

Assembly of Viral Membrane Proteins

J. Krüger and W. B. Fischer*

Institute of Biophotonics, School of Medical Science and Engineering, National Yang-Ming University, 155, Section 2, Li-Nong Street, Taipei 112, Taiwan

Received April 16, 2009

Abstract: The generation of computational models is an alternative route to obtain reliable structures for the oligomeric state of membrane proteins. A strategy has been developed to search the conformational space of all possible assemblies in a reasonable time, taking symmetry considerations into account. The methodology tested on M2 from influenza A, shows an excellent agreement with established structures. For Vpu from HIV-1 a series of conformationally distinct structures are proposed. For the first time a structural model for a fully assembled transmembrane part of 3a from SARS-CoV is proposed.

Introduction

Membrane proteins represent a huge challenge in terms of experimental and computational structure generation. As the proteins are located at the lipid–protein interface their structure is adapted to this special environment. This special environment has to be taken care of by modern structural biological techniques including computational methods.

The biophysical properties of the lipid membrane are imposed by the topology of its constituents and generate a hydrophobic core flanked by two hydrophilic slabs, the hydrophilic head-group regions. Despite its complexity this environment confines the dynamics of the proteins mostly into two dimensions. The confinement reduces the conformational search space in computational methods by one dimension and allows for efficient sampling being a highly exhaustive search otherwise.

Membrane proteins from viruses, such as M2 from influenza A,^{1–3} Vpu from HIV-1,^{4,5} and the more recently discovered 3a protein from SARS-Co virus,⁶ which are known to homo-oligomerize, are used to develop a strategy to generate plausible assemblies on an atomic level.

Plausible oligomers for the transmembrane (TM) part of Vpu⁷ had been suggested using a global search protocol.⁸ In the protocol a limited number of structures are generated and subject to a simulated annealing and energy minimization procedure allowing a significant rearrangement of the initial structures. In a similar approach bundles for M2,⁹ Vpu,^{10–13} and the monomeric part of p7 from HCV¹⁴ had been generated

using simulated annealing combined with short molecular dynamics (MD) simulations. A recent study on M2 from influenza A, glycophorin A, and phospholamban employed a replica exchange approach starting from 16 distinct structures using an implicit membrane approximation.¹⁵ Although the structure optimization algorithm is significantly more sophisticated, still only a partial sampling of the conformational space is possible. The relevance of the monomer conformation for the total energetic was examined among other aspects in a study on the glycophorin dimer.¹⁶ With 324 distinct initial structures only a partial coverage of the conformational space can be assumed. The work of Bowie and co-workers on M2 from influenza A, glycophorin A, and phospholamban as well as other TM proteins evaluates the interaction between two initial TM helices with meticulous Monte Carlo simulations.^{17–19} The models are then duplicated around a central symmetry axis to generate larger assemblies.

Mentioned methods probe only a limited number of bundle conformations. In the approach described in this study the search is extended to cover a fine grained range of distances, helical rotation, and variation in tilt angle covering the whole conformational space of the assembly. With this method several hundreds of thousand conformers are obtained for which the potential energy is then calculated and the bundles are ranked accordingly. M2 is taken as a test case to ‘validate’ the quality of the approach. The study includes a first structural model of the 3a protein from SARS-CoV which has been proposed to have three membrane spanning parts.⁶

* Corresponding author phone: +886 - (0)2 2826 - 7394; fax: +886 - (0)2 2823 - 5460; e-mail: wfischer@ym.edu.tw.

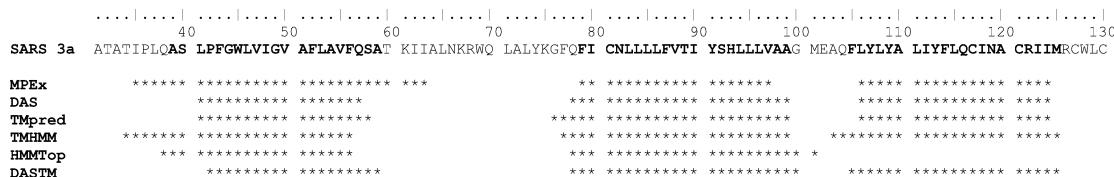


Figure 1. Prediction of the TM parts of 3a from SARS-CoV using different secondary structure prediction programs. The highlighted bold residues reflect the consensus sequence and are used for MD simulations and assembly.

The method can be easily adapted to generate any other membrane protein assembly and thus opens the door for extensive use also in high throughput approaches in proteomics.

Computational Methods

Secondary Structure Prediction and Monomer Modeling. The following ideal helices of M2_{23–43} (SDPLVVAA³⁰ SIIGILHLIL⁴⁰ WIL) (see also refs 20 and 21), Vpu_{1–32} (MQPIPIVAIV¹⁰ ALVVAI²⁰ VVWSIVIIEY³⁰ RK),^{13,22,23} 3a-TM1_{39–59} (AS⁴⁰ LPFGWLVIGV⁵⁰ AFLAVFQSA), 3a-TM2_{79–99} (FI⁸⁰ CNLLLLFVTI⁹⁰ YSHLLLVAAG), and 3a-TM3_{105–125} (FLYLYA¹¹⁰ LIYFLQCINA¹²⁰ CRIIM)⁶ were generated with backbone dihedrals of $\phi = -65^\circ$ and $\psi = -39^\circ$ using the program MOE (Molecular Operation Environment, www.chemcomp.com).

For the prediction of the TM parts of 3a from SARS-CoV different secondary structure prediction programs were used (Figure 1): Membrane Protein Explorer (MPEx, <http://blanco.biomol.uci.edu/mpex/>²⁴), Dense Alignment Surface prediction of TM regions in proteins (DAS, www.enzim.hu/DAS/DAS.html²⁵), TMPred (prediction of transmembrane regions and orientations, www.ch.embnet.org/software/TMPRED_form.html²⁶), TMHMM (prediction of transmembrane helices, www.cbs.dtu.dk/services/TMHMM/²⁷), and HMMTop (prediction of transmembrane helices and topology of proteins, www.enzim.hu/hmmtop/index.html²⁸). They were used with their default setting, and no further adjustments were made. For technical details and algorithm description please refer to the cited literature.

Monomer Equilibration. Prior to any assembly all monomers have been simulated for 10 ns in a fully hydrated POPC bilayer to achieve well equilibrated monomers and to confirm helical stability. The topology for the lipid bilayer (POPC (16:0–18:1 diester PC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) was created on the basis of the parameters of Chandrasekhar et al.²⁹ The stability of the bilayer was confirmed by a 70 ns MD simulation.³⁰

The monomers were inserted into the POPC bilayer, and a stepwise energy minimization and equilibration protocol was used.³⁰

All MD simulations were carried out under GROMACS 3.3.2 with the Gromos96 (ffG45a3) force field. The temperature of the peptide, lipid, and the water molecules were separately coupled to a Berendsen thermostat with a coupling time of 0.1 ps. Full isotropic pressure coupling was applied with a coupling time of 1.0 ps and a compressibility 4.5e-5 bar⁻¹. Long range electrostatics were calculated using the particle-mesh Ewald (PME) algorithm with grid dimensions of 0.12 nm and interpolation order 4. Lennard-Jones and

short-range Coulomb interactions were cut off at 1.4 and 0.8 nm, respectively.

Assembly. For each of the TM parts of the individual proteins the starting structure for the assembly was the average structure of a principal component analysis (PCA) over the backbone atoms of the whole 10 ns equilibrations. PCA was carried out using the program g_covar from the GROMACS-3.3.2 package. The overall rotational and translational motions were removed by fitting the peptide structure of each time frame to the starting structure.

