., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *J. Chem. Phys. 81*, 3684−3690.
44. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N - \log(N)$ method for Ewald sums in large systems, *J. Chem. Phys. 98*, 10089−10092.
45. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981) Interaction models for water in relation to protein hydration, in *Intermolecular Forces* (Pullman, B., Ed.) D. Reidel Publishing Company, Dordrecht, The Netherlands.
46. Faraldo-Gomez, J. D., Smith, G. R., and Sansom, M. S. (2002) Setting up and optimization of membrane protein simulations, *Eur. Biophys. J. 31*, 217−227.
47. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graphics 14*, 33−38.
48. Persistence of Vision Raytracer Pty. Ltd. (2004) Persistence of Vision Raytracer (version 3.6), http://www.povray.org/.

BI060953V