Dawn of a New Era for Membrane Protein Design

Shahin Sowlati-Hashjin, Aanshi Gandhi, and Michael Garton

Institute of Biomedical Engineering, University of Toronto, 164 College Street, Toronto, ON, Canada M5S 3E2

Correspondence should be addressed to Michael Garton; michael.garton@utoronto.ca

1. Introduction

Approximately one-third of all human proteins are located in membranes. However, due to the difficulty in solving membrane protein structures, they have received relatively little attention by the protein design community (see Table 1), as typical protein design pipelines require high-quality structures as a starting point. Of an estimated 6,700 membrane proteins in the human proteome [1], fewer than 200 structures have been solved [2]. This low representation is partly a result of low expression levels in native tissues [3]. For traditional protein structural determination methods such as NMR or diffraction-quality crystals, large and isolated amounts are needed. Proteins can be expressed using bacterial hosts such as E. coli; however, they require sophisticated systems that can selectively deliver membrane proteins to the correct cell membrane [4]. Additionally, due to the hydrophobic nature and lack of flexibility of membrane proteins, isolation methods with detergents generally result in aggregation as a result of misfolding or lower stability, decreasing the quality of solved proteins [3].

A paucity of structures meant that a disproportionate amount of work over the last decade in membrane protein design has been performed using de novo approaches. De novo proteins are traditionally designed using two distinct methods—minimal and rational. A minimal design uses simple polar and hydrophobic rules to drive the design, while rational design uses sequence-structure relationships with computational tools to generate new backbone structures as templates. This is followed by fitting of sequences that are predicted to stabilize them. A de novo design of membrane proteins has been recently—and excellently—reviewed elsewhere ([5]) and will therefore only be lightly referenced in this review.

In 2021, however, there was a seismic shift in the protein structure field, precipitated by dramatic advances in machine learning-driven structure prediction. alphafold2 [6], developed by Google’s DeepMind, used an ingenious deep learning model to provide protein designers with reliable structure predictions for all proteins in the human proteome. In doing so, they imbued the protein design community with unprecedented potential for redesigning membrane proteins and thus designing human cells with new functions. Cell-based therapies are widely predicted to be the next major health care revolution [7], and this step in our ability to design human membrane proteins could contribute significantly to their development. Here, we set out the current state of development in using nature-derived membrane protein scaffolds for design and review established and emerging tools for capitalizing on the sudden wealth of high-quality predicted structures. Enhancing our ability to rationally design new membrane proteins is likely to have considerable impact in many areas of synthetic biology—particularly in sensing and engineering novel multicellular behavior.

1.1. Current Landscape of Natural Scaffold Membrane Protein Design. Early designs of membrane proteins can be traced back to the late 80s and were mainly based on the sequence and structure of known proteins. Based on the
have been deposited in the protein databank and only fewer than 400 beta-barrel membrane proteins (bBMPs) This is due to the lack of solved structures. For example, the application of these techniques to water-soluble proteins. proven highly challenging and is far less developed than structure-based design of membrane spanning proteins has titrating of lysine residues [9]. More sophisticated, rational the protein in the membrane can be controlled by the position of the trend observed for dimerization in water-soluble proteins. Taking advantage of this property, a membrane hemoprotein, ME1, was redesigned with a glycoporphin A to create a heme binding site [15]. Helix association propensity of amino acids in membrane proteins can be rank ordered (Gly > Ala > Val > Ile)—opposite of the trend observed for dimerization in water-soluble structures [16]. Stability and orientation of the transmembrane domain is thus mainly determined by noncovalent interactions, namely, van der Waals and electrostatic (including hydrogen bonds).

However, despite a long history, the design remains challenging. Even coiled coils, which have well-characterized packing rules and have been amenable to design in their water-soluble form for a couple of decades [17], are still largely refractory to the transmembrane design. A redesign of alpha-helical barrels using natural scaffolds is similarly underdeveloped. Examples of design include Cytoysin A—engineered to create a discrete lipid bilayer pore that can translocate DNA [18]. The design of active channels using short synthetic peptides is being energetically developed by the nanopore design community [19].

Membrane proteins are predominantly alpha-helical but transmembrane domains composed of beta-strands are also an important group. These are more commonly seen in bacteria, mitochondria, and chloroplast [20]. They commonly comprise a water-filled pore or solid core with polar side chains instead of a hydrophobic core. A potential term, EZβ, was determined for depth-related hydrophobicity to help redesign such proteins as they have very different structural features to other membrane proteins [20]. Most transmembrane beta-strand sequences alternate between polar and hydrophobic residues to create one hydrophobic face for the lipid membrane environment when folded. By adopting a beta-barrel structure, such that the hydrophilic face creates a polar interior channel, allowing transport of hydrophilic substrates. Transmembrane beta-barrels (TMB) range

| Functional group | Type | Computational design examples |
|------------------|------|-------------------------------|
| Receptor         | Tyrosine kinase | Design of a self-assembling symmetrical β-propeller protein [28] |
|                  | Cytokine receptors | Potent and selective mimics of interleukin-2 and interleukin-15 [29] |
|                  | Serine kinase | Tyrosine and serine kinase-driven protein switches [27] |
| Carrier/transporter | Channel | Transmembrane channel ([30]) |
|                  | Carrier/transporter | Multicentered pathway for transmembrane electron transfer [34] |
| Enzyme           |       | P450 2C3d engineered as soluble dimer [35] |
| Adhesion         |       | N-cadherin [36] |

Table 1: Main functional categories of membrane proteins and their computational design examples.

Hydrophobic packing of amino acids facing the nonpolar acyl chains of lipids in the membrane, polar residues were employed to build the interior part of the channel [8]. This hydrophobic-hydrophilic alternating pattern (H2N-(Leu-Ser-Ser-Leu-Leu-Ser-Leu-Leu-Ser-Leu)3-CONH2) was employed to form ion-conducting channels. It was also shown that replacing a serine with leucine (H2N-(Leu-Ser-Leu-Leu-Leu-Ser-Leu-Leu-Ser-Leu)3-CONH2) would result in proton-selective channels. In another study, it was shown that the overall orientation of the protein in the membrane can be controlled by the positioning of lysine residues [9]. More sophisticated, rational structure-based design of membrane spanning proteins has proven highly challenging and is far less developed than the application of these techniques to water-soluble proteins. This is due to the lack of solved structures. For example, fewer than 400 beta-barrel membrane proteins (bBMPs) have been deposited in the protein databank and only ~60 of these are nonhomologous [9]. This provides extremely limited coverage of the estimated 70,000 bBMPs in nature [10]. Membrane proteins are typically classified by either their structure or function. Each structure type presents different challenges to protein designers. We will first examine the challenges posed by different structure types and then review the major functional types and how these functions have been modified.

1.2. Structural Classification. Membrane protein transmembrane structures are broadly divided into two categories: beta-barrel and alpha-helical. Alpha helix structures are the major class found in both prokaryotic and eukaryotic membranes. They include voltage-gated ion channels, G protein-coupled receptors (GPCR), and ATP-binding cassette (ABC) transporters to name a few. Helix-based transmembrane structures can be further subdivided into helical bundles and helical barrels. Bundles are relatively well characterized, while helical barrels have received much less attention [11]. The design of transmembrane helix bundles has received significant attention, from the identification of a motifs that direct helix-helix assembly and packing [12], to ligand specificity and intracellular signaling engineering of GPCRs [13, 14]. Interactions between transmembrane helices and their surrounding polar groups result in more stable helices in comparison to soluble proteins. Taking advantage of this property, a membrane hemoprotein, ME1, was redesigned with a glycoporphin A to create a heme binding site [15]. Helix association propensity of amino acids in membrane proteins can be rank ordered (Gly > Ala > Val > Ile)—opposite of the trend observed for dimerization in water-soluble structures [16]. Stability and orientation of the transmembrane domain is thus mainly determined by noncovalent interactions, namely, van der Waals and electrostatic (including hydrogen bonds).