The following sequences for each of the TM parts were used for the assembly: M2_{23–43} (SDPLVVAA³⁰ SIIGILHLIL⁴⁰ WIL), Vpu_{8–26} (AIV¹⁰ ALVVAI²⁰ VVWSIV), 3a-TM1_{39–59} (AS⁴⁰ LPFGWLVIGV⁵⁰ AFLAVFQSA), 3a-TM2_{79–99} (FI⁸⁰ CNLLLLFVTI⁹⁰ YSHLLLVAAG), and 3a-TM3_{105–125} (FLYLYA¹¹⁰ LIYFLQCINA¹²⁰ CRIIM). As partial unwinding and strong interaction of the N- and C-terminal residues with the lipid headgroups occurs it was required to shorten the TM parts for the assembly to focus on the main helical core. The truncated residues were not explicitly blocked or protonated and kept neutral.

The helical backbone structure is aligned along the z-axis. The absolute rotational orientation was irrelevant for the following steps, but for each data set the same orientation was used to retain its consistency. The homo-oligomeric assembly was considered to be symmetrical toward the central pore axis (C4, C5 symmetry). Multiple copies of the starting helix were placed in the xy-plane with respect to interhelical distance, relative rotational angle, and tilt toward the z-axis (here also the membrane normal). The construction of either a trimer, tetramer, or pentamer followed basic geometry with interhelical separation angles of 120°, 90°, and 72°, respectively. The influence of the crossing point, here the point where the xy-plane cuts the starting structure, was also evaluated. To cover weak and tight packing interhelical distances in the range from 8 to 12.5 Å were sampled. Due to symmetry all monomers were rotated around their own helical axis in the same sense with respect to the central pore axis. In the case of hetero-oligomers, e.g. 3a from SARS-CoV it was necessary to sample separate rotation angles for each monomer. As there was no absolute orientation of the monomers with respect to the angle, it was chosen arbitrarily but always in the same way to retain the consistency of the specific data set. A further simplification was to use only one uniform interhelical distance for the 3a trimer.

After each positioning, the side chain atoms were reconstructed with a relative orientation considered as the most probable by the rotational library integrated in MOE. After an energy minimization of not more than 5 steps of either/and steepest descend and conjugated gradient the potential

energy was evaluated according to the united-atom Engh-Huber force field³¹ in vacuum without any solvent or lipid present (see the Supporting Information).

To sample the whole conformational space of the bundles each of the degrees of freedom was varied stepwise (interhelical distance 0.05 Å, rotational angle 2°, and tilt 4°). The actual step width for each degree of freedom was evaluated and adjusted with preliminary runs to balance accuracy and performance. For M2 it was possible to limit the angle search to 120° since His-37 and Trp-41 play an important role in the proton conductance through the pore and have to face inward. The tilt search was restricted to 32 to 42° (positive and negative) and a distance restraint between His-37(N_δ) and Trp-41(C_γ) of 3.9 Å was applied, due to experimental evidence.^{32,33} For Vpu and especially 3a less data were available so a more extensive search had to be carried out. Depending on constraints on the search space hundreds of thousand different conformers were created each characterized by the set of three or more degrees of freedom and the corresponding individual energy value. In this way for M2 147620, for Vpu 343900, and for 3a 3686058 conformers were generated. The small step size guaranteed high accuracy in determining local minima on the complex high dimensional energy landscape of the assembly process. Further details of the algorithm including a detailed Entity-Relationship-Model (ERM)³⁴ are available in the Supporting Information.

The simulations were run on a DELL Precision 490n workstation and on facilities of the Paderborn Center for Parallel Computing PC² (<http://wwwcs.uni-paderborn.de/pc2/>). Plots and pictures were generated using xmGrace, VMD, POV-Ray, and MOE.

Results

The multistep method is driven by the full exploration of the conformational space of the assembly of the TM parts. The steps can be described as following:

- (i) **TM Prediction:** based on either experimental results or a series of secondary structure prediction programs.
- (ii) **Equilibration:** 10 ns MD simulations of the monomer in a fully hydrated lipid bilayer and generation of an averaged structure based on PCA analysis.
- (iii) **Assembly:** selection of the core TM spanning part for the bundle assembly and sampling the whole conformational space along the essential degrees of freedom.

Transmembrane Prediction for 3a. Prior to assembly trials the length and amino acid composition of all TM parts of the protein has to be known. The TM parts of M2 and Vpu are experimentally described in the literature (for reviews see refs 35 and 36). 3a from SARS-CoV has been newly discovered, and structural data are still lacking. Therefore a series of TM prediction programs all basing on different algorithms has been used (Figure 1). All predict three TM parts (TM1, TM2, and TM3) of various lengths. The only exception is HMMTop which predicts only the first two parts. The consensus length of the TM parts is calculated to be of 21 amino acids for each of the parts and predicted to have the following sequence: 3a-TM1_{39–59} (AS⁴⁰

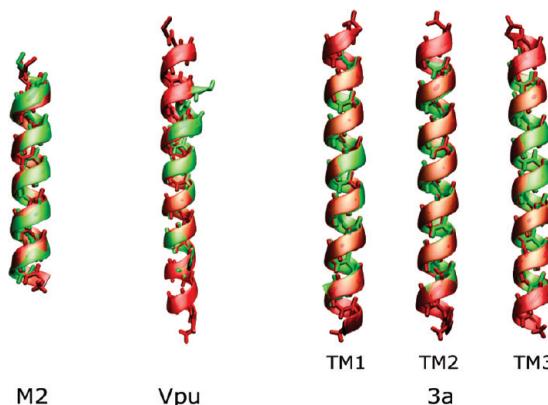


Figure 2. Overlay of the initial ideal helices (red) and the corresponding monomers (green) after MD equilibration which are then used for the assembly protocol.

LPFGWLIGV⁵⁰ AFLAVFQSA), 3a-TM2_{79–99} (FI⁸⁰ CN-LLLLFTI⁹⁰ YSHLLLVA), and 3a-TM3_{105–125} (FLY-YA¹¹⁰ LIYFLQCINA¹²⁰ CRIIM).

Assuming a helical secondary structure the length of the predicted TM parts would correspond to a length of 34 Å which is slightly shorter than a typical lipid thickness of DPPC or POPC with 36 Å. As the protein can develop significant tilt angles it may be the case that further residues interact directly with the lipid, specifically the lipid headgroup.

Equilibration. The monomeric proteins of M2, Vpu, and 3a (each of the three membrane spanning parts separately) had been embedded as ideal helices in a fully solvated POPC bilayer and equilibrated for 10 ns. A principal component analysis (PCA) was carried out on each of the data sets. The eigenvectors of the covariance matrices of positional fluctuation give the direction, while eigenvalues quantify the magnitude of the fluctuation. The average structures derived by this method, which are used for the assembly later, reveal that the helical motif remains intact (Figure 2). Some bending and in the case of Vpu also the development of a kink can be observed (Figure 2, see also ref 30). The deviations from the idealized α-helical starting structure are minor but are expected to have an impact on the packing during pore formation. The Root Mean Square Deviation (rmsd) between the starting structure and the averaged PCA structure based on the C_α-atoms lies within the following range: 0.98 (M2), 0.78 (Vpu), 0.30 (3a TM1), 0.43 (3a TM2), and 0.29 (3a TM3).

At the N- and C-termini minor unwinding can occur, due to strong interaction of polar/charged residues with the lipid headgroup (data not shown). To avoid clashes and artificial bumps during the assembly stage only the core portion of each peptide is used. For the 32 residue Vpu peptide the first and last 6 residues are being omitted, finally using 18 residues. M2 and the three membrane spanning parts of 3a have been used in their original length of 20 or 21 residues.

