

Accepted Manuscript

Computational Studies of Membrane Proteins: Models and Predictions for Biological Understanding

Jie Liang, Hammad Naveed, David Jimenez-Morales, Larisa Adamian, Meishan Lin

PII: S0005-2736(11)00332-4
DOI: doi: [10.1016/j.bbamem.2011.09.026](https://doi.org/10.1016/j.bbamem.2011.09.026)
Reference: BBAMEM 80719

To appear in: *BBA - Biomembranes*

Received date: 27 June 2011
Revised date: 22 September 2011
Accepted date: 26 September 2011



Please cite this article as: Jie Liang, Hammad Naveed, David Jimenez-Morales, Larisa Adamian, Meishan Lin, Computational Studies of Membrane Proteins: Models and Predictions for Biological Understanding, *BBA - Biomembranes* (2011), doi: [10.1016/j.bbamem.2011.09.026](https://doi.org/10.1016/j.bbamem.2011.09.026)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 Computational Studies of Membrane Proteins: Models 2 and Predictions for Biological Understanding

3 Jie Liang, Hammad Naveed, David Jimenez-Morales, Larisa Adamian, and
4 Meishan Lin

5 *Dept of Bioengineering, 851 S. Morgan Street, Room 218, Chicago, IL 60607*

6 **Abstract**

We discuss recent progresses in computational studies of membrane proteins based on physical models with parameters derived from bioinformatics analysis. We describe computational identification of membrane proteins and prediction of their topology from sequence, discovery of sequence and spatial motifs, and implications of these discoveries. The detection of evolutionary signal for understanding the substitution pattern of residues in the TM segments and for sequence alignment are also discussed. We further discuss empirical potential functions for energetics of inserting residues in the TM domain, for interactions between TM helices or strands, and their applications in predicting lipid-facing surfaces of the TM domain. Recent progresses in structure predictions of membrane proteins are also reviewed, with further discussions on calculation of ensemble properties such as melting temperature based on simplified state space model. Additional topics include prediction of oligomerization state of membrane proteins, identification of the interfaces for protein-protein interactions, and design of membrane proteins.

7 *Keywords:* Membrane proteins, bioinformatics, empirical potential
8 function, motifs, ensemble properties, melting temperature, protein-protein

9 interaction, protein design

10 **1. Introduction**

11 Membrane proteins account for about 20% to 30% of all proteins encoded
 12 in a typical genome (1; 2). They play central roles in transport of nutrients
 13 and metabolites, and in signaling of regulatory networks (3; 4; 5). A ma-
 14 jor obstacle in studying membrane proteins is the difficulty in experimental
 15 determination of their three dimensional structures. Computational studies
 16 of membrane proteins can compliment experimental studies and have made
 17 significant strides. In this review, we discuss recent work based on analysis of
 18 sequences and structures of membrane proteins, as well as important under-
 19 standings gained from these studies on the physical processes of membrane
 20 protein assembly. An overview of the scope of studies surveyed in this review
 21 is shown in Fig 1, in the form of a diagram of the central dogma of molecular
 22 biology, in which different aspects where computational studies have made
 23 important contributions are depicted.

24 **2. Identification of Membrane Proteins and Prediction of Their** 25 **Topology**

26 *2.1. Predicting membrane proteins*

27 It was discovered very early on that the presence of stretches of hydropho-
 28 bic residues in a protein sequence is a good indicator that this sequence en-
 29 codes a membrane protein (6). Because most transmembrane helices are
 30 hydrophobic, they appear as periodic stretches of non-polar amino acids
 31 of length 17-25 in the primary sequence. These stretches of hydrophobic

residues cross the lipid membrane multiple times, and are connected by loops containing more polar residues. Such periodicity of hydrophobicity can be easily detected, and early methods for membrane protein prediction are based on calculation of a hydrophobicity index of residues within a window sliding along the protein sequence (6; 7).

A major source of misclassification with this approach is the existence of signal peptides important for targeting proteins for export. Signal peptide contains a hydrophobic region that can easily be mistaken for a transmembrane segment (8). Another source of difficulty is due to C-terminal peptides that are cleaved upon glycosylphosphatidylinositol (GPI)-anchoring, as these peptides are also often hydrophobic (9). An effective solution is to pre-process sequences by deleting signal peptides and cleaved peptides, both can be predicted accurately (10; 11; 12).

Predicting β -barrel membrane proteins is more challenging. Although residues facing the lipid membrane are predominantly hydrophobic, those facing the interior of the barrel can be quite polar (13). Unlike helical membrane proteins, there are no clear stretches of hydrophobic residues in their primary sequences.

2.2. Predicting Topology of Membrane Proteins

Many modern methods for identification of membrane proteins are based on techniques from machine learning and can also predict the topology of membrane proteins. The topology of a membrane protein refers to the number of transmembrane segments and the sidedness of the terminal ends of the protein, namely, whether the *N*- and *C*-end is on the non-translocated side or on the translocated side.

57 The topology of helical membrane protein can be predicted with high ac-
 58 curacy. Most prediction methods are based on processing multiple-sequence
 59 alignment data using machine-learning techniques such as neural networks (14;
 60 15), Hidden Markov models (16; 17; 18; 19; 20), and support vector ma-
 61 chines (21). The well-known “positive-inside” rule (22; 23; 24), namely, Arg
 62 and Lys residues are enriched in loops on the non-translocated side across
 63 the membrane compared to the translocated side, greatly aids in the devel-
 64 opment of these machine learning methods (22; 23; 24). For large scale pre-
 65 diction, recent experimentation using the consensus of many single-sequence
 66 based prediction methods also showed promise, which dispenses with time-
 67 consuming multiple-sequence alignments and are better suited for genome-
 68 scale predictions (25). For β -barrel membrane proteins, despite the lack of
 69 clear hydrophobic stretches of residues in the primary sequences, machine
 70 learning methods can now predict outer membrane proteins also very accu-
 71 rately (see (26) and (27)).

72 An approach alternative to machine learning is to predict membrane pro-
 73 teins and their topology based on physical considerations. This approach
 74 gives more mechanistic insight and is based on the fact that membrane pro-
 75 tein folding and sorting are driven by physical processes. Using the scale
 76 of measured free energy contributions of inserting individual amino acids at
 77 different positions of the TM helices into the endoplasmic reticulum mem-
 78 brane (28), a simple additive free-energy model was used to identify putative
 79 TM helices. Combined with the positive-inside rule, this approach can pre-
 80 dict the topology of α -helical membrane proteins accurately based on physical
 81 principles (29).

82 For β -barrel membrane protein, there are several characteristic observa-
83 tions that can help to determine their topology. First, the periplasmic loops
84 are always short compared to extracellular loops (13), although this may not
85 be true for mitochondrial and chloroplast outer membrane proteins. Second,
86 there is a significant, albeit less dramatic bias in the topological sidedness
87 of the distribution of charged residues. Different from the “positive-inside”
88 rule for helical membrane proteins, there exists an overall “positive-outside”
89 distribution. The extracellular cap region of the β -barrel membrane proteins
90 are disproportionately enriched with positively charged Arg and Lys, which
91 are disfavored in the periplasmic cap region (30). This is likely due to the
92 asymmetric distribution of the two leaflets of the lipid bilayer, in which neg-
93 atively charged lipopolysaccharides (LPS) is enriched in the outer-leaflet of
94 the outer membrane (31). For gram’s negative bacteria, this “positive-out”
95 rule for the outer membrane is consistent and complements the “positive-in”
96 rule for the inner membrane, as both rules implies that positively charged
97 residues are not favored in the periplasmic region.

98 Several computational methods based on machine learning techniques can
99 predict the topology of β -barrel membrane proteins well (32; 33; 34). Built
100 upon earlier results (35), a recent study based on measured physicochemical
101 properties of residues and empirical statistical potential is also shown to have
102 excellent performance in identifying β -barrel membrane proteins (36).

103 3. Motifs in Membrane Proteins

104 3.1. Sequence motifs

105 The GxxxG (or GG4) motif, in which two Gly are separated by three other
 106 residues, was the first sequence motif discovered (37). Originally observed in
 107 glycophorin A, this motif mediates close interaction of TM helices (38; 37). It
 108 is an example of the more general small-xxx-small motif forming helical dimer
 109 interface. Found in many biological systems, this class of motif provides a
 110 general framework for transmembrane helix association (39). Recent studies
 111 greatly broadened our view on the existence of different types of sequence
 112 motifs in membrane proteins, as well as their roles in providing structural
 113 stabilization and in regulating biological signaling (39; 40; 41; 42).

114 Computational discovery of sequence motifs of membrane proteins is a
 115 challenging task. Because the length of a transmembrane segment is short,
 116 there is strong coupling effects between the appearance of residues at one po-
 117 sition and its consequential absence in another position (37; 43; 30; 44). The
 118 discovery of the GxxxG motif is the outcome of an important development,
 119 namely, the formulation of a rigorous statistical treatment of what would be
 120 the expected frequency of various patterns of residues for a given transmem-
 121 brane helix (37). Prior to the study of Senes *et al*, widely used statistical
 122 models such as the Bernoulli/binomial model, the Markovian model, and the
 123 χ^2 model do not account for this finite-size effect (45; 44). This model, sub-
 124 sequently termed as the permutation model (44), enables detection of very
 125 subtle signals even when only limited data are available. Similar permutation
 126 model was also later applied in studying spatial motifs in β -barrel membrane
 127 protein (30). Senes *et al* also introduced a dynamic programming method

that makes it possible to compute efficiently the random distribution and p -values essential for identifying motifs using a database of membrane protein structures (37).

Subsequently, exact formulae for propensities of motifs with arbitrary number of residues under the permutation model were discovered, along with analytical formulae for p -value calculations for several types of sequence motifs (43; 44). An improved model, called *positional null model* that is based on exhaustive permutation but also account for bias of residue at certain positions was also developed (43; 44). Further studies showed that anti-motifs, which are sequence patterns that occur far less than would be expected, also reveal important biological information (30; 43; 44). Applications of these results have lead to the discovery of a large number of sequence motifs and antimotifs in β -barrel membrane proteins (43; 44). For example, the terminal motif YF2 was predicted to be important for recognition by periplasmic chaperon SurA for assisted folding (43), as mutations and deletion of the terminal F residue in PhoE from *E. coli* resulted in impairment of correct assembly of PhoE into the outer membrane (46). The MEMOTIF database contains many computationally derived sequence motifs for α -helical membrane proteins (47). A study of GPCRs using motifs of reduced alphabet of amino acids can be found in (48).

3.2. Spatial motifs

There are strong specific helix-helix and strand-strand interactions that can be detected through computational analysis. Interactions between TM helices and between strands as well as their overall assembly are the structural basis of sequence motifs. A global view on how TM helices interact spatially

153 was obtained in a comprehensive study of interacting helical pairs, in which
 154 pairs of helices were clustered by their shape similarity (49). It was found
 155 that just five clusters accounts for about 74% of all observed interacting
 156 helical pairs. These clusters can be rationalized in simple principles of helix-
 157 helix packing that goes back to Crick (50). The recurring geometric patterns
 158 of helix-helix interactions were organized into a library of spatial motifs of
 159 interacting helical pairs (49). The classification of spatial motifs and the
 160 library of interacting helical pairs lead to important understanding of the
 161 structural organization rule of helix assembly (40). This approach also proved
 162 to be invaluable in predicting membrane protein structures (40).

163 Somewhat similar approach was adopted by Martin *et al* for β -barrel
 164 membrane proteins. From the decomposition of known structures of β -barrel
 165 membrane proteins, a library of four residue fragment were constructed (51).
 166 It was found that there are strong preferences for different fragments to be
 167 located at different regions, and there are also specific preferences for inter-
 168 strand contacts between these fragments (51).

169 Another approach for discovery of spatial motifs of interacting residues
 170 is by comparing the frequency of observed appearance of certain spatial pat-
 171 terns of interacting residues with the frequency of what would be expected
 172 by random chance if there were no specific interhelical or interstrand inter-
 173 actions (52; 30; 44). The *serine-zipper* spatial motif (Fig 2a) was found in
 174 cytochrome c oxidase and in erythropoietin receptor (53; 54), where multi-
 175 ple repeated S-S interacting pairs form a large number of H-bonds (52). The
 176 placement of these small Ser ensure close packing between helices (55; 49).
 177 The *polar clamps* spatial motif (Fig 2b) involves three residues located on two

178 helices, where a residue capable of forming two or more H-bonds is clamped
179 by H-bonds formed with two residues (53). This motif is highly conserved
180 among G-protein coupled receptors, and likely contributed to stability and
181 specificity of the assembly of TM helices (53).

182 A systematic analysis of triplet interactions involving three-residues re-
183 vealed a number of additional spatial motifs, such as A-G-F and A-G-G (Fig
184 2c) (52). These well-defined spatial conformations exist on helices of unre-
185 lated proteins with similar parallel/antiparallel orientation and similar cross-
186 ing angles (52). Often, well-known sequence motifs such as GG4 and AG4
187 participate in these higher order motifs of interaction (52).

188 In β -barrel membrane proteins, Trp and Tyr residues are found to form a
189 frequently occurring motif through non-H-bonded interaction (Fig 2d). The
190 spatial motif *aromatic rescue* consists of interacting G-Y residues and G-F
191 residues across neighboring strands (30). The Tyr residue adopts an un-
192 usual rotamer and covers the backbone of Gly through H-bonding (Fig 2e).
193 This motif stabilizes the protein structure by mitigates the instability Gly
194 causes, as it prevents exposure of the backbone around Gly to solvent, at
195 the same time minimizing exposure of aromatic ring to the solvent (56; 30).
196 Experimental studies on similar motifs in soluble proteins showed that they
197 contribute significantly to protein stability and affect folding dynamics (57).
198 Other spatial motifs found in β -barrel membrane proteins are discussed
199 in (30).

200 4. Patterns of Evolution in Membrane Proteins, Contact Predic- 201 tion, and Functional Classification

202 Both sequence and spatial motifs are products of selection pressure on
203 membrane proteins throughout evolution, either for structural integrity or
204 for biological function. As evolution is a general driving force of biological
205 machineries, we discuss how patterns of evolution of membrane proteins can
206 be detected and how they can be used for biological predictions.

207 *Scoring matrices and patterns of residue substitutions.* An essential compu-
208 tational tool for membrane protein studies is sequence alignment (58; 59),
209 which is used in database searches for homologous proteins. A key compo-
210 nent of sequence alignment is the scoring matrix for quantification of sequence
211 similarity.

212 Standard scoring matrices such as BLOSUM and PAM used in default
213 NCBI sequence alignment were derived from soluble proteins (60; 61), and are
214 inappropriate for membrane protein studies. Overall, membrane proteins are
215 under unique physicochemical constraints, and experience selection pressure
216 very different from that of soluble proteins. The patterns of allowed and
217 forbidden substitutions at different positions of the transmembrane segments
218 are different from that of soluble proteins. Scoring matrices therefore need
219 to be specifically designed to capture the evolutionary pressure experienced
220 by the TM segments.