However, despite a long history, the design remains challenging. Even coiled coils, which have well-characterized packing rules and have been amenable to design in their water-soluble form for a couple of decades [17], are still largely refractory to the transmembrane design. A redesign of alpha-helical barrels using natural scaffolds is similarly underdeveloped. Examples of design include Cytoysin A—engineered to create a discrete lipid bilayer pore that can translocate DNA [18]. The design of active channels using short synthetic peptides is being energetically developed by the nanopore design community [19].

Membrane proteins are predominantly alpha-helical but transmembrane domains composed of beta-strands are also an important group. These are more commonly seen in bacteria, mitochondria, and chloroplast [20]. They commonly comprise a water-filled pore or solid core with polar side chains instead of a hydrophobic core. A potential term, EZβ, was determined for depth-related hydrophobicity to help redesign such proteins as they have very different structural features to other membrane proteins [20]. Most transmembrane beta-strand sequences alternate between polar and hydrophobic residues to create one hydrophobic face for the lipid membrane environment when folded. By adopting a beta-barrel structure, such that the hydrophilic face creates a polar interior channel, allowing transport of hydrophilic substrates. Transmembrane beta-barrels (TMB) range
from 8 strands to 24 strands [21] and can function as both monomers and oligomers. The TMB design has the potential for many applications due to the strong substrate specificity and multiple control points. The design work on TMBs underwent a major recent development in the form of de novo design of experimentally validated transmembrane barrels that do not require assembly machinery or chaperones ([22]). However, the study was limited to 8-strand barrels. A design using a natural structural scaffold has been used to tune oligomerization states by introducing lipid facing strand mutations at protein–protein interaction sites [23].

A design using natural scaffolds has been greatly constrained by the lack of solved structures that represent a tiny fraction of known structures in nature. High-quality structure prediction now potentially allows access to the majority of known transmembrane structures. Using a combination of established and emerging computational methods, it should be possible to investigate the mechanics of these structures at the atomic level and gain a much broader understanding of the features that drive function. This knowledge, coupled with high-throughput screening, removes the major constraint to a transmembrane protein design.

1.3. Functional Classification. Membrane proteins can be grouped into four major categories in the context of function and potential design. These categories are (1) receptors, (2) transporters, (3) enzymes, and (4) proteins involved in adhesion (see Figure 1).

1.3.1. Receptor Design. Receptors have received the most attention in the context of design. G protein-coupled receptors (GPCRs) are the most important membrane receptors in eukaryotes, comprising approximately 35% of all drug targets [45]. They are activated via a diverse array of extracellular molecules that give rise to conformational change of the receptor and release of intracellular signaling molecules. These conformational changes occur in a variety of time scales. For example, retinal isomerization in rhodopsin takes place in a fs-ns time span, while GPCR-G complex dissociation requires seconds to minutes to complete [46]. Thus, a successful design of membrane receptors depends on a thorough understanding of the receptor mechanism of action and the associated structural flexibility, rearrangements, and the time scale necessary for such changes. The role of conformational changes and structural dynamics in function of GPCRs and Type-I transmembrane receptors has been comprehensively discussed in the literature ([47, 48]). There have been many successful attempts to coopt GPCRs as part of synthetic signaling systems. Designer receptors exclusively activated by designer drugs (DREADDs) are the most widely used [14]. Early designs were developed using simple site-directed mutagenesis to modify receptor affinity and selectivity. Later approaches employed directed evolution [49], and structure-guided rational design was also used to make more radical changes to GPCR activation mechanisms [50]. These design methods involve changes at the individual amino acid level—mirroring evolution by natural selection. These methods are distinct from traditional protein engineering approaches that typically fuse unrelated domains to produce chimeric proteins with new functions. Examples of receptors engineered using the fusion approach include chimeric antigen receptors (CARs) [51], synthetic notch receptors (synNotch) [52], modular extracellular sensor architecture (MESA) [53], and transcriptional activation following a resting translocation (TANGO) receptors [54]. CAR and MESA receptors have also been refined and optimized using structural modeling and computational docking [55, 56]. However, due to the lack of solved structures with similar sequences, the development of more variants has been limited.

1.3.2. Transporter Design. Transporters are involved in the movement of a wide variety of molecules across membranes. These include ions, small molecules, and macromolecules such as proteins and nucleic acids. Transport is driven by facilitated diffusion or active transport. From a design and engineering perspective, ion channels have received by far the most attention. This has been largely motivated by a desire to better understand how they function, but also to probe cell biology and physiology [57]. Several studies have adapted ion channels for use as biosensors. For example, green fluorescent protein (GFP) was inserted/fused with an intracellular loop of a voltage gated sodium channel to produce fluorescent signal in response to membrane potential changes [58]. These biosensors are called sodium channel protein-based activity reporting constructs (SPARCs). A huge range of light-activated and ligand-activated ion channels/pumps have also been developed and deployed across a broad range of applications, most notably precipitating seminal advances in neuroscience [59]. In the past few years, there have been promising advances in sensing and sequencing using engineered protein nanopores. Specifically, nanopores have been designed to transport ions across the membrane ([30]) sequence homonucleotides with high fidelity (Van [60]), to read single proteins at single-amino acid resolution [61], and proteasome-nanopores to process single proteins [62].

1.3.3. Membrane Enzyme Design. Membrane enzymes catalyze important chemical processes such as methylation, drug metabolism, ATP synthesis, and hydrolysis. Understanding how enzymes catalyze reactions in membranes is still extremely limited, due in large part to the lack of structures. A systematic search of the literature yields little to no examples of the membrane enzyme design, beyond deletion of the P450 enzyme membrane anchor domain to produce a soluble/secreted version of the enzyme [35].

Challenges with designing membrane proteins are even more profound in the case of membrane enzymes, considering the competition between cavities for enzymatic activity (active site) and tight packing for thermodynamic stability of the protein [63] as creating a cavity could destabilize the enzyme due to the loss of packing. Thus, engineering membrane enzymes require fine-tuning the function and stability of the cavity. In a recent experimental and computational study, a cavity-filling mutation approach was employed to elucidate conformational changes regulating the function
of GlpG, where cavities were shown to be stabilized by interactions with membrane lipids and hence balance between stability and function [63]. This strategy could thus be used to engineer membrane enzymes.

1.3.4. Adhesion Protein Design. Cell adhesion molecules are primarily involved in mediating the interaction between a cell and its neighbors or the extracellular matrix. These proteins are crucial for maintaining tissue structure and

Figure 1: Schematic representation of membrane protein types and their structural examples. Receptor: tyrosine receptor (3NJP) ([37]) and GPCR (6KK1) ([38]). Transporter: potassium ion channel (6FL9) ([39]) and nanopore-hemolysin (7O1Q) [40]. Enzyme: prostaglandin (1PRH) ([41]) and ATP synthase (5TSJ) [42]. Adhesion: gap junction (7JJP) [43], E-cadherin (7STZ) [130]. Figures were made using Protein Imager [44].
function. An early design of adhesion proteins involved rational design to enhance affinity between integrin lymphocyte function-associated antigen-1 (LFA-1) and its ligand: intercellular adhesion molecule (ICAM-1) [64]. More recently, cadherin variants have been designed using computational modeling and validated experimentally as having affinity increased or decreased by up to 2 orders of magnitude [36]. Adhesion proteins often have large extracellular or cytoplasmic domains that can be purified independently of the transmembrane domain and their structures readily solved. Design involving these regions has therefore historically been much more accessible.

2. Structure-Guided Transmembrane Protein Design Tools

A number of useful tools are available to broaden and enrich our understanding of transmembrane protein mechanics from structural investigation. These will be critical for processing the vast number of new protein structures predicted by alpha-fold. Previous examples of redesigning membrane proteins with computational approaches have shown the potential for tuning interactions within membrane proteins and introducing new functionalities with established protocols [46]. In this section, recent advances in computational structural analysis and design with respect to membrane proteins are reviewed and emerging techniques that could be coopted are discussed.