Assembly. The assembly of multiple monomers to form a pore structure has been carried out in the simulation package MOE. Based on its Scientific Vector Language (SVL) existing functions of MOE have been combined that the monomers can be placed in a defined way around the coordinate origin. The distance between the monomers,

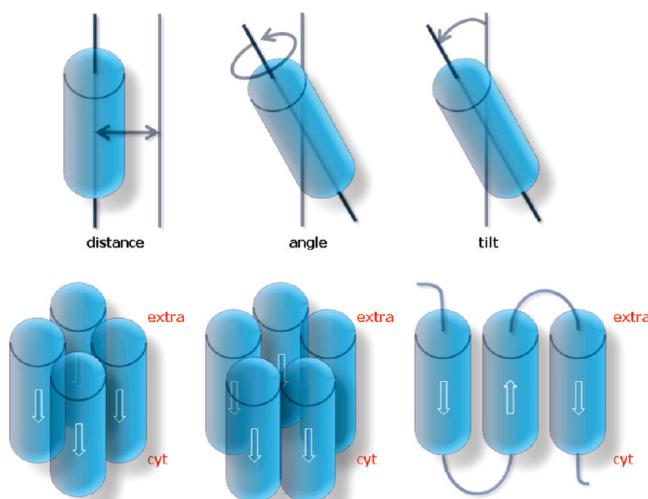


Figure 3. Distance, rotation (angle), and tilt of the TM parts are varied systematically (upper row) to generate e.g. homo tetra- or pentameric bundles of M2 and Vpu and to pack the monomeric trimer of 3a (lower row). White arrows indicate the direction toward the C terminus.

the angle, and the tilt relative to the membrane normal are varied systematically as described in materials and methods (Figure 3).

A major aspect of this approach is to consider homo-oligomers as symmetrical toward their central pore axis (e.g., C4, C5 symmetry). Moreover the dynamics of the monomers are limited to the two-dimensional plane of the lipid bilayer. Therefore it is sufficient to use two degrees of freedom (angle and tilt) to describe the rigid body rotations instead of the usual three Euler angles. These assumptions drastically narrow the search space and enable the creation of reasonable results in an acceptable sampling time, usually hours up to a few days.

Furthermore at this stage it is possible to include restraints, as far as they are known from e.g. NMR experiments, to narrow the search even more.

Distance, Angle, Tilt, and Crossing Point. The distances between packed helices in TM proteins usually show values around 10 Å which is within the range of the 8 to 12.5 Å covered in this study. As each protein has its own characteristics, it is often possible to restrict the distance search to a smaller portion. In the case of the 3a heterotrimer of SARS only one uniform distance was used to simplify the protocol. Preliminary tests have shown that variation in the distance between the three helices is below 0.1 Å and therefore insignificant (data not shown).

The angle for the rotation of each monomer around its own helical axis is sampled for full 360°. Only one value per conformation has to be covered for homo-oligomers, as due to symmetry all monomers are oriented in the same way toward the central pore axis. In cases like M2 from influenza, it is possible to narrow the search space significantly, as it is known from experiments which residues have to be pore lining. His-37 and Trp-41 have been found to play an important role in the proton conductance through the pore^{32,33} by facing inward into the lumen of the pore. This allows narrowing the search by 2/3 to 120°. In the case of hetero-

oligomers for each nonsymmetrical monomer it is required to sample an individual angle, e.g. 3a from SARS.

The tilt describes the orientation of the helical axis toward the membrane normal. As membrane proteins can develop significant tilts up to 50° it is also required to sample this dimension of the conformational space in a sufficient way. One has to distinguish between left handed helices with negative tilt values and right handed with positive tilt values.

Moving the crossing point did not show major influence on M2 or Vpu assembly. Moving it up or down by 2 Å did not affect the position or the depth of minima in the energy landscape. Considering the symmetric pressure profile in lipid bilayers it seems to be very likely that proteins also have to show similar symmetry. Extreme values for the crossing point would lead to a tepee-like conformation, which is unsymmetrical toward the bilayer and only could be created by the presence of rigid extra membrane parts enforcing such an asymmetry. For the proteins studied in this paper this is considered to not be the case.

Influenza-M2. The calculated data for the assembled pore models of M2 indicate one dominant conformation independent of the usage of a distance restraint (Figure 4). This structure is compared to the two available M2 TM part structures derived from NMR data: 1NYJ, which is described as a presumably closed state of M2,³⁷ and 2H95, which represents the open state with a bound channel blocking inhibitor.³⁸

Comparison of the monomers of the structures reveals a high degree of structural overlap based on the calculated rmsd with respect to the C_α atoms in Å (Table 1). The values are in the range of 0.483 to 0.773. The comparison with the ‘closed’ structure (1NYJ) shows lower rmsd values of 0.483 and 0.528 than with the ‘open’ structure (2H95) with rmsd values of 0.773 and 0.735. Internal comparison of the restraint and unrestraint data shows a rmsd value as low as 0.420. It is noteworthy that the two NMR structures differ by 0.789, the highest value in this data set.

All rmsd values for the complete bundles are higher (Table 1). Comparison of the computational generated models and the NMR based structures reveals that the computational models match better with the open ligated NMR model (2.676 with restraints, 2.086 without restraints) than with the closed unligated model (12.765 with restraints, 12.550 without restraints). The difference between the two computed models is the lowest with a value of 1.731 and between the two NMR models the highest with 13.136.

A slight bend found in M2 in the NMR data is reproduced by the computational derived monomer after the 10 ns MD equilibration in POPC (compare Figure 2, further data not shown).

Comparing the equilibrated monomers with experimental structures (1NYJ and 2H95) excellent overlap and very low rmsd values can be observed. When comparing the complete tetrameric pores the values are significantly higher. Taking the size and topology into account, it has to be concluded that rmsd values above 10.0 indicate significant structural differences but not necessarily indicating a dramatically different topology (Figure 5). Although the C_α-rmsd between 1NYJ and the computational models is relatively high with