221 A number of specialized scoring matrices have been developed for helical
222 membrane proteins, including the SLIM and the PHAT matrices. Their ap-
223 plications result in significant improvement in identifying homologs of mem-
224 brane protein (62; 63). These scoring matrices, however, are inappropri-

ate for β -barrel membrane protein studies. As the lipid bilayer of bacterial outer membrane has different composition (*eg.*, the presence of lipopolysaccharides, LPSs), there are significant differences in the selection pressure experienced between helical and barrel membrane proteins. Results of a rigorous test showed that scoring matrices SLIM and PHAT designed for helical membrane proteins misidentified soluble proteins and random sequences as β -barrel membrane proteins (64).

Customized specific scoring matrices can be derived based on a general framework for analyzing amino acid residue substitutions (65). Using a continuous-time Markov process to model amino acid substitution and a Bayesian Monte Carlo estimation algorithm (65), the instantaneous substitution rates of residues in the TM-segments of β -barrel membrane proteins were estimated (64). Scoring matrices specific for different evolutionary time were then derived from the estimated rates (Fig 3), and were shown to have significantly improved sensitivity and specificity in detecting remote homologs of β -barrel membrane proteins (64). As the estimated substitution rates encode probability of exchanges between different residue pairs, they can also be used to suggest design of mutagenesis studies.

A remaining open question is whether evolutionary patterns are sufficiently similar between bacterial and mitochondrial β -barrel membrane proteins, and whether the same scoring matrices would capture their common evolutionary selection pressure. As machineries and mechanisms involved in the assembly of both bacterial and mitochondrial β -barrel membrane protein are quite similar (66; 67; 68), their substitution patterns in the TM strands may be very similar. Further computational study is required to resolve this

250 issue.

251 Another widely used empirical approach to extract evolutionary informa-
 252 tion from sequences is using the method of PSI-BLAST (69). Evolutionary
 253 information of the transmembrane segments are implicitly encoded in search
 254 results, and can be organized into a profile or a position specific weighted
 255 matrices (69). Such information can be effectively used to develop machine
 256 learning method, for example, as input data in the construction of a Hidden
 257 Markov Model or training a neural network for predicting the topology of
 258 transmembrane helices (70; 18).

259 *Lipid binding sites are evolutionarily conserved.* Phospholipid molecules are
 260 not only building blocks of membrane, they also play important roles in influ-
 261 encing the topology, folding, and assembly of membrane proteins, as well as
 262 in modulating their biological functions (71). By estimating the site-specific
 263 ratio of synonymous *vs.* non-synonymous substitution of the underlying DNA
 264 sequences, selection pressure experienced at individual amino acid positions
 265 can be measured (72). It was found that among lipid-facing residues, there
 266 are specific lipid binding sites that are evolutionarily conserved. These in-
 267 clude the cholesterol-binding sites in β_2 -adrenergic receptor and in Na-K
 268 ATPase, the cardiolipin binding site in formate dehydrogenase-N, and the
 269 PG binding site in the KcsA potassium channel (72).

270 *Discovery of packing interaction and predicting functional classes of mem-*
 271 *brane proteins from evolutionary analysis.* If a particular mutation affects the
 272 stability or the function of the protein, another mutation might occur at a
 273 different position to compensate the effects of the original mutation (73; 74).

274 This phenomenon of co-evolution of residues has been exploited for identi-
275 fication of packing interfaces between helices (75; 76) and for detection of
276 residues that mediates gating in voltage-dependent potassium channels (77).

277 Evolutionary information can also help to understand the function and
278 classes of poorly characterized membrane proteins, such as those obtained
279 from large scale genome and meta genome sequencing projects (78; 79). For
280 example, there are now a large number of new sequences homologous to ar-
281 chaeal retinal-containing rhodopsin-like proteins found in marine bacteria,
282 fungi, and unicellular algae (80). However, there is a lack of understand-
283 ing of basic aspects of the biology of these sequences. Structures of known
284 bacterial rhodopsins and evolutionary information contained in the homolo-
285 gous sequences helped to predict and delineate the functional relationship of
286 these rhodopsin-like proteins (81). Although retinal-binding rhodopsins fold
287 in similar structures, the residue make-up of the retinal-binding pockets may
288 be tuned to adapt to different biological functions. Using residue fragments
289 that form the retinal-binding pocket and amino acid substitution matrices
290 derived specifically for the retinal-binding pockets, a relationship tree was
291 obtained that groups rhodopsins by their biological function. This tree char-
292 acterizes well rhodopsins with known functions, and predicts the functions of
293 uncharacterized rhodopsin-like sequences (81). For example, *Gloeobacter vio-*
294 *laceus* rhodopsin was grouped into the same branch as the xantorhodopsin
295 from *Salinibacter ruber*, which uses carotenoids for light harvesting in the
296 blue-green region of the light spectrum (82). Subsequent experimental stud-
297 ies showed that *G. violaceus* rhodopsin indeed binds specifically a carotenoid
298 molecule, which functions as an antenna for light-harvesting (83).

299 5. Hydrophobicity scale from measurement and from calculation

300 The physical forces that hold membrane proteins together are of funda-
301 mental interests (84; 85; 86; 87; 88; 39; 89). Below we discuss experimentally
302 measured hydrophobicity scale and how they are useful for computational
303 studies. We also discuss equivalent scales derived from analysis of structures
304 and sequences of membrane proteins, as well as their applications.

305 5.1. Insertion free energy and hydrophobicity scale

306 Extensive studies have been carried out to measure the free energy of
307 inserting a residue into the lipid membrane. By measuring partitioning of a
308 model helix-forming peptide between water and a reference state, the free en-
309 ergy of helix insertion into a membrane environment is obtained (90; 91; 85).
310 As the environment of inserted helix is important, both membrane center
311 and interface were taken as the reference state in measurements (90; 91; 92).
312 Recently, the free energy contribution from individual amino acid for insert-
313 ing a TM helix into the biological endoplasmic reticulum (ER) membrane via
314 the Sec61 translocon were measured (93; 94; 28). The resulting insertion free
315 energy scale, called the biological hydrophobicity scale or translocon scale,
316 was the first free energy scale for insertion into a biological membrane. Very
317 recently, the first water-to-bilayer transfer free energy scale measured in the
318 context of a native membrane protein and lipid bilayer was reported (95).

319 These experimentally derived insertion free energy scales (or hydrophobic
320 scale) have been used effectively in computational studies of membrane pro-
321 teins. For example, both the Wimley-White whole residue octanol scale and
322 interface scale can be used to accurately predict TM helices in membrane

323 protein (96; 97). The biological hydrophobicity scale was also successfully
 324 used in predicting membrane protein topology, with the topology of 79% of
 325 a set of 123 membrane protein chains predicted correctly, which is better or
 326 comparable to hidden Markov model based methods (29).

327 Similar hydrophobic scales have also been developed computationally
 328 through statistical analysis of known structures of membrane proteins (7; 2;
 329 98; 99; 30; 100; 36). The main idea is to estimate the ratio of the frequency
 330 of observing an amino acid residue in the TM segment *vs* what would be
 331 expected by random chance (101; 102). Similar to experimentally measured
 332 scale, the empirical hydrophobic scale can also be made dependent on the
 333 local helical position of the residue (103; 100), as well as the random model
 334 for expected frequency, which is equivalent to the reference states in experi-
 335 mental studies (102; 43; 44). For example, both computed and measured free
 336 energy costs of embedding Asn and Gln strongly depends on their location in
 337 the TM helix (103). An empirically derived statistical potential function has
 338 been successfully applied in genome wide prediction of membrane proteins,
 339 with test results indicating an accuracy of 99% (35; 36). Such potential func-
 340 tion can also be used to estimate the tilt angle of a TM helix with respect
 341 to the bilayer normal, and to select amino acids in membrane protein design
 342 studies (100).

343 5.2. Predicting lipid facing regions of membrane proteins

344 Empirical hydrophobicity scale can be used to predict lipid-facing regions
 345 of membrane proteins. From structural analysis, it was found that there are
 346 strong preference for certain residues to face the headgroup and the hydro-
 347 carbon core regions of the lipid membrane (99). For example, Lys, Arg, Trp ,

Phe and Leu prefer to face the head-group region of the lipid bilayer instead of facing other helices, whereas Ile, Leu, Phe and Val prefer to face the hydrocarbon core region of the lipid bilayer. Small and polar residues are more likely to be buried inside the helical bundles and are lipophobic. In addition, it was found that Trp is frequently found in the hydrocarbon region, with its side-chain forming extensive interactions with residues on neighboring helices (99). This finding was consistent with subsequent experimental study in which it was found that Trp strongly supports self-assembly of TM helices, especially when placed on the *g*-position of the standard heptad. This position facilitates the side chain of Trp to interacting with neighboring helices (104). Overall, buried or interior-facing residues are significantly more polar and, hence, lipophobic, than the exterior residues (105). This lipophobic effect may play a general role in the folding and assembly of membrane proteins by encouraging the overall aggregation of TM helices, with the final structure determined through more specific interhelical H-bond, packing interactions, and loop constraints (105).

The lipid preferences of residues were quantified as a specialized empirical propensity scale called TMLIP (for TransMembrane helix-LIPid) potential (99). TMLIP was successfully used to predict the orientation of TM helices relative to the phospholipid bilayer (106). Based on a canonical model of the coiled-coil heptad repeat and the combination of the TMLIP propensity and evolutionary information, a computational method called LIPS (LIPid-facing Surface (105) can predict helical surface patches interacting lipid molecules at 88% accuracy. Other studies based on surface propensity scale and evolutionary information also reported excellent re-

sults (77; 107; 108; 109). A recent study that integrates evolutionary profiles and propensities for both membrane exposed residues and solvent exposed residues reported excellent performance (110).

The LIPS method is also useful in detecting inconsistencies in the structures of membrane proteins, such as the two structures of cytochrome b6f complex (105). It has also been used to aid in methods of template-free protein structure prediction (111), as well as in suggesting experimental studies (112). Further development based on TMLIP potentials allowed the development of the RANTS method (for RANking of Transmembrane helices by Solvent accessibility) (106). Predictions made by RANTS have been shown to be useful in designing experiments to identify interior facing residues and important polar interactions in the anion transporter SulP protein family (112).

6. Interactions between helices and between strands

6.1. Physical bases of interhelical and interstrand interactions

Physical forces beyond single body or insertion free energy are also at play in stabilizing membrane proteins. These include multibody interactions involving two or more helices or strands.

Polar interactions. Polar residues buried in the membrane environment likely contribute significantly for maintaining the stability of membrane proteins and their functions (85; 113; 114; 53; 103). Introduction of a single Asn, Asp, Glu or Gln in the TM segment can provide sufficiently strong driving force for helical self-association (113; 114). A survey of known membrane protein structures showed that polar and ionizable residues form extensive

396 H-bond connections between TM helices, as virtually all TM helices form
397 one or more interhelical H-bond (53).

398 Due to extreme experimental difficulty, quantitative assessment of the
399 magnitude of the H-bond energy in the transmembrane environment became
400 available only recently through elegant studies of double-mutant cycle anal-
401 ysis (115). The average energy of side-chain H-bond interactions is found to
402 be modest (-0.6 kcal/mol). It is possible that the unfolded state of mem-
403 brane proteins may already have H-bond largely satisfied through alternative
404 interactions (116), as the polarity of the interior of both membrane and sol-
405 uble proteins are quite similar (99). The apparent contribution of H-bond
406 for specific helical interactions in membrane protein seems to vary signifi-
407 cantly (115).

408 *Other interactions.* Other physical forces important for the assembly of TM
409 helices include side chains packing, overall helical packing with small residues
410 at helical-helical interfaces, aromatic interactions, and salt bridges (98; 117;
411 118; 119; 89; 120; 121). For β -barrel membrane proteins, the classical
412 model of β -strand interactions of β -sheets, in which backbone H-bond, side-
413 chain interactions, and weak H-bond stabilize neighboring strands, works
414 well (122; 123; 30) (Fig 4a-c). The energetic contributions of H-bond and
415 residues in the aromatic girdles of TM strands in the protein OmpA have
416 been measured (124; 125). Recently, it was found that specific interactions
417 between lipid and the TM strands of protein FhuA also provide significant
418 stability to the TM domain (72). Based on the TMSIP potential function
419 and the reduced state space, it was found that strands 7–9 form the most
420 unstable region in the protein FhuA, and strand 8, which runs through the

middle of the LPS-binding site, has the highest energy. These strands are stabilized by binding a lipid molecule.

Mechanical and thermal stability. Inter-helical and inter-strand interactions are also the major source of the mechanical stability of membrane protein, as shown by unfolding experiments of bacteriorhodopsin using atomic force microscopy and single-molecule force spectroscopy (126). In addition, these physical forces can be directly linked to protein stability as measured by the calculated melting temperatures of β -barrel membrane proteins (127) (Fig 4a-c and Fig 5).

6.2. Empirical potential function for biological understanding

A number of empirical potential functions have been developed for helix-helix interactions. In an early study, pairwise potential function based on statistics of known structures was successfully used to predict super-secondary structures of several packed small TM helices (128). Based on an atomic probabilistic model and packing contacts detected through Delaunay triangulation of membrane protein structures, a potential function for helix interaction called MHIP (for membrane helical interfacial pairwise propensity) was developed (98). By combining packing and helix contact analyses, Eilers *et al* developed an interfacial propensity scale for prediction of the relative orientation of TM helices (129). Dobbs *et al* developed a potential function for predicting inter-helical packing based on optimized discrimination of native helix-helix interactions from Monte Carlo generated decoy structures (130). Further development includes distance-based empirical potential that works well in predicting anchoring helix pairs (131). An interhelical contact poten-

445 tial was developed using a reduced alphabet of four amino acid types, which
 446 can discriminate native structures from many decoy conformations (132).
 447 Another empirical pairwise potential function for helical interaction was a
 448 major component of the force field used in the ROSETTA structure predic-
 449 tion method (111; 133; 134).

450 *General mechanisms to stabilize β barrel membrane proteins.* For β -barrel
 451 membrane proteins, an empirical potential function called the TMSIP (for
 452 TransMembrane Strand Interaction Propensity) potential has been devel-
 453 oped based on the canonical interaction model of β -sheet (122; 123; 30). Its
 454 null model is the rigorous permutation model discussed earlier. The TMSIP
 455 potential function can be used to identify weakly stable region in the TM
 456 domain (127). Analysis of these weakly stable regions revealed four general
 457 mechanisms that β -barrel membrane proteins use to stabilize the TM do-
 458 main (Fig 5): the well-known in-plug mechanism, as seen in FhuA (135), in
 459 which an inter-strand loop or a separate domain folds back and plug into
 460 the interior barrel to stabilize the TM barrel; the out-clamp mechanism as
 461 seen in PagP and hemolysin, in which a secondary structural element such
 462 as a helix outside the barrel stabilizes the TM barrel (136; 127), the newly
 463 discovered mechanism of specific lipid binding, in which the unstable region
 464 of the TM barrel is stabilized through specific strong binding with the LPS
 465 lipid molecule (72), as well as the mechanism of protein-protein interactions
 466 with which weakly stable regions are stabilized by another membrane pro-
 467 tein (127).

