ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2001) 219–234

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Towards the design and computational characterization of a membrane protein

Christin T. Choma [a,*], D. Peter Tieleman [b,*,1], David Cregut [c], Luis Serrano [c], Herman J.C. Berendsen [d]

[a] *Cogswell Lab, Department of Chemistry, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180-3590, USA*
[b] *Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada T2N 1N4*
[c] *EMBL, Meyerhofstrasse 1, Heidelberg D-69117, Germany*
[d] *Department of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*

Received 15 December 2000; received in revised form 8 February 2001; accepted 8 February 2001

Abstract

The design of a transmembrane four-helix bundle is described. We start with an idealized four-helix bundle geometry, then use statistical information to build a plausible transmembrane bundle. Appropriate residues are chosen using database knowledge on the sequences of membrane helices and loops, then the packing of the bundle core is optimized, and favorable side chain rotamers from rotamer libraries are selected. Next, we use explicit physical knowledge from biomolecular simulation force fields and molecular dynamics simulations to test whether the designed structure is physically possible. These procedures test whether the designed protein will indeed be α-helical, well packed and stable over a time scale of several nanoseconds in a realistic lipid bilayer environment. We then test a modeling approach that does not include sophisticated database knowledge about proteins, but rather relies on applying our knowledge of the physics that governs protein motions. This independent validation of the design is based on simulated annealing and restrained molecular dynamics simulation in vacuo, comparable to procedures used to refine NMR and X-ray structures.

Keywords: De novo design; Coiled coil; Molecular dynamics; Simulated annealing

1. Introduction

Over the past decade, a variety of small globular, water-soluble proteins have been designed from first principles, with varying degrees of success. These include two-helix [1], three-helix [2] and four-helix bundles [3–5], β-hairpins [6] and mixed α/β structures [7]. Additionally, helical bundle proteins that incorporate prosthetic groups [8,9] or surface catalytic sites [10] have been successfully designed. In the field of synthetic membrane proteins, membrane-soluble peptides have been designed with the purpose of studying peptide–peptide association [11,12], hydrophobicity scales [13,14], helical propensity [15] and helix stability [16,17]. Synthetic peptides have also been designed that aggregate to form ion channels [18–20], inspired by naturally occurring channel forming peptides and viral ion channels such as the Influenza M2 channel [21]. However, compared to water-soluble proteins, our understanding of the factors and forces that control transmembrane helix insertion, stability and association is still incomplete. Consequently, the design of multi-helix transmembrane proteins aimed at testing our overall understanding of membrane protein structure remains a largely untouched area. We do know, however, that the basic requirement for a polypeptide to form a helix in a membrane and traverse the membrane is simply that the sequence be sufficiently hydrophobic [16,22]. We here present our efforts to design a self-associating transmembrane four-helix bundle protein, which we call Membun. The goals of the whole project are three-fold: (i) to see if there is sufficient information available to permit the design of a plausible transmembrane four-helix bundle; (ii) to investigate how molecular dynamics can be developed to

Abbreviations: DMPC, dimyristoylphosphatidylcholine; MALDI-TOF MS, matrix assisted laser desorption ionization time of flight mass spectrometry; RMSD, root mean square deviation; ROP, repressor of primer; SMD, sparse matrix driven algorithm; SA/MD, simulated annealing and molecular dynamics; SPC, simple point charge

*Corresponding authors. Tel.: +1-518-276-2804; fax: +1-518-276-4887.
E-mail addresses: chomac@rpi.edu (C.T. Choma), chomac@rpi.edu, tieleman@ucalgary.ca (D.P. Tieleman).
[1] Tel.: +1-403-220-5434; fax: +1-403-289-9311. The corresponding authors contributed equally to this work.

1093-3263/01/$ – see front matter
PII: S1093-3263(01)00111-5

aid the process of designing membrane proteins; and (iii) to develop strategies for the facile solid phase peptide synthesis and purification of long, hydrophobic peptides. The first two goals are addressed in this paper, the third goal has been addressed previously [73,74].

2. Methods

2.1. Template design

A poly-Ala parallel two-helix coiled-coil template was generated mathematically, essentially as described before [4]. The two helices were designated as 1A and 1B, and the *z*-axis was defined as running down the center of the 1A helix. Using symmetry operators in the InsightII program (Biosym), the 1B helix was rotated 180° along the *y*-axis and then translated a given distance (for example, 8.5 Å) along the *x*-axis. The helices now faced each other across their 'a' and 'd' positions, and were antiparallel. Helix 1A was copied again (2B), rotated 180° along *x* and translated 8.5 Å along the *y*-axis. Helix 1A was copied one more time (2A), rotated 180° along *z*, translated 8.5 Å along the *x* and 8.5 Å along the *y*-axis. Helices 1A and 2A (parallel to each other) were then both translated $(4 \times (-1.5\,\text{Å})) = -6\,\text{Å}$ along the *z*-axis in order to align, in the same plane, the 'd' positions of helices 1A and 2A with the 'a' positions of helices 1B and 2B. This packing of 'a' and 'd' residues is characteristic of antiparallel bundles. The *x*, *y* distances between the centers of helices in the four-helix arrangement was 8.5 Å, and the diagonal distance between helices 1A and 2A (respectively, 1B and 2B) centers was 12 Å. The procedure was repeated to generate bundles with helices separated by 13–15 Å.

2.2. Loop modeling

We used the DgLoop option of the WhatIf [23] program to identify short loop segments in proteins that had the correct geometry for connecting the C-α position of the last residue in helix A with the C-α of the first residue in helix B. Different loop lengths were used and we found that the shortest connection involved four residues, with the best match being the loop in repressor of primer (ROP). This loop was pasted and annealed to helices A and B.