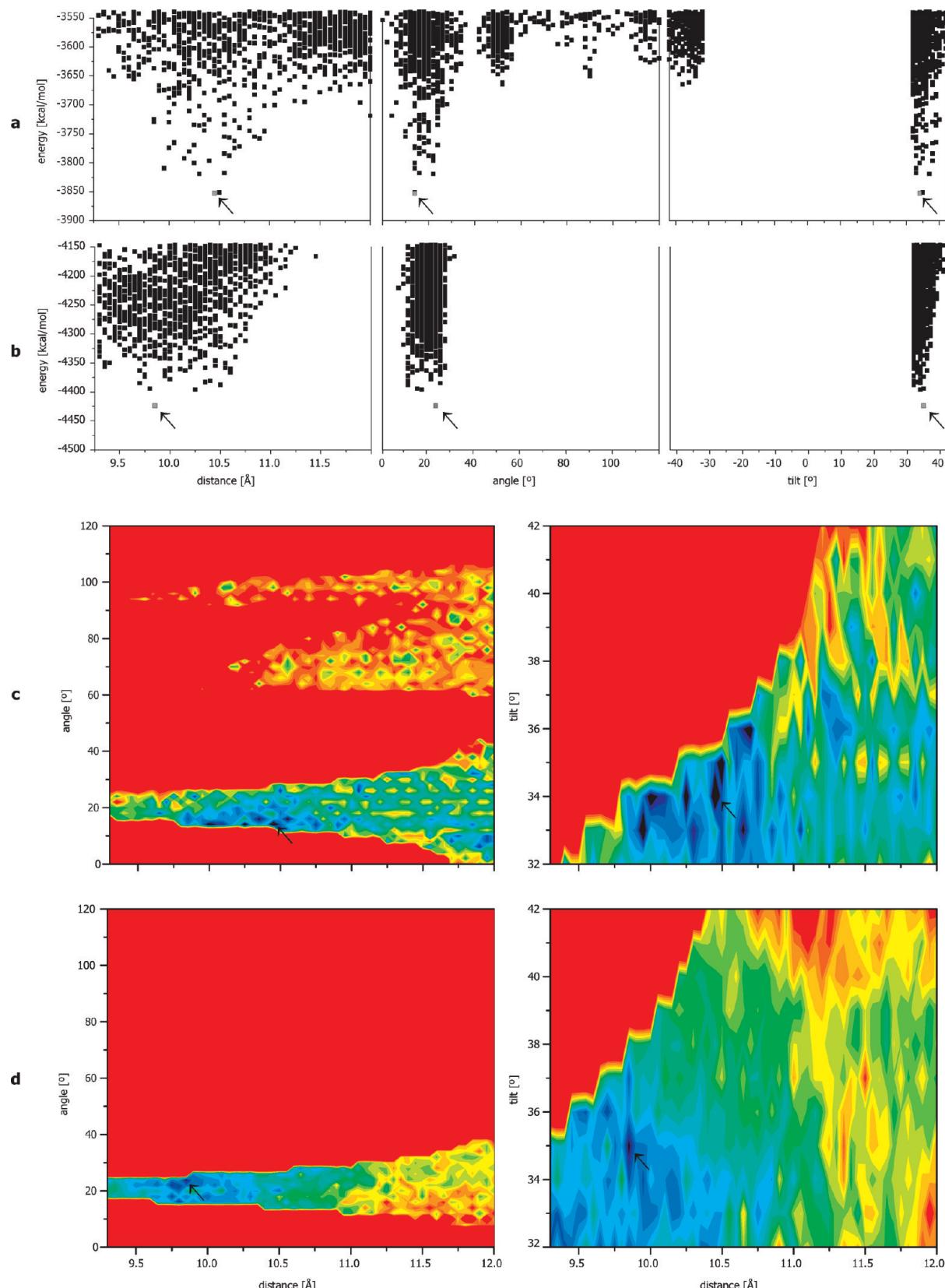


Figure 4. Accumulated energy plots for M2 (a) without restraints and (b) with distance restraint of 3.9 Å between His-37(N_δ) and Trp-41(C_γ). The distance restraint sharpens the energy surface. Comparison of the energy landscapes for the best ranked conformation (arrow) (c) without and (d) with side chain distance restraint for M2 from influenza. The coloring scheme is similar for both cases but covers different energy ranges.

12.765, or respectively 12.550 without constraints, the orientation and shape of the pore are in good agreement. The similarity is even better when compared to the open

NMR structure 2H95 (figure not shown). On the molecular scale all amino acids are in places which are supported by experimental evidence and hypothesis.

Table 1. Crosswise Comparison of the C_{α} -RMSD between Experimental (1NYJ,³⁷ 2H95³⁸) and Computational Structures for the Monomers (Upper Panel) and the Tetrameric Bundles (Lower Panel) of M2 from Influenza

rmsd monomer			
1NYJ	-	-	-
2H95	0.789	-	-
w. ^a restraint	0.483	0.773	-
no restraint	0.528	0.735	0.420
	1NYJ	2H95	w. ^a restraint
			no restraint
rmsd tetramer			
1NYJ	-	-	-
2H95	13.136	-	-
w. ^a restraint	12.765	2.676	-
no restraint	12.550	2.086	1.731
	1NYJ	2H95	w. ^a restraint
			no restraint

^a w = with.

The energy contour plots for the three degrees of freedom, distance, angle and tilt are shown in Figure 4. The gap of data in the energy/tilt plot (Figure 4a,b, right graphs) results from the restriction of the search space to 32 to 42° due to experimental evidence.^{32,33} The minimum for positive tilt values representing right handed bundle assemblies is dominant. In the angle/energy plot four minima can be identified, with one of them containing a structure with the lowest energy of -3860 kJ/mol. Since the other minima are at significantly higher values and located in more narrow minima, they can be rejected as reasonable ‘low-energy’ structures. The first rank structure (-3860 kJ/mol) is found at an interhelical distance of 10.5 Å, a rotational angle of 16° with a tilt of 35°. For the bundle with restraints of 3.9 Å between His-37(N_δ) and Trp-41(C_γ) the first rank structure has an energy of -4425 kJ/mol, interhelical distance of 9.8 Å, and angle and tilt values of 25° and 35°. As the application of a restraint alters the potential energy of these bundles in terms of the force field, the resulting single point energies for the restraint and unrestraint bundle cannot be compared directly. It can be stated that the restraint model allows a slightly tighter packing.

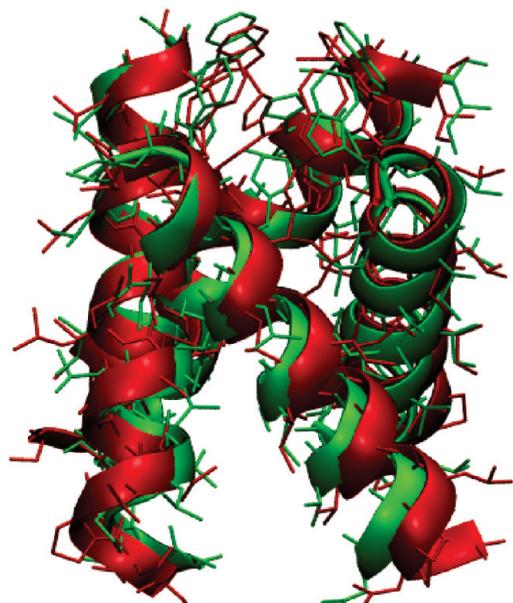


Figure 5. Superposition of 1NYJ (red³⁷) with the best rated model according to the assembly protocol (green).

The three-dimensional color coded energy contour map shows that the energy is only favorable within a narrow range of the angle (Figure 4c,d). This is shown by the sharp valley of lower energy values indicated from green to deep blue. For the tilt the green/blue area representing the low energy structures is less sharply defined as for the angle but clearly localizes all low energy structures on the same spot.

HIV-1-Vpu. For the different energy plots in Figure 6 several minima are observed each corresponding to a different structure. Especially for the angle values (Figure 6a, middle plot) a characteristic pattern with several minima (around 76, 164, 188, 192, and 326°) is observed. The differences in energy are about 75 kcal/mol between the five best structures, each standing representative for the conformers clustering around it (Figure 6b). These five structures adopt tilts of -28, -16, -4, 16, and 24° (Figure 6a, right plot). In contrast to these large differences the interhelical distance between these conformations is relatively small and ranges from 8.55 to 10.30 Å (Figure 6a, left plot).

With respect to the angle the energy surface expresses narrow valleys which are separated by high energy barriers (Figure 6c). In order to pass from one valley to another, the interhelical distance has to be changed by more than 2 Å. With respect to the tilt the low energy regions are wider and shallower, covering a larger range of the tilt (Figure 6c). Thus, changing the tilt is possible over a larger range by only slightly changing the interhelical distance. It is therefore suggested that a possible gating mechanism is rather via changing the tilt than the angle.

Based on the analysis of the various energy plots the five best structures are shown in Figure 6b. Solely based on the energy it is not possible to favor one structure over the other. Also considering structural aspects like the minimum pore radius or side chain hydrophilicity (data not shown) does not lead to any preference as illustrated by the following examples: Model 1 has the bulky tryptophan’s facing into the pore, but with nothing obstructing their move ability to potentially function as gate. In model 4, hydrophilic Ser-24 are facing the lumen of the pore, but they form hydrogen bonds with neighboring carbonyl backbone oxygen’s making it unlikely that they take part in any gating mechanism. Thus, further functional analysis is necessary to evaluate the bundles.

SARS-CoV-3a. The Monomeric Subunit. Currently, there is no structural information available on an atomistic level. For the present investigation each of the membrane spanning part is considered to be helical. Throughout the 10 ns simulation the helicity of the individual helices remains intact.