468 Empirical potential function can also help to gain biological understand-
 469 ing. There are many examples where excellent agreement were reported

470 between results obtained using experimentally derived free energy scale and
471 those obtained using empirical potential function. For example, the mea-
472 sured free energy scales of inserting amino acid residues embedded in a helix
473 into the endoplasmic reticulum agrees well with free-energy profiles derived
474 from statistical analysis of membrane protein structures (28).

475 *Understanding Arg in transmembrane segment from insertion energy and em-*
476 *pirical potential function* . Multiple Arg residues are found in the S4 trans-
477 membrane helix of KvAP ion channel and other related channel proteins,
478 and are likely to be important in sensing membrane depolarization and me-
479 diating channel gating (137; 138). Intuitively, these ionizable residues found
480 in the hydrophobic core of lipid membrane would be energetically costly, and
481 it is important to understand the physical basis of their locations. There is
482 significant discrepancy in free energy of inserting Arg into the hydrophobic
483 core when measured experimentally *vs* when calculated from molecular dy-
484 namics (MD) simulation (93; 139; 140; 141; 95). It was found that extra
485 helices facilitate the retainment of hydration water molecules, which reduces
486 solvation cost significantly (142). It was also suggested that part of the dis-
487 crepancy may be because MD simulation does not account for the tendency
488 of Arg side chain to snorkel towards the membrane-water interface (141). Al-
489 though simulations are carried out using physics based force field, the large
490 number of parameters involved and the difficulty in ensuring full sampling of
491 an equilibrium ensemble of conformations may be sources of non-negligible
492 errors (92; 141).

493 Hydrophobicity scale and empirical potential function can offer signifi-
494 cant insight. According to the analysis of Hristova and Wimley using the

experimentally derived Wimley-White scale (92), less than two Ala to Leu substitutions are required to compensate for one Ala to Arg substitution. It was found that it is easier to insert Arg in the interface region than the core of the bilayer (92).

The occurrence of Arg in hydrophobic core can also be understood through empirical potential function. Important favorable interactions across neighboring helices/strands often exist (98; 53; 103; 132), and such context dependent interactions will significantly modify the overall free energy of the protein. Since measured insertion free energy scales were mostly based on studies designed with single TM helices, the observed occurrence of Arg in the hydrophobic core of natural membrane proteins can be better interpreted with additional consideration incorporating inter-helical and inter-strand interactions.

This can be illustrated by analyzing the energetic consequence of embedding an Arg residue in the TM segment of a β -barrel membrane protein using the empirical potential function TMSIP (Fig 4) (30). Arg in β -barrel membrane proteins facing inside the β -barrel pore is energetically favorable, but very unfavorable when facing the lipid membrane (30). Through interstrand interactions, there are three additional types of interactions that are major contributors to the stability of TM β -strands, namely, strong H-bond between main chain (C-O $\cdots$ H-N), side-chain interaction (R $\cdots$ R) including side chain H-bond, and weak H-bonds between C-O $\cdots$ H-C $_{\alpha}$ (122; 123; 30). According to the recently updated version of the TMSIP scale incorporating additional structural data, Arg can be stabilized by main chain H-bond interactions with Ala, Trp, Val, and Thr, if they are located on appropriate positions of

the neighboring strands (30). Since side chain H-bonds are known to contribute only modestly to the overall stability of membrane proteins (143), the context dependent main chain H-bond interactions are likely the main contributors that modifies the single-body energetics of Arg insertion.

According to TMSIP, Arg is only slightly energetically unfavorable in the extracellular interfacial region, but is highly unfavorable in the hydrophobic core region and the periplasmic interfacial region (30). As Arg residue is inserted from the periplasmic side into the lipid bilayer, favorable main-chain H-bond interactions with Ala, Trp, Val, and Thr located on neighboring strands may compensate for the unfavorable insertion of Arg (Fig 4). This compensation effect would facilitate the translocation of Arg towards the more favorable extracellular interface in β -barrel membrane protein.

Experimentally measured insertion free energy derived from studies of single helix experiments can be regarded as one-body energetics, and the equivalent empirical potential function are hydrophobic scale involving only a single residue and its depth in the membrane environment. An accurate account of the full energetics of residues in the context of a wild type membrane protein needs also incorporate the effects of inter-helical or inter-strand interactions, namely, the two-body interactions. It is possible that higher order cooperative effects may also be relevant (52; 39).

In a recent study, the free energy changes in a wild type membrane protein were measured when an Ala was replaced with each of the 20 amino acids (95). This is the first time such measurements were made in wild type membrane protein placed in a lipid bilayer. Although Trp fluorescence was employed in experimental measurement, and Glu and Asp are mostly likely

in the protonated state at the experimental condition, free energy changes of replacing Ala with the other residue types provide a wealth of quantitative information about membrane protein stability. It was found that Arg substitution incurs only a modest free energy cost (95). Although the analysis of this study was based on a simple one-body additive insertion energy model, interstrand interactions that is context dependent at the host position is likely to be non-negligible in wild type membrane proteins. The wealth of information provided in studies such as (95) can be used for alternative analyses using a statistical mechanical model (127) that considers context dependent interstrand interactions as well as non-native conformations, which works well to account for observed nonlinear and non-additive effects.

7. Predicting Structure of Membrane Protein

7.1. Irregular structures and their prediction

An idealized model of helical membrane proteins is that of an assembly of highly hydrophobic helices connected by loops, with orientations perpendicular to the membrane plane. This is the model upon which many successful hidden Markov model (HMM) methods for topology prediction were based (144). However, recent structures showed that there are many irregular structures. Transmembrane helices are often kinked at varying length and tilt angle (145; 146). In the water-membrane interfacial regions, there may exist amphipathic α -helices parallel to the membrane plane (147; 148). In addition, there exists re-entrant regions that enter and leave the membrane from the same side of the transmembrane region (149).

About 44% of TM helices have kinks, with 35% of which associated with

Pro residue, and others with Ser and Gly at the center of the kink (150; 151). Kinks are likely to be important for membrane protein function, as they provide locations for movement such as hinge bending, and introduces structural diversity even among members of the same protein family. It was suggested that Pro in ancestral proteins may have initiated such kinks (152). TM helices subsequently were stabilized through evolution to an extent that the maintenance of the kinked conformation no longer required the presence of Pro residues (152). Molecular dynamics simulation of single TM helix has been successful in identifying many kinks (151). In a study of 405 TM helices, it was found that 79% of the proline kinks, 59% of the vestigial proline kinks, and 18% of the non-proline helical kinks can be reproduced from 1 ns of MD simulation (151).

A study of the re-entrant regions using the technique of principal component analysis for dimension reduction revealed that these regions have distinct amino acid composition (149). As many re-entrant regions are found in transporters, Gly and Ala are abundantly found in this region (149). In addition, Ser and Thr are also enriched (153). Hidden Markov models developed based on these patterns can now predict the re-entrant regions successfully at 70-75% accuracy (149; 153).

7.2. Comparative three-dimensional model of membrane protein structure

If the structure of a homologous membrane protein exists, comparative or homology structural model can be built based on the template structure (154; 155). This technique has been applied fruitfully to study the G-protein coupled receptors (GPCRs), an important receptor for cellular signal transduction (155; 156). When a template structure is identified and a

quality alignment is obtained, a specialized comparative modeling method
 MEDELLER can identify a reliable core structure, and build a structural
 model by extending the core to other TM region and to the loop region (157).
 This approach showed higher accuracy in modeled structure than generic ho-
 mology modeling methods. For β -barrel membrane proteins, the TMBPRO
 method takes predicted secondary structures and evaluate their overall en-
 ergy to each structural template containing the same number of strands (158).
 Combined with conformational search via simulated annealing for the lowest
 energy alignment of the sequence to the structural template, the confor-
 mation with the lowest overall energy can be taken as the predicted struc-
 ture (158). It is expected that improvement in alignment and detection of
 remote homologs can be obtained through usage of customized scoring matri-
 ces (64; 159). This will allow further leverage of current knowledge of existing
 membrane protein structures, at a rate of about 130 proteins per template
 structure (159). Furthermore, these scoring matrices are found to be useful
 for identifying mitochondria outer membrane proteins in eukaryotes (159).

7.3. Template-free prediction of membrane protein structure

A more challenging task in structure prediction is when there is no known
 structures that can serve as the template structure. That is, none of the
 homologous proteins have known structures. The ROSETTA *de novo* protein
 structure prediction method has been extended to predict structures of helical
 membrane proteins, without the need of a template structure (111; 133; 134),
 although no template-free methods currently exist that can predict structures
 of β -barrel membrane proteins.

Using an empirical potential function that combines van der Waals in-

619 teraction, backbone torsional force, electrostatic interaction, and orientation
 620 dependent H-bond interaction, Barth *et al* developed a method based on
 621 ROSETTA Monte Carlo sampling that can successfully recover the side chain
 622 conformations of membrane proteins, can model distorted TM helices, and
 623 can predict the conformation of glycophorin A interface (133). Further pre-
 624 diction of likely interacting helical pairs with a large sequence separation
 625 was obtained from a carefully constructed library of interacting helical pairs
 626 and the evolutionary profiles of the two helices. With such predicted inter-
 627 helical geometry and co-factor coordination when available to restrict the
 628 conformational space, Barth *et al* successfully predicted three dimensional
 629 structures of a divers set of membrane proteins with different size, topolo-
 630 gies, and biological functions, with excellent results at the level of about 4 Å
 631 in RMSD (134).

632 7.4. *Structure prediction through combined experimental and computational* 633 *studies*

634 Partial experimental information that is insufficient on its own right for
 635 structure determination can be very effective in guiding computational pre-
 636 diction methods towards a much smaller feasible space for conformational
 637 search. An important form of experimental data is coarse grained density
 638 map of cryo-electron microscopy at medium-resolution (7-10 Å), in which
 639 helices are better resolved as rods than strands and loops. By placing pre-
 640 dicted helices into the density rods for helices and adding modeled loops,
 641 the overall structures of helical membrane proteins can be predicted in some
 642 cases with much improved resolution, although this method hinges upon the
 643 correct prediction of helices (160). Combining CryoEM data with evolution-

644 ary information, the C_{α} -trace model of the transmembrane domain of human
645 copper transporter 1 was also successfully constructed (119; 161).

646 Another approach is to integrate experimental mutagenesis data into the
647 structure prediction protocol by biasing the selection of the final model
648 towards those that are consistent with the experimental mutagenesis re-
649 sults. This approach has been applied successfully to predict the struc-
650 ture of the transmembrane domain of the homodimeric BNIP3 (162) and
651 the heterodimeric structure of complete α_{IIb} and β_3 complex (163). How-
652 ever, significant amount of experimental data are required, and therefore
653 this approach is best-suited for well studied membrane proteins. A general
654 theoretical framework to generate protein structures that satisfy different
655 experimentally derived restraints described in (164) may be useful for such
656 tasks.

657 8. Beyond Structure Prediction: Ensemble Properties, Protein- 658 Protein Interactions, and Protein Design

659 Great progresses have been made in predicting structures of membrane
660 proteins. However, many important problems in membrane protein studies
661 require information beyond that of a single native structure. Below we first
662 discuss studies on the ensemble nature of conformations of membrane pro-
663 teins, which is the basis of their thermodynamics properties. We also discuss
664 prediction of oligomerization state and protein-protein interactions. In addi-
665 tion, we discuss future development in protein design, in which computational
666 studies will likely make significant contributions.

667 *8.1. Ensemble Nature of Membrane Protein Structure and Their Thermody-*
 668 *namic Properties*

669 There are many important questions beyond the knowledge of a single
 670 predicted structure. For example, do membrane proteins exist in multiple
 671 conformations (Fig 4)? What are their associated probabilities? How ther-
 672 modynamic properties can be calculated from ensemble properties of confor-
 673 mations? How do dynamic transitions occur among these conformations and
 674 how such changes may contribute to the observed biological functions?

675 In the study of β -barrel membrane proteins, progresses have been made
 676 in addressing some of these questions (127). Because of the relatively regular
 677 pattern in strand interactions, the conformational space of TM strands can
 678 be effectively modeled using a simplified state-space model (30). By assuming
 679 a reduced conformational space in which each strand can slide up or down
 680 for a total of 7 positions, one can enumerate all possible conformations and
 681 calculate the energy value for each conformation. Thermodynamic properties
 682 of the transmembrane domain can then be computed (127). Fig 5 depicts
 683 one such thermodynamic property, namely, the relative melting temperature
 684 calculated for the TM domains of a number of β -barre membrane proteins.

685 *Role of nonnative and alternative conformations.* It is important to consider
 686 non-native conformations in computing thermodynamic properties of mem-
 687 brane proteins. Although it was not immediately obvious why TM strands
 688 would not always adopt the ground state conformation, as it would be very
 689 costly to break all the H-bonds to move up or down to a different register,
 690 experimental results on PagP showed that there can be significant conforma-
 691 tional change when different detergent is present (165). In fact, alternative

conformations with low energy may serve as obligate on-pathway transient states (166).

Recent studies in helical membrane protein demonstrated the flexible nature of transmembrane helices, which contain many kinks, bulges, and re-entrant loops (149; 167; 150). Furthermore, the spatial close proximity among newly synthesized TM helices during co-translational insertion to membrane suggests that there may exist interhelical interactions even in the early stage of membrane protein folding (168). For example, several experimentally determined TM helices in Glt_{ph} glutamate transporter were found not to have lowest free energy of insertion in wild type protein, and the segment with both measured and predicted lowest free energy has significant position displacement compared to the wild type protein (168). These findings suggest that TM helices may shift positions dramatically during the folding and oligomerization process, which may be important for bringing functionally important polar residues into places.

Overall, the population of alternative conformational states may play important roles in determining the final native structure and function of membrane protein, and in ensuring the overall stability and robustness of the cell machineries in which membrane proteins are important components.

8.2. Protein-Protein Interactions

A genomic scale survey of domain combinations of helical membrane proteins suggested that membrane proteins exist mostly as single domains, and oligomerization within the membrane may be the general mechanism for membrane proteins to gain new biological functions (169; 170). For GPCRs, characterizing their oligomerization state is of considerable importance (171).

717 Computational docking and molecular dynamics simulations have been ap-
718 plied in gaining insight into the oligomerization state and in delineating the
719 protein-protein interface (see ref (171) for a recent review).

720 The oligomerization state of β -barrel membrane proteins can be accu-
721 rately predicted computationally (127). Based on the TMSIP empirical po-
722 tential function and the reduced conformational state model, it was found
723 that the average deviation in energy of the unstable strands from the mean of
724 all strands serves as an excellent predictor of the overall oligomerization state
725 of the membrane protein. In a leave-one out blind test of 25 non-homologous
726 β -membrane proteins, in which each of the protein is taken in turn for testing,
727 while the remaining 24 proteins used for model construction, excellent results
728 are obtained in predicting the oligomeric state. As subsequently realized that
729 protein FhuA can exist in dimeric form, the predictions of the oligomeriza-
730 tion state for these 25 β -barrel membrane proteins are 100% accurate with
731 100% specificity (172; 127). These predictions are robust, as the outcome
732 does not depend on specific choice of structures used in the construction of
733 the energy function. Furthermore, as structural information is not essential
734 for such predictions, the oligomerization state can also be predicted quite
735 successful even when only sequence information is employed (127): The ac-
736 curacy and specificity are 96% and 94% when only sequence information is
737 used (127), respectively, with the consideration that protein FhuA indeed
738 form a dimer (172).