2.3. Rotamer analysis

WhatIf was used to check the amino acids in our sequence for the most frequently-occurring rotamers found in α-helices. We used the Scan3D and Sequen options to search the preferred rotamers of amino acids in the middle of a 13-residue helix. We then used the Scnsts option to do a statistical analysis of the different χ angles for this residue at this position. These numbers were compared with the rotamer preferences found in the Ponder and Richards [24] library.

2.4. Side chain packing and rotamer compatibility

To optimize the packing of the selected residues, we used the sparse matrix driven (SMD) algorithm [25]. This algorithm explores all allowable rotamers for every amino acid in the sequence on a fixed backbone and employs a rotamer library extended from that of Ponder and Richards [24]. SMD searches the protein structure for pairs of side chain–side chain interactions that have an interaction energy above a chosen threshold (generally taken as 1 kcal/mol). The combinatorial problem of searching for ideal side chain rotamers is restricted to these residues, thus permitting a quick solution close to the optimal conformation for each residue. Strongly interacting residues are grouped into clusters, then an exhaustive search is performed on the residues in each cluster. The combination of these single cluster solutions leads to local minima, and then a global solution is approached by introducing weighted information from neighboring clusters. Global information from neighboring clusters is fed back to the individual clusters, and the rotamers are duly changed (using only rotamers from the rotamer library) to reflect that interaction energy. The process is repeated in an iterative manner until no improvement in the overall conformational energy is obtained between two successive steps.

2.5. Energy minimization

Minimization was performed using the AMBER 4.1 program [26], initially with the backbone fixed, and then repeated on the fully unrestrained structure. No major problems with the structure were noted. The solvent-accessible surface of the structure was checked by rolling a 1.4 Å radius ball (the radius of a water molecule) over the entire minimized structure to check for poor side chain packing.

2.6. Simulated annealing and restrained molecular dynamics

The simulated annealing protocol used the X-Plor program with the CHARMM param19 force field [27] based on the method described by Nilges and Brunger [28,29]. Similar protocols have been used in studies on membrane channel models by Sansom et al. [30]. The protocol consists of four steps.

1. Generate a C-α template of the structure. Initially, all atoms in a residue are superimposed on the C-α atom. This C-α template embodies most of the assumptions made about the structure. We used two different templates for both ROP and Membun. The first was just the C-α co-ordinates of the crystal structure for ROP and the model for Membun (the SA1 systems in Table 1). This can be considered an upper limit on the accuracy of this method. For the second template, we generated an antiparallel four-helix bundle of straight helices, assuming an ideal geometry of 1.5 Å helical rise per residue and

Table 1
Overview of the simulations[a]

System name	Description
MEMBUN	The Membun design in a DMPC bilayer
ROP	Crystal structure of ROP in water
ROPlip	Crystal structure of ROP in a DMPC bilayer
ROPSA1	An SA/MD model based on a template consisting of the actual positions of the C-α atoms in the crystal structure of ROP in water
ROPSA2	An SA/MD model based on a template consisting of the positions of the C-α atoms of four straight, uncoiled, α-helices
MembunSA1	An SA/MD model based on a template of the C-α atoms of the original design of MEMBUN, in a DMPC bilayer
MembunSA2	An SA/MD model based on a template of the C-α atoms of four straight α-helices of MEMBUN in a DMPC bilayer

[a] The first group describes simulations based on the design of Membun and on the crystal structure of ROP, and the second group describes simulations based on the structures resulting from the simulated annealing modeling procedure.

an inter-helix distance of 8.4 Å. The four helices were rotated such that the hydrophobic side for ROP and the H-bond network side for Membun face inward (the SA2 systems in Table 1). The vertical orientation in both cases is determined by the inter-helical loops. In the case of Membun, the disulfide bond is present during the entire modeling phase.

2. The C-α atoms are kept fixed during a simulated annealing run starting at 1000 K. Initially, the force constants for all bonded interactions are scaled down to a fraction of their normal strength. In the course of a 2 ps run at 1000 K, first the bonds and angles, and with a small lag, the dihedrals and improper dihedrals, are increased in strength until they reach their normal values. When the bonded terms have reached their normal values, a small repulsive van der Waals term is introduced and the system is cooled to 500 K. The resulting structure is energy minimized. This process is repeated five times, resulting in five structures.
3. Harmonic restraints are imposed on all C-α atoms of the helices in the five structures. These restraints are gradually relaxed as the system is cooled from 500 to 300 K. A weak restraining potential prevents the helices from moving too far apart. In addition, backbone hydrogen bonds are strengthened by weak asymmetric restraints. Electrostatic interactions are also slowly introduced during this stage. Repeating this procedure five times created an ensemble of 25 structures.
4. The 25 structures are simulated in vacuum for 5 ps with the full potential form, including restraints on inter-helix distance, restraints on the backbone helical hydrogen bonds, and using a distance-dependent dielectric constant. The 25 resulting structures are averaged by best fitting them all on one structure and calculating the average position of the co-ordinates. The resulting structure was energy minimized and used as the starting point for a long MD run.