In contrast to M2 and Vpu three different angles have to be considered, since there is no symmetry due to the different amino acid composition of each membrane spanning part (TM1, TM2, and TM3). The plot distance versus energy shows an almost linear decrease in energy when the helices approach each other (Figure 7a, first plot). In the range 10.4 to 10.7 Å a minimum is observed. Plotting the angle TM1 versus energy, similar to Vpu several minima, distinct from each other are observed with a dominant minimum at 150° (Figure 7a, second plot). For angle TM2 a single minimum around 0° is observed (Figure 7a, third plot). Rotating TM3

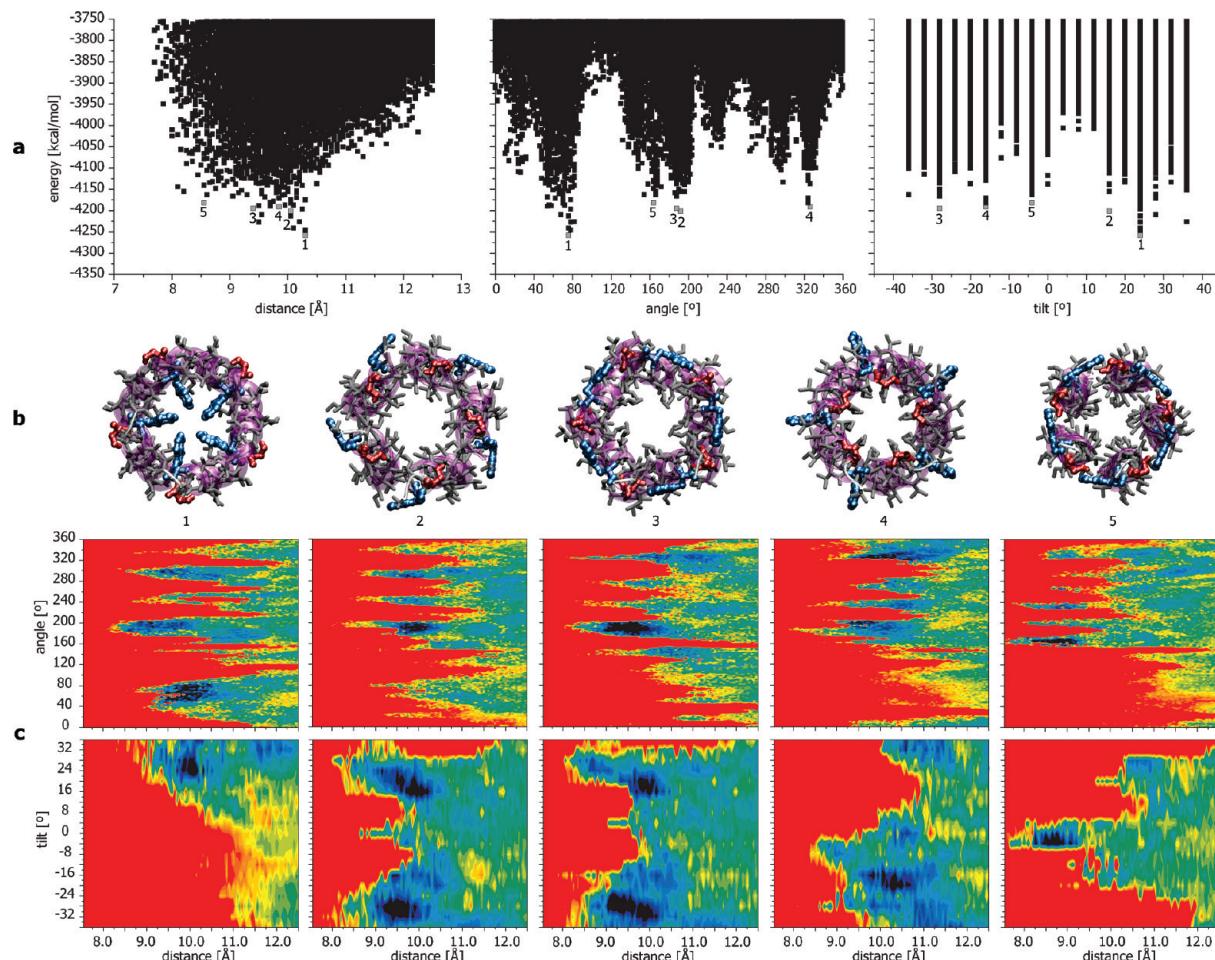


Figure 6. (a) Accumulated plots for distance, rotation, and tilt for Vpu from HIV-1. The contour of the plots indicates the probability of a ‘good’ conformation. (b) Structural model of the five best models. Trp-23 is highlighted in blue and Ser-24 in red. (c) Energy landscapes for the five most probable conformers of Vpu form HIV-1.

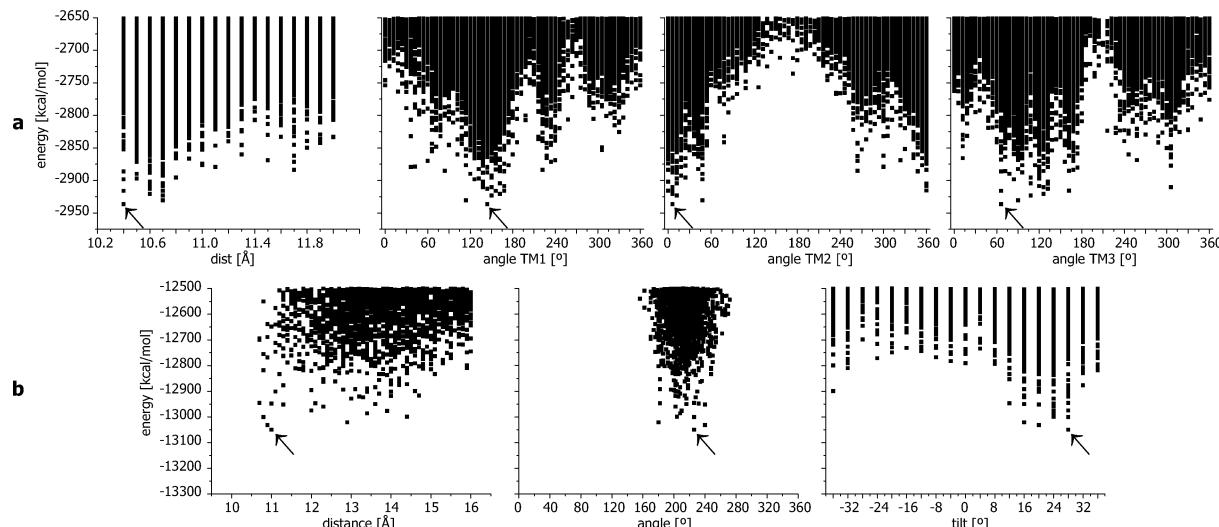


Figure 7. (a) Accumulated plots for distance and rotation for each of the three membrane spanning parts for the monomer of 3a from SARS-CoV. The contour of the plots indicates the probability of a ‘good’ conformation. (b) Energy plots for the assembly of four subunits of SARS-CoV-3a forming a full pore. Only one clear minimum with respect to the angle is observed. The energy values for the best ranked model (see Figure 8) are marked with an arrow.

proposes a range from 50 to 150° which results in low energy structures (Figure 7a, fourth plot).

Analyzing the dependency on the tilt it has been observed that only small values of $0 \pm 2^\circ$ occur for the low energy structures (plot not shown). Larger values as observed for

M2 and Vpu seem to be unlikely as they would decrease the tight packing for the three membrane spanning parts of the monomer.