739 The interface of protein-protein interaction for β -barrel membrane pro-
740 teins can also be predicted (127). Based on the observation that the protein-
741 protein interface is enriched with weakly stable strands, interfaces can be

742 predicted either with the knowledge of the structure where high accuracy
743 can be achieved, or with sequence information only where accuracy is slightly
744 degraded (Fig 6). Another approach based on the machine learning method
745 of random forest can also predict residues located in the protein-protein in-
746 terface accurately (173).

747 Success in predicting the oligomerization state and in identifying protein-
748 protein interaction interface in the TM domain will likely reveal novel in-
749 sight into the mechanism of many membrane proteins. For β -barrel mem-
750 brane proteins, mutations can be suggested that would strongly affect the
751 oligomerization state (Fig 6, inset). It is conceivable that protein-protein in-
752 terface for eukaryotic membrane proteins can also be predicted, and mutants
753 with different oligomerization behavior can be engineered. For example, the
754 eukaryotic proteins VDAC found in mitochondria oligomerizes during the
755 induction of apoptosis (174). Predicted oligomerization site on VDAC can
756 aid in experimental design of studies to identify key residues involved in
757 VDAC oligomerization. Such investigations will be important for studying
758 the underlying mechanism of apoptosis (174; 175).

759 8.3. Design and engineering of membrane proteins

760 *De novo design of membrane proteins and inhibitors.* *De novo* protein de-
761 sign and protein engineering aim to produce proteins with new or enhanced
762 activity and stability. Although significant progress has been made in recent
763 years (176), there are only a limited number of reported successes in *de novo*
764 membrane protein design. The most promising approach is to extend compu-
765 tational methods used for the design of globular proteins. This approach lead
766 to the successful design of a four helix bundle membrane protein engineered

767 to bind two Fe(II/III) diphenylporphyrins in a bis-His geometry. This de-
 768 signed membrane protein forms a channel capable of transmembrane electron
 769 transfer (177). There has also been significant progress in the design of small
 770 peptides that target the transmembrane proteins and inhibit protein-protein
 771 interactions in the TM domain (178). Anti- α_{IIb} peptide that targets the
 772 transmembrane domain of the α subunit of the integrin $\alpha_{IIb}\beta_3$ disrupts the
 773 heteromeric helix-helix interactions. The specificity of the designed anti- α_{IIb}
 774 was validated both *in vitro* and *in vivo* (178; 179).

775 *Engineering stability and oligomerization state of membrane proteins.* As
 776 more structures of membrane proteins become available, improved under-
 777 standing of their organizational principles has led to efforts in engineering
 778 membrane proteins with improved protein stability (180). For example, a
 779 metal binding site was engineered in the mastoparan X protein, an am-
 780 phiphilic α -helix that is too short to form a stable helix in water. This
 781 newly acquired metal binding ability stabilizes the helical structure of the
 782 protein, and increased the binding and lysis ability of the protein to the
 783 membrane (181). Longer transmembrane regions were also engineered for
 784 the β -barrel membrane protein FhuA to match the hydrophobic cores of
 785 thick polymeric membranes, with the goal for targeted drug delivery (182).

786 There have also been successes in engineering stability of oligomerized
 787 membrane proteins. Using a statistical potential function, mutations that
 788 would stabilize or destabilize the dimeric interface of GPCRs were predicted
 789 based on a *de novo* designed rhodopsin homodimer model (183). These pre-
 790 dictions compared favorably with experimental studies (183). Computational
 791 study on β -barrel membrane protein has also suggested that oligomers form

792 primarily due to the instability of monomers. Such oligomerization can be
793 altered by mutations that stabilize or destabilize the monomeric form of the
794 β -barrel membrane protein (127).

795 *Geometry and selectivity.* Success has also been reported on engineering the
796 geometry of β -barrel membrane protein. Most β -barrel membrane proteins
797 consist of an even number of strands, and β -hairpins are often thought as
798 the basic repeating unit (184; 185). It is plausible that the evolution of β -
799 barrel membrane proteins are based on the modularity of hairpin duplication
800 and oligomeric assembly of these hairpins (185). Indeed, bacterial toxin α -
801 hemolysin and the multidrug efflux system TolC forms β -barrel membrane
802 protein upon oligomerization once multiple hairpins are inserted into the lipid
803 membrane (186; 187). Arnold *et al* constructed an artificial β -barrel mem-
804 brane protein by duplicating the sequence of 8-strand OmpX. The resulting
805 protein has a pore size of that of a 16-strand porin based on single-channel
806 conductance measurements (185).

807 Pores with specially constructed filters have been successfully engineered
808 to control the flow of ions and metabolites through the membrane bilayer.
809 β -barrel membrane protein OmpF, which is slightly cation-selective due to
810 the -1 net charge in the filter region, has been converted into Ca^{2+} -selective
811 channel by carefully mutating two Args located in the constriction zone to
812 Glu (188). Similarly, aquaporin-1 filter was engineered to enhance proton
813 conductance computationally and the results were subsequently confirmed by
814 experiments (189). Dynamics of reconstituted native plugged FhuA channels
815 in an ion-conducting state have been studied by adding 4M urea on the cis
816 side, which reversibly unfolds the plug domain and open an ion-conducting

817 pathway that mimics the TonB dependent channel (190). Mutants of OmpF
818 whose extracellular loops were deleted one at a time were also engineered to
819 be pH insensitive (191).

820 It is likely that the pace of designing membrane protein will accelerate,
821 and many more novel membrane proteins with desirable biophysical proper-
822 ties and novel or enhanced functions will be made.

823 9. Conclusion

824 We have summarized key aspects of computational studies of membrane
825 proteins, including bioinformatics prediction of membrane proteins and their
826 topology, the discovery and implication of sequence and spatial motifs, mem-
827 brane protein evolution and the substitution patterns of amino acids in the
828 TM domain, as well as the modeling of the underlying physical forces through
829 empirical potential function. We have also discussed recent successes in
830 structure prediction and in protein-protein interactions prediction, as well
831 as progress in characterization of ensemble properties of membrane proteins.
832 We believe that computational studies based on both the underlying physical
833 forces as well as bioinformatics analysis of evolutionary signal will continue to
834 make important contributions in understanding and manipulating membrane
835 proteins that compliments experimental investigations.

836 10. Acknowledgment

837 We thank Drs. Patrick Barth, Ann Delcour, Robert Eisenberg, Alan
838 Finkelstein, Dennis Gessman, Nathan Joh, Linda Kenney, Renhao Li, Stephan
839 Nussberger, Varda Shoshan-Barmatz, Alessandro Senes, Ann Stanley, and

840 Bill Wimley for very helpful discussions. This work was supported by NIH
841 grants GM079804 and GM086145, NSF grants DBI 1062328 and DMS-0800257,
842 and ONR grant N00014-09-1-0028.

843 References

- 844 [1] E. Wallin, G. von Heijne, Genome-wide analysis of integral membrane
845 proteins from eubacterial, archaean, and eukaryotic organisms, *Protein*
846 *Science: A Publication of the Protein Society* 7 (4) (1998) 1029–1038,
847 PMID: 9568909. doi:10.1002/pro.5560070420.
- 848 [2] I. Arkin, A. Brunger, Statistical analysis of predicted transmembrane
849 alpha-helices., *Biochim Biophys Acta* 1429 (1) (1998) 113–128.
- 850 [3] L. DeFelice, Transporter structure and mechanism., *Trends Neurosci*
851 27 (6) (2004) 352–359. doi:10.1016/j.tins.2004.04.007.
- 852 [4] W. Cho, R. Stahelin, Membrane-protein interactions in cell signaling
853 and membrane trafficking., *Annu Rev Biophys Biomol Struct* 34 (2005)
854 119–151. doi:10.1146/annurev.biophys.33.110502.133337.
- 855 [5] G. von Heijne, The membrane protein universe: what's out there and
856 why bother?, *J Intern Med* 261 (6) (2007) 543–557. doi:10.1111/j.1365-
857 2796.2007.01792.x.
- 858 [6] J. Kyte, R. F. Doolittle, A simple method for displaying the hydro-
859 pathic character of a protein, *Journal of Molecular Biology* 157 (1)
860 (1982) 105–132, PMID: 7108955.
861 URL <http://www.ncbi.nlm.nih.gov/pubmed/7108955>

- 862 [7] G. von Heijne, Membrane protein structure prediction, hydrophobicity
863 analysis and the positive-inside rule, *J. Mol. Biol.* 225 (1992) 487–494.
- 864 [8] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Identification
865 of prokaryotic and eukaryotic signal peptides and prediction of their
866 cleavage sites, *Protein Engineering* 10 (1) (1997) 1–6, PMID: 9051728.
867 URL <http://www.ncbi.nlm.nih.gov/pubmed/9051728>
- 868 [9] P. Orlan, A. K. Menon, Thematic review series: lipid posttransla-
869 tional modifications. GPI anchoring of protein in yeast and mammalian
870 cells, or: how we learned to stop worrying and love glycopospholipids,
871 *Journal of Lipid Research* 48 (5) (2007) 993–1011, PMID: 17361015.
872 doi:10.1194/jlr.R700002-JLR200.
873 URL <http://www.ncbi.nlm.nih.gov/pubmed/17361015>
- 874 [10] J. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction
875 of signal peptides: SignalP 3.0., *J Mol Biol* 340 (4) (2004) 783–795.
876 doi:10.1016/j.jmb.2004.05.028.
- 877 [11] G. Poisson, C. Chauve, X. Chen, A. Bergeron, FragAnchor: a large-
878 scale predictor of glycosylphosphatidylinositol anchors in eukaryote
879 protein sequences by qualitative scoring, *Genomics, Proteomics Bioin-*
880 *formatics / Beijing Genomics Institute* 5 (2) (2007) 121–130, PMID:
881 17893077. doi:10.1016/S1672-0229(07)60022-9.
- 882 [12] A. Pierleoni, P. L. Martelli, R. Casadio, PredGPI: a GPI-anchor
883 predictor, *BMC Bioinformatics* 9 (2008) 392, PMID: 18811934.
884 doi:10.1186/1471-2105-9-392.

- 885 [13] G. E. Schulz, beta-Barrel membrane proteins, *Current Opinion in*
886 *Structural Biology* 10 (4) (2000) 443–447, PMID: 10981633.
887 URL <http://www.ncbi.nlm.nih.gov.proxy.cc.uic.edu/pubmed/10981633>
- 888 [14] B. Rost, P. Fariselli, R. Casadio, Topology prediction for helical trans-
889 membrane proteins at 86% accuracy., *Protein Sci* 5 (8) (1996) 1704–
890 1718. doi:10.1002/pro.5560050824.
- 891 [15] P. L. Martelli, P. Fariselli, R. Casadio, An ENSEMBLE machine
892 learning approach for the prediction of all-alpha membrane proteins,
893 *Bioinformatics (Oxford, England)* 19 Suppl 1 (2003) i205–211, PMID:
894 12855459.
895 URL <http://www.ncbi.nlm.nih.gov/pubmed/12855459>
- 896 [16] G. E. Tusnady, I. Simon, The HMMTOP transmembrane topology
897 prediction server, *Bioinformatics (Oxford, England)* 17 (9) (2001) 849–
898 850, PMID: 11590105.
899 URL <http://www.ncbi.nlm.nih.gov/pubmed/11590105>
- 900 [17] A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting
901 transmembrane protein topology with a hidden markov model: ap-
902 plication to complete genomes, *Journal of Molecular Biology* 305 (3)
903 (2001) 567–580, PMID: 11152613. doi:10.1006/jmbi.2000.4315.
904 URL <http://www.ncbi.nlm.nih.gov/pubmed/11152613>
- 905 [18] H. Viklund, A. Elofsson, Best alpha-helical transmembrane pro-
906 tein topology predictions are achieved using hidden markov mod-
907 els and evolutionary information, *Protein Science: A Publication*

- 908 of the Protein Society 13 (7) (2004) 1908–1917, PMID: 15215532.
909 doi:10.1110/ps.04625404.
910 URL <http://www.ncbi.nlm.nih.gov/pubmed/15215532>
- 911 [19] L. Käll, A. Krogh, E. L. L. Sonnhammer, A combined trans-
912 membrane topology and signal peptide prediction method, Journal
913 of Molecular Biology 338 (5) (2004) 1027–1036, PMID: 15111065.
914 doi:10.1016/j.jmb.2004.03.016.
915 URL <http://www.ncbi.nlm.nih.gov/pubmed/15111065>
- 916 [20] L. Käll, A. Krogh, E. L. L. Sonnhammer, An HMM posterior decoder
917 for sequence feature prediction that includes homology information,
918 Bioinformatics (Oxford, England) 21 Suppl 1 (2005) i251–257, PMID:
919 15961464. doi:10.1093/bioinformatics/bti1014.
920 URL <http://www.ncbi.nlm.nih.gov/pubmed/15961464>
- 921 [21] T. Nugent, D. T. Jones, Transmembrane protein topology prediction
922 using support vector machines, BMC Bioinformatics 10 (2009) 159,
923 PMID: 19470175. doi:10.1186/1471-2105-10-159.
924 URL <http://www.ncbi.nlm.nih.gov/pubmed/19470175>
- 925 [22] G. von Heijne, The distribution of positively charged residues in bac-
926 terial inner membrane proteins correlates with the trans-membrane
927 topology, EMBO J. 5 (1986) 3021–3027.
- 928 [23] G. von Heijne, Y. Gavel, Topogenic signals in integral membrane pro-
929 teins., Eur J Biochem 174 (4) (1988) 671–678.