2.1. Membrane Protein Molecular Dynamics Simulations.

Molecular dynamics (MD) simulations of membrane proteins are generally computationally more resource intensive due to the requirement for phospholipid molecules (lipid bilayer membrane). Until very recently, multimicrosecond MD studies on such systems were only feasible using a coarse-grained approach. Currently, microsecond-long MD simulations on large protein-membrane systems (10^5 - 10^6 atoms) are achievable in the span of weeks. In this section, previous MD studies on the function and structure of select membrane proteins are reviewed.

2.2. MD Simulation Setup for Membrane Protein Systems.

Modeling of membrane proteins requires protein embedding in a lipid bilayer. The VMD package [65] can be used to set up such systems for the NAMD simulation package [66]; nevertheless, its database for phospholipids is limited to only POPC and POPE. GROMACS [67] is also capable of building a membrane-protein system for MD simulations and requires force field parameters to be added for any given lipid molecule. However, CHARMM-GUI ([68]) is probably the most versatile, user friendly, and automated tool for MD simulation input preparation, not only for membrane and membrane protein systems but also for soluble proteins, polymers, nanodiscs, nanomaterials, micelles, monolayers, etc. To our knowledge, CHARMM-GUI provides the most comprehensive set of parameters for lipids required for modeling a variety of membranes and generates inputs for most of the widely used MD packages such as CHARMM, Amber, Gromacs, NAMD, and Desmond, using CHARMM, AMBER, and OPLS force fields. It also provides options for modification of N- and C-termini, protonation states of ionizable amino acids, set up disulfide bridges, and mutations during the early steps of the system setup. One particular challenge in the preparation of membrane protein systems is the orientations of proteins in the membrane. To address this issue, the OPM database [69] has been developed. This database includes more than 1,200 membrane proteins representing about 3,800 Protein Data Bank entries. Considering the hydrophobic and electrostatic interactions of the proteins with both water (polar) and lipid (nonpolar) environments, the PPM 2.0 and 3.0 (positioning of proteins in membranes) [70] optimize the positions of protein within the lipid bilayer. CHARMM-GUI uses this database and PPM server and automatically sets up the protein in the lipid bilayer of interest.

2.3. MD Studies of Membrane Proteins.

MD simulations on membrane proteins have been employed to study a wide variety of systems and properties such as structural analysis of the protein in lipid environment [71], linking structure to function [72], oligomeric state of membrane proteins [73], role of specific amino acids in the function [74], transport function through change of state [75], permeation [76], and selectivity mechanisms [77], to name a few. For a detailed evaluation of simulation protocols for membrane proteins, see reference [78]. In this section, we highlight a few notable membrane protein MD studies.

2.3.1. Case Study #1: Conformational Changes in ABC Transporter.

ABC transporter facilitates import-export of ligands across the membrane using adenosine triphosphate (ATP). The translocation of a ligand mediated by ABC requires a large-scale conformational change in the transmembrane domains (TMDS), converting the cytoplasm open-state to the periplasm open-state and vice-versa. The free energy required for such a process is provided by the ATP binding, hydrolysis, and dissociation of the product at the nucleotide-binding domains (NBDs).

ABC transporters have been the subject of several computational and computational-experimental studies ([75, 79, 80]). A comprehensive multimicrosecond-long MD (100 simulations of at least 500 ns length) study combined with electron paramagnetic resonance (EPR) were employed to elucidate the conformational transitions of the heterodimeric ABC exporter (TM287/TM288) [79]. It was shown that the protein’s state changes from the cytoplasm open-state to the periplasm open-state through an occluded intermediate. The occluded and periplasm open-state structures obtained with MD simulations were in excellent agreement with the corresponding crystal structures. The transport mechanism elucidated by the MD simulations involves cooperative motions of both transmembrane and nucleotide-binding domains. The occlusion state ensures that the cytoplasm and periplasm sides are not simultaneously open (Figures 2(a)–2(c)). This study reveals that while vital for the function, conformational changes required for switching between the states involves a considerable energy barrier to overcome.
Figure 2: Examples of membrane protein structures in lipid membranes. (a–c) Three states of ABC transporter (a) outward-facing, (b) occluded, and (c) inward-facing structures. (d) Structure of a gap junction (Cx50, PDB ID: 7JJP) embedded in two lipid bilayers. (e) Extracellular and (f) intracellular views of aquaporin-1 (AQP1) [81].
2.3.2. Case Study #2: Selectivity Mechanism in Aquaporins. Aquaporins and aquaglyceroporins are a family of integral membrane proteins that function in a tetrameric form and are responsible for the selective translocation of solutes across the cell membranes. In a comprehensive MD study, the selectivity and permeability of human AQP1 (hAQP1) (Figures 2(e) and 2(f)) and GlpF as members of the two aquaporin subfamilies for some small molecules including O$_2$, CO$_2$, NH$_3$, glycerol, urea, and water were examined [81]. In order to obtain the permeation barrier across the protein, the potential of mean force (PMF) was calculated using the umbrella sampling technique for all the solutes. For a detailed overview of the process, see references [82, 83]. From the MD results, a selectivity site for uncharged/nonpolar solutes in both aquaporin and aquaglyceroporin was identified (called ar/R). This region is narrow and hydrophilic in hAQP1. The interaction between water molecules and polar amino acids in this region can be replaced by more favorable interactions between the amino acids and a polar solute. However, larger molecules such as urea and glycerol are also filtered through steric clashes they cause with this narrow passage. The ar/R site is wider in GlpF and also includes more nonpolar amino acids, which makes GlpF permeable to a broader range of molecules such as polyols. It was suggested that the permeability of AQP1 is controlled by a rather simple filter mechanism, where large molecules are excluded through steric occlusion and the ar/R region is selective for small and hydrophilic molecules. GlpF on the other hand is less selective and can effectively translocate all molecules considered in the set except for urea. It was concluded that water-pore (and not solute-pore) interactions combined with steric effects control the selectivity and the energy barriers in AQP1 and GlpF.

2.3.3. Case Study #3: Oligomerization of NanC. It is estimated that ~25% of the cell membrane is occupied by membrane proteins [84]. Such a crowded environment promotes the oligomerization of membrane proteins; however, it has been shown that lateral interactions between the proteins are not solely driven by the lipid raft [85]. Oligomerization has been reported for many membrane proteins [86]. In fact, oligomerization motifs have been identified in several membrane proteins [87]. The lateral interactions and dimerization pattern of membrane proteins have been investigated for a bacterial outer membrane protein, NanC, responsible for transportation of sialic acid. With the aid of coarse-grained MD simulations and umbrella sampling technique, association energy landscape of two NanC monomers with four different orientational configurations were examined in a lipid bilayer [88]. The obtained energies range between -66 and -45 kJ/mol, depending on the shape complementarity of the proteins. It was also shown that the PMF shows nearly linear dependency to the buried surface area.

2.3.4. Case Study #4: Redesigning FhuA Nanochannel. Redesigning and engineering of proteins to enhance and/or alter their function is of great interest. For example, separation of racemic mixtures into pure enantiomers has been shown to be carried out by reengineered proteins. Specifically, a redesigned beta-barrel membrane protein, FhuAF4, derived from FhuA, was shown to be able to resolve D- and L-arginine mixtures [89]. Steered MD simulations (SMD) were used to find the molecular basis of the enantiopreference of the FhuAF4 for L-arginine. It was shown that while D-arginine interacts with two residues (S134 and T146) in the selectivity filter region of the FhuA4, L-arginine freely passes through the channel and only forms hydrogen bonds with S134. Thus, the obtained preference of FhuA4 for L-arginine is a result of its ability to form more interactions with D-arginine.