In summary, it appears that the angle values of TM1 150°, TM2 0°, and TM3 50–150° define the location of the global

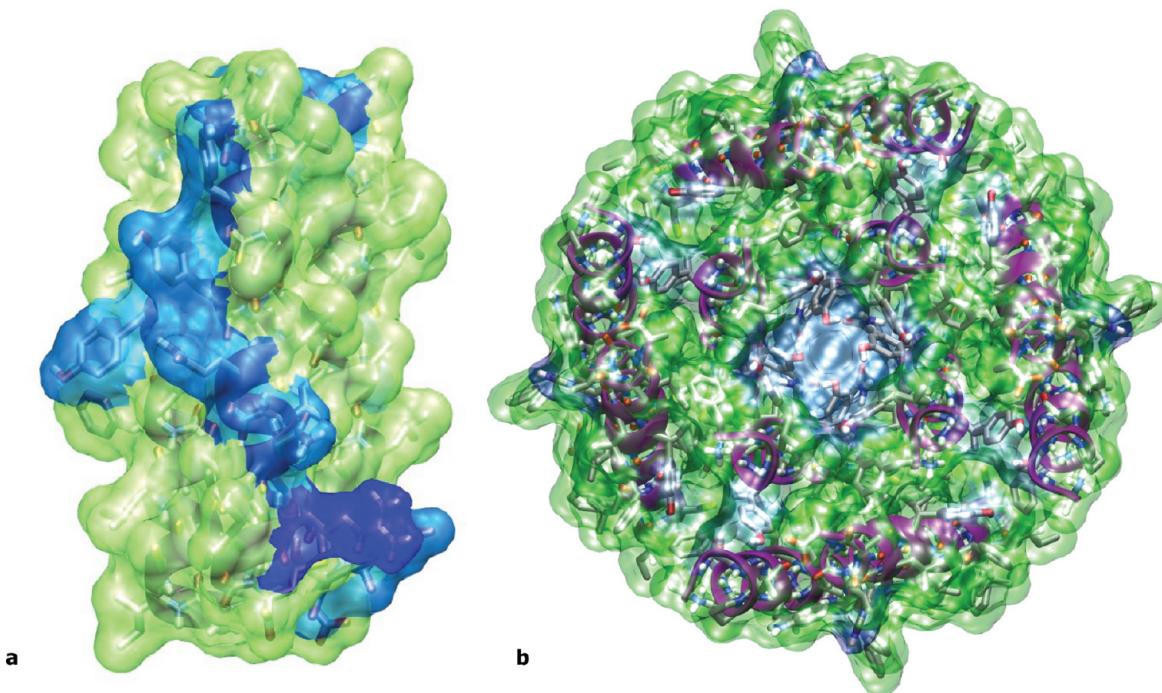


Figure 8. (a) Color coded representation of the best structure of the membrane spanning part of 3a form SARS-CoV. Polar and hydrophilic residues are shown with a light blue surface, ionic residues are shown with a dark blue surface, and hydrophobic residues are shown in green. The hydrophilic stripe (Tyr-109, Tyr-113, Gln-116, Asn-119, and Arg-122) along the structural model can be considered as putative pore lining. (b) Structural model of the fully assembled pore of SARS-CoV-3a. The hydrophilic stripe observed on one subunit lines the pore.

minimum for the monomer of 3a from SARS. The global minimum structure shows a clustering of hydrophilic residues alongside of the monomer assembly. Mapping of these residues at the outside of the ‘best structure’ of the monomeric trimer indicates a line of Tyr-109, Tyr-113, Gln-116, Asn-119, and Arg-122 in TM3 forming a hydrophilic stripe stretching over the whole TM part (Figure 8a). Other hydrophilic residues such as Thr-89 and Ser-92 of TM2 are buried within the bundle forming hydrogen bonds with neighboring backbone carbonyl oxygen within the same helix. Rotating these residues toward the outside of the monomeric trimer would result in an unfavorable ‘high energy’ structure. The hydrophilic stripe could be the pore lining part of the ion channel.

Tetramer. To push the assembly protocol to its limits it has been attempted to assemble four trimeric subunits to form a full structural model of the pore of SARS-CoV-3a. The energy contour plots show one minimum at about 26° for the tilt and at 200° for the angle (Figure 7b). This structure corresponds to a tetramer of SARS-CoV-3a where the outer side of the bundle shows a clustering of hydrophobic amino acids, while the putative pore harbors the hydrophilic residues (Figure 8b). In this model Trp-45 and His-93 form a corona at the cytoplasmic side possibly helping anchoring the bundle in the membrane. In addition a series of hydrophobic phenylalanine point outward, facing the hydrophobic tails of surrounding lipid molecules.

Discussion

Validity of the Approach. The assumption is that viral channel forming proteins are produced in the ER as a single

unity. Assembly is a consecutive step which then leads to the functional bundles. Between production and assembly the protein has to adopt an equilibrated monomeric structure which then forms the basis of the assembly. At the current state we assume an axial symmetry toward the center of the pore, which has to be adapted to by the average global minimum structure. Therefore this method considers equilibrium structures and does not offer insight into the kinetic pathways leading to a bundle assemble.

The monomer, built as an ideal helix, is due to a 10 ns MD simulation embedded into a fully hydrated POPC lipid bilayer. This system is considered to be a reasonable representation for a membrane environment, although some experimental measurements were published based on notably different conditions. The equilibration reveals helices which can be significantly bend or kinked.³⁰ This has been reported for other simulations on viral channel forming proteins^{9,39} and has been attributed to the electrostatic interactions at the end of the helices with the lipid head groups. These highly flexible residues may not reflect the global minimum situation and thus using the ‘core’ TM part is reasonable searching for the bundle structure.

The assembly can be described as a positioning of the monomers with respect to their backbone atoms and a consequent positioning of the side chains. In order to release stress a short minimization is done and the potential energy is calculated. The spatial resolution of positioning with respect to distance, angle, and tilt is extremely small covering finely grained the potential energy surface.

In another search algorithm,^{7,40} similar positioning is done prior to a simulated annealing protocol. In that study the CNS

Software^{41,42} has been used including the united-atom OPLS parameter set. Large rearrangements of the protein are allowed, while the sampling of the conformational space is very limited. In a modified version the same software has been used with a combined simulated annealing (SA) and short MD simulation protocol.^{9–14,43} The conformational search space was narrowed based on the assumption that hydrophilic residues should face the lumen of the pore. In both approaches, the number of potential bundle candidates has then been restricted to be below 300 and 30, respectively.

The replica exchange approach employed by Brooks and co-workers¹⁵ is also limited in the number of starting conformations but has a largely improved chance to reach local minima. Nevertheless only a small part of the conformational space can be sampled by this method. The role of the monomer conformation for the total energetic was put into focus by the group of Lazaridis.¹⁶ The possibility of alternative conformation was explicitly considered. The SA/MD approach used in this study is based on a rotational angle with a stepwidth of 20°.

To this date it remains unclear how monomeric proteins migrate from the ER to their point of action. It has to be considered that different proteins follow different pathways. Bowie and co-workers followed the two-stage-model^{44,45} to generate accurate initial dimeric assemblies. But the interaction interface does not have to remain the same for higher oligomeric assemblies. As it is still not known how these higher oligomeric states are reached on a biological and kinetic level it has to be assumed that significant conformational rearrangement occur. We also follow the two-stage model in this study by carefully equilibrating the monomeric subunits. The brief unconstrained energy minimization of the assemblies implicitly takes rearrangements into account, although no statements about kinetic pathways can be made.

The method presented here has its strength in its fine grained sampling. The energy landscape around the proposed equilibrium conformations of the assemblies is mapped thoroughly. Structure optimization steps as mentioned above could easily be implemented into the protocol but are not desired. In comparison to the present approach, any excessive SA or MD simulation steps would smooth out and thereby obscure the fine structure of the energy landscape, consequently missing out low energy structures. Although the energy minimization used in our approach is very short and might contain only partially relieved conformations the high conformational sampling density ensures that the energetically ‘best’ conformations are identified. The decision to carry out this study with a pure united-atom vacuum force field (Engh-Huber) tries to balance accuracy and computational costs. The usage of implicit membrane models as done by others^{15,16} does not represent a provable improvement in accuracy at this stage since it will also not account for specific interactions with the lipid head groups or central water columns filling the pores. Therefore the approach with the least number of assumptions was followed using a reasonable protein vacuum force field.

