Modeling the structure of SARS 3a transmembrane protein using a minimum unfavorable contact approach

S RAMAKRISHNA, SILADITYA PADHI and U DEVA PRIYAKUMAR*

Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad 500 032, India
e-mail: deva@iiit.ac.in

Abstract. 3a is an accessory protein from SARS coronavirus that is known to play a significant role in the proliferation of the virus by forming tetrameric ion channels. Although the monomeric units are known to consist of three transmembrane (TM) domains, there are no solved structures available for the complete monomer. The present study proposes a structural model for the transmembrane region of the monomer by employing our previously tested approach, which predicts potential orientations of TM α-helices by minimizing the unfavorable contact surfaces between the different TM domains. The best model structure comprising all three α-helices has been subjected to MD simulations to examine its quality. The TM bundle was found to form a compact and stable structure with significant intermolecular interactions. The structural features of the proposed model of 3a account for observations from previous experimental investigations on the activity of the protein. Further analysis indicates that residues from the TM2 and TM3 domains are likely to line the pore of the ion channel, which is in good agreement with a recent experimental study. In the absence of an experimental structure for the protein, the proposed structure can serve as a useful model for inferring structure-function relationships about the protein.

Keywords. Membrane protein modeling; ion channel; transmembrane helices; viroporin; molecular dynamics; SARS 3a.

1. Introduction

The outbreak of severe acute respiratory syndrome (SARS) in China and subsequently in other countries in the beginning of this century has necessitated an understanding of the underlying molecular mechanisms.[1–3] The causative agent of this atypical form of pneumonia has been identified as SARS-associated coronavirus (SARS-CoV).[4,5] Although this virus has a positive stranded RNA genome that is characteristic of coronaviruses, it has been shown to be phylogenetically different from other coronaviruses.[6,7] The genome has 14 open reading frames (ORFs), which code for structural proteins like the spike glycoprotein, the envelope protein, the membrane protein, and the nucleocapsid protein.[4] While the former three constitute the envelope of the virus, the latter is complexed with genomic RNA to form the nucleocapsid.[8,9] A protein important in the infection cycle of the virus is the replicase polyprotein, which is processed to yield proteases, RNA-dependent polymerase and RNA helicase.[7]

Apart from the proteins mentioned above, SARS-CoV encodes a number of proteins, the functions of which are still not understood. One of the proteins that has recently been characterized is 3a, a 274 amino-acid transmembrane protein.[6,7] It is coded by the ORF 3a gene, which is present in between the genes coding for the spike and the envelope proteins.[6,7,10] The protein has three transmembrane (TM) domains on the N-terminal side, and a long cytoplasmic domain on the C-terminal side (figure 1).[6] Monomeric forms of this protein have been shown to form dimers via a disulfide linkage; two dimers, in turn, can associate non-covalently to form a homotetrameric complex.[11] The possible role of this tetrameric complex is to function as an ion channel that promotes viral release.[12] The channel has been shown to be sensitive towards potassium ions, and can be inhibited by barium ions.[11] Atomistic details of the structure of the channel, however, are yet to be elucidated.

Although membrane proteins contribute to about ~25% of the open reading frames in most of the genomes, the structure-function relationship of only a very small fraction of these proteins is understood.[13] An extensive understanding of the membrane protein folding mechanism is important.[14,15] The majority of membrane proteins are formed by the transmembrane (TM) α-helices.[16] Membrane protein folding has been proposed to proceed in two stages, namely the insertion
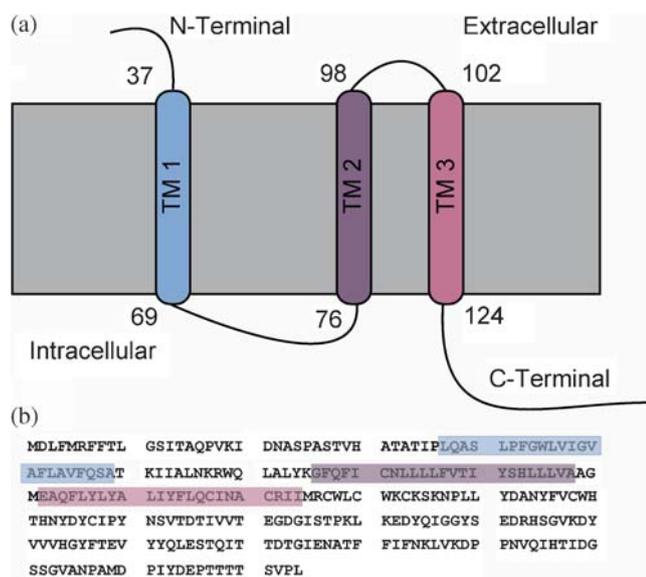
Figure 1. (a) Schematic representation of the three transmembrane domains of SARS 3a. (b) Sequence of the SARS 3a protein. Residues that are part of TM1, TM2 and TM3 are highlighted using similar colors in (a).

of stable TM α-helices into a lipid bilayer, and the assembly of these helices into a helix bundle.[17,18] Hydrophobic interactions are likely to play a key role in driving the insertion of the transmembrane α-helices into lipid bilayers. Taking into consideration the difficulties in experiments, computational approaches are expected to play a major role in understanding the structure-function relationships of the membrane proteins. The study of the interaction between membrane proteins and membrane lipid bilayers is the key step in insertion of the membrane proteins into the lipid bilayer.[19,20] The study of the energy landscape of the membrane proteins is important for the assembly of individual helices in forming the helix bundle.[21] The interactions between the proteins and the lipid bilayer has been studied using all-atom molecular dynamics (MD) simulations, which involve several thousands of non-bonded interactions at every time step for such large systems, making simulation time a major constraint.[22–28] Hence, the use of implicit membrane method has turned out to be very efficient in modeling transmembrane phenomena.[29,30]

From the earlier studies of Fischer and coworkers, it has been proposed that, out of the three transmembrane domains (TM1, TM2 and TM3), either TM2 or TM3 is likely to line the interior of the SARS 3a ion channel;[31,32] the results are based on two different TM domain assembly algorithms. Recently, they have experimentally reconstituted the TM domains and placed them both individually and in equal mixtures of all the possible combinations of the two of each of

the TM domains (1:1 ratio of TM1:TM2, TM1:TM2 and TM2:TM3 domains) in an artificial lipid bilayer to investigate the ion channel activity.[33] Based on the Goldmann-Hodgkin-Katz constant field approach,[34,35] they have measured the permeability ratio for all the compositions between the two sides of artificial lipid bilayer. The results showed that TM2:TM3 domains show higher values of current as compared with the individual TM domains and TM1:TM2 and TM1:TM3 mixtures. As far as the structure of the protein is concerned, there are still no structures available for the protein, thereby limiting an understanding of the processes that the protein can possibly be involved in. In the present study, a structural model has been proposed for the TM region of the 3a monomer using a method that optimizes favorable contacts between transmembrane helical bundles. The approach which has previously been used to successfully predict the structure of a number of proteins,[36] is employed here to propose a structural model for 3a. The model is shown to be stable in a membrane environment, and is able to account for a number of properties of the protein reported in previous experimental studies.