2.3.5. Case Study #5: Rational Design of Aerolysin Nanopores. In a recent study, a set of mutated aerolysin nanopores were rationally designed for detection of a range of biomolecules [90]. It was revealed that the selectivity and sensing function of the aerolysin is governed by electrostatic properties of the pore as well as the pore diameter. It was shown that the pore diameter could be adjusted (from 0.5 to 1.5 nm) by site-directed mutagenesis. This in turns, allows for detection of a broad range of molecules by aerolysin. To validate the models, MD simulations were carried out on the aerolysin pores, and the ion current across the pore was calculated. The electrostatic potential at 100 mV voltage was calculated and compared to those obtained experimentally. The measured pore current was shown to agree with MD prediction, hence proving the molecular explanations.

With the invention of alphafold and recent advances in both computer hardware and MD simulation algorithms, it is expected to witness a new wave of MD studies on the membrane proteins in the next few years.

3. Channel Characterization Tools

Channel proteins govern many physiological functions and allow for transportation of moieties necessary for cell survival. Depending on their function, the physical properties of membrane proteins vary and channels assume different states to control the transportation function. For example, permeation in ion channels is controlled via energy barrier regulation or gates along the pore. Of particular interest are the physicochemical properties of channels among the different families of membrane proteins. Some important properties include (but not limited to) channel width and length, hydrophobicity, hydrophathy, polarity of lining amino acids, lipophilicity, solubility, and water number density. Traditionally, the channels were examined by size-based approaches, which provides useful information about the permeation pathway. However, gates that are not based on steric occlusion of the pathway cannot be identified by this method. In this section, some of the available and more widely used tools for the analysis of membrane protein properties, including channel properties, will be introduced.

3.1. HOLE. Developed based on the first pore analysis study ([91]), the HOLE program was designed to analyze and visualize the holes along the ion channels and was introduced in 1996 ([92]). HOLE descriptions can be found at http://www
Computational protein design tools typically involve backbone sampling, scoring functions, sequence optimization, and functional site design [98]. Naturally existing protein function is dependent on the stability of its backbone, which dictates the shape. Optimizing sequences of naturally existing proteins generally try to maintain backbone conformation and select compatible amino acid side chains. These methods, most notably Rosetta and FoldX, include scoring functions which rank designs based on sequence-structure combinations, energy terms, and statistical-based methods from known structures. Applying these techniques to membrane protein design, it has been shown that synthetic membrane proteins can be generated that fold and have good stability ([31, 98, 99]). They are able to accurately predict membrane compatible orientations of amphipathic aromatic residues with positively charged residues on the cytoplasmic side [98]. Scoring membrane proteins has been particularly challenging due to the lack of previously solved transmembrane proteins in the Protein Database Bank. The huge
Figure 3: Output examples of membrane protein analysis tools. (a) Structure of 5-TH3 receptor (PDB ID: 4PIR) ([97]). (b) Radius profile produced by CHAP. (c) Pore represented by red spheres proportional to the diameter by PoreWalker. Blue and orange represent pore lining residues and atoms, respectively. (d) Cross-section of the protein (PoreWalker). (e, f) Side and top views of the channel. (g) Pore properties (free radius, hydrophobicity, and hydropathy calculated along the pore by MoleOnline). While visualization methods and outputs vary for these methods, they all demonstrate the nature of the channel and capture its important features. It is recommended that the user carefully reads the documentations of each program and decides which one may be most suitable for a particular system.
number of new predicted protein structures is now available may lead to significant improvements in accuracy.

4.1. Calculating the Free Energy of Binding. Free energy calculations can be used to accurately predict the contribution of individual amino acids to protein function and as such can be used to assist protein design. Thermodynamic integration (TI) and free energy perturbation (FEP) are currently among the most accurate computational techniques for free energy calculation. They have been consistently shown to match experimental binding free energy values with minimal error [100–102]. Drug–target, protein–protein [103], protein–DNA [104], and protein–peptide [105] interaction energies have all been accurately predicted using TI. Until recently, performing anything more than a very small number of TI calculations was prohibitively resource intensive. Usually, it was necessary to run very short simulations and severely limit the number of replicates. However, the recent arrival of GPU accelerated TI [106] has dramatically increased capabilities in this regard. It is now possible to test dozens of mutations over significant time scales and perform many replicates. Similarly, fast growth TI (FGTI) [107] has been shown to provide reliable results. Given these advances, it should now be possible to calculate protein–protein binding affinities between two membrane proteins, though this has yet to be demonstrated. Application of these highly accurate techniques to membrane proteins would allow high confidence characterization and design.

It is timely to emphasize that while binding affinities for protein–ligand and protein–protein systems can be experimentally determined for soluble proteins, in many cases, membrane protein systems do not possess soluble parts and lipid membranes are required for their stability and function and hence experimentally determining the binding affinity is not possible with current methods.

4.2. Design Using Artificial Neural Networks. Machine learning and particularly deep learning are leading to major breakthroughs across a broad range of sectors, particularly image classification and natural language processing. These successes have now been extended to solving major challenges in computational biology, alphafold being perhaps the most significant to date. Protein design using generative models such as generative adversarial networks (GANs) are still a very nascent field, with just a handful of experimental validated methodologies published [108–110]. It should be noted that none of these methodologies claim to generate new or improved proteins, but simply demonstrate that new polypeptide sequences can be generated that successfully fold and function on a par with their wild type counterparts. Generative models are typically trained on multiple sequence alignments of thousands of protein variants. Variants are typically evolutionarily related orthologs and/or from deep mutational scanning. The latter has the advantage that negative data can be used to train. It is certainly feasible that these techniques can be employed in the generation of new membrane proteins. Combining generative models with alphafold-derived structure prediction is likely to be especially powerful.

5. Integrating Computational Design with High-Throughput Functional Selection

5.1. Directed Evolution. Directed evolution is done by diversifying the native protein sequence and filtering the designs with the most promising function ([111]). Diversification methods can occur through random mutation or by targeted approaches (see Figures 4). Instead of mutating the entire protein sequence which generates an unmanageable number of variants, focused mutagenesis can be used to limit sequence space to library sizes below commercially synthesizable limits [112]. Knowledge of the structure is invaluable for determining which residues to vary and may also allow informed restriction of residue type—both of which help to minimize sequence space. In addition to some of the characterization methods mentioned above (HOLE, PoreWalker, etc.), there are other computational methods such the computed atlas of surface topography of proteins (CASTp) that have been used to guide rational design. This software identifies pockets and voids in 3D protein structures for ligand binding, DNA interaction, and enzymatic activity (Yajie [111]). Another computational tool, CAVER 3.0 calculates and clusters tunnels and channel pathways to characterize individual pathways, their time evolutions, and gating mechanisms [113]. A more in depth review describing many other computational tools for structural features was recently published [114]. An example of computational design-driving protein design is the acetylcholine receptor alpha-1 subunit, a ligand-gated ion channel which was redesigned to be a water-soluble protein allowing structurally characterized with NMR [115]. With the information obtained from these tools, protein design regions can be carefully selected according to morphological, geometrical, and chemical constraints. For instance, restricting design for only hydrophobic residues in membrane protein’s core would allow a manageable library size that is tailored for the specific application.