As shown for M2 the presented strategy delivers results matching experimental findings fairly well. Also for Vpu, bundles are found which are similar to the ones suggested

earlier based on aforementioned protocols. Therefore it is concluded that a pre-equilibrated monomeric structure is an important step to achieve a good starting position to work on the assembly and that a small step size in positioning is adequate to cover essential aspects of protein assembly.

The Models Proposed. M2. There has been considerable evidence in the literature that His-37 and Trp-41 participate in the proton conductance of M2.^{32,46,56} Therefore they have to be accessible to water meaning that any conformation where they point into the surrounding lipid can be safely excluded. Limiting the search by 2/3 is a considerable speed up and avoids the potential risk of creating false positive results. Furthermore the employment of NMR based distance restraints has been probed. It has to be noted that they were measured on an unligated M2 pore. It was found that their usage does not improve the already good quality of the proposed structure in this limited search space. But it can be stated that the usage ‘sharpens’ the energy landscapes. For other proteins the possibility to use experimentally derived restraints might turn out to be more significant.

Comparison of the monomeric and bundle structures from this study with NMR based structures 1NYJ and especially 2H95 shows an excellent agreement, verifying the validity of the approach (Table 1). Describing the two published structures as closed and open states cannot be justified on the basis of our results. Also with the recently published X-ray structure 3BKD⁴⁷ on the level of the monomers a good agreement is observed with rmsd values of 1.330–1.495. For comparison, the rmsd among the individual helices within the crystallographic unit cell of 3BKD spreads over a considerable range (0.214–1.130). This indicates that the bended, helical monomer conformation seems to be stable under various conditions, as it is also found in a recent MD study.⁴⁸

A comparison of the bundles delivers a smaller deviation between the computational and the NMR bundle model 1NYJ (2.676) than with the NMR bundle model 2H95 (12.765). For the later the deviation is in the same range as between the two NMR models (13.136). A comparison of the computational model with the bundles within 3BKD was not carried out as their pore conformation is reminiscent of an open umbrella, which most likely represents an energetically costly conformation within a lipid environment.

The computational model is generated without any structural bias induced by the presence of ligands or crystallization agents and therefore may be seen as a very plausible model. Further studies need to be done to associate a model with an ‘open’ or ‘closed’ pore.

Vpu. Models 1, 2, 3, and 5 show the hydrophilic residue Ser-24 pointing to the outside of the bundle or been buried between two subunits. Trp-23, the only other hydrophilic residue in the TM part of Vpu, is facing outward and/or interacting with neighboring subunits. These structures suggest the lumen of the pore to be a widely hydrophobic stretch. Model 4 is similar to the one used so far in MD simulations with the TM part of Vpu,^{11,49,50} whereas model 1 corresponds to a model which has been suggested earlier.^{7,10} In the former model Ser-24 points into the pore, while in the latter it is Trp-23. At this stage based on the

energy values none of the models can be preferred above any other. This needs further functional *in silico* evaluation such as longer MD simulations to assess the stability of the bundles and simulation of ion permeation through the respective pores. In the present study Ser-24 forms an intrahelical hydrogen bond with the backbone carbonyl oxygen of Ile-20 in all models.

The conformational transition between two models requires variation of the distance, angle, and tilt ‘walking on the energy surface’. With respect to the energy plots for the angle (Figure 6c, upper row) a huge energy barrier would have to be crossed while changing the angle and increasing the distance by more than 2 Å to move from one minimum to another. This makes it very unlikely that conformational transitions include huge rotational movements of the individual subunits. Focusing on models 2, 3, and 5 it seems to be more likely that the tilt changes, while the distance and angle would only be varied to a minor degree (Figure 6c, lower row). It can be proposed that models 2 and 3 are ion conducting and can change their conformation toward model 5 a potentially closed state. The models 2 and 3 would represent alternative conducting states, which would be in good agreement with the experimental finding of multiple conductance states for Vpu.^{51–53} More than just one model could contribute to the functioning of Vpu. The results further underline the flexibility of the TM part of the protein.³⁰

SARS-CoV. The sequence based TM prediction is an established technique. Nevertheless some deviations between the different protocols can be observed. Creating a consensus⁵⁴ between the six different techniques used in this study leads to a robust and reliable prediction.

It is noteworthy that the Cys-bridges reported to link the subunits are located in the extramembrane part of the protein, presumably not directly affecting the TM part. The assembly of the monomeric unit results in a profound model with hydrophilic residues clustering on one side. The pore assembly into a putative bundle leads to the first structural model of 3a from SARS with residues such as tyrosines, glutamine, asparagines, and arginine from TM3 lining the lumen. This motif is rather unusual as more commonly serines and tyrosines are suggested for pore lining residues in channels.⁵⁵

To screen the ‘whole’ conformational space of an assembly represents an auspicious approach for both experimentally and computationally based studies. It leads to reliable structural models, helps to avoid structural pitfalls, and opens insights into mechanistic details of the mode of action and can help revealing alternative conformations. The fine grained full search approach is the most direct route for exploring the conformational space of a protein assembly. By simplifying and considering the symmetry of the studied proteins a significant confinement of the search space can be made, without biasing toward a certain result. This enables the resolvability of the search in an acceptable sampling time.

The quality of the constructed structural models does not rank behind any experimental technique. Based on this method alternative configurations of pentameric Vpu have been shown, and a novel pore lining motif is suggested for the bundle model of 3a. Regarding the topology of all of

the studied energy landscapes, it has to be concluded that conformational transitions from any open to closed states would have to take place by variation of the tilt and not the angle. As a further quality check it is recommended to do functional studies on the most plausible models suggested, e.g. assessing the bundle stability in a lipid environment with consequent multins MD simulations or to do cross mutations verifying explicit interactions between the monomers in the models.

Acknowledgment. W.B.F. thanks the NYMU and the government of Taiwan for financial support (Aim of Excellence Program). This work was supported by the National Science Council of Taiwan (NSC). J.K. acknowledges a fellowship granted by the Alexander von Humboldt-Foundation and the NSC. We thank the Paderborn Center for Parallel Computing PC² (<http://wwwcs.uni-paderborn.de/pc2/>) for providing computer time.

Supporting Information Available: Detailed description of the placement algorithm including an ERM for all consecutive steps of the assembly. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Winter, G.; Fields, S. *Nucleic Acids Res.* **1980**, 8, 1965.
- Allen, H.; McCauley, J.; Waterfield, M.; Gethering, M. *Virology* **1980**, 107, 548.
- Lamb, R. A.; Choppin, P. W. *Virology* **1981**, 112, 729.
- Strebel, K.; Klimkait, T.; Martin, M. A. *Science* **1988**, 241, 1221.
- Cohen, E. A.; Terwilliger, E. F.; Sodroski, J. G.; Haseltine, W. A. *Nature* **1988**, 334, 532.
- Lu, W.; Zheng, B.-J.; Xu, K.; Schwarz, W.; Du, L.; Wong, C. K. L.; Chen, J.; Duan, S.; Deubel, V.; Sun, B. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, 103, 12540.
- Kukol, A.; Arkin, I. T. *Biophys. J.* **1999**, 77, 1594.
- Adams, P. D.; Arkin, I. T.; Engelman, D. M.; Brünger, A. T. *Nature Struct. Biol.* **1995**, 2, 154.
- Forrest, L. R.; Kukol, A.; Arkin, I. T.; Tieleman, D. P.; Sansom, M. S. *Biophys. J.* **2000**, 78, 55.
- Cordes, F.; Kukol, A.; Forrest, L. R.; Arkin, I. T.; Sansom, M. S. P.; Fischer, W. B. *Biochim. Biophys. Acta* **2001**, 1512, 291.
- Cordes, F. S.; Tustian, A.; Sansom, M. S. P.; Watts, A.; Fischer, W. B. *Biochemistry* **2002**, 41, 7359.
- Lemaitre, V.; Ali, R.; Kim, C. G.; Watts, A.; Fischer, W. B. *FEBS Lett.* **2004**, 563, 75.
- Kim, C. G.; Lemaitre, V.; Watts, A.; Fischer, W. B. *Anal. Bioanal. Chem.* **2006**, 386, 2213.
- Patargias, G.; Zitzmann, N.; Dwek, R.; Fischer, W. B. *J. Med. Chem.* **2006**, 49, 648.
- Bu, L.; Im, W.; Brooks, C. L., III *Biophys. J.* **2007**, 92, 854.
- Mottomal, M.; Zhang, J.; Lazaridis, T. *Proteins* **2006**, 62, 996.
- Kim, S.; Chamberlain, A. K.; Bowie, J. U. *J. Mol. Biol.* **2003**, 329, 831.