2. Computational

2.1 *Methods*

The three transmembrane α-helices of SARS 3a, TM1 (LQAS[40] LPFGWLVIGV[50] AFLAVFQSA), TM2 (GFQFI[80] CNLLLLFVTI[90] YSHLLLVA), and TM3 (EAQFLYLYA[110] LIYFLQCINA[120] CRII) were modeled as ideal α-helix structures using the Sybyl 7.2 program.[37] To these three TM domains, three additional residues in the C- and N-terminal ends were included for modeling the structures. Hence the number of residues in each of the TM domains is $23 + 3 + 3 = 29$. A detailed description of the methodology followed for MD simulations of the individual TM domains in implicit and explicit lipid bilayer environment can be found in the Supporting Information. The three TM domains were assembled in such a way that unfavorable contacts between the domains were minimized (see Results and Discussion for details). The assembled protein was modeled in an implicit membrane environment using the generalized Born model with a switching function (GBSW).[38,39] A thickness of 33 Å was used for the membrane hydrophobic core, with a smoothing length of 0.4 Å. This was followed by addition of loops connecting the TM domains using the CHARMM program,[40] which uses CHARMM force field parameters to create an initial model of the loop.[41,42]

Initial minimization and equilibration was performed in an implicit membrane environment, so as to allow the loops to relax, followed by extensive equilibration in an explicit lipid bilayer. All simulations were performed using the CHARMM (for implicit membrane environment) and NAMD (for explicit membrane environment) programs with the CHARMM22 protein force field with CMAP corrections.[40–43] Trajectory analysis was performed using the CHARMM program,[40] and structural images were rendered using the visual molecular dynamics (VMD) program.[44]

3. Results and Discussion

3.1 *Behavior of the individual TM domains*

The behavior of the individual TM domains in a membrane environment was investigated prior to the assembly of the three TM domains, so as to obtain insights into the intramolecular and protein-lipid interactions

that could possibly stabilize these TM domains. The intramolecular hydrogen bonds between the residues of the individual TM domains were calculated throughout the simulations in both implicit and explicit membranes with a hydrogen bond donor (D)-acceptor (A) distance cut-off of 3.5 Å and a D-H. . .A angle cut-off of 135°. Figures 2(a) and 2(b) represent the probability distribution of intramolecular hydrogen bonds of all the TM domains in implicit and explicit membrane simulations, respectively. The average number of intramolecular hydrogen bonds is less for TM1 and TM3 when compared to the TM2 domain in implicit membrane simulations. In the case of the TM3 domain, the predominant occurrence of hydrophilic residues (in an otherwise hydrophobic membrane environment) results in disruption of the helical structure, leading to less number of hydrogen bonds. On the other hand, the presence of a kink in the TM1 domain (see figure S1 in Supporting Information) explains the relatively low number of hydrogen bonds in TM1.

The hydrogen bonding ability of the TM domains with the lipid/water environment was examined by com-
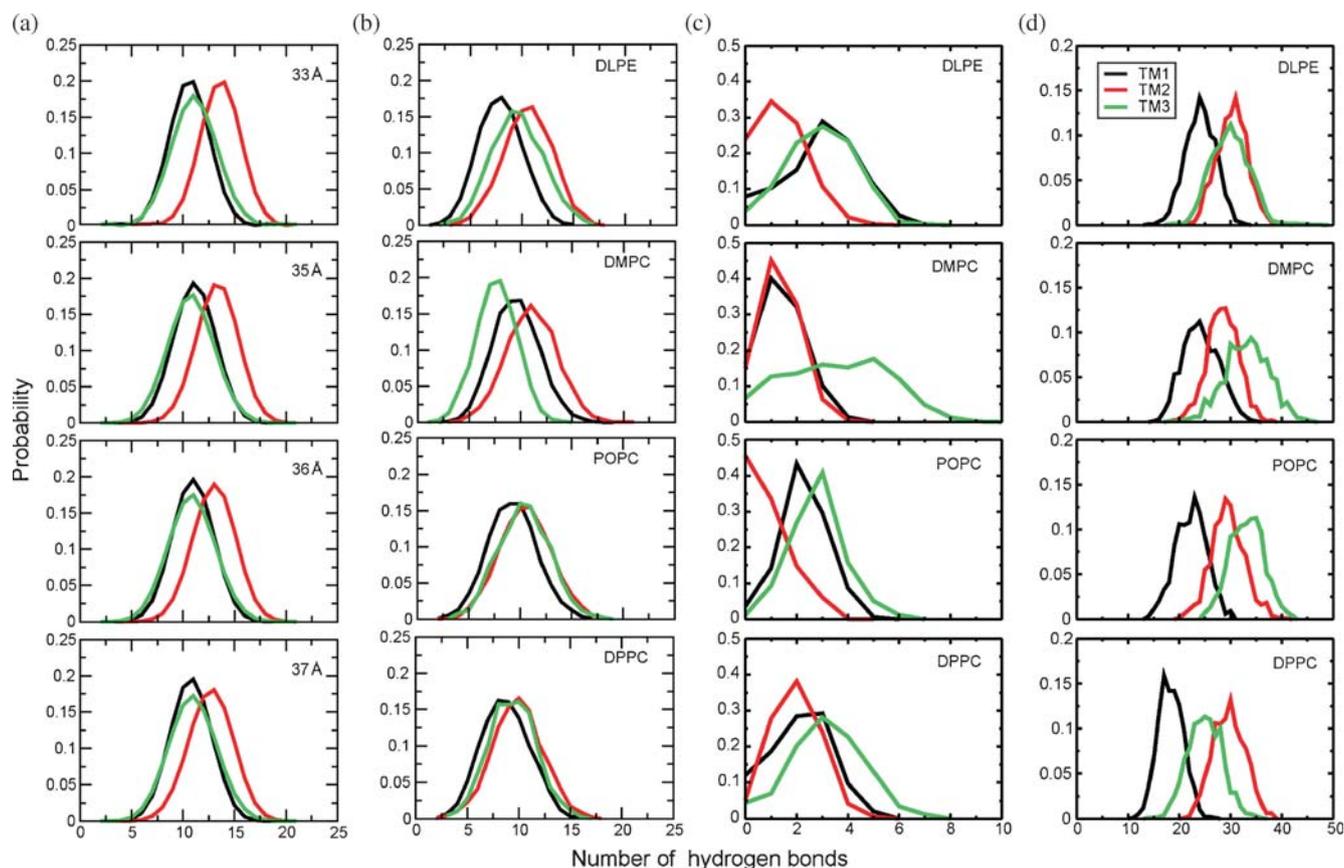