2.7. Molecular dynamics

For ROP we used the 1.7 Å resolution crystal structure [31] (simulations ROPlip, ROP) or SA models (simulations ROPSA1, ROPSA2), and for Membun we used the designed model as described above (simulation MEMBUN) or SA models (MembunSA1, MembunSA2). The initial dimyristoylphosphatidylcholine (DMPC) bilayer was constructed from an equilibrated bilayer of 64 DPPC lipids [32]. The last two carbons of both tails were removed and the two bilayer leaflets were translated 0.4 nm toward each other. The resulting DMPC bilayer was simulated for 50 ps, after which a cylindrical space was created by removing 11 lipids per leaflet and applying a repulsive cylindrical potential. Membun was placed in the hole, with the Tyr and Trp residues flanking the core of the protein positioned at the acyl chain–glycerol interface. Water was re-added to the system using standard procedures. Two sodium ions were added to the system, replacing water molecules at the positions with the lowest Coulomb potential, thus, making the system electrically neutral. The resulting system, containing 10,786 atoms, was energy minimized and simulated for 50 ps with harmonic position restraints on the protein, followed by a production run of 2 ns. The same bilayer structure was used as the starting point for ROPlip, the simulation of ROP in DMPC. The ROP crystal structure was superimposed (backbone atoms) on the Membun structure (a total of 10,745 atoms). This system was energy minimized and simulated for 50 ps with harmonic position restraints on the heavy atoms in the protein. This restrained the protein co-ordinates while allowing the lipids and water molecules to relax; the system was then simulated for an additional 2 ns without restraints. The models obtained from simulated annealing were used in additional molecular dynamics simulations. For these simulations, a structure from the simulation of the Membun design in DMPC was used as a starting point. In each case, a 50 ps position restraint run was conducted before the production runs (2 ns for MembunSA1, 7 ns for MembunSA2).

All simulations were done with the Gromacs suite of programs [33].[2] The GROMOS87 force field, with minor changes, was used for the proteins. The lipid parameters were taken from Berger et al. [34]. In all bilayer simulations, the same simulation parameters and force field were used as in previous simulations (of alamethicin), except that the long-range cut-off was only 1.4 nm due to the smaller

[2] URL: http://md.chem.rug.nl/~gmx.

box size in the current simulations [35]. The ROP crystal structure and simulated annealing models were solvated in simple point charge (SPC) water [36], with eight Na^+ counter ions to make the system neutral (18,497 atoms, 2 ns for ROP in water; 16,511 atoms, 1 ns for ROPSA1; 19,487 atoms, 1.8 ns for ROPSA2). For ROP in water, a twin-range cut-off of 1.0/1.8 nm was used. In all simulations, bonds were constrained using LINCS [37], with a 2 fs time step. Temperature and pressure was controlled using the weak coupling method, with $T = 300\,\mathrm{K}$, $\tau = 0.1\,\mathrm{ps}$, water, protein, and lipids separately; $p = 1\,\mathrm{bar}$, $\tau = 1.0\,\mathrm{ps}$, isotropic for solution simulations, x, y and $z = 1\,\mathrm{bar}$ independently for bilayers simulations [38].

3. Results

3.1. General design considerations

The strategy for the design of the transmembrane bundle reflected the fact that there are both essential differences and similarities in the factors affecting membrane protein and water-soluble protein stability and folding [39]. The most important difference between the two classes of proteins is that unlike in water, helices are intrinsically stable in membranes. This is because the polar amide bonds of the H-bonded backbone are shielded from the non-polar environment. Another essential difference is that helical content in membranes is relatively indifferent to amino acid sequence as long as there is a high proportion of hydrophobes. A third difference is that the exteriors of membrane proteins are hydrophobic, whereas the surfaces of water-soluble proteins are hydrophilic. However, the two classes of proteins are similar in that the interiors of all proteins are comparable in hydrophobicity and packing density [40]. Our overall design strategy was based on procedures developed for the design of water-soluble coiled coils and bundles [41] while taking into account known differences between water- and membrane-soluble proteins. The four-helix bundle was designed to be a homodimer, with the two helices of the monomer having different sequences.

Water-soluble coiled coils are characterized by a seven-residue (heptad) repeat designated a-b-c-d-e-f-g, where the residues at the 'a' and 'd' positions are generally hydrophobic [42] and the other residues are relatively hydrophilic. Interactions between helices occur predominantly at the 'a' and 'd' residues, and conservative substitutions at these positions can critically affect the oligomerization and stability of the helices [43]. The sensitivity of membrane proteins to substitutions at these positions is less clear, but there is evidence from studies on glycophorin [44] and detergent-soluble versions of GCN4 [12,45,46] that the 'a' and 'd' positions likewise critically affect the association of transmembrane helices. In particular, the GCN4 model peptide studies [12,45] showed the decisive importance of hydrogen bonding at a single Asn residue at an 'a' position in the detergent-soluble dimer. In the present design, a hydrogen bond network was designed between 'a' and 'd' residues in the bundle segment spanning the hydrocarbon core region of the lipid bilayer. The network was designed to prevent the helices from rotating along the longitudinal axes, and to favor the formation of a four-helix bundle over other oligomerization states.

3.2. Design of poly-alanine four-helix bundle templates

The optimal length of a transmembrane helix is between 20–30 residues, with an average length of 26 residues [47]. As this is longer than the length required to span a membrane, most transmembrane helices extend into the aqueous solvent flanking the bilayer. A 31-residue poly-alanine α-helix (dihedral angles $\Phi = -57^\circ$, $\Psi = -47^\circ$) was generated using the Biopolymer module in InsightII (Biosym/Molecular Simulations Inc.). The ends of the helices were later incorporated into loops, leaving a helix length of 27 residues. This α-helix was further twisted into a 3.5 residue per turn helix in order to align the residues on the backbone every seven positions [4].

Like water-soluble proteins, antiparallel orientations of helices are more common in membrane proteins than are parallel orientations; in addition, antiparallel membrane helices pack closer together [47]. A series of four antiparallel four-helix bundles with interhelical distances ranging from 12 to 15 Å were built using symmetry operators [4]. We later chose the bundle template that provided the best packing of the interior side chains (12 Å). In accordance with the statistical preference found in helical membrane proteins [47], the helices were prevented from diverging from each other by tilting them 20° to each other along the bundle axis and twisting the bundle into a left-handed supercoil with a pitch of 190 Å [4].