- 930 [24] Y. Gavel, J. Steppuhn, R. Herrmann, G. von Heijne, The 'positive-
931 inside rule' applies to thylakoid membrane proteins., *FEBS Lett* 282 (1)
932 (1991) 41–46.
- 933 [25] A. Hennerdal, A. Elofsson, Rapid membrane protein topology pre-
934 diction, *Bioinformatics (Oxford, England)* 27 (9) (2011) 1322–1323,
935 PMID: 21493661. doi:10.1093/bioinformatics/btr119.
936 URL <http://www.ncbi.nlm.nih.gov/pubmed/21493661>
- 937 [26] M. M. Gromiha, M. Suwa, Discrimination of outer membrane proteins
938 using machine learning algorithms, *Proteins* 63 (4) (2006) 1031–1037,
939 PMID: 16493651. doi:10.1002/prot.20929.
940 URL <http://www.ncbi.nlm.nih.gov/pubmed/16493651>
- 941 [27] J. Hu, C. Yan, A method for discovering transmembrane beta-
942 barrel proteins in gram-negative bacterial proteomes, *Computational*
943 *Biology and Chemistry* 32 (4) (2008) 298–301, PMID: 18467177.
944 doi:10.1016/j.compbiolchem.2008.03.010.
945 URL <http://www.ncbi.nlm.nih.gov/pubmed/18467177>
- 946 [28] T. Hessa, N. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-
947 Bader, I. Nilsson, S. White, G. von Heijne, Molecular code for
948 transmembrane-helix recognition by the sec61 translocon., *Nature*
949 450 (7172) (2007) 1026–1030. doi:10.1038/nature06387.
- 950 [29] A. Bernsel, H. Viklund, J. Falk, E. Lindahl, G. von Heijne, A. Elof-
951 sson, Prediction of membrane-protein topology from first principles,
952 *Proceedings of the National Academy of Sciences of the United*

- 953 States of America 105 (20) (2008) 7177–7181, PMID: 18477697.
 954 doi:10.1073/pnas.0711151105.
 955 URL <http://www.ncbi.nlm.nih.gov/pubmed/18477697>
- 956 [30] R. Jackups Jr, J. Liang, Interstrand pairing patterns in beta-barrel
 957 membrane proteins: the positive-outside rule, aromatic rescue, and
 958 strand registration prediction., *J Mol Biol* 354(4) (2005) 979–93.
- 959 [31] H. Nikaido, T. Nakae, The outer membrane of Gram-negative bacteria.,
 960 *Adv Microb Physiol* 20 (1979) 163–250.
- 961 [32] P. G. Bagos, T. D. Liakopoulos, I. C. Spyropoulos, S. J. Hamod-
 962 rakas, PRED-TMBB: a web server for predicting the topology of
 963 beta-barrel outer membrane proteins, *Nucleic Acids Research* 32 (Web
 964 Server issue) (2004) W400–W404, PMID: 15215419 PMCID: 441555.
 965 doi:10.1093/nar/gkh417.
- 966 [33] A. Randall, J. Cheng, M. Sweredoski, P. Baldi, TMBpro: secondary
 967 structure, beta-contact and tertiary structure prediction of transmem-
 968 brane beta-barrel proteins, *Bioinformatics (Oxford, England)* 24 (4)
 969 (2008) 513–520, PMID: 18006547. doi:10.1093/bioinformatics/btm548.
- 970 [34] N. K. Singh, A. Goodman, P. Walter, V. Helms, S. Hayat, TMBHMM:
 971 a frequency profile based HMM for predicting the topology of trans-
 972 membrane beta barrel proteins and the exposure status of transmem-
 973 brane residues, *Biochimica Et Biophysica Acta* 1814 (5) (2011) 664–
 974 670, PMID: 21426944. doi:10.1016/j.bbapap.2011.03.004.
 975 URL <http://www.ncbi.nlm.nih.gov/pubmed/21426944>

- 976 [35] W. Wimley, Toward genomic identification of beta-barrel membrane
977 proteins: composition and architecture of known structures., *Protein*
978 *Sci* 11 (2) (2002) 301–312. doi:10.1110/ps.29402.
- 979 [36] J. F. Thomas C, W. C. Wimley, A highly accurate statistical ap-
980 proach for the prediction of transmembrane beta-barrels, *Bioinformat-*
981 *ics* (Oxford, England) 26 (16) (2010) 1965–1974, PMID: 20538726.
982 doi:10.1093/bioinformatics/btq308.
983 URL <http://www.ncbi.nlm.nih.gov/pubmed/20538726>
- 984 [37] A. Senes, M. Gerstein, D. Engelman, Statistical analysis of amino acid
985 patterns in transmembrane helices: the GxxxG motif occurs frequently
986 and in association with beta-branched residues at neighboring posi-
987 tions., *J Mol Biol* 296 (3) (2000) 921–936. doi:10.1006/jmbi.1999.3488.
- 988 [38] M. Lemmon, J. Flanagan, H. Treutlein, J. Zhang, D. Engelman, Se-
989 quence specificity in the dimerization of transmembrane alpha-helices.,
990 *Biochemistry* 31 (51) (1992) 12719–12725.
- 991 [39] A. Senes, D. E. Engel, W. F. DeGrado, Folding of helical membrane
992 proteins: the role of polar, GxxxG-like and proline motifs, *Curr Opin*
993 *Struct Biol.* 14 (2004) 465–479.
- 994 [40] D. Moore, B. Berger, W. DeGrado, Protein-protein interactions in
995 the membrane: sequence, structural, and biological motifs., *Structure*
996 16 (7) (2008) 991–1001. doi:10.1016/j.str.2008.05.007.
- 997 [41] J. Herrmann, J. Panitz, S. Unterreitmeier, A. Fuchs, D. Frish-
998 man, D. Langosch, Complex patterns of histidine, hydroxylated

- 999 amino acids and the GxxxG motif mediate high-affinity transmem-
1000 brane domain interactions., *J Mol Biol* 385 (3) (2009) 912–923.
1001 doi:10.1016/j.jmb.2008.10.058.
- 1002 [42] B. Berger, D. Kulp, L. Span, J. DeGrado, P. Billings, A. Senes, J. Ben-
1003 nett, W. DeGrado, Consensus motif for integrin transmembrane he-
1004 lix association., *Proc Natl Acad Sci U S A* 107 (2) (2010) 703–708.
1005 doi:10.1073/pnas.0910873107.
- 1006 [43] R. Jackups Jr, S. Cheng, J. Liang, Sequence motifs and antimotifs in
1007 beta-barrel membrane proteins from a genome-wide analysis: the Ala-
1008 Tyr dichotomy and chaperone binding motifs, *J Mol Biol* 363(2) (2006)
1009 611–23.
- 1010 [44] R. Jackups, Jr., J. Liang, Combinatorial analysis for se-
1011 quence and spatial motif discovery in short sequence fragments.,
1012 *IEEE/ACM Trans Comput Biol Bioinform* 7 (3) (2010) 524–536.
1013 doi:10.1109/TCBB.2008.101.
- 1014 [45] S. Robin, F. Rodophe, S. Schbath, DNA, Words and Models: Statistics
1015 of Exceptional Words, Cambridge University Press, 2005.
- 1016 [46] Struyvé, M. and Moons, M. and Tommassen, J., Carboxy-terminal
1017 phenylalanine is essential for the correct assembly of a bacterial outer
1018 membrane protein, *J Mol Biol.* 218 (1991) 141–148.
- 1019 [47] A. Marsico, K. Scheubert, A. Tuukkanen, A. Henschel, C. Winter,
1020 R. Winnenburg, M. Schroeder, MeMotif: a database of linear motifs in

- 1021 alpha-helical transmembrane proteins., *Nucleic Acids Res* 38 (Database
1022 issue) (2010) D181–9. doi:10.1093/nar/gkp1042.
- 1023 [48] R. Gangal, K. Kumar, Reduced alphabet motif methodology for GPCR
1024 annotation., *J Biomol Struct Dyn* 25 (3) (2007) 299–310.
- 1025 [49] R. Walters, W. DeGrado, Helix-packing motifs in membrane pro-
1026 teins., *Proc Natl Acad Sci U S A* 103 (37) (2006) 13658–13663.
1027 doi:10.1073/pnas.0605878103.
- 1028 [50] F. H. C. Crick, The packing of α -helices: simple coiled-coils, *Acta Crys-*
1029 *tallographica* 6 (8-9) (1953) 689–697. doi:10.1107/S0365110X53001964.
1030 URL <http://dx.doi.org/10.1107/S0365110X53001964>
- 1031 [51] J. Martin, A. de Brevern, A. Camproux, In silico local structure ap-
1032 proach: a case study on outer membrane proteins., *Proteins* 71 (1)
1033 (2008) 92–109. doi:10.1002/prot.21659.
- 1034 [52] L. Adamian, R. Jackups, T. A. Binkowski, J. Liang, Higher-order in-
1035 terhelical spatial interactions in membrane proteins, *J Mol Biol.* 327
1036 (2003) 251–272.
- 1037 [53] L. Adamian, J. Liang, Interhelical hydrogen bonds and spatial motifs
1038 in membrane proteins: polar clamps and serine zippers, *Proteins.* 47
1039 (2002) 209–218.
- 1040 [54] W. Ruan, V. Becker, U. Klingmuller, D. Langosch, The interface be-
1041 tween self-assembling erythropoietin receptor transmembrane segments

- 1042 corresponds to a membrane-spanning leucine zipper., J Biol Chem
1043 279(5) (2004) 3273–9.
- 1044 [55] W. Ruan, E. Lindner, D. Langosch, The interface of a membrane-
1045 spanning leucine zipper mapped by asparagine-scanning mutagenesis.,
1046 Protein Sci 13(2) (2004) 555–9.
- 1047 [56] E. Hutchinson, R. Sessions, J. Thornton, D. Woolfson, Determinants
1048 of strand register in antiparallel β -sheets of proteins., Protein Sci 7(11)
1049 (1998) 2287–300.
- 1050 [57] J. S. Merkel, L. Regan, Aromatic rescue of glycine in β sheets, Fold
1051 Des. 3 (1998) 449–455.
- 1052 [58] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic
1053 local alignment search tool, Journal of Molecular Biology 215 (3) (1990)
1054 403–410. doi:10.1006/jmbi.1990.9999.
- 1055 [59] M. Larkin, G. Blackshields, N. Brown, R. Chenna, P. McGet-
1056 tigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm,
1057 R. Lopez, J. Thompson, T. Gibson, D. Higgins, Clustal W and
1058 Clustal X version 2.0., Bioinformatics 23 (21) (2007) 2947–2948.
1059 doi:10.1093/bioinformatics/btm404.
- 1060 [60] M. Dayhoff, R. Schwartz, B. Orcutt, Atlas of Protei Sequence and
1061 Structure, vol.5, suppl. 3., National Biomedical Research Foundation,
1062 Washington, D.C, 1978, Ch. A model of evolutionary change in pro-
1063 teins, pp. pp. 345–352.

- 1064 [61] S. Henikoff, J. Henikoff, Amino acid substitution matrices from protein
1065 blocks., *Proc Natl Acad Sci U S A* 89 (22) (1992) 10915–10919.
- 1066 [62] P. Ng, J. Henikoff, S. Henikoff, PHAT: a transmembrane-specific sub-
1067 stitution matrix. predicted hydrophobic and transmembrane., *Bioin-*
1068 *formatics* 16 (9) (2000) 760–766.
- 1069 [63] T. Muller, S. Rahmann, M. Rehmsmeier, Non-symmetric score matri-
1070 ces and the detection of homologous transmembrane proteins., *Bioin-*
1071 *formatics* 17 Suppl 1 (2001) S182–9.
- 1072 [64] D. Jimenez-Morales, L. Adamian, J. Liang, Detecting remote ho-
1073 mologues using scoring matrices calculated from the estimation of
1074 amino acid substitution rates of beta-barrel membrane proteins,
1075 *Conference Proceedings: ... Annual International Conference of the*
1076 *IEEE Engineering in Medicine and Biology Society. IEEE Engineering*
1077 *in Medicine and Biology Society. Conference 2008* (2008) 1347–1350,
1078 PMID: 19162917. doi:10.1109/IEMBS.2008.4649414.
1079 URL <http://www.ncbi.nlm.nih.gov.proxy.cc.uic.edu/pubmed/19162917>
- 1080 [65] Y. Tseng, J. Liang, Estimation of amino acid residue substitution rates
1081 at local spatial regions and application in protein function inference:
1082 A Bayesian Monte Carlo approach., *Mol. Biol. Evol.* 23(2) (2006) 421–
1083 436.
- 1084 [66] D. Walther, D. Papic, M. Bos, J. Tommassen, D. Rapaport, Signals
1085 in bacterial beta-barrel proteins are functional in eukaryotic cells for

- targeting to and assembly in mitochondria., *Proc Natl Acad Sci U S A* 106 (8) (2009) 2531–2536. doi:10.1073/pnas.0807830106.
- [67] D. Walther, M. Bos, D. Rapaport, J. Tommassen, The mitochondrial porin, VDAC, has retained the ability to be assembled in the bacterial outer membrane., *Mol Biol Evol* 27 (4) (2010) 887–895. doi:10.1093/molbev/msp294.
- [68] J. Tommassen, Assembly of outer-membrane proteins in bacteria and mitochondria., *Microbiology* 156 (Pt 9) (2010) 2587–2596. doi:10.1099/mic.0.042689-0.
- [69] S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs., *Nucleic Acids Res* 25 (17) (1997) 3389–3402.
- [70] B. Rost, R. Casadio, P. Fariselli, C. Sander, Transmembrane helices predicted at 95% accuracy., *Protein Sci* 4 (3) (1995) 521–533. doi:10.1002/pro.5560040318.
- [71] W. Dowhan, M. Bogdanov, Lipid-dependent membrane protein topogenesis., *Annu Rev Biochem* 78 (2009) 515–540. doi:10.1146/annurev.biochem.77.060806.091251.
- [72] L. Adamian, H. Naveed, J. Liang, Lipid-binding surfaces of membrane proteins: evidence from evolutionary and structural analysis., *Biochim Biophys Acta* 1808 (4) (2011) 1092–1102. doi:10.1016/j.bbamem.2010.12.008.

- 1109 [73] D. Altschuh, A. M. Lesk, A. C. Bloomer, A. Klug, Correlation of co-
1110 ordinated amino acid substitutions with function in viruses related to
1111 tobacco mosaic virus, *Journal of Molecular Biology* 193 (4) (1987) 693–
1112 707, PMID: 3612789.
- 1113 [74] D. Altschuh, T. Vernet, P. Berti, D. Moras, K. Nagai, Coordinated
1114 amino acid changes in homologous protein families, *Protein Engineer-*
1115 *ing* 2 (3) (1988) 193 –199. doi:10.1093/protein/2.3.193.
- 1116 [75] S. J. Fleishman, V. M. Unger, M. Yeager, N. Ben-Tal, A calpha
1117 model for the transmembrane alpha helices of gap junction intercel-
1118 lular channels, *Molecular Cell* 15 (6) (2004) 879–888, PMID: 15383278.
1119 doi:10.1016/j.molcel.2004.08.016.
- 1120 [76] A. Fuchs, A. J. Martin-Galiano, M. Kalman, S. Fleishman,
1121 N. Ben-Tal, D. Frishman, Co-evolving residues in membrane pro-
1122 teins, *Bioinformatics (Oxford, England)* 23 (24) (2007) 3312–3319.
1123 doi:10.1093/bioinformatics/btm515.
- 1124 [77] S. J. Fleishman, O. Yifrach, N. Ben-Tal, An evolutionarily conserved
1125 network of amino acids mediates gating in voltage-dependent potas-
1126 sium channels, *Journal of Molecular Biology* 340 (2) (2004) 307–318,
1127 PMID: 15201054. doi:10.1016/j.jmb.2004.04.064.
1128 URL <http://www.ncbi.nlm.nih.gov/pubmed/15201054>
- 1129 [78] O. Beja, L. Aravind, E. Koonin, M. Suzuki, A. Hadd, L. Nguyen, S. Jo-
1130 vanovich, C. Gates, R. Feldman, J. Spudich, E. Spudich, E. DeLong,

- 1131 Bacterial rhodopsin: evidence for a new type of phototrophy in the
1132 sea., *Science* 289 (5486) (2000) 1902–1906.
- 1133 [79] J. Gilbert, C. Dupont, Microbial metagenomics: beyond the genome.,
1134 *Ann Rev Mar Sci* 3 (2011) 347–371.
- 1135 [80] L. Brown, K. Jung, Bacteriorhodopsin-like proteins of eubacteria
1136 and fungi: the extent of conservation of the haloarchaeal proton-
1137 pumping mechanism., *Photochem Photobiol Sci* 5 (6) (2006) 538–546.
1138 doi:10.1039/b514537f.
- 1139 [81] L. Adamian, Z. Ouyang, Y. Tseng, J. Liang, Evolutionary patterns
1140 of retinal-binding pockets of type I rhodopsins and their functions.,
1141 *Photochem Photobiol* 82(6) (2006) 1426–35.
- 1142 [82] S. Balashov, E. Imasheva, V. Boichenko, J. Anton, J. Wang,
1143 J. Lanyi, Xanthorhodopsin: a proton pump with a light-
1144 harvesting carotenoid antenna., *Science* 309 (5743) (2005) 2061–2064.
1145 doi:10.1126/science.1118046.
- 1146 [83] E. Imasheva, S. Balashov, A. Choi, K. Jung, J. Lanyi, Recon-
1147 stitution of *gloeobacter violaceus* rhodopsin with a light-harvesting
1148 carotenoid antenna., *Biochemistry* 48 (46) (2009) 10948–10955.
1149 doi:10.1021/bi901552x.
- 1150 [84] S. White, W. Wimley, Hydrophobic interactions of peptides with mem-
1151 brane interfaces., *Biochim Biophys Acta* 1376 (3) (1998) 339–352.