Figure 2. (a) Probability distribution of intra-helical hydrogen bonds of TM domains in implicit membrane simulations. (b) Probability distribution of intra-helical hydrogen bonds of TM domains in explicit membrane simulations. (c) Probability distribution of the number of hydrogen bonds between TM domains and lipid molecules in the explicit membrane simulations. (d) Probability distribution of the number of hydrogen bonds between TM domains and water molecules in the explicit membrane simulations.

puting the number of hydrogen bonds formed between the two. Figure 2(c) gives the probability distribution of protein-lipid hydrogen bonds for all the TM domains in various explicit membranes, and figure 2(d) gives the probability distribution for protein-water hydrogen bonds in all the explicit membrane systems. The average number of protein-lipid hydrogen bonds varies as TM3 > TM2 > TM1 in general. As far as protein-water hydrogen bonds are concerned, TM1 forms less number of hydrogen bonds with water molecules in comparison with TM2 or TM3, suggesting that it is very unlikely that TM1 will face the hydrophilic pore (in an oligomeric form of the protein). TM3 forms more number of intermolecular hydrogen bonds (protein-lipid and protein-water) due to the presence of more number of hydrophilic residues compared to TM1 and TM2. The unusually high number of hydrogen bonds for TM3 in DMPC can be attributed to the unwinding of the helix in the N-terminal end, leading to favorable interactions between the unfolded part of TM3 with lipid heads and water. In general, the results indicate a propensity of the TM3 domain to form hydrogen bonds with the external environment, suggesting that this domain could possibly face the water-filled pore.

The tilt angle of the TM domains with respect to the membrane normal in different membrane environments is shown in figure S2 in SI. The TM domains are seen to exhibit different tilt angle values in the different membrane environments. Since the membrane models have different thicknesses, the hydrophobic mismatch arising from the occurrence of the hydrophobic part of the protein in the solvent region varies in the different models, leading to different tilt angles being seen across these models.

3.2 *Modeling the TM protein as a bundle*

When the three TM domains form a bundle within the lipid bilayer, the hydrophilic residues of the TM domains will minimize the exposure of these residues to the hydrophobic tails of the lipids by forming hydrogen bonds with each other (i.e., by burying the hydrophilic residues in the inter-helical interface), or by facing the pore lumen. Therefore, the probable structure of the bundle is expected to have maximum number of hydrophilic residues interacting with each other in the interface of the three α-helices. Furthermore, the orientation of hydrophilic residues in the bundle is expected to be such that the number of hydrophilic residues facing the exterior hydrophobic environment is minimal. The method used here, which is based on an analysis of the solvent accessible surface areas (SASA) of each of the residues and their relative orientations, has been successfully used to model the structures of a number of transmembrane proteins.[36]

The SASA of each of the side chains in the three TM domains was calculated and normalized with respect to the residue with the highest SASA value (W45). The orientations of the side chain of each of these residues were determined based on the position of the center of mass of their constituent non-hydrogen atoms and by taking a projection on a plane perpendicular to the helical axis. Polar plots of the average SASA values and the orientations that are represented in the projection plane perpendicular to the helical axis are depicted in figure 3. Each line in the plots corresponds to a residue in the TM domain, with the length of the line representing the SASA value and the angle formed between any two lines representing their relative orientations in
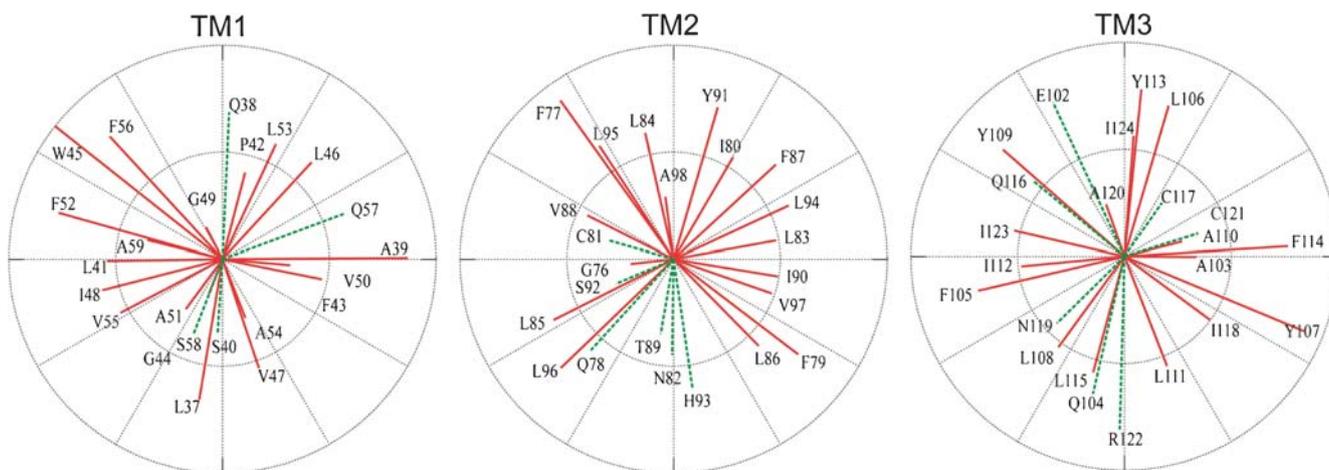

Figure 3. Polar plots representing the SASA values (Å²) and the relative orientation of all the residues that form a part of the three TM domains (hydrophilic and hydrophobic residues are given in green and red, respectively).