3.3. Choosing appropriate residues

Natural four-helix homodimers form functional and stable proteins. We designed Membun to be an antiparallel homodimer consisting of two copies of two different helices, A and B (Fig. 1A). A homodimeric target structure was chosen so that the synthesis of the protein would be feasible by solid phase peptide synthesis. Since a typical bilayer has a hydrophobic core of approximately 25–30 Å flanked on each side by a polar headgroup region of approximately 10–15 Å [48], transmembrane helices will experience three environments: aqueous solvent, polar headgroup and hydrocarbon core. Accordingly, each 31-residue helix was divided into three distinct regions (Fig. 1B). Our choice of residues for packing each region of the bundle was guided by the statistical occurrences of amino acids found in these three regions in natural transmembrane helices [49], and also on model peptide studies [14,15].

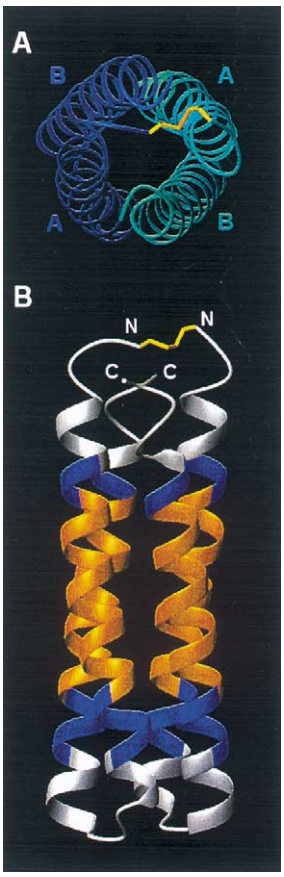


Fig. 1. (A) Top view of the C-α ribbon representation of Membun, showing the two monomers (cyan and blue) which form a four-helix coiled coil. (B) Side view showing the three distinct regions along the helix axes: white, the N- and C-termini of the α-helices which interact with the solvent; blue, interfacial region containing surface aromatic residues which interact with the phospho head groups of the bilayer; orange, the central core of the bundle which contacts the hydrocarbon tails of the phospholipids. The N- and C-termini of both monomers are indicated. The disulfide bridge linking the two N-termini is colored yellow.

In the four-helix bundle, the N- and C-termini of the helices were designed to interact with the aqueous solvent. Residues appropriate for this region were hydrophobes (Ala, Leu and Val) for the buried positions, and small, partly hydrophobic polar residues (Thr, Ser) for interfacial positions [14,15]. An acid residue was positioned at the N-termini of the helices, and a positively charged residue at the C-termini, in order to stabilise the helices through charge interaction with the helix dipole [50]. In addition, the terminal charges could favor a defined orientation of the four-helix bundle through favorable electrostatic interactions with the membrane–water interface [51,52]. To reinforce the specific orientation of the two A–B monomers in the bundle, a network of salt bridges was designed between residues at the end of helix A and the beginning of helix B in each monomer.

Moving longitudinally towards the center of the bundle, the next layers of the bundle interact with the phospholipid head group region of the bilayer, and thus form the boundary between the hydrophilic exterior and hydrophobic interior of the membrane. In natural membrane helices, aromatic ‘floaters’ or ‘anchors’ are observed in this region [53–55]. The aromatic side chains face the lipid, and their purpose is believed to be to fix the helix longitudinally in the membrane, and to buffer the protein from varying bilayer thickness caused by heterogeneous lipid content [56]. To provide asymmetry to the bundle similar to that observed in natural proteins, one aromatic ring was composed of four Tyr, and the other of four Trp. Polar residues able to make H-bonds to the phospholipid heads (Gln and Asn) were placed at positions ‘b’, ‘c’ and ‘f’. Ala and Val were chosen for positions ‘e’ and ‘g’ based on stereochemical packing considerations.

The central heptads of the bundle (four ‘a’ and ‘d’ layers) were taken to reside in the hydrocarbon region of the bilayer. The ‘a’ and ‘d’ residues were chosen to be hydrophilic and capable of H-bond formation, while the membrane side

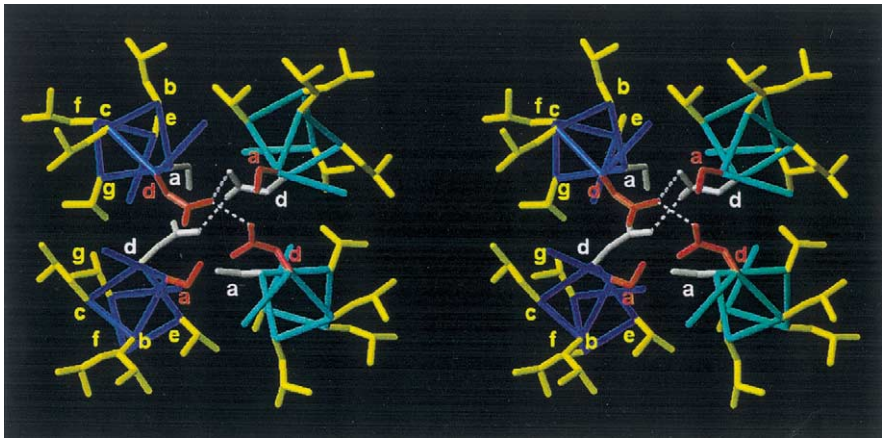


Fig. 2. Stereo view of a slice of the central region (residues 15–21, 45–51) of the bundle, showing one heptad. Side chains of residues belonging to the top ‘a’, ‘d’ layer are colored pink while side chains of residues belonging to the bottom ‘a’, ‘d’ layer are colored white. The ‘a’ and ‘d’ positions facing the interior of the bundle are occupied by Ser and Asn residues, respectively, and hydrogen bonds between these residues are indicated. The ‘b’, ‘c’ and ‘f’ positions are occupied by Leu residues (yellow). Backbones of the two monomers are colored accordingly.