- 1152 [85] S. White, W. Wimley, Membrane protein folding and stability: phys-
1153 ical principles., *Annu Rev Biophys Biomol Struct* 28 (1999) 319–365.
1154 doi:10.1146/annurev.biophys.28.1.319.
- 1155 [86] J. Popot, D. Engelman, Helical membrane protein folding, stability,
1156 and evolution., *Annu Rev Biochem* 69 (2000) 881–922.
- 1157 [87] J. Liang, Experimental and computational studies of the determinants
1158 of membrane protein folding, *Curr. Opin. Chem. Biol.* 6 (2002) 878–
1159 884.
- 1160 [88] D. Engelman, Y. Chen, C. Chin, A. Curran, A. Dixon, A. Dupuy,
1161 A. Lee, U. Lehnert, E. Matthews, Y. Reshetnyak, A. Senes, J. Popot,
1162 Membrane protein folding: beyond the two stage model., *FEBS Lett*
1163 555 (1) (2003) 122–125.
- 1164 [89] J. Bowie, Solving the membrane protein folding problem., *Nature*
1165 438 (7068) (2005) 581–589. doi:10.1038/nature04395.
- 1166 [90] W. Wimley, T. Creamer, S. White, Solvation energies of amino acid
1167 side chains and backbone in a family of host-guest pentapeptides., *Bio-*
1168 *chemistry* 35 (16) (1996) 5109–5124. doi:10.1021/bi9600153.
- 1169 [91] W. Wimley, S. White, Experimentally determined hydrophobicity scale
1170 for proteins at membrane interfaces., *Nat Struct Biol* 3 (10) (1996) 842–
1171 848.
- 1172 [92] K. Hristova, W. Wimley, A look at arginine in membranes., *J Membr*
1173 *Biol* 239 (1-2) (2011) 49–56.

- 1174 [93] T. Hessa, H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson,
1175 I. Nilsson, S. White, G. von Heijne, Recognition of transmembrane
1176 helices by the endoplasmic reticulum translocon., *Nature* 433 (7024)
1177 (2005) 377–381. doi:10.1038/nature03216.
- 1178 [94] T. Hessa, S. White, G. von Heijne, Membrane insertion of a
1179 potassium-channel voltage sensor., *Science* 307 (5714) (2005) 1427.
1180 doi:10.1126/science.1109176.
- 1181 [95] C. Moon, K. Fleming, From the cover: Side-chain hydrophobic-
1182 ity scale derived from transmembrane protein folding into lipid bi-
1183 layers., *Proc Natl Acad Sci U S A* 108 (25) (2011) 10174–10177.
1184 doi:10.1073/pnas.1103979108.
- 1185 [96] S. Jayasinghe, K. Hristova, S. White, Energetics, stability, and pre-
1186 diction of transmembrane helices., *J Mol Biol* 312 (5) (2001) 927–934.
1187 doi:10.1006/jmbi.2001.5008.
- 1188 [97] C. Snider, S. Jayasinghe, K. Hristova, S. H. White, MPEx: a tool
1189 for exploring membrane proteins, *Protein Science* 18 (12) (2009)
1190 2624–2628. doi:10.1002/pro.256.
1191 URL <http://onlinelibrary.wiley.com/doi/10.1002/pro.256/abstract>
- 1192 [98] L. Adamian, J. Liang, Helix-helix packing and interfacial pairwise in-
1193 teractions of residues in membrane proteins, *J Mol Biol.* 311 (2001)
1194 891–907.
- 1195 [99] L. Adamian, V. Nanda, W. DeGrado, J. Liang, Empirical lipid propen-

- 1196 sities of amino acid residues in multispan alpha helical membrane pro-
1197 teins., *Proteins* 59(3) (2005) 496–509.
- 1198 [100] A. Senes, D. Chadi, P. Law, R. Walters, V. Nanda, W. DeGrado, E(z),
1199 a depth-dependent potential for assessing the energies of insertion of
1200 amino acid side-chains into membranes: derivation and applications to
1201 determining the orientation of transmembrane and interfacial helices.,
1202 *J Mol Biol* 366 (2) (2007) 436–448. doi:10.1016/j.jmb.2006.09.020.
- 1203 [101] S. Miyazawa, R. Jernigan, Residue-residue potentials with a favor-
1204 able contact pair term and an unfavorable high packing density term,
1205 for simulation and threading., *J Mol Biol* 256 (3) (1996) 623–644.
1206 doi:10.1006/jmbi.1996.0114.
- 1207 [102] X. Li, J. Liang, Computational algorithms for protein structure predic-
1208 tion, Springer, 2006, Ch. Knowledge-based energy functions for com-
1209 putational studies of proteins.
- 1210 [103] J. Lear, H. Gratkowski, L. Adamian, J. Liang, W. DeGrado, Position-
1211 dependence of stabilizing polar interactions of asparagine in transmem-
1212 brane helical bundles., *Biochemistry* 42(21) (2003) 6400–7.
- 1213 [104] A. Ridder, P. Skupjen, S. Unterreitmeier, D. Langosch, Tryptophan
1214 supports interaction of transmembrane helices., *J Mol Biol* 354 (4)
1215 (2005) 894–902. doi:10.1016/j.jmb.2005.09.084.
- 1216 [105] L. Adamian, J. Liang, Prediction of transmembrane helix orientation
1217 in polytopic membrane proteins., *BMC Struct Biol* 6 (2006) 13.

- 1218 [106] L. Adamian, J. Liang, Prediction of buried helices in multispan
1219 alpha helical membrane proteins., *Proteins* 63 (1) (2006) 1–5.
1220 doi:10.1002/prot.20874.
- 1221 [107] T. Beuming, H. Weinstein, A knowledge-based scale for the analysis
1222 and prediction of buried and exposed faces of transmembrane domain
1223 proteins., *Bioinformatics* 20(12) (2004) 1822–35.
- 1224 [108] Z. Yuan, F. Zhang, M. Davis, M. Boden, R. Teasdale, Predicting the
1225 solvent accessibility of transmembrane residues from protein sequence.,
1226 *J Proteome Res* 5 (5) (2006) 1063–1070. doi:10.1021/pr050397b.
- 1227 [109] Y. Park, S. Hayat, V. Helms, Prediction of the burial status of trans-
1228 membrane residues of helical membrane proteins., *BMC Bioinformatics*
1229 8 (2007) 302. doi:10.1186/1471-2105-8-302.
- 1230 [110] K. Illergard, S. Callegari, A. Elofsson, MPRAP: an accessibility pre-
1231 dictor for a-helical transmembrane proteins that performs well in-
1232 side and outside the membrane., *BMC Bioinformatics* 11 (2010) 333.
1233 doi:10.1186/1471-2105-11-333.
- 1234 [111] V. Yarov-Yarovoy, J. Schonbrun, D. Baker, Multipass membrane pro-
1235 tein structure prediction using rosetta., *Proteins* 62 (4) (2006) 1010–
1236 1025. doi:10.1002/prot.20817.
- 1237 [112] F. Leves, M. Tierney, S. Howitt, Polar residues in a conserved mo-
1238 tif spanning helices 1 and 2 are functionally important in the SulP
1239 transporter family., *Int J Biochem Cell Biol* 40 (11) (2008) 2596–2605.
1240 doi:10.1016/j.biocel.2008.05.007.

- 1241 [113] F. Zhou, M. Cocco, W. Russ, A. Brunger, D. Engelman, Interhelical
1242 hydrogen bonding drives strong interactions in membrane proteins.,
1243 Nat Struct Biol 7(2) (2000) 154–60.
- 1244 [114] C. Choma, H. Gratkowski, J. Lear, W. DeGrado, Asparagine-mediated
1245 self-association of a model transmembrane helix., Nat Struct Biol 7(2)
1246 (2000) 161–6.
- 1247 [115] N. Joh, A. Min, S. Faham, J. Whitelegge, D. Yang, V. Woods,
1248 J. Bowie, Modest stabilization by most hydrogen-bonded side-chain
1249 interactions in membrane proteins., Nature 453 (7199) (2008) 1266–
1250 1270. doi:10.1038/nature06977.
- 1251 [116] J. Bowie, Membrane protein folding: how important are hy-
1252 drogen bonds?, Curr Opin Struct Biol 21 (1) (2011) 42–49.
1253 doi:10.1016/j.sbi.2010.10.003.
- 1254 [117] S. Smith, M. Eilers, D. Song, E. Crocker, W. Ying, M. Groesbeek,
1255 G. Metz, M. Ziliox, S. Aimoto, Implications of threonine hydrogen
1256 bonding in the glycophorin a transmembrane helix dimer., Biophys J
1257 82 (5) (2002) 2476–2486. doi:10.1016/S0006-3495(02)75590-2.
- 1258 [118] S. Faham, D. Yang, E. Bare, S. Yohannan, J. Whitelegge, J. Bowie,
1259 Side-chain contributions to membrane protein structure and stability.,
1260 J Mol Biol 335 (1) (2004) 297–305.
- 1261 [119] S. Fleishman, S. Harrington, R. Friesner, B. Honig, N. Ben-Tal, An au-
1262 tomatic method for predicting transmembrane protein structures using
1263 cryo-em and evolutionary data., Biophys J 87(5) (2004) 3448–59.

- 1264 [120] B. North, L. Cristian, X. Fu Stowell, J. Lear, J. Saven, W. DeGrado,
1265 Characterization of a membrane protein folding motif, the Ser zip-
1266 per, using designed peptides., J Mol Biol 359 (4) (2006) 930–939.
1267 doi:10.1016/j.jmb.2006.04.001.
- 1268 [121] Y. Zhang, D. Kulp, J. Lear, W. DeGrado, Experimental and com-
1269 putational evaluation of forces directing the association of trans-
1270 membrane helices., J Am Chem Soc 131 (32) (2009) 11341–11343.
1271 doi:10.1021/ja904625b.
- 1272 [122] M. Wouters, P. Curmi, An analysis of side chain interactions and pair
1273 correlations within antiparallel β -sheets: the differences between back-
1274 bone hydrogen-bonded and non-hydrogen-bonded residue pairs., Pro-
1275 teins 22(2) (1995) 119–31.
- 1276 [123] B. Ho, P. Curmi, Twist and shear in β -sheets and β -ribbons., J Mol
1277 Biol 317(2) (2002) 291–308.
- 1278 [124] H. Hong, G. Szabo, L. Tamm, Electrostatic couplings in OmpA ion-
1279 channel gating suggest a mechanism for pore opening., Nat Chem Biol
1280 2 (11) (2006) 627–635. doi:10.1038/nchembio827.
- 1281 [125] H. Hong, S. Park, R. Jimenez, D. Rinehart, L. Tamm, Role of aro-
1282 matic side chains in the folding and thermodynamic stability of inte-
1283 gral membrane proteins., J Am Chem Soc 129 (26) (2007) 8320–8327.
1284 doi:10.1021/ja068849o.
- 1285 [126] F. Oesterhelt, D. Oesterhelt, M. Pfeiffer, A. Engel, H. Gaub,

- 1286 D. Muller, Unfolding pathways of individual bacteriorhodopsins., Sci-
1287 ence 288 (5463) (2000) 143–146.
- 1288 [127] H. Naveed, R. Jackups Jr, J. Liang, Predicting weakly stable regions,
1289 oligomerization state, and protein-protein interfaces in transmembrane
1290 domains of outer membrane proteins, Proc Nat Aca Sci USA In press
1291 (2009) doi:10.1073/pnas.0902169106.
- 1292 [128] D. Jones, Successful ab initio prediction of the tertiary structure of NK-
1293 lysin using multiple sequences and recognized supersecondary struc-
1294 tural motifs., Proteins Suppl 1 (1997) 185–191.
- 1295 [129] M. Eilers, A. B. Patel, W. Liu, S. O. Smith, Comparison of helix in-
1296 teractions in membrane and soluble α -bundle proteins, Biophys J. 82
1297 (2002) 2720–2736.
- 1298 [130] H. Dobbs, E. Orlandini, R. Bonaccini, F. Seno, Optimal potentials for
1299 predicting inter-helical packing in transmembrane proteins., Proteins
1300 49 (3) (2002) 342–349. doi:10.1002/prot.10229.
- 1301 [131] C. Wendel, H. Gohlke, Predicting transmembrane helix pair configura-
1302 tions with knowledge-based distance-dependent pair potentials., Pro-
1303 teins 70 (3) (2008) 984–999. doi:10.1002/prot.21574.
- 1304 [132] A. Sodt, T. Head-Gordon, Driving forces for transmembrane
1305 alpha-helix oligomerization., Biophys J 99 (1) (2010) 227–237.
1306 doi:10.1016/j.bpj.2010.03.071.
- 1307 [133] P. Barth, J. Schonbrun, D. Baker, Toward high-resolution pre-
1308 diction and design of transmembrane helical protein structures.,

- 1309 Proc Natl Acad Sci U S A 104 (40) (2007) 15682–15687.
1310 doi:10.1073/pnas.0702515104.
- 1311 [134] P. Barth, B. Wallner, D. Baker, Prediction of membrane
1312 protein structures with complex topologies using limited con-
1313 straints., Proc Natl Acad Sci U S A 106 (5) (2009) 1409–1414.
1314 doi:10.1073/pnas.0808323106.
- 1315 [135] B. van den Berg, The FadL family: unusual transporters for un-
1316 usual substrates., Curr Opin Struct Biol 15 (4) (2005) 401–407.
1317 doi:10.1016/j.sbi.2005.06.003.
- 1318 [136] G. Huysmans, S. Radford, D. Brockwell, S. Baldwin, The N-
1319 terminal helix is a post-assembly clamp in the bacterial outer
1320 membrane protein pagp., J Mol Biol 373 (3) (2007) 529–540.
1321 doi:10.1016/j.jmb.2007.07.072.
- 1322 [137] Y. Jiang, A. Lee, J. Chen, V. Ruta, M. Cadene, B. Chait, R. MacK-
1323 innon, X-ray structure of a voltage-dependent k⁺ channel., Nature
1324 423 (6935) (2003) 33–41. doi:10.1038/nature01580.
- 1325 [138] V. Ruta, J. Chen, R. MacKinnon, Calibrated measurement of gating-
1326 charge arginine displacement in the KvAP voltage-dependent k⁺ chan-
1327 nel., Cell 123 (3) (2005) 463–475. doi:10.1016/j.cell.2005.08.041.
- 1328 [139] S. Dorairaj, T. Allen, On the thermodynamic stability of a charged
1329 arginine side chain in a transmembrane helix., Proc Natl Acad Sci U S
1330 A 104 (12) (2007) 4943–4948. doi:10.1073/pnas.0610470104.