Table 1. List of hydrophilic and hydrophobic residues in the different clusters identified for TM1, TM2 and TM3.

TM domain	Cluster	Hydrophilic residues in cluster	Hydrophobic residues in cluster
TM1	C1	SER40, GLN57, SER58	ALA39, PHE43, GLY44, VAL47, VAL50, ALA51, ALA54
TM1	C2	GLN38, GLN57	ALA39, PRO42, PHE43, LEU46, VAL50, LEU53
TM2	C1	GLN78, CYS81, ASN82, THR89, SER92, HIS93	GLY76, LEU85, VAL88, LEU96
TM3	C1	GLU102, GLN116, CYS117, CYS121	LEU106, TYR109, ALA110, TYR113, ALA120, ILE124
TM3	C2	GLN104, ASN119, ARG122	LEU108, LEU111, LEU115, LEU118

the α-helix. The lines are colored based on the nature of the amino acid residue (hydrophobic residues in red and hydrophilic residues in green). The next step is to arrange the three polar plots so that there is a maximum overlap between the hydrophilic residues between each of the three pairs of the TM domains (TM1-TM2, TM2-TM3, and TM3-TM1). Based on the relative orientations of the hydrophilic residues, they were clustered into groups that can potentially interact with other TM domains. Accordingly, two clusters
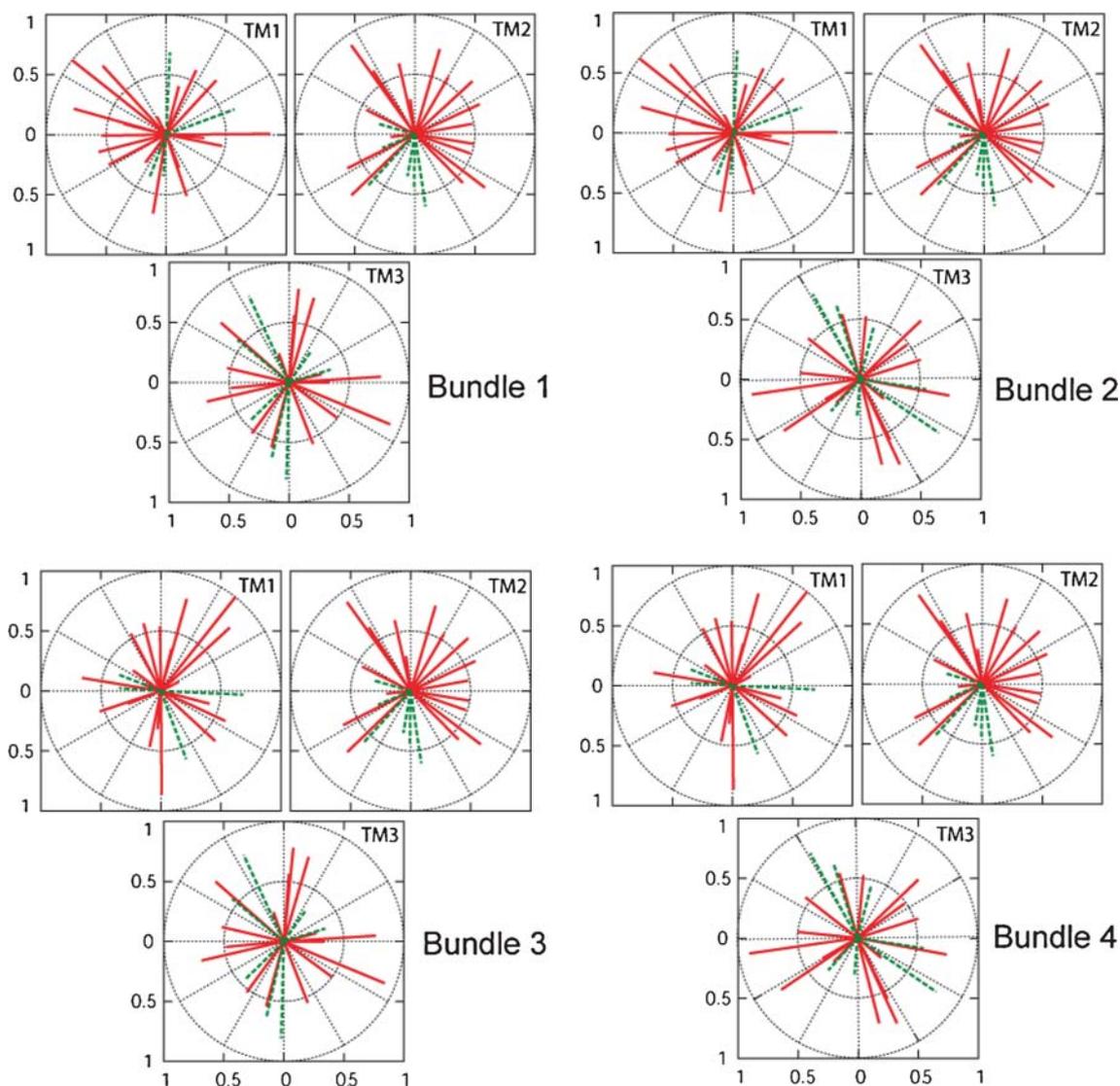
Figure 4. Arrangement of the individual polar plots (figure 3) to form four potential TM domain bundles (1, 2, 3, and 4) with respect to varying orientation of residues.

each of hydrophilic residues were identified for TM1 and TM3, and one cluster for TM2. The lists of the hydrophilic and hydrophobic residues that form each of these groups are given in table 1. It must be noted that Q57 was included to be a part of both the clusters of TM1. Since the side chains of the hydrophilic residues of TM3 form two distinct sets of orientations, there is no such overlap. There exist four unique ways of arranging the three TM domains (2, 1, and 2 different ways for TM1, TM2 and TM3, respectively) to form the bundle. Figure 4 depicts all the four potential structures of bundles. The individual TM domains in each of these bundles are arranged following these