For membrane proteins specifically, directed evolution has been used to improve GPCR stability in detergents and increased expression levels—the two limitations affecting membrane protein studies [116]. A library of neurotensin receptors (NTRI) was created with error-prone PCR, induced within the inner membrane of E. coli, and introduced with fluorescent agonists to select variants with maintained binding affinities [116]. These were isolated with fluorescence-activated cell sorting (FACS) and were redesigned and screened for four rounds. Using computational tools to optimize sequence design of a poor resolution structure, statistical probabilities were calculated to limit residue mutations at each position [116]. This method involves ex mammalian systems where the proteins are purified and evolved in vitro. Since purification and production of membrane proteins remains a challenge, methods that can be done within mammalian systems offer a great advantage. Such an example is the viral evolution of genetically actuating sequence (VEGAS) system. This uses the RNA alphavirus Sindbis, to link viral vector replication with the diversification of the target protein under selective pressures
Their proof of concepts showed the ability to evolve tetracycline transactivator (tTa) to avoid inhibition by doxycycline, GPCR to activate endogenous signaling pathways, and nanobodies to selectively activate GPCRs to activate serum response [49]. Other mammalian-based methods such as TRACE [24] and CRISPR-X [117] rely on cytidine deaminase which introduces point mutations by deaminating cytosine to uracil resulting in three different mutagenic processes, known as somatic hypermutation (SHM). CRISPR-X uses dead Cas9 (dCas9) and a single guide RNA (sgRNA) with hairpin-binding sites to recruit MS2 proteins fused with AIDΔ to initiate hypermutation [117]. They were able to show this system evolved spectra-shifted variants of GFP and mutated PSMB5 against its inhibitor, bortezomib. However, this method is limited by where the sgRNA binds to; therefore, there are tight genomic windows where mutations can occur. TRACE was developed with this limitation in mind and inserts T7 promoter and uses a T7 RNAP linked with a cytidine deaminase to generate mutations 2 kb downstream of the promoter [118]. They showed this system to shift the fluorescent spectra of blue fluorescent protein to GFP and also evolve mitogen-activated protein kinase 1 (MAP2K1) to be resistant to pharmacological inhibitors [24]. Although these systems have not been used to evolve membrane proteins, we see no reason for future work not to be applied with these methods. Additional systems for directed evolution in mammalian cells that are not specific to membrane proteins are captured in this review [119].

After successful diversification, libraries of up to a billion gene variants can be generated, and high-throughput screening strategies can be used to select for functional variants ([111]). Functional assays that can be performed as high-throughput screens allow for tractable evaluation of these vast libraries (see Figures 5). To achieve the most accurate results, membrane proteins should be tested in an environment closest to that of the final application. Common selection methods are discussed below and are grouped based on similar mechanisms, i.e., fluorescence based, thermal scanning, and electrophysiology assays.

Fluorescence-based assays work by measuring the emission of fluorophore upon interaction with the protein target of interest. This method of detecting protein–protein interactions and ligand binding has high sensitivity, is noninvasive to the sample, and has relatively fast response times [120]. Functions of transporters such as ion channels have also been characterized using an indirect measure of membrane potential or ion concentration changes based on fluorescence changes [121]. In recent years, the introduction of voltage-sensitive dyes and ion-specific fluorescent probes has increased sensitivity and ability to screen variants in a high-throughput manner. Voltage-sensitive dyes are enzyme substrates designed to fluoresce upon enzymatic cleavage in cells [122]. Fluorescent probes have been developed to be ion-specific in order to quantify the channel activity [121]. One specific example is the FluxOR assay which uses thallium-affinity probes that are activated as extracellular thallium enters...
the cell through potassium channels and the signal is interpreted on a plate reader [123]. This serves as an indirect measure of potassium channel activity and is able to highlight inhibitors of the potassium channel. Fluorescence-based assays can also be used to characterize enzymatic membrane protein functions such as the one developed in 2019 [124]. Using apolipoprotein n-acyltransferase transferred onto biotin and click chemistry to conjugate a fluorescent group to the fatty acid, they were able to show product formation using fluorescent signals with the presence of enzymes compared to without. Previously, fluorescence polarization was used to study interactions between proteins and molecules. This is a binding assay which uses the change in emission light when a fluorescent molecule is unbound and bound to a larger molecule [125]. This has been used in the past to discover inhibitors and activators for GPCRs [126].

In addition to functional changes, designed protein stability is also important to evaluate. One method to measure stability is thermal scanning which measures protein unfolding with a fluorescent marker. To integrate this approach as a high-throughput screen, GFP-thermal shift assays were used to quantify ligand binding to solute carrier (SLC) transporters [127]. A method in 2009 used thermal scanning with lipophilic dyes to bind to the protein in the unfolded state and thus fluoresce at room temperatures [128]. This method can be used to indirectly measure the strength of protein–ligand binding based on which temperature the fluorescence signal peaks—indicating when the protein has unfolded. MaMTH luciferase interaction assay is a split-ubiquitin assay that uses a transcription factor release to drive expression of luciferase or GFP [129]. Another example of this is the TANGO assay which links GPCR activation to TEV protease cleavage of its cognate site and consequent release of a transcription factor to the nucleus to activate reporter gene expression [13]. To use either assay, the C terminus of the target protein gets fused with the recognition site (ubiquitin or the TEV cleavage site) and its corresponding transcription factor.

Electrophysiology methods allow for the tracking of ions through voltage-sensitive channels, and the historical gold standard is the patch-clamp method. More recently, automated electrophysiology assays have been developed that offer a promising alternative. IonWorks and its derivatives have shown success in single-cell recording along with population patches. The SyncroPatch 384PE is another device that can clamp 384 cells in parallel to record around 20,000 data points a day [103]. Taken altogether, these high-throughput methodologies offer the membrane protein

---

**Figure 5:** Graphical representation of the high-throughput methods. (a) Fluorescent-based assays that use a probe that fluoresces upon interaction with target protein of interest. (b) Thermal scanning assays use the natural denaturing of proteins to expose binding sites which fluorescent probes bind to. Using the change in fluorescence signal, the apparent temperature can be determined which provides insight on the stability of the designed protein. (c) Gene expression assays use a recognition site that is coupled with a transcription factor. Upon ligand binding, the recognition site triggers the transcription factor to activate downstream reporter genes. (d) Automated patch clamp methods allow for screening of 384 cells in parallel capturing individual electrical activity in a high-throughput manner.
designer an array of effective tools to screen variant libraries generated using structure-guided computational design.

6. Conclusions and Outlook

Transmembrane protein design holds much potential for a broad range of advancements in synthetic biology and biotechnology. Despite decades of development however, the field is still woefully underexplored. This has almost entirely been due to the paucity of solved structures, meaning that rules of assembly and packing have been difficult to elucidate. An absence of natural scaffolds and design rules meant that large swathes of membrane protein types have received little to no attention by protein designers. With the recent advent of high-quality structure prediction in the form of alphafold, this barrier has been very significantly lowered. The field is now poised to systematically investigate and elucidate structure and function at the atomistic level. To do this, a range of established and emerging tools are at the disposal of investigators in the field. Deploying these tools to dynamically model and characterize predicted structures will yield an unprecedented number of natural scaffolds for design. It is likely that the structural coverage engendered by systematic investigation of predicted structures will provide membrane protein assembly insights that will further catalyze design efforts in this area. It will also provide invaluable training data for emerging methods for protein design using artificial neural networks. With this advancement, future challenges that remain include developing systems for directed evolution within mammalian cells that include negative selection pressures. As synthetic biology and cell-based therapies begin to gain traction with regulators, protein designers are now very well positioned to explore this new frontier of cell–cell interaction.

Data Availability

Any data used to generate the figures and support the text of this review are available on request made by email to the corresponding author: michael.garton@utoronto.ca

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Shahin Sowlati-Hashjin and Aanshi Gandhi are equal authors.

Acknowledgments

The authors wish to acknowledge the Canadian Institute for Health Research (CIHR) funding that supported this work.

