



## Review

Hydrogen bond dynamics in membrane protein function <sup>☆</sup>Ana-Nicoleta Bondar <sup>a,\*</sup>, Stephen H. White <sup>b,\*</sup><sup>a</sup> Theoretical Molecular Biophysics, Freie Universität Berlin, Department of Physics, Arnimallee 14, 14195 Berlin, Germany<sup>b</sup> Department of Physiology and Biophysics and Center for Biomembrane Systems, University of California at Irvine, Irvine, CA, 92697-4560, USA

## ARTICLE INFO

## Article history:

Received 21 September 2011  
 Received in revised form 22 November 2011  
 Accepted 30 November 2011  
 Available online 8 December 2011

## Keywords:

Membrane protein structure  
 Hydrogen bond  
 Membrane protein dynamics  
 Lipid–protein interactions

## ABSTRACT

Changes in inter-helical hydrogen bonding are associated with the conformational dynamics of membrane proteins. The function of the protein depends on the surrounding lipid membrane. Here we review through specific examples how dynamical hydrogen bonds can ensure an elegant and efficient mechanism of long-distance intra-protein and protein–lipid coupling, contributing to the stability of discrete protein conformational substates and to rapid propagation of structural perturbations. This article is part of a Special Issue entitled: Protein Folding in Membranes.

© 2011 Elsevier B.V. All rights reserved.

## Contents

1. Introduction . . . . .	942
2. Networks of “static” hydrogen bonds in membrane protein crystal structures. . . . .	942
3. Hydrogen bond interactions can be extracted from MD simulations . . . . .	945
4. The dynamics of inter-helical H-bonding networks can be very complex . . . . .	945
5. The lipid membrane affects membrane protein H-bonding interactions. . . . .	946
6. Conclusions . . . . .	947
Acknowledgements . . . . .	949
References . . . . .	949

## 1. Introduction

Membrane proteins (MPs) function in the complex environment of the lipid membrane. There, protein atoms interact not only with each other and with water, but also with atoms of the surrounding lipid molecules. Hydrogen-bonding (H bonding) interactions – which form between a hydrogen atom covalently bound to an electronegative atom, and another electronegative atom – are of paramount importance for the assembly, structure, and functioning of membrane proteins. Discussions of MP H bonds tend to focus on the energetics of H bonding in membrane protein folding [1] and structure formation [2]. Here we focus instead on the role of H bonds in MP functions. In particular, we consider how MP function depends upon networks of H bonds and their dynamics. We discuss examples,

largely from our laboratories, of H-bonding networks within functionally different  $\alpha$ -helical membrane proteins and the complexity of the networks arising from H bonds between helices and between helices and the lipid membrane. Finally, we consider the relationship between hydrogen bonding and the conformational dynamics of membrane proteins.

## 2. Networks of “static” hydrogen bonds in membrane protein crystal structures

An analysis of a relatively small set of 134 TM helices in 13  $\alpha$ -helical protein structures indicated that almost each TM helix was connected via H bonding to the most proximal helix [3]; albeit helices with more than one H-bonding cluster were also observed, most inter-helical H bonds observed were between two amino acids from the two helices. Because each helix is likely H-bonded to a nearby helix then, even if each pair has only one H bond, a more complex picture arises in which the TM helices of the protein are interconnected

<sup>☆</sup> This article is part of a Special Issue entitled: Protein Folding in Membranes.

\* Corresponding authors. Tel.: +49 30 838 53583; fax: +49 30 838 56510.

E-mail address: [nbondar@zedat.fu-berlin.de](mailto:nbondar@zedat.fu-berlin.de) (A.-N. Bondar).

via direct or indirect H bonds (Fig. 1A). The dynamic breaking and forming of H bonds, allowing the TM helices to exchange H bonding partners, can only add to the complexity of this picture. Do such dynamical inter-helical H bonded networks exist? We show here that such dynamical interconnections do exist and, when present, they likely participate in controlling the conformational dynamics of TM proteins. We depict in Fig. 2 examples of different classes of membrane proteins in which we have identified networks of H bonds that interconnect TM helices. These inter-helical H-bond networks ensure coupling of key structural elements to remote regions of the protein.

The AHA2 plasma membrane proton pump is a member of the P-type ATPase family. Proteins from this family couple the hydrolysis of adenosine triphosphate (ATP) with large-scale conformational changes and pumping of cations across the membrane. In AHA2, the central proton donor and acceptor group on TM9 D684 is within hydrogen-bonding distance from N106 of TM2 (Fig. 2A). TM2 also contains D92, one of the acidic groups proposed as putative proton release groups [4]. The close interaction between D92 and E808 (N. Bondar, work in progress) ensures coupling of TM2/TM9 to the TM7–TM8–TM9 segment of the protein. That is, the inter-helical H-bonding connections ensure that the protonation states of D684 and D92 can be relayed to remote distances in the protein, and that changes in protein structure and dynamics can be relayed to the local environment of the proton transfer groups, modulating their electrostatic environment.

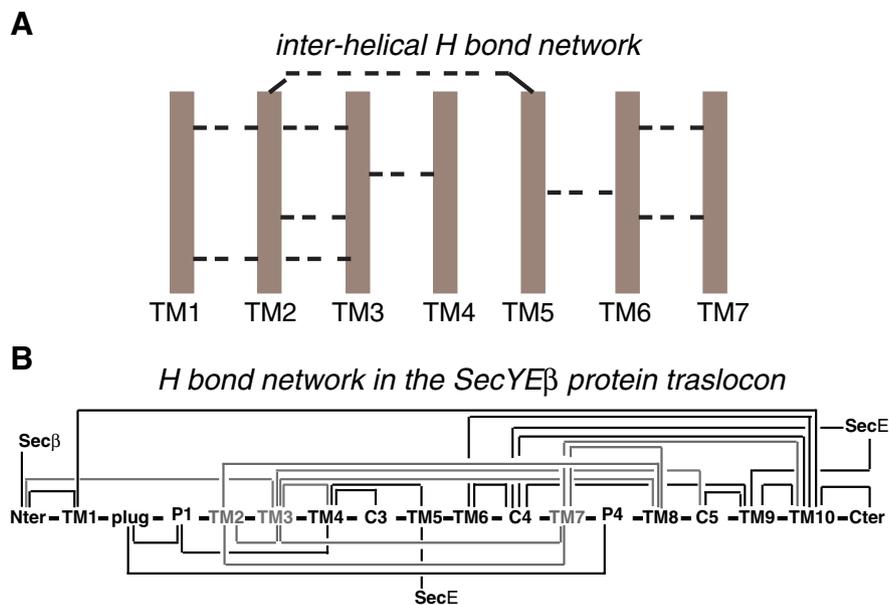
SecY $\beta$  is the protein translocation channel (translocon) found in the plasma membrane of prokaryotes. There, SecY $\beta$  is the central component of a larger secretion machinery that ensures that newly-synthesized proteins targeted to the SecY pathway are either secreted into the periplasm or inserted into the plasma membrane. The selection process appears to be based on biophysical principles of partitioning between the membrane and the translocon [5–7]. Release of TM helices into the lipid membrane is thought to occur via the lateral opening of helices TM2 and TM7 (Fig. 2B) [8–10]. We have found that SecY $\beta$  has a remarkable network of inter-helical H bonds: no fewer than 70 H bonds that interconnect different structural elements of

the translocon, and each of the TM helices has at least one inter-helical H bond (Fig. 1B) [11]. Importantly, the gate helices TM2 and TM7 are interconnected via H bonding with each other, and with TM3 (Figs. 1B, 2B). The presence of this central cluster of H bonds, and the extensive inter-helical H bonding of the translocon, couple the heart of the translocon to the remaining parts of the machinery.