steps: (a) the three TM domains are kept parallel to each other and are arranged such that the three centers of mass of the backbone atoms of the three TM domains form the vertices of an equilateral triangle, and the centroid of this triangle is located (with the distance between the TM domains being such that any two TM domains are as close together as possible without any bad contacts); (b) the centers of mass of the hydrophilic residues forming a cluster are computed for each of the three TM domains; (c) the individual TM domains are then rotated so that the center of mass of the hydrophilic cluster of each of the TM domains lies on the line connecting the vertex and centroid of the
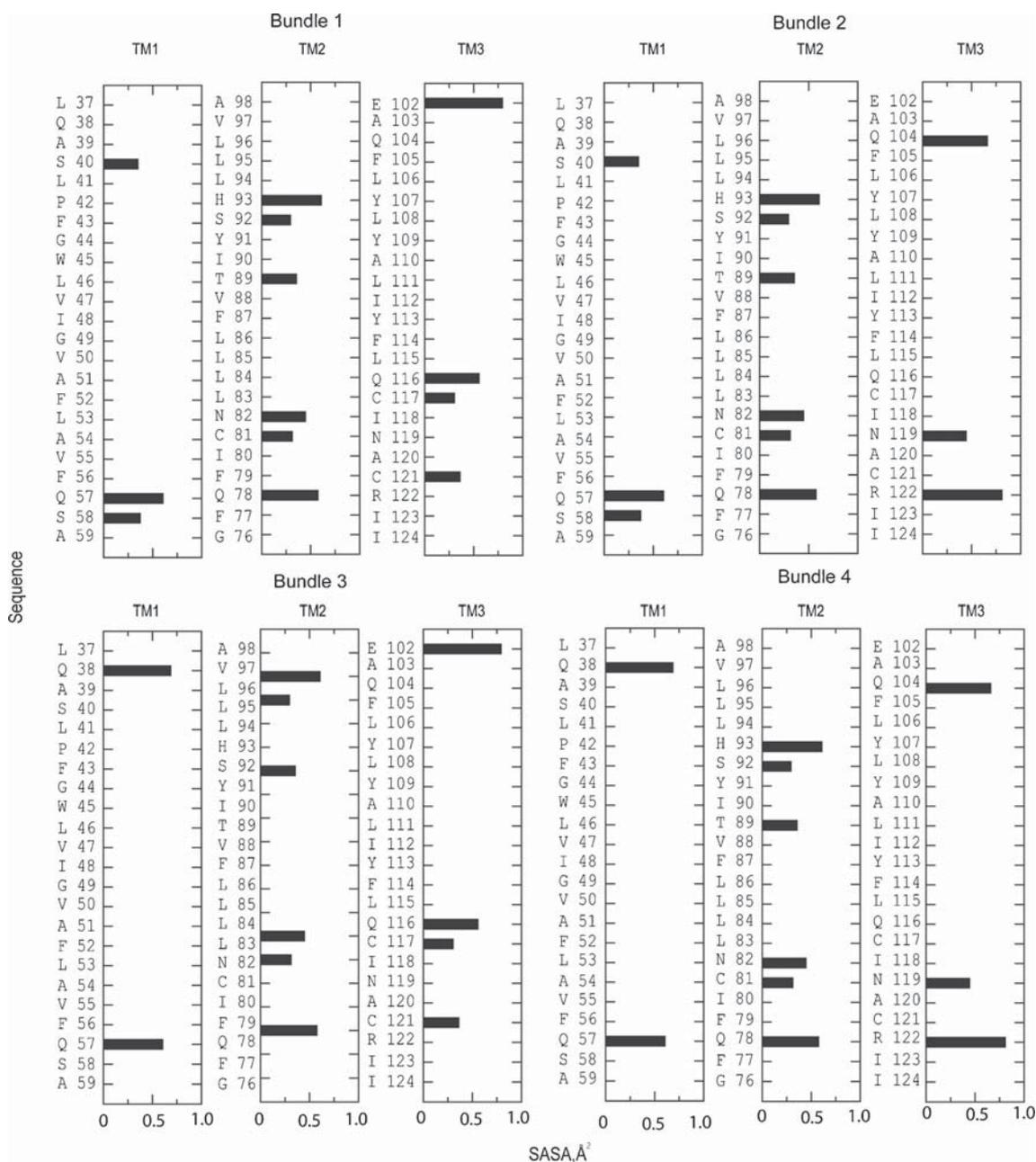

Figure 5. Sequence versus SASA of the overlapping hydrophilic residues in the four potential bundles.
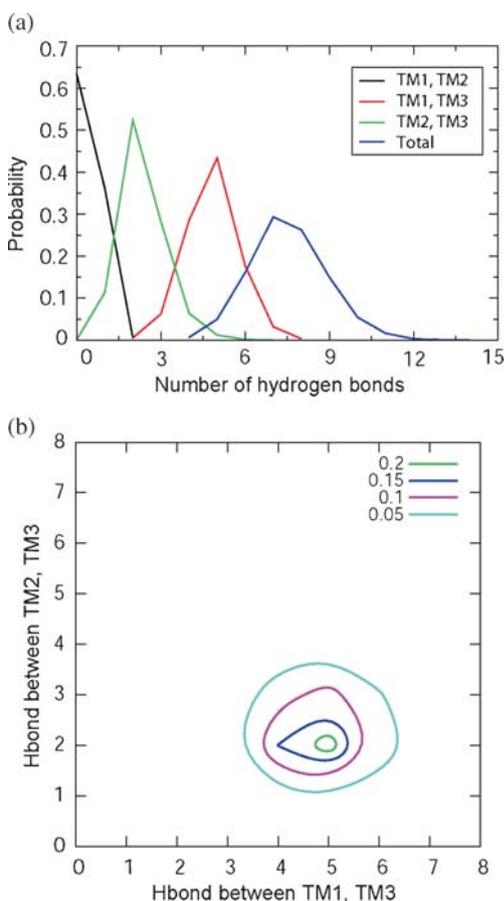
Figure 6. (a) Probability distribution of the number of hydrogen bonds between each pair of TM domains in bundle 2. (b) Contour plot of the probability distribution of the number of hydrogen bonds between TM1 and TM3, and TM2 and TM3.

triangle. Inter-helical hydrophilic interactions depend also on their positions along the helical axis in addition to their orientations. This was analyzed by examining the SASA of the residues lying in the interface of the three helices (figure 5). Models 2 and 4 were found to exhibit maximum interaction of hydrophilic residues between any two given TM domains, whereas in models 1 and 3, some of the interacting residue pairs were found to be a combination of hydrophilic and hydrophobic residues. Based on this, models 2 and 4 are proposed to be potential structures for the TM domain bundle for SARS 3a protein. Model 2 was further subjected to MD simulations to validate this structure, and to examine the inter-TM domain interactions that stabilize this bundle.