- 1331 [140] J. MacCallum, W. Bennett, D. Tieleman, Partitioning of amino acid
1332 side chains into lipid bilayers: results from computer simulations and
1333 comparison to experiment., *J Gen Physiol* 129 (5) (2007) 371–377.
1334 doi:10.1085/jgp.200709745.
- 1335 [141] E. Schow, J. Freites, P. Cheng, A. Bernsel, G. von Heijne, S. White,
1336 D. Tobias, Arginine in membranes: the connection between molecular
1337 dynamics simulations and translocon-mediated insertion experiments.,
1338 *J Membr Biol* 239 (1-2) (2011) 35–48.
- 1339 [142] A. Johansson, E. Lindahl, Protein contents in biological mem-
1340 branes can explain abnormal solvation of charged and polar
1341 residues., *Proc Natl Acad Sci U S A* 106 (37) (2009) 15684–15689.
1342 doi:10.1073/pnas.0905394106.
- 1343 [143] N. Joh, A. Min, S. Faham, J. Whitelegge, D. Yang, V. Woods, J. Bowie,
1344 Modest stabilization by most hydrogen-bonded side-chain interactions
1345 in membrane proteins., *Nature* 453 (7199) (2008) 1266–1270.
- 1346 [144] E. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for
1347 predicting transmembrane helices in protein sequences., *Proc Int Conf*
1348 *Intell Syst Mol Biol* 6 (1998) 175–182.
- 1349 [145] G. von Heijne, Proline kinks in transmembrane alpha-helices., *J Mol*
1350 *Biol* 218 (3) (1991) 499–503.
- 1351 [146] C. Papaloukas, E. Granseth, H. Viklund, A. Elofsson, Estimating the
1352 length of transmembrane helices using z-coordinate predictions., *Pro-*
1353 *tein Sci* 17 (2) (2008) 271–278. doi:10.1110/ps.073036108.

- [147] K. Hristova, W. Wimley, V. Mishra, G. Anantharamiah, J. Segrest, S. White, An amphipathic alpha-helix at a membrane interface: a structural study using a novel X-ray diffraction method., *J Mol Biol* 290 (1) (1999) 99–117. doi:10.1006/jmbi.1999.2840.
- [148] E. Granseth, G. von Heijne, A. Elofsson, A study of the membrane-water interface region of membrane proteins., *J Mol Biol.* 346 (2005) 377–385.
- [149] H. Viklund, E. Granseth, A. Elofsson, Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes., *J Mol Biol* 361 (3) (2006) 591–603. doi:10.1016/j.jmb.2006.06.037.
- [150] A. Kauko, K. Illergard, A. Elofsson, Coils in the membrane core are conserved and functionally important., *J Mol Biol* 380 (1) (2008) 170–180. doi:10.1016/j.jmb.2008.04.052.
- [151] S. Hall, K. Roberts, N. Vaidehi, Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction., *J Mol Graph Model* 27 (8) (2009) 944–950. doi:10.1016/j.jmglm.2009.02.004.
- [152] S. Yohannan, S. Faham, D. Yang, D. Grosfeld, A. Chamberlain, J. Bowie, A C alpha-H...O hydrogen bond in a membrane protein is not stabilizing., *J Am Chem Soc* 126 (8) (2004) 2284–2285. doi:10.1021/ja0317574.

- 1376 [153] C. Yan, J. Luo, An analysis of reentrant loops., *Protein J* 29 (5) (2010)
1377 350–354. doi:10.1007/s10930-010-9259-z.
- 1378 [154] L. Forrest, C. Tang, B. Honig, On the accuracy of homology model-
1379 ing and sequence alignment methods applied to membrane proteins.,
1380 *Biophys J* 91 (2) (2006) 508–517. doi:10.1529/biophysj.106.082313.
- 1381 [155] Y. Zhang, M. Devries, J. Skolnick, Structure modeling of all identified
1382 G protein-coupled receptors in the human genome., *PLoS Comput Biol*
1383 2 (2) (2006) e13. doi:10.1371/journal.pcbi.0020013.
- 1384 [156] D. Goldfeld, K. Zhu, T. Beuming, R. Friesner, Successful predic-
1385 tion of the intra- and extracellular loops of four G-protein-coupled
1386 receptors., *Proc Natl Acad Sci U S A* 108 (20) (2011) 8275–8280.
1387 doi:10.1073/pnas.1016951108.
- 1388 [157] S. Kelm, J. Shi, C. Deane, MEDELLER: homology-based coordinate
1389 generation for membrane proteins., *Bioinformatics* 26 (22) (2010) 2833–
1390 2840. doi:10.1093/bioinformatics/btq554.
- 1391 [158] A. Randall, J. Cheng, M. Sweredoski, P. Baldi, TMBpro: secondary
1392 structure, beta-contact and tertiary structure prediction of transmem-
1393 brane beta-barrel proteins., *Bioinformatics* 24 (4) (2008) 513–520.
1394 doi:10.1093/bioinformatics/btm548.
- 1395 [159] D. Jimenez-Morales, J. Liang, Pattern of amino acid substitutions in
1396 transmembrane domains of β -barrel membrane proteins for detecting
1397 remote homologs in bacteria and mitochondria, *PLoS ONE*.

- 1398 [160] S. Lindert, R. Staritzbichler, N. Wotzel, M. Karakas, P. Stewart,
1399 J. Meiler, EM-fold: De novo folding of alpha-helical proteins guided
1400 by intermediate-resolution electron microscopy density maps., *Struc-*
1401 *ture* 17 (7) (2009) 990–1003. doi:10.1016/j.str.2009.06.001.
- 1402 [161] M. Schushan, Y. Barkan, T. Haliloglu, N. Ben-Tal, C(alpha)-trace
1403 model of the transmembrane domain of human copper transporter 1,
1404 motion and functional implications., *Proc Natl Acad Sci U S A* 107 (24)
1405 (2010) 10908–10913. doi:10.1073/pnas.0914717107.
- 1406 [162] D. Metcalf, P. Law, W. DeGrado, Mutagenesis data in the automated
1407 prediction of transmembrane helix dimers., *Proteins* 67 (2) (2007) 375–
1408 384.
- 1409 [163] D. Metcalf, D. Kulp, J. Bennett, W. DeGrado, Multiple approaches
1410 converge on the structure of the integrin alphaIIb/beta3 transmem-
1411 brane heterodimer., *J Mol Biol* 392 (4) (2009) 1087–1101.
- 1412 [164] M. Lin, H. Lu, R. Chen, J. Liang, Generating properly weighted en-
1413 semble of conformations of proteins from sparse or indirect distance
1414 constraints., *J Chem Phys* 129(9) (2008) 094101.
- 1415 [165] J. Cuesta-Seijo, C. Neale, M. Khan, J. Moktar, C. Tran, R. Bishop,
1416 R. Pomes, G. Prive, PagP crystallized from SDS/cosolvent reveals
1417 the route for phospholipid access to the hydrocarbon ruler., *Structure*
1418 18 (9) (2010) 1210–1219. doi:10.1016/j.str.2010.06.014.
- 1419 [166] M. Lin, J. Zhang, H. Lu, R. Chen, J. Liang, Constrained proper sam-

- 1420 pling of conformations of transition state ensemble of protein folding.,
1421 J Chem Phys 134 (7) (2011) 075103. doi:10.1063/1.3519056.
- 1422 [167] A. Elofsson, G. von Heijne, Membrane protein structure: pre-
1423 diction versus reality., Annu Rev Biochem 76 (2007) 125–140.
1424 doi:10.1146/annurev.biochem.76.052705.163539.
- 1425 [168] A. Kauko, L. Hedin, E. Thebaud, S. Cristobal, A. Elofsson,
1426 G. von Heijne, Repositioning of transmembrane alpha-helices dur-
1427 ing membrane protein folding., J Mol Biol 397 (1) (2010) 190–201.
1428 doi:10.1016/j.jmb.2010.01.042.
- 1429 [169] U. Lehnert, Y. Xia, T. Royce, C. Goh, Y. Liu, A. Senes, H. Yu,
1430 Z. Zhang, D. Engelman, M. Gerstein, Computational analysis of mem-
1431 brane proteins: genomic occurrence, structure prediction and helix in-
1432 teractions., Q Rev Biophys 37 (2) (2004) 121–146.
- 1433 [170] Y. Liu, M. Gerstein, D. Engelman, Transmembrane protein domains
1434 rarely use covalent domain recombination as an evolutionary mech-
1435 anism., Proc Natl Acad Sci U S A 101 (10) (2004) 3495–3497.
1436 doi:10.1073/pnas.0307330101.
- 1437 [171] L. Simpson, B. Taddese, I. Wall, C. Reynolds, Bioinformatics and
1438 molecular modelling approaches to GPCR oligomerization., Curr Opin
1439 Pharmacol 10 (1) (2010) 30–37. doi:10.1016/j.coph.2009.11.001.
- 1440 [172] K. Locher, J. Rosenbusch, Oligomeric states and siderophore binding of
1441 the ligand-gated FhuA protein that forms channels across escherichia
1442 coli outer membranes., Eur J Biochem 247 (3) (1997) 770–775.

- 1443 [173] A. Bordner, Predicting protein-protein binding sites in membrane pro-
1444 teins., *BMC Bioinformatics* 10 (2009) 312. doi:10.1186/1471-2105-10-
1445 312.
- 1446 [174] N. Keinan, D. Tyomkin, V. Shoshan-Barmatz, Oligomerization of the
1447 mitochondrial protein voltage-dependent anion channel is coupled to
1448 the induction of apoptosis., *Mol Cell Biol* 30 (24) (2010) 5698–5709.
1449 doi:10.1128/MCB.00165-10.
- 1450 [175] V. Shoshan-Barmatz, N. Keinan, S. Abu-Hamad, D. Tyomkin,
1451 L. Aram, Apoptosis is regulated by the VDAC1 n-terminal region
1452 and by VDAC oligomerization: release of cytochrome c, AIF and
1453 Smac/Diablo., *Biochim Biophys Acta* 1797 (6-7) (2010) 1281–1291.
1454 doi:10.1016/j.bbabi.2010.03.003.
- 1455 [176] I. Samish, C. MacDermaid, J. Perez-Aguilar, J. Saven, Theoretical and
1456 computational protein design., *Annu Rev Phys Chem* 62 (2011) 129–
1457 149.
- 1458 [177] I. Korendovych, A. Senes, Y. Kim, J. Lear, H. Fry, M. Therien,
1459 J. Blasie, F. Walker, W. DeGrado, De novo design and molecular as-
1460 sembly of a transmembrane diporphyrin-binding protein complex., *J*
1461 *Am Chem Soc* 132 (44) (2010) 15516–15518.
- 1462 [178] H. Yin, J. Slusky, B. Berger, R. Walters, G. Vilaire, R. Litvinov,
1463 J. Lear, G. Caputo, J. Bennett, W. DeGrado, Computational design of
1464 peptides that target transmembrane helices., *Science* 315 (5820) (2007)
1465 1817–1822.

- 1466 [179] G. Caputo, R. Litvinov, W. Li, J. Bennett, W. DeGrado, H. Yin, Com-
1467 putationally designed peptide inhibitors of protein-protein interactions
1468 in membranes., *Biochemistry* 47 (33) (2008) 8600–8606.
- 1469 [180] C. Sarkar, I. Dodevski, M. Kenig, S. Dudli, A. Mohr, E. Hermans,
1470 A. Pluckthun, Directed evolution of a G protein-coupled receptor for
1471 expression, stability, and binding selectivity., *Proc Natl Acad Sci U S*
1472 *A* 105 (39) (2008) 14808–14813.
- 1473 [181] R. Signarvic, W. DeGrado, Metal-binding dependent disruption of
1474 membranes by designed helices., *J Am Chem Soc* 131 (9) (2009) 3377–
1475 3384.
- 1476 [182] N. Muhammad, T. Dworeck, M. Fioroni, U. Schwaneberg, Engineering
1477 of the E. coli Outer Membrane Protein FhuA to overcome the Hy-
1478 drophobic Mismatch in Thick Polymeric Membranes., *J Nanobiotech-*
1479 *nology* 9 (1) (2011) 8.
- 1480 [183] M. Taylor, H. Fung, R. Rajgaria, M. Filizola, H. Weinstein, C. Floudas,
1481 Mutations affecting the oligomerization interface of G-protein-coupled
1482 receptors revealed by a novel de novo protein design framework., *Bio-*
1483 *phys J* 94 (7) (2008) 2470–2481.
- 1484 [184] G. Schulz, The structure of bacterial outer membrane proteins.,
1485 *Biochim Biophys Acta* 1565 (2) (2002) 308–317.
- 1486 [185] T. Arnold, M. Poynor, S. Nussberger, A. Lupas, D. Linke, Gene dupli-
1487 cation of the eight-stranded beta-barrel OmpX produces a functional

- 1488 pore: a scenario for the evolution of transmembrane beta-barrels., J
1489 Mol Biol 366 (4) (2007) 1174–1184.
- 1490 [186] L. Song, M. Hobaugh, C. Shustak, S. Cheley, H. Bayley, J. Gouaux,
1491 Structure of staphylococcal alpha-hemolysin, a heptameric transmem-
1492 brane pore., Science 274 (5294) (1996) 1859–1866.
- 1493 [187] V. Koronakis, A. Sharff, E. Koronakis, B. Luisi, C. Hughes, Crystal
1494 structure of the bacterial membrane protein TolC central to multidrug
1495 efflux and protein export., Nature 405 (6789) (2000) 914–919.
- 1496 [188] H. Miedema, A. Meter-Arkema, J. Wierenga, J. Tang, B. Eisenberg,
1497 W. Nonner, H. Hektor, D. Gillespie, W. Meijberg, Permeation proper-
1498 ties of an engineered bacterial OmpF porin containing the EEEE-locus
1499 of Ca²⁺ channels., Biophys J 87 (5) (2004) 3137–3147.
- 1500 [189] H. Li, H. Chen, C. Steinbronn, B. Wu, E. Beitz, T. Zeuthen, G. Voth,
1501 Enhancement of proton conductance by mutations of the selectivity
1502 filter of aquaporin-1., J Mol Biol 407 (4) (2011) 607–620.
- 1503 [190] E. Udho, K. Jakes, S. Buchanan, K. James, X. Jiang, P. Klebba,
1504 A. Finkelstein, Reconstitution of bacterial outer membrane tonb-
1505 dependent transporters in planar lipid bilayer membranes., Proc Natl
1506 Acad Sci U S A 106 (51) (2009) 21990–21995.
- 1507 [191] A. Basle, R. Qutub, M. Mehrazin, J. Wibbenmeyer, A. Delcour, Dele-
1508 tions of single extracellular loops affect pH sensitivity, but not voltage
1509 dependence, of the escherichia coli porin ompf., Protein Eng Des Sel
1510 17 (9) (2004) 665–672.