of the helices (positions 'b', 'c' and 'f') were made totally hydrophobic (Leu) (Fig. 2). The H-bond network was composed of Ser and Asn because they are small residues, and therefore the number of possible side chain rotamers will be small. Additionally, manual modeling (see below) indicated that Asn would give better helix–helix packing than Gln. Most importantly, the judicious introduction of Asn and Ser residues allows hydrogen bonds to be formed not only with residues in the same layer, but also with residues in adjacent layers.

3.4. Helix capping

Residues at the N- and C-termini of α-helices are defined [57] in the following manner:

N″-N′-Ncap-N1-N2-N3 . . .

where Ncap is the first residue of the helix.

C3-C2-C1-Ccap-C′-C″ . . .

where Ccap is the last residue of the helix.

A capping-box motif, where reciprocal H-bonds are formed between Ncap and N3 [58], was introduced at the N-terminus of helix A with the sequence Ser–Pro–Asp–Gln. Statistically, this sequence is found frequently at the beginning of α-helices [59,60] and has been found experimentally to stabilize and nucleate helical conformation [61]. Ser and Gln (at positions Ncap and N3, respectively) form the capping box, whereas Asp in position N2 adds a negative charge at the positively charged N-terminus of the helix dipole. In this motif, the side chain O_γ-group of the Ser side chain makes an H-bond with the backbone NH of Gln, while at the same time, the O_γ of Gln makes an H-bond with the backbone NH of Ser [62]. At the N-terminus of helix B, Gly33 was chosen as the N-capping residue, with Asp34 making a favorable interaction with the helix dipole. Helix A has a Gly at the Ccap position, which is a preferred residue at this location [63]. At the C-terminus of helix B, Asn61 caps the C-terminus and Pro62 ensures that the helix does not elongate. Lys residues were introduced at the C-terminus of the helices to compensate for the negative charge of the helix dipole (Fig. 3).

3.5. Linker design

A disulfide bridge was introduced between the two identical monomers comprising the bundle, so the loops joining helices A and B were on the same side of the bundle (Fig. 1). We searched the fragment database of WhatIf [23] using the SCAN3D option to find appropriate sequences for linking helices A and B, and to terminate the helices. ROP is a natural, water-soluble four-helix bundle [31], composed of two identical dimers, with the same helix orientation and packing as our model bundle. The loop region of a hyperstable mutant of ROP [63] contains a point mutation (Asp to Gly) in the loop. This loop was incorporated into our model bundle as its geometry is ideal for joining the C-terminus of helix A to the N-terminus of helix B. The loop sequence in the mutant ROP is Glu–Leu*–Gly*–Ala*–Asp*–Glu ('*' denotes the loop). We replaced the Leu by a polar residue since in our model this residue is solvent exposed, while in ROP it is buried. The final sequence of our loop connection was Asn*–Gly*–Gly*–Asp*–Ala. The second Gly of the loop was introduced to increase loop flexibility and to allow some leeway in the distance between the two connected helices. The sequence and heptad positions of each residue in the final helix–loop–helix monomer peptide was:

					f	g	a	b	c	d	e
Cys–	Gly–	Gly–	Ser–	Pro–	Asp–	Gln–	Val–	Trp–	Leu–	Asn–	Val–
f	g	a	b	c	d	e	f	g	a	b	c
Leu–	Val–	Ser–	Leu–	Leu–	Asn–	Val–	Leu–	Val–	Ser–	Leu–	Tyr–
d	e	f	g								c
Thr–	Ala–	Gln–	Lys–	Ala–	Lys–	Asn–	Gly–	Gly–	Asp–	Ala–	Glu–
d	e	f	g	a	b	c	d	e	f	g	a
Leu–	Glu–	Asn–	Ala–	Val–	Tyr–	Leu–	Asn–	Ala–	Leu–	Val–	Ser–
b	c	d	e	f	g	a	b	c	d	e	f
Leu–	Leu–	Asn–	Ala–	Leu–	Val–	Ser–	Leu–	Trp–	Thr–	Ala–	Lys–
Asn–	Pro–	Gly									

3.6. Side chain rotamer selection

The requirement for 'parallel' packing of side chains in position 'a' and 'perpendicular' packing in position 'd' [43] restricts the χ_1 angles (in the ideal case) to 180° and −63°, respectively. To check the absoluteness of this requirement, we measured the χ_1 angles of all 'a' and 'd' position side chains in ROP. To within approximately 20° of the ideal, 5 of 12 'a' position side chains and 8 of 12 'd' position residues obeyed this restriction; there was wide variation observed in the χ_1 angles that did not comply. We thus concluded that significant flexibility in permitted χ_1 angles would be tolerated by the antiparallel structure, and that our major restriction on the side chain rotamers would be that they adopt a preferred (low energy) rotamer conformation.