References

[1] M. S. Almén, K. J. V. Nordström, R. Fredriksson, and H. B. Schiöth, “Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin,” BMC Biology, vol. 7, no. 1, p. 50, 2009.
[2] D. Bausch-Fluck, U. Goldmann, S. Müller et al., “The in silico human surfaceome,” Proceedings of the National Academy of Sciences of the United States of America, vol. 115, no. 46, pp. E10988–E10997, 2018.
[3] P. Aguilar, J. Manuel, and J. Saven, “Computational design of membrane proteins,” Structure, vol. 20, no. 1, pp. 5–14, 2012.
[4] E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata, “Overcoming the challenges of membrane protein crystallography,” Current Opinion in Structural Biology, Carbohydrates and glycoconjugates/Biophysical methods, vol. 18, no. 5, pp. 581–586, 2008.
[5] A. A. Vorobjieva, “Principles and methods in computational membrane protein design,” Journal of Molecular Biology, vol. 433, no. 20, article 167154, 2021.
[6] J. Jumper, R. Evans, A. Pritzel et al., “Highly accurate protein structure prediction with alphafold,” Nature, vol. 596, no. 7873, pp. 583–589, 2021.
[7] D. Cyranoski, “How human embryonic stem cells sparked a revolution,” Nature, vol. 555, no. 7697, pp. 428–430, 2018.
[8] J. D. Lear, Z. R. Wasserman, and W. F. DeGrado, “Synthetic amphipathic peptide models for protein ion channels,” Science, vol. 240, no. 4856, pp. 1177–1181, 1988.
[9] W. Tian, M. Lin, K. Tang, J. Liang, and H. Naveed, “High-resolution structure prediction of β-barrel membrane proteins,” Proceedings of the National Academy of Sciences of the United States of America, vol. 115, no. 7, pp. 1511–1516, 2018.
[10] A. F. Roumia, K. D. Tsirigos, M. C. Theodoropoulou, I. A. Tamposis, S. J. Hamodrakas, and P. G. Bagos, “OMPdb: a global hub of beta-barrel outer membrane proteins,” Frontiers in Bioinformatics, vol. 1, article 646581, p. 1, 2021.
[11] A. Niitsu, J. W. Heal, K. Fauland, A. R. Thomson, and D. N. Woolfson, “Membrane-spanning α-helical barrels as tractable protein-design targets,” Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 372, no. 1726, p. 20160213, 2017.
[12] H. Yin, J. S. Slusky, B. W. Berger et al., “Computational design of peptides that target transmembrane helices,” Science, vol. 315, no. 5820, pp. 1817–1822, 2007.
[13] W. K. Kroese, M. F. Sassano, X.-P. Huang et al., “PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome,” Nature Structural & Molecular Biology, vol. 22, no. 5, pp. 362–369, 2015.
[14] Y. Pei, S. Dong, and B. L. Roth, “Generation of designer receptors exclusively activated by designer drugs (DREADDs) using directed molecular evolution,” Current Protocols in Neuroscience, vol. 50, no. 1, pp. 4.33.1–4.33.25, 2010.
[15] J. M. Cordova, P. L. Noack, S. A. Hilcove, J. D. Lear, and G. Ghirlanda, “Design of a functional membrane protein by engineering a heme-binding site in glycoporphin A,” Journal of the American Chemical Society, vol. 129, no. 3, pp. 512–518, 2007.
[16] A. Acharya, V. Rishi, and C. Vinson, “Stability of 100 homo and heterotypic coiled–Coila–a’ pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R),” Biochemistry, vol. 45, no. 38, pp. 11324–11332, 2006.
[17] J. Walshaw, J. M. Shipway, and D. N. Woolfson, “Guidelines for the assembly of novel coiled-coil structures: alpha-sheets
and alpha-cylinders,” Biochemical Society Symposium, vol. 68, no. 68, pp. 111–123, 2001.

[18] L. Franceschini, M. Soskine, A. Biesemans, and G. Maglia, “A nanopore machine promotes the vectorial transport of DNA across membranes,” Nature Communications, vol. 4, no. 1, p. 2415, 2013.

[19] M. Pazo, G. Salluce, I. Lostalé-Seijo et al., “Short oligoalanine helical peptides for supramolecular nanopore assembly and protein cytosolic delivery,” RSC Chemical Biology, vol. 2, no. 2, pp. 503–512, 2021.

[20] J. A. Stapleton, T. A. Whitehead, and V. Nanda, “Computational redesign of the lipid-facing surface of the outer membrane protein OmpA,” Proceedings of the National Academy of Sciences, vol. 112, no. 31, pp. 9632–9637, 2015.

[21] K. R. Vinothkumar and R. Henderson, “Structures of membrane proteins,” Quarterly Reviews of Biophysics, vol. 43, no. 1, pp. 65–158, 2010.

[22] A. A. Vorobieva, P. White, B. Liang et al., “De novo design of transmembrane β-barrels,” Science, vol. 371, no. 6531, p. eabc8182, 2021.

[23] H. Naveed, D. Jimenez-Morales, J. Tian, V. Pasupuleti, L. J. Kenney, and J. Liang, “Engineered oligomerization state of OmpF protein through computational design decouples oligomer dissociation from unfolding,” Journal of Molecular Biology, vol. 419, no. 1–2, pp. 89–101, 2012.

[24] K.-Y. M. Chen, D. Keri, and P. Barth, “Computational design of G protein-coupled receptor allosteric signal transductions,” Nature Chemical Biology, vol. 16, no. 1, pp. 77–86, 2020.

[25] M. Young, T. Dahoun, B. Sokrat et al., “Computational design of orthogonal membrane receptor-effector switches for rewiring signaling pathways,” Proceedings of the National Academy of Sciences of the United States of America, vol. 115, no. 27, pp. 7051–7056, 2018.

[26] P. Popov, Y. Peng, L. Shen et al., “Computational design of thermostabilizing point mutations for G protein-coupled receptors,” Elife, vol. 7, article e43729, 2018.

[27] N. B. Woodall, Z. Weinberg, J. Park et al., “De novo design of tyrosine and serine kinase-driven protein switches,” Nature Structural & Molecular Biology, vol. 28, no. 9, pp. 762–770, 2021.

[28] A. R. V. Voet, H. Noguchi, C. Addy et al., “Computational design of a self-assembling symmetrical β-propeller protein,” Proceedings of the National Academy of Sciences of the United States of America, vol. 111, no. 42, pp. 15102–15107, 2014.

[29] D. A. Silva, S. Yu, U. Y. Ulge et al., “De novo design of potent and selective mimics of IL-2 and IL-15,” Nature, vol. 565, no. 7738, pp. 186–191, 2019.

[30] C. Xu, P. Lu, T. M. Gamal El-Din et al., “Computational design of transmembrane pores,” Nature, vol. 585, no. 7823, pp. 129–134, 2020.

[31] P. Lu, D. Min, F. DiMaio et al., “Accurate computational design of multipass transmembrane proteins,” Science, vol. 359, no. 6379, pp. 1042–1046, 2018.

[32] K. Shimizu, B. Mijidjorj, M. Usami et al., “De novo design of a nanopore for single-molecule detection that incorporates a β-hairpin peptide,” Nature Nanotechnology, vol. 17, no. 1, pp. 67–75, 2022.

[33] C. N. Bedbrook, K. K. Yang, A. J. Rice, V. Gradinaru, and F. H. Arnold, “Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization,” PLoS Computational Biology, vol. 13, no. 10, article e1005786, 2017.

[34] I. V. Korendovych, A. Senes, Y. H. Kim et al., “De novo design and molecular assembly of a transmembrane diporphyrin-binding protein complex,” Journal of the American Chemical Society, vol. 132, no. 44, pp. 15516–15518, 2010.

[35] C. Von Wachenfeldt, T. H. Richardson, J. Cosme, and E. F. Johnson, “Microsomal P450 2C3 is expressed as a soluble dimer in _Escherichia coli_ following modifications of its N-terminus,” Archives of Biochemistry and Biophysics, vol. 339, no. 1, pp. 107–114, 1997.

[36] J. Vendome, K. Felsovalyi, H. Song et al., “Structural and energetic determinants of adhesive binding specificity in type I cadherins,” Proceedings of the National Academy of Sciences of the United States of America, vol. 111, no. 40, pp. E4175–E4184, 2014.

[37] C. Lu, L.-Z. Mi, M. J. Grey et al., “Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor,” Molecular and Cellular Biology, vol. 30, no. 22, pp. 5432–5443, 2010.

[38] Y. Xu, Y. Wang, Y. Wang et al., “Mutagenesis facilitated crystallization of GLP-1R,” IUCrJ, vol. 6, no. 6, pp. 996–1006, 2019.

[39] H. Hu, A. Nemezcz, C. Van Renterghem et al., “Crystal structures of a pentameric ion channel gated by alkaline PH show a widely open pore and identify a cavity for modulation,” Proceedings of the National Academy of Sciences, vol. 115, no. 17, pp. E3959–E3968, 2018.