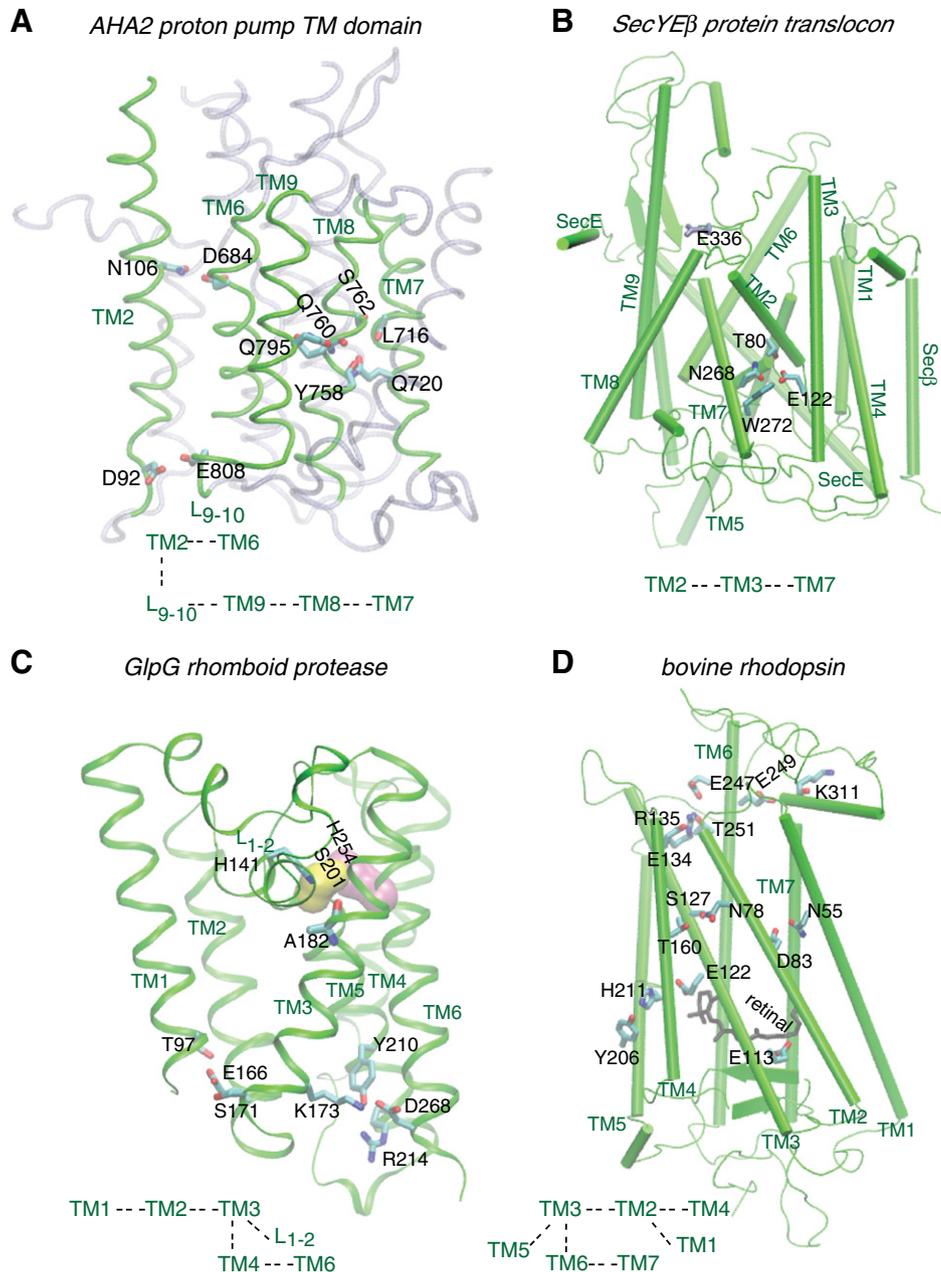
The GlpG rhomboid protease from *Escherichia coli* is a model system for understanding intramembrane proteolysis, a fascinating process in which a membrane-embedded protein cleaves other TM segments [12–18]. Although the function of GlpG is completely different from that of SecY, there is a remarkable symmetry in their mechanism of action: whereas SecY must open a lateral helical gate to release TM substrates into the surrounding lipid membrane, in GlpG the lateral gate helix TM5 opens towards the membrane to admit the TM substrates for docking and cleavage. That is, both SecY and GlpG are helix-gated membrane proteins.

The flexible TM5 gate helix of GlpG is not involved in inter-helical H bonding, but it has been noted that it connects to TM2 via hydrophobic interactions [19–22]. On the other hand, TM2 is part of an H-bonding cluster with TM1 and TM3; TM3 is further connected to TM4 and TM6, which carry the catalytic groups, and to loop L<sub>12</sub> (Fig. 2C)—a loop that may play important structural [23] and lipid-sensing roles [22]. TM3 of GlpG could thus be seen as having a similar role as that of TM3 in SecY (Fig. 1B), serving as a key node in the network of H bonds that couple different regions of the protein.

Bovine rhodopsin is a prototype for the rhodopsin family of G-protein-coupled receptors (GPCRs). The protein is covalently bound to the 11-*cis* retinal cofactor via a protonated Schiff base; absorption of light by the retinal chromophore triggers conformational changes of the protein that ultimately lead to the active state of the receptor, in which binding to the G protein occurs. About 70% of the GPCRs of the rhodopsin family contain the sequence (D/E)R(Y/W) on TM3 [24]. Amino acids E134 and R135 of this conserved motif form the so-called ionic lock with TM6 amino acids E247 and T251 (Fig. 2D). The ionic lock stabilizes the inactive conformation of the GPCR [24,25], and affects the energetics of the transition between the active and inactive states of bovine rhodopsin [26]. TM7 connects to TM6



**Fig. 1.** Hydrogen bonds (H bonds) interconnect TM helices. (A) Schematic representation of TM H-bond interconnections with TM helices depicted as cartoons, and H bonds as dashed lines. One or more H bonds can interconnect pairs of TM helices. Note that TM2 is a connectivity node; it interconnects the TM1–TM2–TM3–TM4 and TM5–TM6–TM7 segments of the protein. (B) Interconnectivity map of the SecY $\beta$  protein translocon. Note that TM3 connects the gate helices TM2 and TM7 to other regions of the translocon. Panel B is modified from Bondar et al. [11].