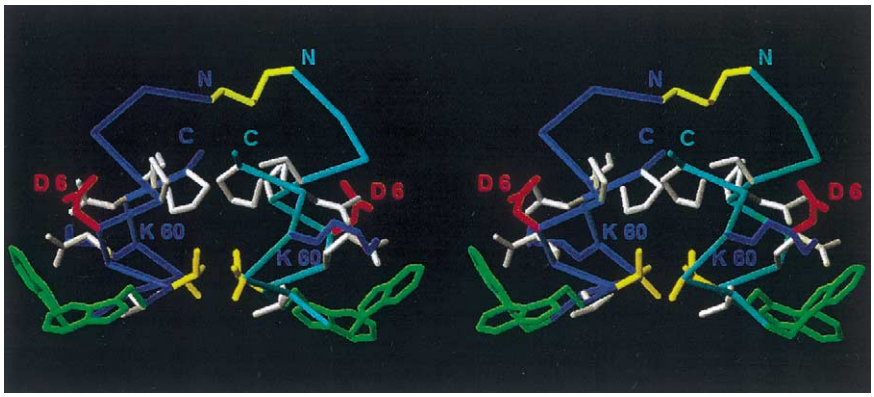


Fig. 3. Stereo view of the aqueous and interface regions at the top of the bundle. Trp residues (green) were positioned in the interface region to act as "floaters" in order to orient the bundle perpendicular to the membrane plane. An Asp (negatively charged, red) and a Lys (positively charged, blue) were introduced at the N- and C-termini of the first and second helix, respectively, to stabilize the helix dipoles. White, polar and Pro residues; yellow, hydrophobic residues. The disulfide bridge connecting the two N-termini is also shown in yellow at the top of the figure.

We used the Ponder and Richards [24] rotamer library for our initial, manual fitting of side chain residues, as this library has withstood the test of time, and it is the library in InsightII. To check that the rotamer preferences used in our model were compatible with the helical conformation, we used the rotamer library of WhatIf [23], which allows secondary structure rotamer preferences to be searched. For each amino acid, we identified the rotamers most likely to occur in α-helices, and found that most rotamers identified by WhatIf corresponded to the most preferred rotamer for each amino acid in the Ponder and Richards [24] library. In some instances, the rotamer chosen by WhatIf corresponded to the second most preferred rotamer in the library. Note that both the rotamer library and the databases searched by WhatIf are based on water-soluble proteins, not membrane proteins. However, at present there is no evidence that water- and membrane-soluble proteins have different preferred rotamers. Visual inspection and manual manipulation of the model indicated that the chosen residues were compatible with good interior packing of the bundle and H-bond network formation. The SMD algorithm, which allows amino acid side chain conformations to be rapidly optimized for a given backbone conformation [25], was then used to check for global compatibility.

3.7. Energy minimization

We applied van der Waals surfaces to successive layers of the bundle and looked for overlaps and holes in the packing. We also looked at the H-bond network, identified unsatisfied potential H-bonds, and corrected the deficiencies either by manual rotation of side chains or by replacing appropriate residues. The final stage of the design process was limited energy minimization on the bundle using the AMBER 4.1 program [26]. No holes in the helix–helix packing interfaces were observed.

3.8. Simulated annealing and molecular dynamics (SA/MD)

The design of our membrane protein was validated using two molecular dynamics approaches. In the first approach, Membun was simulated in a DMPC lipid bilayer environment without any restraints [64,65]. This approach is likely to be useful for exposing major problems with the design. For example, two models of alamethicin channels differing in their charge states of pore-lining glutamates gave rise to distinctly different structures after simulation, in agreement with results from continuum electrostatics calculations [35]. The alamethicin result demonstrates that simulations do have some predictive ability to distinguish between valid and incorrect models. Similar simulations have also been used successfully to explore different models for the Influenza A M2 transmembrane tetramer [66].

In the second approach, we used simulated annealing and restrained molecular dynamics (SA/MD) in an attempt to predict the structure of Membun, given the sequence and a number of built-in structural constraints. These constraints include enforcing helicity, and putting an upper limit on the distance that two helices within the bundle are allowed to move apart [28,30]. The procedure results in an ensemble of protein models, generated in vacuum. Some of these models can then be used for further simulations in the more realistic environment of a lipid bilayer, without additional restraints on helicity and inter-helix distance. The choice of which models to use for further simulations is to some extent arbitrary: possible candidates might include the most symmetrical model, the one closest to the average structure of the bundle, or an energy minimized structure based on the average structure of the set of models.

Structurally, Membun is similar to ROP, and has the same topology. A high-resolution structure of ROP is available, making detailed molecular dynamics simulations possible, and also providing an indication of the motions that occur in

four-helix bundles. In addition, the SA/MD procedure can be tested on ROP in order to assess how closely the true structure can be approximated if we start with a deliberately incomplete structure. Finally, we can use ROP to determine what might happen when a water-soluble protein is placed into an inappropriate environment, such as a lipid bilayer. Table 1 gives an overview of all the molecular dynamics simulations conducted in this study.

3.9. Challenging Membun: molecular dynamics simulations

A number of structural properties that change over the course of a simulation provide insights into the stability of a protein. These properties include the root mean square deviation (RMSD), secondary structure, number of hydrogen bonds, the radius of gyration and the solvent accessible surface. Each of these properties were studied for both ROP and Membun, and in addition, essential dynamics analyses were used to judge whether Membun behaves like a protein in a dynamical sense.

The RMSD curve for ROP in water (Fig. 4A) is typical for molecular dynamics simulations of proteins, with a fast initial rise and, after a few hundred picoseconds, a stable RMSD of ca. 0.15 nm for the C-α atoms and 0.25 nm for all atoms. Membun has an even lower RMSD for the C-α atoms in the bilayer, compared to ROP in water (Fig. 4B). Surprisingly, the difference between the RMSD for all atoms and just the C-α atoms is much smaller for Membun than for ROP, suggesting that the conformational freedom of the side chains is much more limited in Membun than in ROP. If ROP, with its typical hydrophilic exterior, is placed inside a lipid bilayer, the RMSD does not stabilise and keeps increasing to much higher values than for ROP in water. Although this is only one result, it does indicate that simulations have some power to distinguish between a "normal" and an unfavorable situation.