[40] J. Wu, T. B. Blum, D. P. Farrell, F. DiMaio, J. P. Abrahams, and J. Luo, “Cryo-electron microscopy imaging of Alzheimer’s amyloid-beta 42 oligomer displayed on a functionally and structurally relevant scaffold,” Angewandte Chemie International Edition, vol. 60, no. 34, pp. 18680–18687, 2021.

[41] D. Picot, P. J. Loll, and R. Michael Garavito, “The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1,” Nature, vol. 367, no. 6460, pp. 243–249, 1994.

[42] R. B. Davies, C. Smits, A. S. W. Wong et al., “Cryo-EM analysis of a domain antibody bound rotary ATPase complex,” Journal of Structural Biology, vol. 197, no. 3, pp. 350–353, 2017.

[43] J. A. Flores, B. G. Haddad, K. A. Dolan et al., “Connexin-46/50 in a dynamic lipid environment resolved by CryoEM at 1.9 Å,” Nature Communications, vol. 11, no. 1, p. 4331, 2020.

[44] G. Tomasello, I. Armenia, and G. Molla, “The protein imager: A full-featured online molecular viewer interface with server-side HQ-rendering capabilities,” Bioinformatics, vol. 36, no. 9, pp. 2909–2911, 2020.

[45] K. Srimam and P. A. Insel, “G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs?”, Molecular Pharmacology, vol. 93, no. 4, pp. 251–258, 2018.

[46] G. G. Gregorio, M. Masureel, D. Hilger et al., “Single-molecule analysis of ligand efficacy in β2AR-G-protein activation,” Nature, vol. 547, no. 7661, pp. 68–73, 2017.

[47] L. M. P. Chatagnier, N. Leloup, B. J. C. Janssen, and B. J. C. Janssen, “Structural perspectives on extracellular recognition and conformational changes of several type-I transmembrane receptors,” Frontiers in Molecular Biosciences, vol. 7, 2020.
antigen-1-dependent adhesion,” *The Journal of Biological Chemistry*, vol. 281, no. 8, pp. 5042–5049, 2006.

[65] W. Humphrey, A. Dalke, and K. Schulten, “VMD: visual molecular dynamics,” *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[66] J. C. Phillips, D. J. Hardy, J. D. C. Maia et al., “Scalable molecular dynamics on CPU and GPU architectures with NAMD,” *The Journal of Chemical Physics*, vol. 153, no. 4, article 044130, 2020.

[67] M. J. Abraham, T. Murtola, R. Schulz et al., “GROMACS: high performance molecular simulations through multilevel parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1–2, pp. 19–25, 2015.

[68] J. Lee, X. Cheng, J. M. Swails et al., “CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMP simulations using the CHARMM36 additive force field,” *Journal of Chemical Theory and Computation*, vol. 12, no. 1, pp. 405–413, 2016.

[69] M. A. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, “OPM: orientations of proteins in membranes database,” *Bioinformatics*, vol. 22, no. 5, pp. 623–625, 2006.

[70] M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, and A. L. Lomize, “OPM database and PPM web server: resources for positioning of proteins in membranes,” *Nucleic Acids Research*, vol. 40, no. D1, pp. D370–D376, 2012.

[71] K. Goossens and H. De Winter, “Molecular dynamics simulations of membrane proteins: an overview,” *Journal of Chemical Information and Modeling*, vol. 58, no. 11, pp. 2193–2202, 2018.

[72] S. Shi, C. Pang, S. Ren et al., “Molecular dynamics simulation of TMEM16A channel: linking structure with gating,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1864, no. 1, pp. 183777, 2022.

[73] A. Ardalan, S. Sowlati-Hashjin, H. Oduwoye et al., “Biphasic proton transport mechanism for uncoupling proteins,” *The Journal of Physical Chemistry B*, vol. 125, no. 32, pp. 9130–9144, 2021.

[74] A. Ardalan, S. Sowlati-Hashjin, S. O. Uwumarenogie et al., “Functional oligomeric forms of uncoupling protein 2: strong evidence for asymmetry in protein and lipid bilayer systems,” *The Journal of Physical Chemistry B*, vol. 125, no. 1, pp. 169–183, 2021.

[75] C. A. Hutter, M. H. Timachi, L. M. Hürlimann et al., “The extracellular gate shapes the energy profile of an ABC exporter,” *Nature Communications*, vol. 10, no. 1, pp. 2260, 2019.

[76] J. Wong-ekkabut and M. Kerttunen, “Molecular dynamics simulation of water permeation through the alphahemolysin channel,” *Journal of Biological Physics*, vol. 42, no. 1, pp. 133–146, 2016.

[77] P. Pongprayoon, O. Beckstein, C. L. Wei, and M. S. P. Sansom, “Simulations of anion transport through OprP reveal the molecular basis for high affinity and selectivity for phosphate,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 51, pp. 21614–21618, 2009.

[78] J. Wong-ekkabut and M. Kerttunen, “Assessment of common simulation protocols for simulations of nanopores, membrane proteins, and channels,” *Journal of Chemical Theory and Computation*, vol. 8, no. 8, pp. 2905–2911, 2012.

[79] H. Göddeke, M. Hadi Timachi, C. A. J. Hutter et al., “Atomistic mechanism of large-scale conformational
transition in a heterodimeric ABC exporter,” Journal of the American Chemical Society, vol. 140, no. 13, pp. 4543–4551, 2018.

[80] J.-F. St-Pierre, A. Bunker, T. Róg, M. Karttunen, and N. Mousseau, “Molecular dynamics simulations of the bacterial ABC transporter SAV1886 in the closed form,” The Journal of Physical Chemistry B, vol. 116, no. 9, pp. 2934–2942, 2012.

[81] J. S. Hub and B. L. de Groot, “Mechanism of selectivity in aquaporins and aquaglyceroporins,” Proceedings of the National Academy of Sciences of the United States of America, vol. 105, no. 4, pp. 1198–1203, 2008.

[82] J. Küstner, “Umbrella sampling,” WIREs Computational Molecular Science, vol. 1, no. 6, pp. 932–942, 2011.

[83] Q. Liao, “Chapter four-enhanced sampling and free energy calculations for protein simulations,” in Progress in Molecular Biology and Translational Science, B. Strodel and B. Barz, Eds., vol. 170 of Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly, pp. 177–213, Academic Press, 2020.

[84] A. D. Dupuy and D. M. Engelman, “Protein area occupancy at the center of the red blood cell membrane,” Proceedings of the National Academy of Sciences of the United States of America, vol. 105, no. 8, pp. 2848–2852, 2008.

[85] J. Spector, S. Zakharov, Y. Lill, O. Sharma, W. A. Cramer, and K. Ritchie, “Mobility of BtuB and OmpF in the _Escherichia coli_ outer membrane: implications for dynamic formation of a translocon complex,” Biophysical Journal, vol. 99, no. 12, pp. 3880–3886, 2010.

[86] C. Govaerts, “Lipids can make them stick together,” Trends in Biochemical Sciences, vol. 42, no. 5, pp. 329–330, 2017.

[87] P. J. Bond and M. S. P. Sansom, “Insertion and assembly of membrane proteins via simulation,” Journal of the American Chemical Society, vol. 128, no. 8, pp. 2697–2704, 2006.

[88] T. A. Dunton, J. E. Goose, D. J. Gavaghan, M. S. P. Sansom, and J. M. Osborne, “The free energy landscape of dimerization of a membrane protein, NanC,” PLoS Computational Biology, vol. 10, no. 1, article e1003417, 2014.

[89] D. Anand, G. V. Dhoke, J. Gehrmann et al., “Chiral separation of l/1-arginine with whole cells through an engineered FhuA nanochannel,” Chemical Communications, vol. 55, no. 38, pp. 5431–5434, 2019.

[90] C. Cao, N. Cirauqui, M. J. Marcaida et al., “Single-molecule sensing of peptides and nucleic acids by engineered amyloid nanopores,” Nature Communications, vol. 10, no. 1, p. 4918, 2019.