**Fig. 2.** Examples of networks of H bonds that interconnect TM helices of different classes of membrane proteins. Only selected H-binding amino acids sidechains are depicted explicitly. Each panel is accompanied by a schematic representation of the inter-helical connections mediated by H bonds. (A) The AHA2 P-type plasma membrane proton pump (PDB ID: 3B8C). For simplicity, TM helices that do not participate in the H bonds depicted explicitly are shown as transparent gray ribbons. The dashed line between L9-10 and TM9 indicates that those two structural elements are linked together. (B) The central cluster in the SecYE $\beta$  protein translocon (PDB ID: 1RHZ). E122 of TM3 mediates a cluster of H bonds that involve amino acids of the gate helices TM2 and TM7; E122 is highly conserved as Glu in archaea and eukarya, and present mostly as Gln in bacteria [11]. (C) The rhomboid intramembrane protease (PDB ID: 2IRV). The catalytic groups Ser201 and His254 are shown as yellow and purple surfaces, respectively. (D) Bovine rhodopsin with the retinal cofactor shown as black bonds (PDB ID: 1U19). E134 and R135 are part of the conserved E(D)RY motif. At 3.8–3.9 Å, the distances between the E122 and H211 sidechains and between E249 sidechain and the K311 amide group are somewhat long for an H bond, but in a flexible protein environment that distance could easily sample H-bonding (see, e.g., Fig. 3C). The molecular graphics images in panels A–D were prepared using the VMD software [52] based on published crystal structures [4,9,19,53]. The simulations of SecYE $\beta$  and GlpG were performed using the CHARMM [54] force field parameters for the protein [55] and lipid [56] atoms, and the TIP3P water model [57]. The length of the bonds involving H atoms are constrained using the SHAKE algorithm [58], the short-range interactions are cut-off at 12 Å using a switching function between 8 Å and 12 Å, and the Coulomb interactions are computed using the smooth particle mesh Ewald summation [59,60]. Langevin dynamics were used to maintain the temperature constant at 300 K (POPC lipids) or 310 K (POPE lipids), and a Nosé–Hoover thermostat [61,62] for keeping the pressure at 1 bar. After an initial equilibration with weak harmonic constraints (2 kcal mol<sup>-1</sup> to 5 kcal mol<sup>-1</sup>) and an integration step of 1 fs, we switched off all harmonic constraints and used the reversible multiple time-step algorithm [63,64] with integration time-steps of 1 fs for the bonded-forces, 2 fs for the short-range non-bonded forces, and 4 fs for the long-range electrostatic forces. MD simulations of GlpG were based on the crystal structure of Ben Shem et al. [19]; the simulation systems comprised ~160,000 atoms (~500 lipid molecules, solvent water, and ions for charge neutrality). In the MD simulation of the SecYEG translocon we used the crystal structure of Van den Berg et al. [9] for the protein atoms, and a patch of 475 POPC lipids (217,820 atoms including solvent water and ions for charge neutrality).

[26], which connects to TM3; TM3 is also part of a network involving TM5, TM2, TM1, and TM4 (Fig. 2D). It thus appears that TM3, which harbors not only amino acids from the ionic lock, but also the E113 counterion of the protonated retinal Schiff base, is interconnected to other regions of the protein via H bonding (Fig. 2D). Changes in H

bonding between TM3 and TM5 have been associated with the formation of the active state of rhodopsin [27].

The above discussion of H-bond networks that interconnect critical regions of the protein to other structural elements is largely based on the analysis of static crystal structures. Although the H

bonds observed in crystal structures are undoubtedly important in the function of the protein in the native membrane environment, the dynamics of complex H-bonded networks and how these networks may respond to perturbations (such as mutations or variations in the lipid membrane environment) cannot be determined from crystallographic structures without the help of molecular dynamics (MD) simulations. In what follows, we will use the results from MD simulations of the SecYE $\beta$  protein translocon and the GlpG rhomboid protease to illustrate the complex dynamics of inter-helical networks of H bonds, the dependence of the H-bond dynamics on the lipid membrane environment, and how extensive H-bonding networks can control the conformational states of membrane proteins. We begin with a brief introduction to MD simulations.

### 3. Hydrogen bond interactions can be extracted from MD simulations

Membrane proteins in their native membrane environments are not static as observed in crystal structures. Rather, they are dynamic. MD simulations are crucial for understanding H-bond dynamics, because they allow us to extend the observational range of the experiments by reconstructing the physiological lipid membrane environment of solvated membrane proteins, and consequently to investigate dynamics in atomic detail. A molecular mechanics (classical) MD simulation consists of solving numerically the classical equations that describe the motions of all particles of the system—the protein, lipid, and water atoms, and ions [28]. One obtains from simulations trajectories that describe the evolution in time of the coordinates of all the atoms of the system (generally at room temperature) for a finite period of time. The resulting MD trajectories can be used to dissect interactions between atoms. Currently, typical all-atom simulation times are 50–100 ns long, but microsecond times can be achieved for small systems at elevated temperatures using modest processors [29]. Emerging processor technology [30] will soon make it possible to achieve routinely 10  $\mu$ s time scales and beyond for membrane proteins [31,32] and millisecond time scales for small soluble proteins [33,34].

Starting with the crystallographic coordinates, the first step in the simulation of a membrane protein is to insert it into a lipid bilayer composed of the lipids of interest. To perform the insertion, the centers of mass of the bilayer system and the protein are made to coincide by aligning the protein's transmembrane principal normal to the bilayer, followed by removal of lipids that overlap the protein. The system must then be equilibrated. In short, the protein is first restrained to allow the bilayer to relax around it. Then the restraints are slowly removed until the system runs freely. Besides the composition, the temperature and pressure of system are held constant using standard algorithms. The system is allowed to run until it is well equilibrated as judged by the stability of the simulation box with periodic boundary conditions. To illustrate the kind of information on H bond dynamics that can be gleaned from MD simulations, we consider two examples from our laboratories: prolonged MD simulations of the SecYEG protein translocon [11] and of the GlpG intramembrane protease [22]. Those papers should be consulted for the technical details of the simulations. A brief description of the protocol used for MD simulations of GlpG and SecY is given in the legend of Fig. 2.

### 4. The dynamics of inter-helical H-bonding networks can be very complex

In the crystal structure of the *Methanococcus jannaschii* SecY protein translocon [9], amino acids T80, E122 and N268 of the central H-bonding cluster are interconnected via H-bonds characterized by distances of 2.7 Å (T80:E122) and 3.2 Å (T80:N268) (Fig. 2B). The static distance between W272 and E122 (4.4 Å) is somewhat long for an H bond. The MD simulations of the translocon at room temperature

revealed a much more complex picture [11]: only one of the H bonds of the TM2–TM3–TM7 cluster, formed by sidechains of T80 and N268, is stable throughout the entire simulation (Fig. 3A). The dynamics of the other sidechain:sidechain distances have a stable pattern in which H bonds are broken and reformed (Fig. 3B–D). The breaking and reforming of an H bond within this cluster can take between picoseconds to nanoseconds and tens of nanoseconds. W272 H bonds transiently to E122 (Fig. 3C). E122 appears as a key player in this cluster; it can have two H bonds with N268 (one for each carboxylic oxygen atom), and it also H bonds with T80 and W272. The complexity of the interactions mediated by E122 is reflected in the existence of two or even three sub-conformers of SecY characterized by different distances between E122 and N268 (Fig. 4B), E122 and W272 (Fig. 4C), and between E122 and T80 (Fig. 4E). Because the breaking and forming of the H bonds is relatively fast – that is, the energy barriers separating the sub-conformers are small and can be overcome easily – the overall structure of the protein remains stable; indeed, in MD simulations at room temperature the root-mean squared distances of the TM region of SecY relative to the starting crystal-structure coordinates was stable at  $\sim 2$  Å [11].