Secondary structure was calculated with Dictionary of Protein Secondary Structure (DSSP) [67]. Both ROP and Membun have stable secondary structures over 2 ns (Fig. 5). In both cases, minor fluctuations occur at the termini of each helix, but there is no systematic change in structure. Because both ROP and Membun are dimeric, we have an internal consistency check: the structures of both monomers must be similar. The only noticeable difference between the two monomers in Membun is found in the loop joining the two helices of each monomer, but small differences in loops are common. In ROP, small fluctuations in structure occur in both monomers, particularly at the N-terminus. However, the overall designed secondary structure of Membun is stable during the 2 ns of the simulation.

An indication of the degree of side chain packing of a protein is given by the radius of gyration. The total radius of gyration of ROP in water remains constant during the

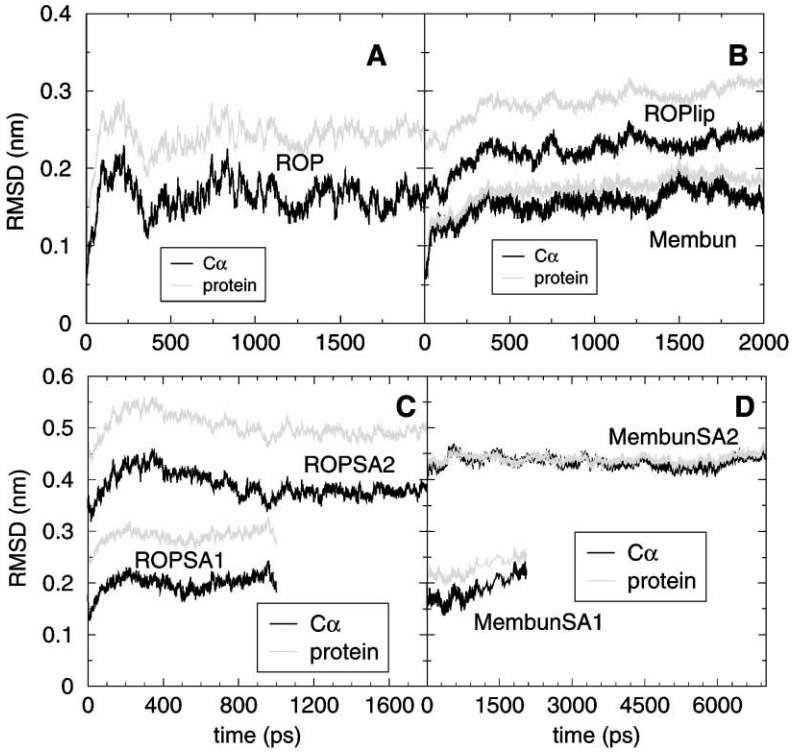


Fig. 4. (A) RMSD as a function of time for ROP in water, fitted on the crystal structure C-α atoms. (B) RMSD as a function of time, fitted on the C-α atoms of the crystal structure for ROP and the Membun model, both in DMPC. (C) RMSD of the simulated annealing models ROPSA1 and ROPSA2 as a function of time with respect to the ROP crystal structure. (D) RMSD of the simulated annealing models of Membun (MembunSA1 and MembunSA2) as a function of time with respect to the Membun model.

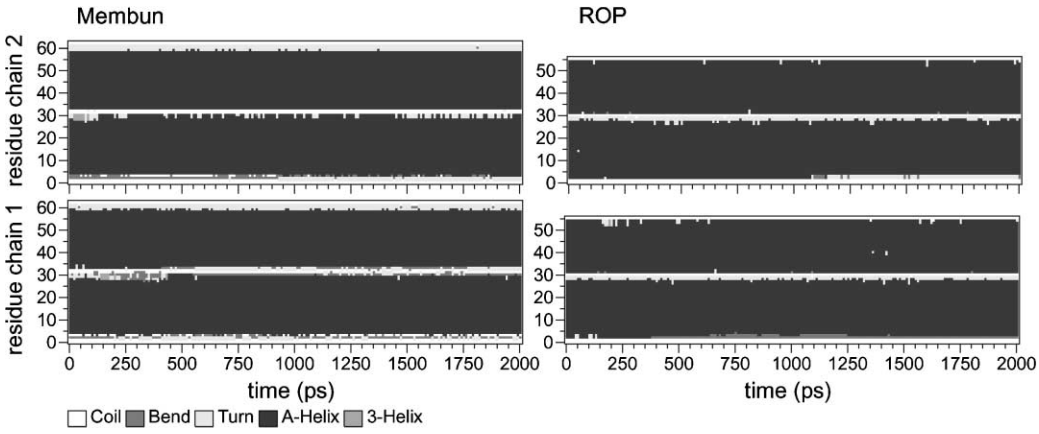


Fig. 5. Secondary structure as a function of time for (A) Membun in DMPC; (B) ROP in water.

simulation, whereas for Membun in a DMPC bilayer the radius of gyration decreases by almost 2%. This is probably due to changes in the side chains at the Membun exterior, due to the lipid and water environment. If ROP is placed in a bilayer, its radius of gyration increases slightly, from 1.44 to 1.48 nm, indicating that the protein expands compared to its structure in water. This is mostly due to significant changes in the structure of the hydrophilic side chains that are exposed to the hydrophobic lipid interior. The hydrophilic and hydrophobic solvent accessible areas changed less than a few percent for both ROP in water and Membun in DMPC (data not shown).