3.3 *Initial validation of the predicted structure of model 2*

A putative structure of the bundle was modeled by placing the three dimensional structures of the individual

TM domains as close to each other as possible and by avoiding obvious non-physical contacts. The relative orientations of the TM domains with respect to each other were decided based on the plot (Bundle 2) given in figure 4. The resultant model system was then placed in an implicit membrane environment and was aligned along the membrane normal. The simulation protocol used for this simulation was similar to that used for the implicit membrane simulations of the individual TM domains (see Methods and Supplementary Information). Inspection of the trajectories of this MD simulation indicated that the relative orientations of the three α-helices within the bundled structure remain very similar to the initial structure, indicating that the predicted structure of the bundle is reasonably stable. Most of the residues in the interface of the individual TM domains are hydrophobic. The hydrophilic residues that lie in the interface of the α-helices are expected to form hydrogen bonds with each other. The hydrophobic packing and hydrogen bond interactions were examined by calculating the SASA of the hydrophobic contacts, and by analyzing probability distributions of the hydrogen bonds formed between each pair of the individual TM domains, respectively. Most of the hydrophilic residues in the interface were found to form hydrogen bonds and the number of such interactions was found to vary from 4 to 12 (figure 6(a)). Among the three α-helical pairs, more hydrogen bonds are formed between TM1 and TM3 compared to the other two, with the number being the lowest between TM1 and TM2. The low number of hydrogen bonds between TM1 and TM2 can be attributed to the absence of a sufficient number of polar residues in the TM1/TM2 interface. The association between TM1 and TM2 is instead driven by hydrophobic packing, with a large number of hydrophobic residues lying in the interface. The exterior of the bundle should be dominated by hydrophobic residues rather than hydrophilic residues for favorable interactions with the hydrophobic lipid tails. To evaluate this criterion, SASA values of hydrophobic side chains and hydrophilic side chains that form the exterior of the bundle were calculated. Notably, the SASA of the hydrophobic residues were found to be more than four times that of the hydrophilic residues that form the exterior of the TM bundle (figure 7(a)). In the interface, hydrophobic-hydrophobic, and hydrophilic-hydrophilic contacts are expected to contribute to the stability of the bundle due to their packing and polar interactions, respectively. However, contacts between hydrophilic and hydrophobic residues in the interface are undesirable. These three possible natures of the contacts between interfacial residues were quantified using SASA calculations (figure 7(b)). The SASA of
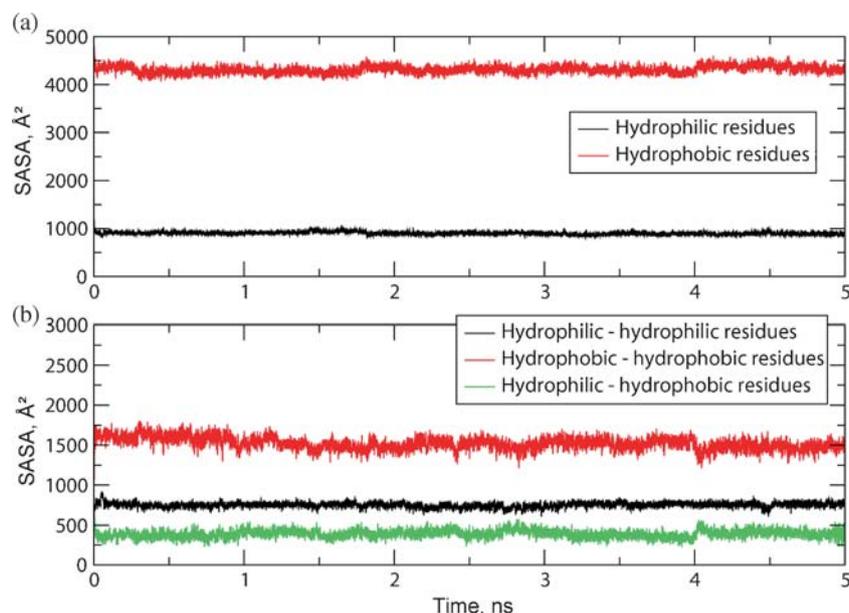
Figure 7. (a) SASA of hydrophilic and hydrophobic residues of bundle 2 against simulation time. (b) Contact area between hydrophilic and hydrophilic residues, hydrophobic and hydrophobic residues, and hydrophobic and hydrophilic residues versus simulation time in bundle 2.

favorable contacts were found to be much larger compared to the unfavorable contacts, further supporting the good quality of the structure.

3.4 *Addition of loops and further validation*

Loops were added to the equilibrated bundle to obtain a complete monomer, with the addition of loops expected to stabilize the distance between the different TM domains. The structure was initially equilibrated in an implicit membrane in order to allow the loops and the TM domains to relax. The backbone RMSD of the bundle over the course of the implicit membrane equilibration shows rapid stabilization of the structure within the first few picoseconds (figure S3 in SI). This equilibrated structure was then modeled in a fully explicit lipid bilayer environment in order to validate the stability and integrity of the model in a more realistic environment. Figure 8(a) shows the RMSD of the protein backbone over the course of simulations in the explicit membrane, and figure 8(b) shows the distance between the centers of mass of the three TM domains during these simulations. The results suggest retention of the structure of the bundle, with slight adjustment of the distance between the TM domains.

As mentioned previously, the unfavorable contacts between unlike amino acid side chains were comparatively smaller than the favorable contacts. Visual inspection of the structural model indicates that the hydrophilic residues N82, T89, and H93 of TM2 domain, which were originally interacting with the hydrophobic residues of TM1 and TM3 (see figure 4), are found to be present in the exterior of the bundle after the MD simulation. Figures 9(a) and 9(b) show the top and side views, respectively, of the model structure of the bundle obtained after the MD simulation in explicit lipid bilayer. These three hydrophilic residues of TM2 lie in one side of the helix to form a hydrophilic strip. Additionally, the hydrophilic residues C117 and C121 of TM3 also lie on the same side of the exterior of the bundle structure. These two sets of amino acids form a complete hydrophilic stripe on one side of the 3a protein, and are likely to constitute the pore lining of the ion channel after oligomerization. A recent study by Fischer and coworkers indicated that TM2 and TM3 are likely to line the pore, since a TM2:TM3 composition in an artificial lipid bilayer was found to conduct maximum current due to its ion channel activity, as compared to other combinations of TMDs.[33] This indicates good agreement of the structure with this experimental observation, further validating the adequacy of the model.