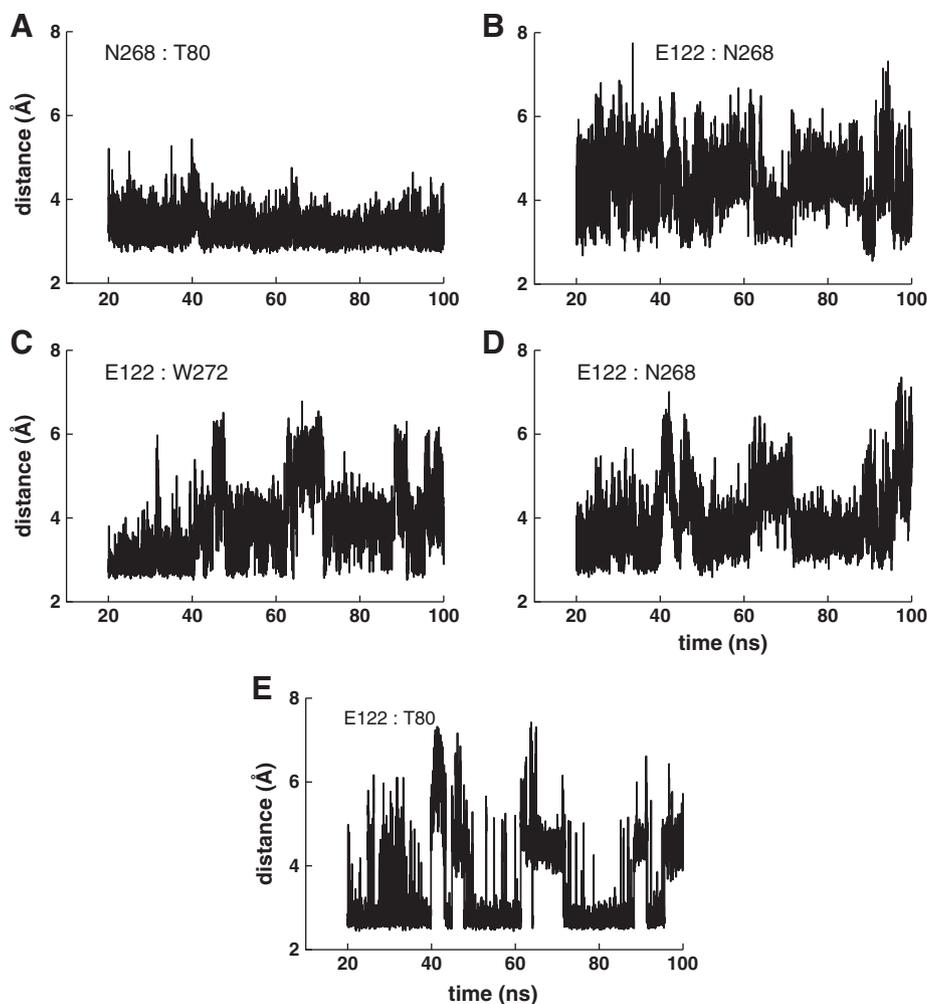
The central role of E122 is supported by our observations that this amino acid is highly conserved as Glu in archaea, as Gln or Glu in eukarya, and largely as Gln in bacteria [11]. Within the data set used for the sequence analyses, E122 is highly conserved. E122 is never replaced by an Asp sidechain, suggesting that a long sidechain at position 122 is required for mediating inter-helical H bonds. The replacement of E122 by Gln in bacterial SecY is accompanied by the absence of the N268 sidechain [11]. One would thus expect that the dynamics of the TM2–TM3–TM7 helices in bacterial SecY to be different from that of the archaeal SecY discussed here.

The complex pattern of H bonding dynamics of the TM2–TM3–TM7 cluster in SecY could not have been foreseen from the crystal structure alone. Furthermore, it appears that the dynamics of the inter-helical H bonds can depend on the lipid interactions.

In extensive MD simulations of GlpG, we observed that the dynamics of the H-bonded network that interconnects helices TM1, TM2, and TM3 (Fig. 2C) is different in lipid membranes composed of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine (POPC) compared to 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylethanolamine (POPE): the TM2:TM3 H bond mediated by E166 and S171 is stable at  $2.6 \pm 0.1$  Å in POPE, whereas in POPC it breaks and reforms rapidly throughout the 35 ns MD, being present only for  $\sim 50\%$  of the time.

The observation that the dynamics of the inter-helical H bonds can depend on the composition of the lipid membrane is important, because there is increasing evidence from experiments on various membrane proteins that changes in the lipid membrane composition can affect protein function, or that a protein can function only if certain lipids are present. For example, the *E. coli* GlpG rhomboid protease is active when reconstituted in PE lipids, but not in PC [35]. Proper functioning of the secondary multidrug transporter LmrP requires H bonding between a surface-exposed Asp and the PE lipid membrane, and is incompatible with PC lipids [36]. Anionic lipids enhance significantly binding of the SecY translocation channel to the SecA motor of the translocase machinery [37], and direct lipid:protein interactions are certainly involved in the recognition of TM helices by the translocon [5–7]. The energetics of the transition between the active and inactive states of visual rhodopsin depends on the composition of the lipid membrane [38]. Lipids can have conserved binding sites on the protein surface [39], and bind tightly to these specific sites [40].

The mechanisms by which the lipid membrane composition affects the functioning of the membrane protein are not yet entirely clear. Because changes in inter-helical H bonding can be associated with protein conformational transitions, differences in the dynamics of inter-helical H bonds [22] could contribute to the observed effects of the lipid membrane composition on membrane protein function.



**Fig. 3.** The dynamics of inter-helical H bonding can be highly complex. Illustration of the dynamics of the inter-helical H bonds in the central cluster (TM2–TM3–TM7) of the SecY protein translocon. The dynamics of the H bonds are monitored here by the time-dependent distances between the heavy atoms of the H bond donor and acceptor groups. All distances are reported in Å. Panels B and D illustrate H bonding of N268 to the two carboxyl oxygen atoms of E122. Distances  $\leq 3.5$  Å are considered here as H-bonding distances. We used for this analysis the last 80 ns of the trajectory. See Fig. 2B for a molecular picture of the central H-bonding cluster. The complex dynamics of the inter-helical H bonds in wild-type SecY was discussed in [11]. The protonation state of His amino depends on the local electrostatic environment [65]. In the MD simulation of the wild-type translocon from Bondar et al. [11], we modeled all His amino acids in the N $\delta$ 1 tautomeric state; that simulation was prolonged to 49.3 ns. The time-series of H bonds presented here are from a new  $\approx 100$  ns trajectory in which the His amino acids were modeled in the N $\delta$ 2 tautomeric state; the complex dynamics of the inter-helical H bonds in the TM2–TM3–TM7 cluster depicted here are consistent with those observed in [11].

The question then is why would the dynamics of inter-helical H bonds – or of the entire protein – depend on the lipid membrane? The location of the GlpG inter-helical H bonds relatively close to the helix termini – and thus close to the lipid headgroup region – could be used as an argument to suggest that simple electrostatic effects are important. That is, different lipid headgroups would create a different electrostatic environment for the inter-helical H bonds. Although it certainly is true that different lipid headgroups would provide different electrostatic environments, the effects of the lipid membrane composition on protein dynamics, and in particular on its H-bonding interactions, can be rather complex.