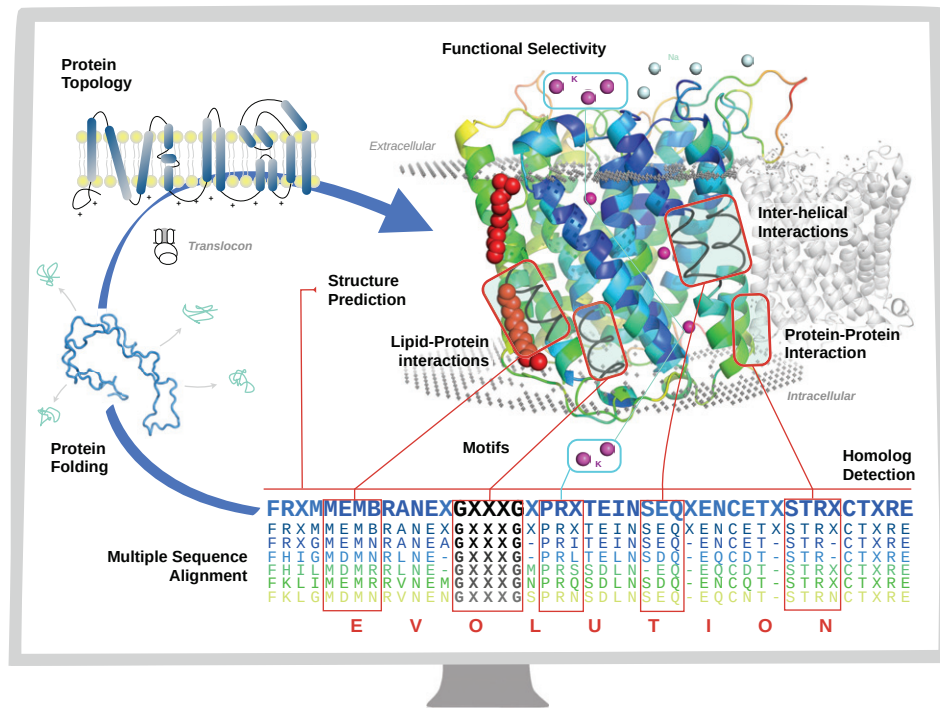


Figure 1: The central dogma of molecular biology for membrane proteins (blue arrow) through the monitor of a computer. The chain of amino acids folds through the mediation of the translocon to a final stable low energy structure, with a specific topology (Section 2.2). The goal of structure prediction is to derive the three-dimensional structure of a membrane protein from its sequence (Section 7). The assembly of the helices in the transmembrane domains is facilitated by interhelical interactions (Section 6) via sequence and spatial motifs (Section 3), as well as protein lipid interactions (Sections 5.2 and 4). Membrane proteins also participate in protein-protein interactions (Section 8.2) for biological functions. The evolutionary relationship between membrane proteins can be detected through (multiple) sequence alignment, for which an evolutionary model of substitutions of residues in the transmembrane domain is essential in deriving specialized scoring matrices for alignment and for detection of homologs (Section 4). With significant understanding of the organizing principles of membrane proteins, computational studies can be carried out to design membrane proteins with desired properties such as functional selectivity (Section 8.3).

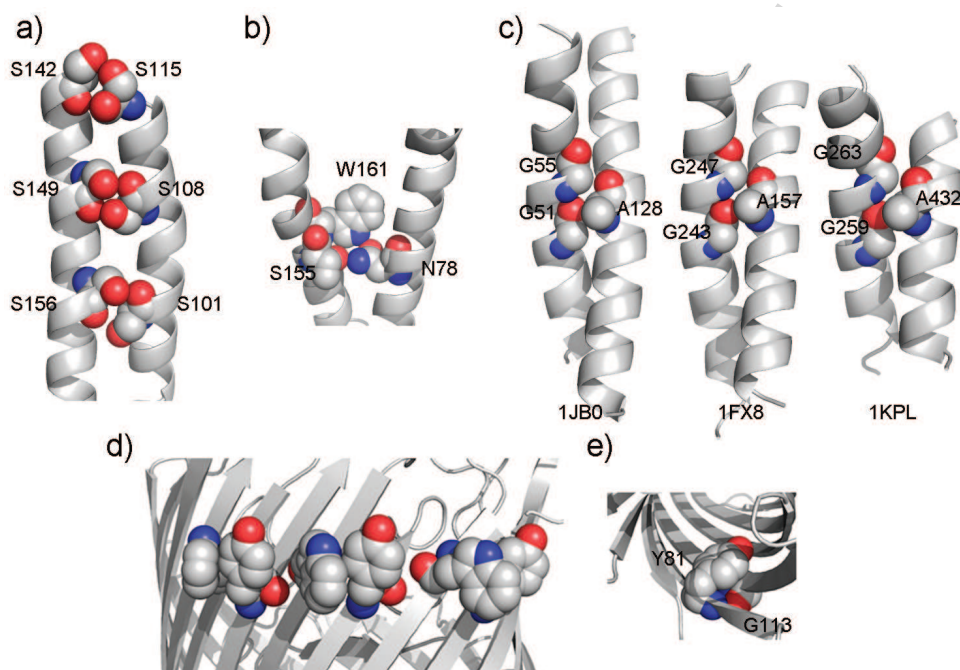


Figure 2: Spatial motifs in α -helical and β -barrel membrane proteins: a) The serine zipper in bovine cytochrome c oxidase (helices III and IV), in which H-bonds are formed between S101-S156, S108-S149, and S115-S142. b) A polar clamp in bovine rhodopsin formed by residues W161 and T160 from helix IV and by N78 from helix II. The side chain of W161 is positioned such that its NE1 atom forms an H-bond with the OD1 atom from N78, while the OG1 oxygen from T160 is H-bonded to one of ND hydrogens of N78. c) The A-G-G triplet spatial motif interacting with GG4 sequential motif. Three helical pairs are from unrelated proteins (1jb0: photosystem I; 1fx8: glycerol conducting channel; 1jpl: Clc chloride channel), but all have similar parallel helical orientation with similar crossing angle values between -33 and -48 degrees. d) An instance of the WY non-H-bonded interaction motif in LamB. The aromatic side-chains of Trp and Tyr show considerable contact interaction. e) An instance of the GY strong H-bonded interaction motif in NspA. The protein has been tilted to show the motif on the internal side of the barrel. The aromatic side-chain of Tyr interacts with the Gly residue on the adjacent strand. This is an example of “aromatic rescue” (adapted from (30)).

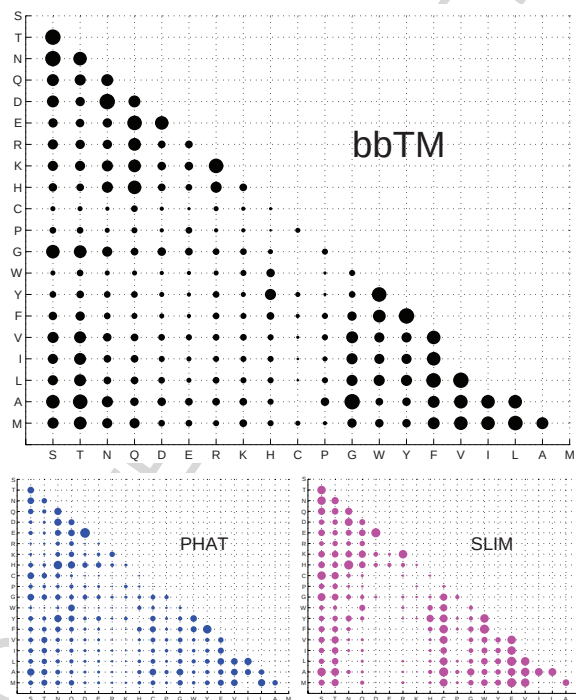


Figure 3: The scoring matrices representing the substitution probability between different residues in transmembrane segments of β -barrel membrane proteins (bbTM) and α -helical membrane proteins (according to the PHAT and SLIM matrices) (62; 63). The size of a bubble is roughly proportional to the probability of substitution between the two corresponding residues (adapted from (64)).

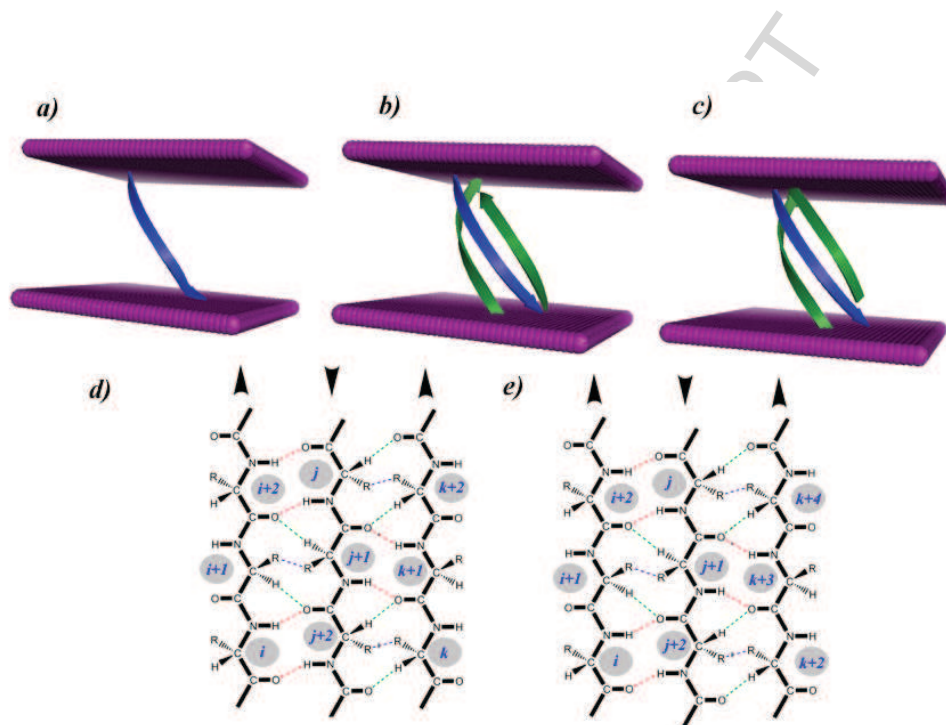


Figure 4: Interactions of transmembrane strands in β -barrel membrane proteins and energetics of embedding Arg in the transmembrane domain. a) A single β -strand inserted in the membrane bilayer. Both experimental and computational potential functions show that the insertion energy of amino acid varies with their depth in the bilayer. b) Three strands inserted in the bilayer. Although the experimentally measured insertion scales derived from single helix experiments are insightful, a complete picture of the energetics requires considering interactions with neighboring strands/helices. c) An alternate conformation of the same three strands as shown in b). A good computational model can assess how prevalent each conformation is, and can estimate the associated probability. d) and e) show the pairwise interactions by the TMSIP potential function according to the β -sheet canonical model (122; 123) for the strands shown in b) and c), respectively. Strong H-bonds between C-O $\cdots$ H-N, weak H-bonds between C-O $\cdots$ H-C $_{\alpha}$, and side-chain interactions are shown.

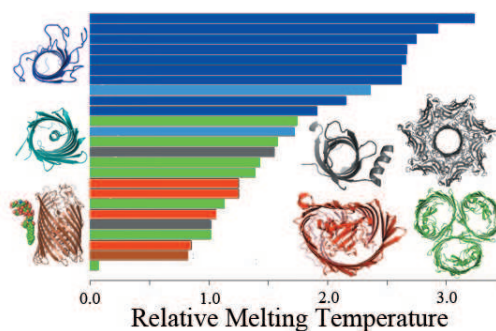


Figure 5: The relative melting temperature of the transmembrane domains of 25 β -barrel membrane proteins can be calculated by enumerating all possible conformations in a reduced state space. Monomers that are stable without in-plugs and out-clamps, *e.g.*, OmpA are shown in dark blue. Monomers stabilized by small in-plugs, *e.g.*, NalP are shown in light blue. Monomers stabilized by out-clamps are represented by PagP and α -hemolysin (grey). β -barrels that require oligomerization for stability, *e.g.*, ScrY are shown in green. Monomers stabilized by large in-plugs *e.g.* FptA are shown in red. β -barrel membrane proteins can also have specific protein-lipid interactions, *e.g.*, FhuA (brown) that increase protein stability. All stable monomers tend to have higher relative melting temperature and group towards the top of the graph (adapted from (127)).

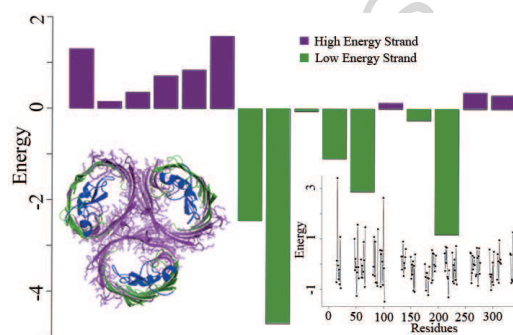


Figure 6: The β -barrel membrane protein OmpF exists as a trimer, with strands 1-5 and 16 forming the protein-protein interaction (PPI) interface. The expected energy of the trans-membrane domains of each of the β -strands is calculated using the TMSIP statistical potential function. The consecutive strands 1-6 and 15-16 have high expected energy and coincide with the real PPI interface of the protein. Here high energy strands are also termed as weakly stable. The accuracy of identifying the β -strands located in the PPI interface in a data set of 25 non-redundant β -barrel membrane proteins is 78% using structural information and 66% using sequence information only. The right inset plots the contribution of each residue to the stability of the protein. This can be used to suggest mutagenesis studies that aim to change the stability of the protein (adapted from (127)).

- >Prediction of membrane proteins and topology,
- >Discovery of sequence and spatial motifs, detection of evolutionary signal,
- >Empirical potential functions,
- >Structure predictions and ensemble properties,
- >Prediction of protein-protein interactions