- (18) Faham, S.; Yang, D.; Bare, E.; Yohannan, S.; Whitelegge, J. P.; Bowie, J. U. *J. Mol. Biol.* **2004**, *335*, 297.
- (19) Bowie, J. U. *Nature* **2005**, *438*, 581.
- (20) Duff, K. C.; Kelly, S. M.; Price, N. C.; Bradshaw, J. P. *FEBS Lett.* **1992**, *311*, 256.
- (21) Kovacs, F. A.; Cross, T. A. *Biophys. J.* **1997**, *73*, 2511.
- (22) Lemaitre, V.; Kim, C. G.; Fischer, D.; Lam, Y. H.; Watts, A.; Fischer, W. B. In *Viral membrane proteins: structure, function and drug design*; Fischer, W. B., Ed.; Kluwer Academic/Plenum Publishers: New York, 2005; Vol. 1, p 187.
- (23) Candler, A.; Featherstone, M.; Ali, R.; Maloney, L.; Watts, A.; Fischer, W. B. *Biochim. Biophys. Acta* **2005**, *1716*, 1.
- (24) Jaysinghe, S.; Hristova, K.; Wimley, W.; Snider, C.; White, S. H. 2006. <http://blanco.bimol.uci.edu/mpex>.
- (25) Cserzo, M.; Eisenhaber, F.; Eisenhaber, B.; Simon, I. *Protein Eng.* **2002**, *15*, 745.
- (26) Hofmann, K.; Stoffel, W. *Biol. Chem.* **1993**, *374*, 166.
- (27) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. *J. Mol. Biol.* **2001**, *305*, 567.
- (28) Tusnády, G. E.; Simon, I. *Bioinformatics* **2001**, *17*, 849.
- (29) Chandrasekhar, I.; Kastenholz, M.; Lins, R. D.; Oostenbrink, C.; Schuler, L. D.; van Gunsteren, W. F. *Eur. Biophys. J.* **2003**, *32*, 67.
- (30) Krüger, J.; Fischer, W. B. *J. Comput. Chem.* **2008**, *29*, 2416.
- (31) Engh, R. A.; Huber, R. *Acta Crystallogr. Sect. A: Found. Crystallogr.* **1991**, *47*.
- (32) Hu, J.; Fu, R.; Nishimura, K.; Zhang, L.; Zhou, H.-X.; Busath, D. D.; Vijayvergiya, V.; Cross, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6865.
- (33) Wang, J.; Kim, S.; Kovacs, F.; Cross, T. A. *Protein Sci.* **2001**, *10*, 2241.
- (34) Chen, P. P.-S. *ACM Trans. Database Systems* **1976**, *1*, 9.
- (35) Fischer, W. B.; Sansom, M. S. P. *Biochim. Biophys. Acta* **2002**, *1561*, 27.
- (36) *Viral membrane proteins: structure, function and drug design*; Fischer, W. B., Ed.; Kluwer Academic/Plenum Publisher: New York, 2005; Vol. 1, p 291.
- (37) Nishimura, K.; Kim, S.; Zhang, L.; Cross, T. A. *Biochemistry* **2002**, *41*, 13170.
- (38) Hu, J.; Asbury, T.; Achuthan, S.; Bertram, R.; Quine, J. R.; Fu, R.; Cross, T. A. *Biophys. J.* **2007**, *92*, 4335.
- (39) Fischer, W. B.; Forrest, L. R.; Smith, G. R.; Sansom, M. S. P. *Biopolymers* **2000**, *53*, 529.
- (40) Kukol, A.; Adams, P. D.; Rice, L. M.; Brunger, A. T.; Arkin, I. T. *J. Mol. Biol.* **1999**, *286*, 951.
- (41) Brunger, A. T. *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*; Yale University Press: New Haven, CT, 1992.
- (42) Brunger, A.; Adams, P.; Clore, G.; Gros, W.; Gross-Kunstleve, R.; Jiang, J.; Kuszewski, J.; Nilges, M.; Pannu, N.; Read, R.; Rice, L.; Simonson, T.; Warren, G. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 905.
- (43) Kerr, I. D.; Doak, D. G.; Sankararamakrishnan, R.; Breed, J.; Sansom, M. S. P. *Protein Eng.* **1996**, *9*, 161.
- (44) Popot, J.-L.; Engelman, D. M. *Biochemistry* **1990**, *29*, 4031.
- (45) Engelman, D. M.; Chen, Y.; Chin, C.-N.; Curran, A. R.; Dixon, A. M.; Dupuy, A. D.; Lee, A. S.; Lehnert, U.; Matthews, E. E.; Reshetnyak, Y. K.; Senes, A.; Popot, J.-L. *FEBS Lett.* **2003**, *555*, 122.
- (46) Tian, C.; Tobler, K.; Lamb, R. A.; Pinto, L. H.; Cross, T. A. *Biochemistry* **2002**, *41*, 11294.
- (47) Stouffer, A. L.; Acharya, R.; Salom, D.; Levine, A. S.; Di Constanzo, L.; Soto, C. S.; Tereshko, V.; Nanda, V.; Stayrook, S.; DeGrado, W. F. *Nature* **2008**, *451*, 596.
- (48) Yi, M.; Cross, T. A.; Zhou, H.-X. *J. Phys. Chem. B* **2008**, *112*, 7977.
- (49) Grice, A. L.; Kerr, I. D.; Sansom, M. S. P. *FEBS Lett.* **1997**, *405*, 299.
- (50) Moore, P. B.; Zhong, Q.; Husslein, T.; Klein, M. L. *FEBS Lett.* **1998**, *431*, 143.
- (51) Schubert, U.; Bour, S.; Ferrer-Montiel, A. V.; Montal, M.; Maldarelli, F.; Strebel, K. *J. Virol.* **1996**, *70*, 809.
- (52) Mehnert, T.; Lam, Y. H.; Judge, P. J.; Routh, A.; Fischer, D.; Watts, A.; Fischer, W. B. *J. Biomol. Struct. Dyn.* **2007**, *24*, 589.
- (53) Mehnert, T.; Routh, A.; Judge, P. J.; Lam, Y. H.; Fischer, D.; Watts, A.; Fischer, W. B. *Proteins* **2008**, *70*, 1488.
- (54) Cuthbertson, J. M.; Doyle, D. A.; Sansom, M. S. P. *Prot. Eng. Des. Sel.* **2005**, *18*, 295.
- (55) Akabas, M. H.; Kaufmann, C.; Archdeacon, P.; Karlin, A. *Neuron* **1994**, *13*, 919.
- (56) Schnell, J. R.; Chou, J. J. *Nature* **2008**, *451*, 591.

CT900185N