### 5. The lipid membrane affects membrane protein H-bonding interactions

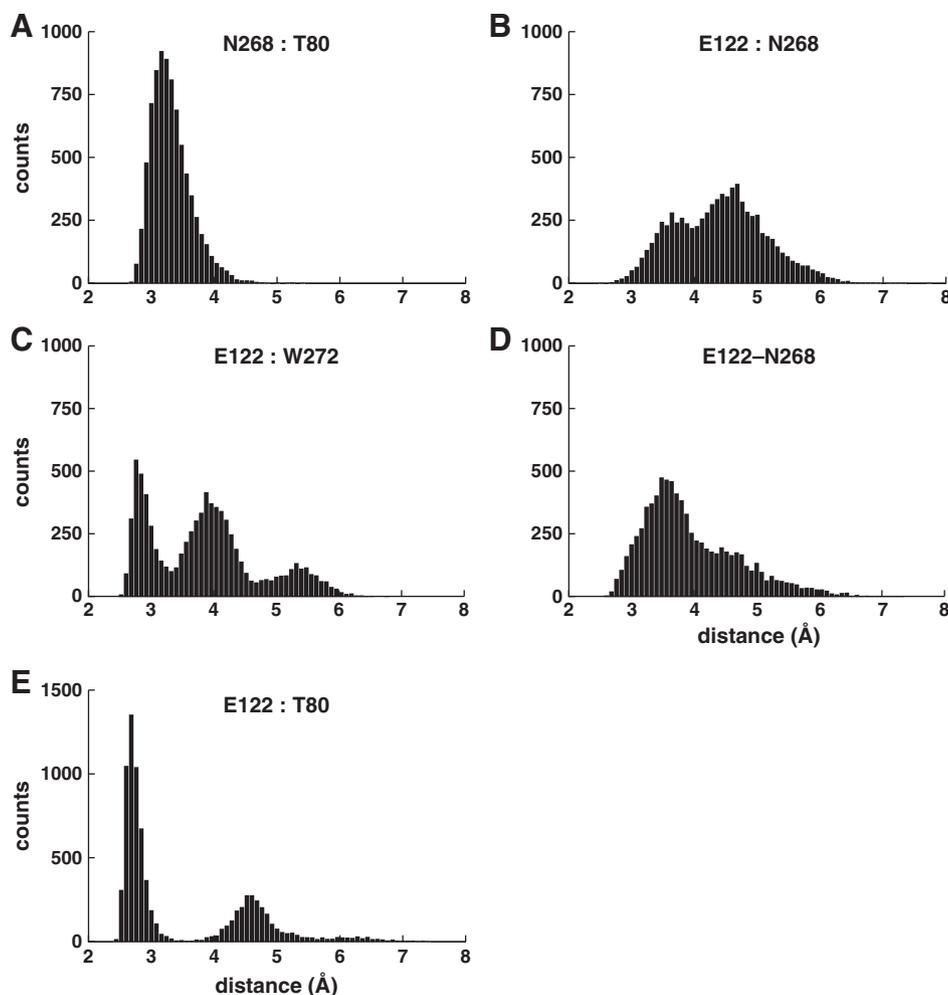
Several excellent reviews of the mechanisms by which lipids influence protein function have been published in the last several years [41–43]. These reviews noted, for example, that different lipid membrane compositions could imply differences in the macroscopic properties of the membrane (viscosity, phase transition, lateral pressure), but also differences in specific protein:protein and protein:lipid interactions [41]. The MD simulations reviewed by Jensen and Mouritsen

[42] illustrate at atomic detail how changes in the lipid headgroups influence the formation of a water wire inside the GlpF channel.

The close coupling between GlpG and the lipid membrane that is necessary for docking and cleaving the TM substrate makes GlpG a challenging model system for understanding the general physical principles of how lipids modulate protein function. We use here GlpG to illustrate how complex the molecular picture of protein:lipid interactions can be when one accounts for dynamics.

Crystal structure analyses in which electron densities for detergent and/or lipids could be observed have provided valuable glimpses into the possible interactions of GlpG with the membrane [44]. The electron densities observed by Wang et al. [44] indicated that the membrane is very thin close to the protease. Vinothkumar [45] has shown in atomic detail how lipids adapt to the surface features of GlpG to match its varying hydrophobic thickness. Indeed, significant nonuniform thinning of the membrane close to the protease was revealed by detailed MD simulations of GlpG in hydrated lipid membranes [22].

The thinning of the membrane (Fig. 5) occurs as the lipid molecules mold to the small hydrophobic thickness and rather unusual shape of GlpG (Figs. 2C, 5). Although thinning of the membrane is observed with both POPC and POPE lipids [22], there are significant



**Fig. 4.** Distribution of sidechain distances in the TM2–TM3–TM7 cluster of the SecY translocon. Histograms were calculated using the time-series presented in Fig. 3, for the range of values between 2 Å and 10 Å, with a bin size of 0.08 Å (i.e., 100 bins). All distances are reported in Å. Note that the distribution of the N268:T80 distances is slightly skewed towards non-H-bond distances; the skew towards large distances is more pronounced for the E122:N268 distance in panel D.

differences in how POPC and POPE interact with GlpG, and the H-bond dynamics of GlpG depends on the lipid membrane environment. The difference could be explained by the fact that the critical loop  $L_{12}$  and the cap loop thought to control access to the catalytic site contain polar amino acids located at the lipid headgroup interface.

E134 of  $L_{12}$  H bonds tightly to a POPE lipid (Fig. 5A); as a result of the E134:POPE H bond, water molecules cannot penetrate into the lipid bilayer. In POPC, however, the E134:lipid headgroup interaction is water-mediated, and water molecules move deeper into the membrane, where they replace the Y138:K132 H bond with protein:water interactions. As a result,  $L_{12}$  is locally less structured in POPC than in POPE. A dependence on the lipid membrane environment was also observed for lipid H bonding of other amino acids located on the GlpG surface [22].

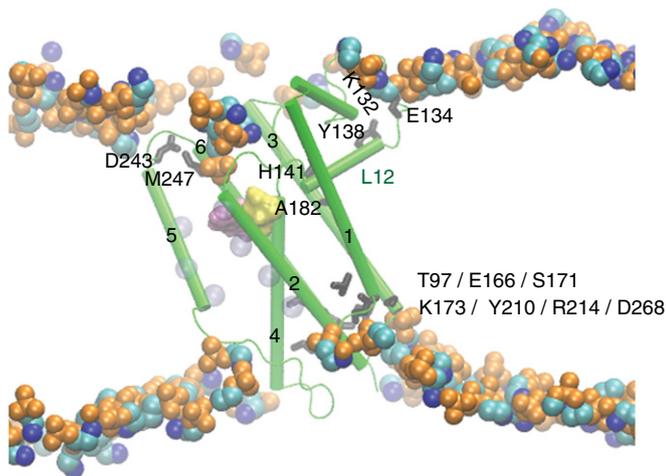
The H-bonding connectivity between  $L_{12}$  and TM3 implies that the structure and dynamics of the loop is coupled to that of TM3, and can affect the dynamics of the H-bonding network mediated by TM3 (Fig. 2C). This was indeed observed in the MD simulations [22]. Because TM3 H bonds to TM2, which contributes to the substrate-docking site, the extensive coupling via H bonds mediated by TM3 and the sensitivity of  $L_{12}$  to the lipid environment could represent (or be part of) the mechanism by which the protein is tightly coupled to the lipid membrane. That is, the protein has at least one structural element that can H bond to both lipid and protein groups. H bonding to the lipid ensures that the local protein structure and dynamics depends on the composition of the lipid membrane; the intra-protein H

bond couples the structure and dynamics at the protein:lipid interface to remote regions of the protein.