4. Conclusions

The present study proposes a model structure for a full monomeric unit of the SARS 3a protein, based on an
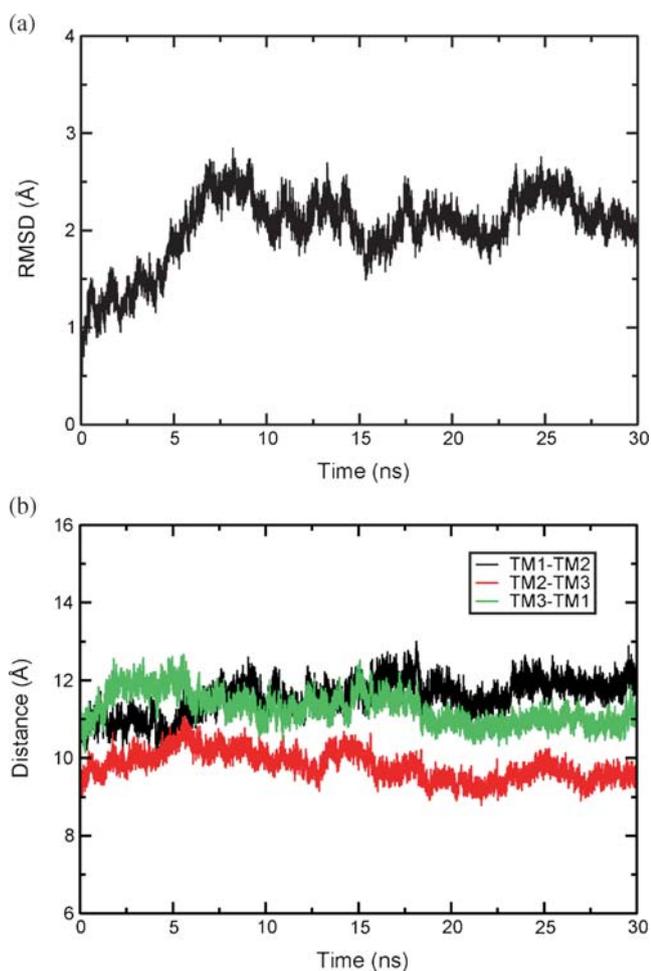
Figure 8. (a) Backbone RMSD of the bundle over the course of equilibration in an explicit membrane environment. (b) Distance between the centers of mass of the three TM domains during the simulation.
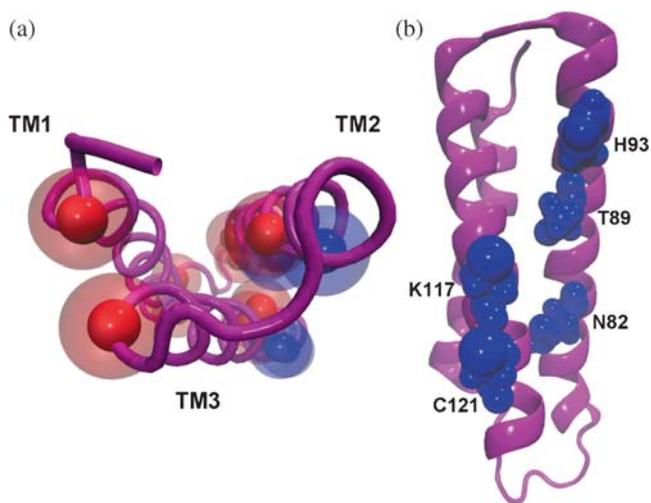

Figure 9. Structure of bundle 2 after the simulation. (a) Top view. (b) Side view. Interfacial hydrophilic residues between TM domains are presented as red spheres, and hydrophilic residues which are present outside the bundle and form a hydrophilic stripe, are shown as blue spheres.

approach that assembles individual TM domains such that unfavorable contacts are minimized. The three TM domains were initially investigated independently to gain insights into the properties of the individual TM domains, and the results suggest that the TM3 domain is likely to face the pore, rather than the lipid bilayer. A previously tested method based on the residue composition, orientation and SASA of the TM domains is employed by which TM bundles can be modeled starting from individual α-helices. The model proposed by the method has been validated using MD simulations and by comparison with available experimental data. The model structure was found to possess a hydrophilic stripe along one side of the protein formed by the hydrophilic residues of TM2 and TM3. This is in excellent agreement with a recent experimental report, which showed that the TM2 and TM3 domains are likely to line the pore of the ion channel. The structural model provides insights into residues that are likely to face the pore, and thereby play a role in ion channel activity. Future studies will exploit the model that is proposed here in the modeling of the oligomeric state of SARS 3a and in the understanding of its activity.

Supplementary Information

Details of the simulation methodology are given as Supplementary Text. Table S1 shows the magnitude of the force constant used for the different restraints during the six-step equilibration scheme. Figure S1 shows a schematic illustration of kink angle, as well as the time series for the kink angle over the course of implicit and explicit membrane simulations. Figure S2 shows the tilt angle of the TM domains with respect to the membrane normal in the different membrane environments. Figure S3 shows the backbone RMSD of the bundle during the implicit membrane simulations. Supplementary Information is available at www.ias.ac.in/chemsci.

Acknowledgments

References

1. Poutanen S M, Low D E, Henry B, Finkelstein S, Rose D, Green K, Tellier R, Draker R, Adachi D, Ayers M, Chan A K, Skowronski D M, Salit I, Simor A E, Slutsky A S, Doyle P W, Krajden M, Petric M, Brunham R C and McGeer A J 2003 *N. Engl. J. Med.* **348** 1995

2. Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt G M, Ahuja A, Yung M Y, Leung C B, To K F, Lui S F, Szeto C C, Chung S and Sung J J 2003 *N. Engl. J. Med.* **348** 1986

3. Tsang K W, Ho P L, Ooi G C, Yee W K, Wang T, Chan-Yeung M, Lam W K, Seto W H, Yam L Y, Cheung T M, Wong P C, Lam B, Ip M S, Chan J, Yuen K Y and Lai K N 2003 *N. Engl. J. Med.* **348** 1977

4. Drosten C, Günther S, Preiser W, van der Werf S, Brodt H R, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier R A, Berger A, Burguière A M, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra J C, Müller S, Rickerts V, Stürmer M, Vieth S, Klenk H D, Osterhaus A D, Schmitz H and Doerr H W 2003 *N. Engl. J. Med.* **348** 1967