[91] O. S. Smart, J. M. Goodfellow, and B. A. Wallace, “The pore dimensions of gramicidin A,” Biophysical Journal, vol. 65, no. 6, pp. 2455–2460, 1993.

[92] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, and M. S. Sansom, “HOLE: a program for the analysis of the pore dimensions of ion channel structural models,” Journal of Molecular Graphics, vol. 14, no. 6, pp. 354–360, 1996.

[93] M. Pellegrini-Calace, T. Maiwald, and J. M. Thornton, “PorEWalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure,” PLoS Computational Biology, vol. 5, no. 7, article e1000440, 2009.

[94] L. Pravda, D. Sehnal, D. Toušek et al., “MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update),” Nucleic Acids Research, vol. 46, no. W1, pp. W368–W373, 2018.

[95] T. Nugent and D. T. Jones, “Membrane protein orientation and refinement using a knowledge-based statistical potential,” BMC Bioinformatics, vol. 14, no. 1, p. 276, 2013.

[96] G. Klesse, S. Rao, M. S. P. Sansom, and S. J. Tucker, “CHAP: a versatile tool for the structural and functional annotation of ion channel pores,” Journal of Molecular Biology, vol. 431, no. 17, pp. 3353–3365, 2019.

[97] G. Hassaine, C. Deluz, L. Grasso et al., “X-ray structure of the mouse serotonin 5-HT1 receptor,” Nature, vol. 512, no. 7514, pp. 276–281, 2014.

[98] X. Pan and T. Kortemme, “Recent advances in de novo protein design: principles, methods, and applications,” Journal of Biological Chemistry, vol. 296, article 100558, 2021.

[99] M. Mravic, J. L. Thomaston, M. Tucker, P. E. Solomon, L. Liu, and W. F. DeGrado, “Packing of apolar side chains enables accurate design of highly stable membrane proteins,” Science, vol. 363, no. 6434, pp. 1418–1423, 2019.

[100] I. V. Khavrutskii and A. Wallqvist, “Improved binding free energy predictions from single-reference thermodynamic integration augmented with Hamiltonian replica exchange,” Journal of Chemical Theory and Computation, vol. 7, no. 9, pp. 3001–3011, 2011.

[101] M. Lawrenz, R. Baron, Y. Wang, and J. Andrew McCammon, “Independent-trajectory thermodynamic integration: a practical guide to protein-drug binding free energy calculations using distributed computing,” Methods in Molecular Biology, vol. 819, pp. 469–486, 2012.

[102] P.-C. Su and M. E. Johnson, “Evaluating thermodynamic integration performance of the new amber molecular dynamics package and assess potential halogen bonds of enoyl-ACP reductase (FabI) benzimidazole inhibitors,” Journal of Computation Chemistry, vol. 37, no. 9, pp. 836–847, 2016.

[103] D. J. Cole, E. Rajendra, M. Roberts-Thomson et al., “Interrogation of the protein–protein interactions between human BRCA2 BRC repeats and RAD51 reveals atomistic determinants of affinity,” PLoS Computational Biology, vol. 7, no. 7, article e1002096, 2011.

[104] F. R. Beierlein, G. Geoff Kneale, and T. Clark, “Predicting the effects of basepair mutations in DNA-protein complexes by thermodynamic integration,” Biophysical Journal, vol. 101, no. 5, pp. 1130–1138, 2011.

[105] Panel, F. V. Nicolas, E. J. Fuentes, and T. Simonson, “Accurate PDZ/peptide binding specificity with additive and polarizable free energy simulations,” Biophysical Journal, vol. 114, no. 5, pp. 1091–1102, 2018.

[106] T.-S. Lee, H. Yuan, B. Sherborne, Z. Guo, and D. M. York, “Toward fast and accurate binding affinity prediction with PmmmdGTI: an efficient implementation of GPU-accelerated thermodynamic integration,” Journal of Chemical Theory and Computation, vol. 13, no. 7, pp. 3077–3084, 2017.

[107] V. Gapsys, S. Michielsens, D. Seeliger, and B. L. de Groot, “Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan,” Angewandte Chemie International Edition, vol. 55, no. 26, pp. 7364–7368, 2016.

[108] F. Grisoni, C. S. Neuhaus, G. Gabernet, A. T. Müller, J. A. Hiss, and G. Schneider, “Designing anticancer peptides by constructive machine learning,” ChemMedChem, vol. 13, no. 13, pp. 1300–1302, 2018.

[109] D. Repecka, V. Jauniskis, L. Karpus et al., “Expanding functional protein sequence spaces using generative adversarial
networks,” *Nature Machine Intelligence*, vol. 3, no. 4, pp. 324–333, 2021.

[110] A. Strokach, D. Becerra, C. Corbi-Verge, A. Perez-Riba, and P. M. Kim, “Fast and flexible protein design using deep graph neural networks,” *Cell Systems*, vol. 11, no. 4, pp. 402–411.e4, 2020.

[111] Y. Wang, P. Xue, M. Cao, Y. Tianhao, S. T. Lane, and H. Zhao, “Directed evolution: methodologies and applications,” *Chemical Reviews*, vol. 121, no. 20, pp. 12384–12444, 2021.

[112] P.-S. Huang, S. E. Boyken, and D. Baker, “The coming of age of _de novo_ protein design,” *Nature*, vol. 537, no. 7620, pp. 320–327, 2016.

[113] E. Chovancova, A. Pavelka, P. Benes et al., “CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures,” *PLoS Computational Biology*, vol. 8, no. 10, article e1002708, 2012.

[114] C. E. Sequeiros-Borja, B. Surpeta, and J. Brezovsky, “Recent advances in user-friendly computational tools to engineer protein function,” *Briefings in Bioinformatics*, vol. 22, no. 3, article bbaa150, 2021.

[115] T. Cui, D. Mowrey, V. Bondarenko et al., “NMR structure and dynamics of a designed water-soluble transmembrane domain of nicotinic acetylcholine receptor,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1818, no. 3, pp. 617–626, 2012.

[116] C. A. Sarkar, I. Dodevski, M. Kenig et al., “Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 39, pp. 14808–14813, 2008.

[117] G. T. Hess, L. Frésard, K. Han et al., “Directed evolution using DCas9-targeted somatic hypermutation in mammalian cells,” *Nature Methods*, vol. 13, no. 12, pp. 1036–1042, 2016.

[118] H. Chen, S. Liu, S. Padula et al., “Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor,” *Nature Biotechnology*, vol. 38, no. 2, pp. 165–168, 2020.

[119] S. J. Hendel and M. D. Shoulders, “Directed evolution in mammalian cells,” *Nature Methods*, vol. 18, no. 4, pp. 346–357, 2021.

[120] X. Fang, Y. Zheng, Y. Duan, Y. Liu, and W. Zhong, “Recent advances in design of fluorescence-based assays for high-throughput screening,” *Analytical Chemistry*, vol. 91, no. 1, pp. 482–504, 2019.

[121] H.-b. Yu, M. Li, W.-p. Wang, and X.-l. Wang, “High throughput screening technologies for ion channels,” *Acta Pharmacologica Sinica*, vol. 37, no. 1, pp. 34–43, 2016.

[122] S. Dube, H. Dube, N. B. Green et al., “In vivo delivery and activation of masked fluorogenic hydrolase substrates by endogenous hydrolases in C. elegans,” *Chembiochem*, vol. 18, no. 18, pp. 1807–1813, 2017.

[123] D. W. Beacham, T. Blackmer, M. O’Grady, and G. T. Hanson, “Cell-based potassium ion channel screening using the FluxOR™ assay,” *Journal of Biomolecular Screening*, vol. 15, no. 4, pp. 441–446, 2010.

[124] K. Nozeret, A. Boucharlat, F. Agou, and N. Buddelmeijer, “A sensitive fluorescence-based assay to monitor enzymatic activity of the essential integral membrane protein apolipo-protein N-acyltransferase (Lnt),” *Scientific Reports*, vol. 9, no. 1, p. 15978, 2019.