Tight H-bond-mediated coupling between the protein and the surrounding lipid membrane was also observed in the case of the SecY protein translocon [11]. Amino acid E336 is located on the cytoplasmic tip of TM8 (Fig. 2B), where it participates in an H-bond cluster that includes TM2, TM3, and TM8 [11]. It has been shown by experiments that mutating to Arg the corresponding E382 in yeast increases the translocation of protein segments with more positively charged ends [46]. Direct changes in the electrostatic interactions between SecY and the translocating protein segments could contribute to the observed changes in peptide translocation [46]. But simulations on the E336R mutant of the *M. jannaschii* translocon indicated that R336 H bonds to a lipid headgroup instead of participating in the TM2–TM3–TM8 H bond network; the structure and internal solvation of the mutant are different from those of the wild-type translocon [11]. The presence of the extensive H-bonded network interconnecting the TM helices and the loops of the translocon (Fig. 1B) likely explains why the protein structure and water interactions change when an H bond from an inter-helical H bond cluster is replaced with a lipid H bond.

## 6. Conclusions

Inter-helical H bonding of TM membrane proteins can ensure an elegant and efficient mechanism for long-distance coupling within



**Fig. 5.** Lipid and intra-protein H bond coupling in GlpG. The active site groups S201 and H254 are shown as yellow and purple surfaces, respectively. Loop L<sub>12</sub> H bonds to the POPE lipid membrane via E134, and to the protein via the L<sub>12</sub>-H141:TM3-A182 H bond. TM3 further connects via H bonding to TM1, TM2, TM4, and TM6 (see also Fig. 2C). TM2 is connected to the gate helix TM5 via hydrophobic interactions; these interactions are illustrated schematically by the transparent blue van der Waals spheres on TM2 and TM5, which represent the C $\alpha$  atoms of amino acids L161, W157, and F153 TM2, and L229, F232, W236 on TM5. The sidechain:backbone interaction between D243 and M247 is significantly more stable in POPE than POPC lipids [22]. In a POPC lipid bilayer the H bond between the Y138 sidechain and the K132 backbone is broken as water molecules penetrate deeper into the lipid membrane. Fig. 5 is modified from Bondar et al. [22].

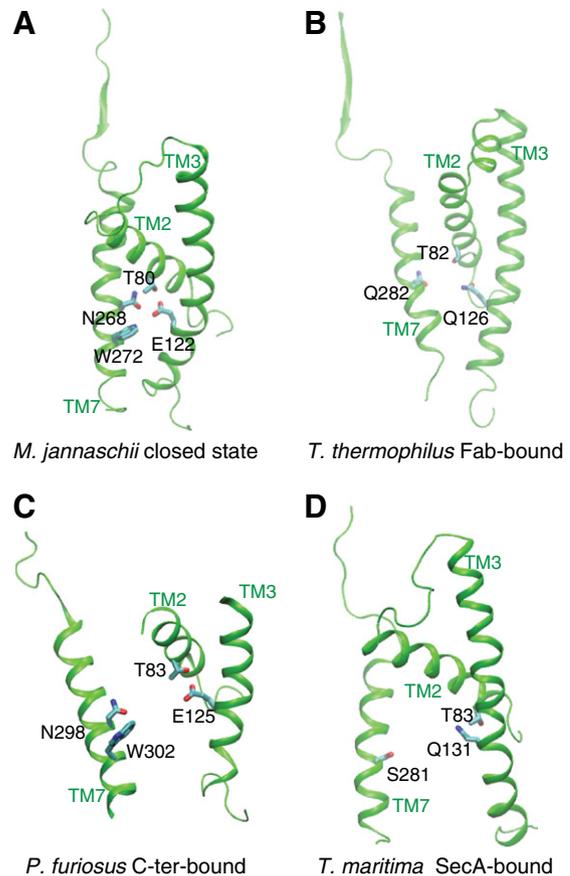
the protein. The coupling can be easily extended to the surrounding lipid membrane via a structural element that H bonds both to the membrane and to the protein.

The H bonds interconnecting TM helices can have a complex dynamics that can be assessed with MD simulations, but not from visual inspection of a static crystal structure. Inter-helical H bonds can be very stable, or can have a stable pattern of breaking and reforming with time scales ranging from picoseconds to nanoseconds and tens of nanoseconds (Fig. 3). An important question that emerges from the observation of such H-bonding patterns is how the breaking and reforming of H bonds on the nanosecond time scale relate to the large-scale, slow global structural rearrangements that may be associated with protein function. Based on the analysis discussed here, we suggest that the clusters of H bonds that inter-connect TM helices of membrane proteins contribute significantly to controlling protein conformation.

In the absence of perturbations, the clusters of H bonds help stabilize protein conformation. Although an H bond may break and reform rapidly, without a high energetic cost, during the time that that particular H bond is broken the H-bonding partners H bond with other groups within the cluster. For example, while T80 and N268 are engaged in a stable interaction (Figs. 3A, 4A) E122 interacts mostly with N268 at time ~65 ns (Fig. 3B), and with T80 at time ~80 ns (Fig. 3E). The overall protein structure is maintained.

Conformational dynamics is essential for enzyme function. Changes in the preferred geometry occur along the enzyme reaction cycle — for example, upon binding of a ligand. Importantly, the enzyme can sample conformations similar to those in the ligand-bound state even in the absence of the ligand; binding of the ligand would then simply shift the enzyme's conformational equilibrium towards the active, ligand-bound form [47]. The slow collective motion of lid opening may be facilitated by fast ps–ns dynamics at local hinge sites of the enzyme [48]. As discussed below, we think that networks of inter-helical H bonds with distinct conformational modes (as the example in Fig. 4) may presage shifts in the population of the conformational states sampled along the protein functional cycle.

SecY helices TM2 and TM7 are expected to undergo motions that would allow opening of a lateral gate towards the lipid membrane [8–10]. One would thus expect that at least some of the inter-helical H bonds of TM2 and TM7 (Figs. 1B, 2B and 3) would break when the translocon opens towards the membrane. TM3, which H bonds to both TM2 and TM7, H bonds with additional regions of the translocon (Fig. 1B); the extensive H bonds of TM3 would indicate that large-scale motions of TM3 are unlikely to accompany lateral opening of the translocon. Since the H bonds between TM3 and TM7 are relatively weak (Figs. 3B–D, 4B–D), and in the closed state of the translocon TM3-E122, TM7-N268 and TM7-W272 side-chains already sample conformations in which the E122:N267 and E122:W272 are too long for TM3:TM7 inter-helical H bonds, one could expect that lateral opening of the translocon could involve breaking of the TM3:TM7 H bond. Breaking of the TM3:TM7 H bond would mean enhanced dynamics of TM7, and thus a de-stabilization of the TM2:TM7 H bond (Fig. 4A) with a shift towards a conformer in which TM7 is free of H bonds with TM2 and TM3, while the TM3:TM2 H bond (Fig. 4E) may still be present.



**Fig. 6.** H bonding in the central TM2–TM3–TM7 cluster of SecY is coupled to SecY's conformation. The TM2, TM3, and TM7 helices with selected amino acid residues are depicted for the *M. jannaschii* SecY translocon in the closed state [9] (A), and for the structures thought to represent SecY open to various extents: the Fab-bound *T. thermophilus* [50] (B), the archaeal *P. furiosus* bound to the C-terminal fragment of another SecY copy [51] (C), and the SecA-bound *T. maritima* translocon [49] (D). Note that in the various open structures TM2 and TM3 remain relatively close to each other, whereas in the structures depicted in panels B–D, TM7 moves away from TM2/TM3. *M. jannaschii* E122 is conserved as Glu in the archaeal *P. furiosus* SecY (panel C), but replaced by a Gln in the bacterial *T. thermophilus* and *T. maritima* translocons (panels B&D). S281 of *T. maritima* SecY is part of an array of Ser/Thr groups along TM7 [11]. In the closed state of the *T. maritima* SecY, the short sidechain of S281 and the presence of a Gln instead of a Glu at position 131 could make H bonding to T83/Q131 weaker than in the corresponding *M. jannaschii* cluster. That is, the extent to which the translocon opens and the kinetics of translocon opening may be different in archaeal vs. bacterial translocons.