5. Ksiazek T G, Erdman D, Goldsmith C S, Zaki S R, Peret T, Emery S, Tong S, Urbani C, Comer J A, Lim W, Rollin P E, Dowell S F, Ling A E, Humphrey C D, Shieh W J, Guarner J, Paddock C D, Rota P, Fields B, DeRisi J, Yang J Y, Cox N, Hughes J M, LeDuc J W, Bellini W J and Anderson L J 2003 *N. Engl. J. Med.* **348** 1953

6. Marra M A, Jones S J, Astell C R, Holt R A, Brooks-Wilson A, Butterfield Y S, Khattra J, Asano J K, Barber S A, Chan S Y, Cloutier A, Coughlin S M, Freeman D, Girn N, Griffith O L, Leach S R, Mayo M, McDonald H, Montgomery S B, Pandoh P K, Petrescu A S, Robertson A G, Schein J E, Siddiqui A, Smailus D E, Stott J M, Yang G S, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth T F, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples G A, Tyler S, Vogrig R, Ward D, Watson B, Brunham R C, Krajden M, Petric M, Skowronski D M, Upton C and Roper R L 2003 *Science* **300** 1399

7. Rota P A, Oberste M S, Monroe S S, Nix W A, Campagnoli R, Icenogle J P, Peñaranda S, Bankamp B, Maher K, Chen M H, Tong S, Tamin A, Lowe L, Frace M, DeRisi J L, Chen Q, Wang D, Erdman D D, Peret T C, Burns C, Ksiazek T G, Rollin P E, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Günther S, Osterhaus A D, Drosten C, Pallansch M A, Anderson L J and Bellini W J 2003 *Science* **300** 1394

8. Lai M and Holmes K 2001 In *Fields Virology* D M Knipe and P M Howley (eds.) (Philadelphia: Lippincott-Willams and Wilkins) p. 1163

9. Holmes K V and Enjuanes L 2003 *Science* **300** 1377

10. Snijder E J, Bredenbeek P J, Dobbe J C, Thiel V, Ziebuhr J, Poon L L, Guan Y, Rozanov M, Spaan W J and Gorbalenya A E 2003 *J. Mol. Biol.* **331** 991

11. Lu W, Zheng B J, Xu K, Schwarz W, Du L, Wong C K, Chen J, Duan S, Deubel V and Sun B 2006 *Proc. Natl. Acad. Sci. USA* **103** 12540

12. Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 *Nucleic Acids Res.* **8** 235

13. Lear J D, Stouffer A L, Gratkowski H, Nanda V and Degrado W F 2004 *Biophys. J.* **87** 3421

14. Senes A, Engel D E and DeGrado W F 2004 *Curr. Opin. Struct. Biol.* **14** 465

15. Nath D 2005 *Nature* **438** 577

16. Popot J L and Engelman D M 1990 *Biochemistry* **29** 4031

17. Popot J L and Engelman D M 2000 *Annu. Rev. Biochem.* **69** 881

18. Feller S E, Gawrisch K and Woolf T B 2003 *J. Am. Chem. Soc.* **125** 4434

19. Lindahl E and Sansom M S P 2008 *Curr. Opin. Struct. Biol.* **18** 425

20. White S H and Wimley W C 1999 *Annu. Rev. Biophys. Biomol. Struct.* **28** 319

21. Johansson A C and Lindahl E 2006 *Biophys. J.* **91** 4450

22. MacCallum J L, Bennett W F D and Tieleman D P 2007 *J. Gen. Physiol.* **129** 371

23. MacCallum J L, Bennet W F D and Tieleman D P 2008 *Biophys. J.* **94** 3393

24. Biggin P C and Sansom M S 1999 *Biophys. Chem.* **76** 161

25. Forrest L R and Sansom M S P 2000 *Curr. Opin. Struct. Biol.* **10** 174

26. Domene C, Bond P J and Sansom M S P 2003 *Adv. Protein Chem.* **66** 159

27. Nymeyer H, Woolf T B and Garcia A E 2005 *Proteins* **59** 783

28. Woolf T B, Zuckerman D M and Lu N D 2004 *J. Mol. Graph.* **22** 359

29. Efremov R G, Nolde D E, Konshina A G, Syrtcev N P and Arseniev A S. 2004 *Curr. Med. Chem.* **11** 2421

30. Krüger J and Fischer W B 2009 *J. Chem. Theory Comput.* **5** 2503

31. Hao-Jen H and Fischer W B 2011 *J. Mol. Model.* **18** 501

32. Goldman D E 1943 *J. Gen. Physiol.* **27** 37

33. Chien T H, Chiang Y, Chen C P, Henklein P, Hanel K, Hwang I S, Willbold D and Fischer W 2013 *Biopolymers* **99** 628

34. Hodgkin A L and Katz B 1949 *J. Physiol.* **108** 37

35. Jäckle J 2007 *J. Theor. Biol.* **249** 445

36. Padhi S, Ramakrishna S and Priyakumar U D 2015 *J. Comput. Chem.* **36** 539

37. The Sybyl Software, version 7.2 2006 (St Louis, MO, USA: Tripos Inc.)

38. Im W, Lee M S and Brooks III C L 2003 *J. Comput. Chem.* **24** 1691

39. Im W M, Feig M and Brooks III C L 2003 *Biophys. J.* **85** 2900

40. Brooks B R, Brooks III C L, MacKerell A D Jr, Nilsson L, Petrella R J, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner A R, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor R W, Post C B, Pu J Z, Schaefer M, Tidor B, Venable R M, Woodcock H L, Wu X, Yang W, York D M and Karplus M 2009 *J. Comput. Chem.* **30** 1545

41. MacKerell A D Jr, Bashford D, Bellott M, Dunbrack R L, Evanseck J D, Field M J, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau F T, Mattos C, Michnick S, Ngo T, Nguyen D T, Prodhom B, Reiher W E, Roux B, Schlenkrich M, Smith J C, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D and Karplus M 1998 *J. Phys. Chem. B* **102** 3586

42. MacKerell A D Jr, Feig M and Brooks III CL 2004 *J. Comput. Chem.* **25** 1400

43. Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kale L and Schulten K 2005 *J. Comput. Chem.* **26** 1781

44. Humphrey W, Dalke A and Schulten K 1996 *J. Molec. Graphics* **14** 33