Our proposal that the pattern of H bonding in the central TM2–TM3–TM7 predicts qualitatively conformational changes associated with translocon opening appears to be supported by inspection of crystal structures thought to represent snapshots of the translocon along its opening path (Fig. 6)—the bacterial *Thermotoga maritima* translocon in its open SecA-bound state [49], the Fab-bound pre-open *Thermus thermophilus* translocon [50], and the *Pyrococcus furiosus* translocon structure solved from a crystal in which the C-terminal  $\alpha$ -helical region of one SecY copy is bound to the cytoplasmic region of another SecY copy [51]. In these three structures binding of the ligand (SecA, Fab segment, or C-terminal region of another translocon) appears associated with changes in how the TM2, TM3, and TM7 helices interact with each other: TM7 is away from TM2 and TM3, but TM2 remains relatively close to TM3. The distance between the groups corresponding to the *M. jannaschii* T80:E122 H bond are 3.4 Å in *T. maritima* (T83:Q131), 4.7 Å in *T. thermophilus* (T82:Q126), and 3.3 Å in *P. furiosus* (T83:E125).

The dynamics of the inter-helical H bonds appear to be coupled to the overall conformational dynamics of the protein. Marginally stable H bonds (that is, bonds that rapidly break and reform) could contribute to the structural and dynamical stability of the protein in the absence of perturbations. Once the membrane protein is perturbed, however, these H bonds may be rapidly rearranged to help stabilize a new conformation. The perturbation could be binding of a substrate, mutation, or changes in the lipid membrane composition.

## Acknowledgements

This research was supported in part by grants GM-74637 from the National Institute of General Medical Sciences and GM-86685 from NIGMS and the National Institute of Neurological Disorders and Stroke (S.H.W.), a Marie Curie International Reintegration Award IRG-276920 (A.-N.B), and an allocation of computer time from the National Science Foundation through TeraGrid resources. We are indebted to our collaborators Prof. Douglas J. Tobias, Dr. Coral del Val, and Dr. Alfredo J. Freitas, for many valuable discussions on the protein translocon and the GlpG intramembrane protease, and Mr. Joseph Farran for excellent technical support.

## References

- [1] J.U. Bowie, Membrane protein folding: how important are hydrogen bonds? *Curr. Opin. Struct. Biol.* 21 (2011) 42–49.
- [2] S.H. White, How hydrogen bonds shape membrane protein structure, *Adv. Protein Chem.* 72 (2005) 157–172.
- [3] L. Adamian, J. Liang, Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers, *Proteins* 47 (2002) 209–218.
- [4] B.P. Pedersen, M.J. Buch-Pedersen, J.P. Morth, M.G. Palmgren, P. Nissen, Crystal structure of the plasma membrane proton pump, *Nature* 450 (2007) 1111–1114.
- [5] S.H. White, G. Von Heijne, How translocons select transmembrane helices, *Ann. Rev. Biophys. Chem.* 37 (2008) 23–42.
- [6] S. Jaud, M. Fernández-Vidal, I. Nillson, N.M. Meindl-Beinker, N.C. Hübner, D.J. Tobias, G. von Heijne, S.H. White, Insertion of short transmembrane helices by the Sec61 translocon, *Proc. Natl. Acad. Sci. USA* 106 (2009) 11588–11593.
- [7] K. Öjemalm, T. Higuchi, Y. Jiang, Ü. Langel, I. Nilsson, S.H. White, H. Suga, G. von Heijne, Apolar surface area determines the efficiency of translocon-mediated membrane-protein integration into the endoplasmic reticulum, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) E359–E364.
- [8] K. Plath, W. Mothes, B.M. Wilkinson, C.J. Stirling, T.A. Rapoport, Signal sequence recognition in posttranslational protein transport across the yeast ER membrane, *Cell* 94 (1998) 795–807.
- [9] B. Van den Berg, W.M. Clemons Jr., I. Collinson, Y. Modis, E. Hartmann, S.C. Harrison, T.A. Rapoport, X-ray structure of a protein-conducting channel, *Nature* 427 (2004) 36–44.
- [10] D.J.F. Du Plessis, G. Berrelkamp, N. Nouwen, A.J.M. Driessen, The lateral gate of SecYEG opens during protein translocation, *J. Biol. Chem.* 284 (2009) 15805–15814.
- [11] A.-N. Bondar, C. del Val, J.A. Freitas, D.J. Tobias, S.H. White, Dynamics of SecY translocons with translocation-defective mutations, *Structure* 18 (2010) 847–857.
- [12] M.S. Brown, J.L. Goldstein, The SREBP Pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor, *Cell* 89 (1997) 331–340.
- [13] G.A. McQuibban, S. Saurya, M. Freeman, Mitochondrial membrane remodelling regulated by a conserved rhomboid protease, *Nature* 423 (2003) 537–541.
- [14] G. Struhl, I. Greenwald, Presenilin is required for activity and nuclear access to Notch in *Drosophila*, *Nature* 398 (1999) 522–525.
- [15] S. Urban, J.R. Lee, M. Freeman, *Drosophila* Rhomboid-1 defines a family of putative intramembrane serine proteases, *Cell* 107 (2001) 173–182.
- [16] A. Weihofen, K. Binns, M.K. Lemberg, K. Ashman, B. Martoglio, Identification of signal peptide peptidase, a presenilin-type aspartic protease, *Science* 296 (2002) 2215–2218.
- [17] M.S. Wolfe, R. Kopan, Intramembrane proteolysis: theme and variations, *Science* 305 (2004) 1119–1123.
- [18] M. Lal, M. Caplan, Regulated intramembrane proteolysis: signaling pathways and biological functions, *Physiology* 26 (2011) 34–44.
- [19] A. Ben-Shem, D. Fass, E. Bibi, Structural basis for intramembrane proteolysis by rhomboid serine proteases, *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007) 462–466.
- [20] Y. Wang, Y. Zhang, Y. Ha, Crystal structure of a rhomboid family intramembrane protease, *Nature* 444 (2006) 179–183.
- [21] Z. Wu, N. Yan, L. Feng, A. Oberstein, H. Yan, R.P. Baker, L. Gu, P.D. Jeffrey, S. Urban, Y. Shi, Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry, *Nature Struct. Mol. Biol.* 13 (2006) 1084–1091.
- [22] A.-N. Bondar, C. del Val, S.H. White, Rhomboid protease dynamics and lipid interactions, *Structure* 17 (2009) 395–405.
- [23] R.P. Baker, K. Young, L. Feng, Y. Shi, S. Urban, Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate, *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007) 8257–8262.
- [24] B.K. Kobilka, X. Deupi, Conformational complexity of G-protein-coupled receptors, *Trends Pharmacol. Sci.* 28 (2007) 397–406.
- [25] H.-J. Kim, C. Altenbach, R.L. Thurmond, H.G. Khorana, W.L. Hubbell, Structure and function in rhodopsin: rhodopsin mutants with a neutral amino acid at E134 have a partially activated conformation in the dark state, *Proc. Natl. Acad. Sci. U.S.A.* 94 (1997) 14273–14278.
- [26] R. Vogel, M. Mahalingam, S. Lüdeke, T. Huber, F. Siebert, T.P. Sakmar, Functional role of the “ionic lock”—an interhelical hydrogen-bond network in family A heptahelical receptors, *J. Mol. Biol.* 380 (2008) 648–655.
- [27] A.B. Patel, E. Crocker, P.J. Reeves, E.V. Getmanova, M. Eilers, H.G. Khorana, S.O. Smith, Changes in interhelical hydrogen bonding upon rhodopsin activation, *J. Mol. Biol.* 347 (2005) 803–812.
- [28] A.R. Leach, *Molecular Modelling: Principles and Applications*, 2nd ed. Pearson Education Ltd., Harlow, 2001.
- [29] J.P. Ulmschneider, J.C. Smith, S.H. White, M.B. Ulmschneider, In silico partitioning and transmembrane insertion of hydrophobic peptides under equilibrium conditions, *J. Am. Chem. Soc.* 133 (2011) 15487–15495.
- [30] D.E. Shaw, M.M. Deneroff, R.O. Dror, J.S. Kuskin, R.H. Larson, J.K. Salmon, C. Young, B. Batson, J.K. Bowers, J.C. Chao, M.P. Eastwood, J. Gagliardo, J.P. Grossman, C.R. Ho, D.J. Lerardi, I. Kolossváry, J.L. Klepeis, T. Layman, C. McLeavey, M.A. Moraes, R. Mueller, E.C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S.C. Wang, Anton, A special-purpose machine for molecular dynamics simulation, *Proc. 34th Annu. Internat. Sym. Computer Architect*, 2007.
- [31] D.M. Rosenbaum, C. Zhang, J.A. Lyons, R. Holl, D. Aragao, D.H. Arlow, S.G.F. Rasmussen, H.-J. Choi, B.T. DeVree, R.K. Sunahara, P.S. Chae, S.H. Gellman, R.O. Dror, D.E. Shaw, W.I. Weis, M. Caffrey, P. Gmeiner, B.K. Kobilka, Structure and function of an irreversible agonist- $\beta_2$  adrenoceptor complex, *Nature* 469 (2011) 236–240.
- [32] R.O. Dror, D.H. Arlow, P. Maragakis, T.J. Mildorf, A.C. Pan, H. Xu, D.W. Borhani, D.E. Shaw, Activation mechanism of the  $\beta_2$ -adrenergic receptor, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2011) 18684–18689.
- [33] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers, Atomic-level characterization of the structural dynamics of proteins, *Science* 330 (2010) 341–346.
- [34] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold, *Science* 334 (2011) 517–520.
- [35] S. Urban, M.S. Wolfe, Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity, *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005) 1883–1888.
- [36] P. Hakizimana, M. Masureel, B. Gbaguidi, J.-M. Ruyschaert, C. Govaerts, Interactions between phosphatidylethanolamine headgroup and LmrP, a multidrug transporter: a conserved mechanism for protein gradient sensing? *J. Biol. Chem.* 283 (2008) 9369–9376.
- [37] M. Alami, K. Dalal, B. Leij-Garolla, S.G. Sligar, F. Duong, Nanodiscs unravel the interaction between the SecYEG channel and its cytosolic partner SecA, *EMBO J.* 26 (2007) 1995–2004.
- [38] M.F. Brown, Modulation of rhodopsin function by properties of the membrane bilayer, *Chem. Phys. Lipids* 73 (1994) 159–180.
- [39] L. Adamian, H. Naveed, J. Liang, Lipid-binding surfaces of membrane proteins: evidence from evolutionary and structural analysis, *Biochim. Biophys. Acta* 1808 (2011) 1092–1102.
- [40] H. Palsdottir, C. Hunte, Lipids in membrane protein structures, *Biochim. Biophys. Acta* 1666 (2004) 2–18.
- [41] A.G. Lee, How lipids affect the activities of integral membrane proteins, *Biochim. Biophys. Acta* 1666 (2004) 62–87.
- [42] M.Ø. Jensen, O.G. Mouritsen, Lipids do influence protein function: the hydrophobic matching hypothesis revisited, *Biochim. Biophys. Acta* 1666 (2004) 205–226.
- [43] O.S. Andersen, R.E. Koeppe, Bilayer thickness and membrane protein function: an energetic perspective, *Annu. Rev. Biophys. Biomol. Struct.* 36 (2007) 107–130.

- [44] Y. Wang, Y. Ha, Open-cap conformation of intramembrane protease GlpG, *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007) 2098–2102.
- [45] K.R. Vinothkumar, Structure of rhomboid protease in a lipid environment, *J. Mol. Biol.* 407 (2011) 232–247.
- [46] T. Junne, T. Schwede, V. Goder, M. Spiess, Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology, *J. Biol. Chem.* 282 (2007) 33201–33209.
- [47] J.A. Hanson, K. Duderstadt, L.P. Watkins, S. Bhattacharyya, J. Brokaw, J.-W. Chu, H. Yang, Illuminating the mechanistic roles of enzyme conformational dynamics, *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007) 18055–18060.
- [48] K.A. Henzler-Wildman, M. Lei, V. Thai, S.J. Kerns, M. Karplus, D. Kern, A hierarchy of timescales in protein dynamics is linked to enzyme catalysis, *Nature Letters* 450 (2007) 913–916.
- [49] J. Zimmer, Y. Nam, T.A. Rapoport, Structure of a complex of the ATPase SecA and the protein-translocation channel, *Nature* 455 (2008) 936–943.
- [50] T. Tsukazaki, H. Mori, S. Fukai, R. Ishitani, T. Mori, N. Dohmae, A. Perederina, Y. Sugita, D.G. Vassylyev, K. Ito, O. Nureki, Conformational transition of Sec machinery inferred from bacterial SecYE structures, *Nature* 455 (2008) 988–991.
- [51] P.F. Egea, R.M. Stroud, Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes, *Proc. Natl. Acad. Sci. U.S.A.* 107 (2010) 17182–17187.
- [52] W. Humphrey, W. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [53] T. Okada, M. Sugihara, A.-N. Bondar, M. Elstner, P. Entel, V. Buss, The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure, *J. Mol. Biol.* 342 (2004) 571–583.
- [54] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics, *J. Comput. Chem.* 4 (1983) 187–217.
- [55] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102 (1998) 3586–3616.
- [56] S.E. Feller, A.D. MacKerell Jr., An improved empirical potential energy function for molecular simulations of phospholipids, *J. Phys. Chem. B* 104 (2000) 7510–7515.
- [57] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926–935.
- [58] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, *J. Comput. Phys.* 23 (1977) 327–341.
- [59] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [60] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.* 103 (1995) 8577–8593.
- [61] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant-pressure molecular-dynamics algorithms, *J. Chem. Phys.* 101 (1994) 4177–4189.
- [62] S.E. Feller, Y. Zhang, R.W. Pastor, B.R. Brooks, Constant pressure molecular dynamics simulation: the Langevin piston method, *J. Chem. Phys.* 103 (1995) 4613–4621.
- [63] H. Grubmüller, H. Heller, A. Windemuth, K. Schulten, Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions, *Mol. Simul.* 6 (1991) 121–142.
- [64] M. Tuckerman, B.J. Berne, Reversible multiple time scale molecular dynamics, *J. Chem. Phys.* 97 (1992) 1990–2001.
- [65] K.L. Baran, M.S. Chimenti, J.L. Schlessman, C.A. Fitch, K.J. Herbst, B.E. Garcia-Moreno, Electrostatic effects in a network of polar and ionizable groups in Staphylococcal nuclease, *J. Mol. Biol.* 379 (2008) 1045–1062.