A characteristic property of proteins is that they possess a small number of collective degrees of freedom that account for the majority of the protein's motions [68,69]. The essential dynamics analysis of a protein trajectory gives a description of large-scale collective motions in the protein. Such motions could involve distinct domains within the protein [70], or less obvious combinations of, e.g., secondary structure elements. Mathematically, the analysis calculates the covariance matrix of the fluctuations of all co-ordinates. This matrix is diagonalized, yielding eigenvectors and eigenvalues. The eigenvectors describe the directions of the collective modes of motion of the protein, and the eigenvalues describe the amplitude of these modes. Most of the motions of a protein can be described by a set of the 10 or 20 collective co-ordinates with the largest amplitude. The cumulative total of the first N eigenvalues shows how much of the total motion is accounted for by the N corresponding eigenvectors. This total is plotted in Fig. 6 for Membun and ROP. The precise shape of the curve is less relevant; considerable fluctuations occur between different proteins and even different runs of the same protein, and simulations do not sample all the conformations that are accessible to proteins in a few nanoseconds. However, it is clear that Membun, like ROP, requires only a few dozen eigenvectors to describe most of its motions. From this point of view, Membun behaves like a native protein.

The secondary and tertiary structures of a protein impose significant constraints on the protein's possible motions. It can be expected that a designed protein does not have the same degree of intrinsic constraints as a native protein, since these constraints are formed by many contributing factors that are difficult to pinpoint. A lack of powerful constraints on the structure becomes visible in an essential dynamics analysis as a collective mode of motion that occurs in part of a simulation, but not in other parts. In a native protein, this dichotomy should not be possible, because the collective modes of a protein are an intrinsic property of its dynamics. The major collective modes should be reproducible in different simulations, if the simulations are long enough to accurately define the essential subspace, the space spanned

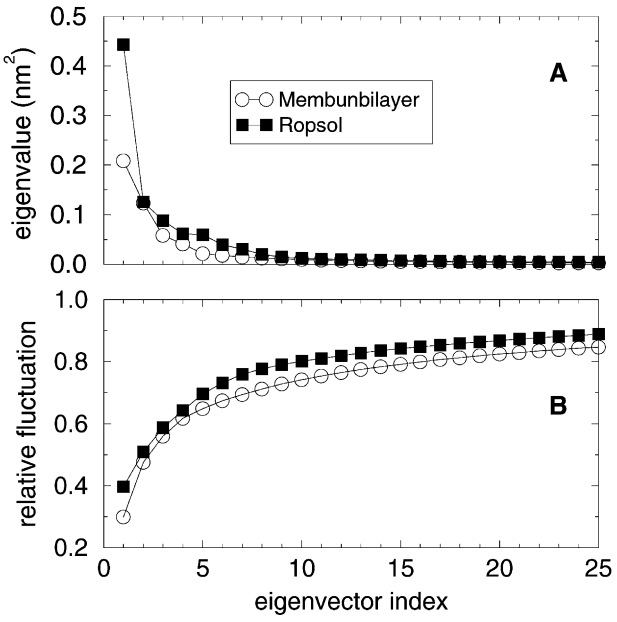


Fig. 6. (A) Eigenvalues of the first 25 eigenvectors for Membun and ROP. (B) Fraction of the total atomic fluctuation in the protein (y-axis) that is described by the first N (x-axis) eigenvectors.

by the 10 or 20 collective motions with the largest amplitude [71,72]. To analyze the occurrence of major collective modes of motion that are not reproducible between the first and second halves of a simulation, we can calculate an inner product matrix of the eigenvectors from the essential dynamics analysis on the two halves of a simulation. If the motion described by the first eigenvector of simulation one (in this case the first nanosecond of a 2 ns simulation) is the same as the motion that is described by the first eigenvector of simulation two, the inner product between the two eigenvectors is 1. If the motions described by these eigenvectors are completely different, the inner product is 0. If we calculate the inner product for each of the eigenvectors of simulation one with each of the eigenvectors of simulation two, we can judge the similarity of the collective motions between the two simulations. To be similar, the inner product of eigenvector *N* of simulation one with eigenvector *N* of simulation two should be high, but the inner product of eigenvector *N* of simulation one with eigenvector *M* of simulation two should be low. This is conveniently visualized as a matrix, where the grey value of a matrix cell (*N*, *M*) corresponds to the value of the inner product of the eigenvector *N* from simulation 1 and the eigenvector *M* from simulation 2. Peaks far away from the diagonal indicate that there are collective motions that occur in simulation 1 but not in simulation 2. The results of this analysis are given in Fig. 7A for ROP and 7B for Membun. In both cases, there are no major peaks off the diagonal for the most important eigenvectors (those with the lowest index). From this point of view, Membun behaves like a protein, with well-defined collective motions. The results also suggest that the dynamics of Membun might be more restricted than for ROP. This agrees with the generally low RMSD values and stable secondary structure and radius of gyration of Membun. An intriguing question for future work will be to identify the most critical structural requirements in a model or native protein necessary for obtaining results similar to those observed for Membun or ROP.

3.10. Simulated annealing and subsequent simulations

For the SA runs, we used two sets of initial assumptions for both ROP and Membun. In the first set, it was assumed that the C-α positions were known from the Membun design or the ROP crystal structure. In the second set, we assumed only that ROP and Membun form four-helix bundles with approximately parallel helical axes and the correct face of each helix faces the interior of the bundle. Details of the assumptions are described in Section 2.

The primary result from the SA runs is four sets of 25 structures, two sets for Membun and two for ROP. A random selection of these structures is plotted in Fig. 8A for ROP and in Fig. 8B for Membun. If the C-α atoms are used as a template, the overall coiled-coil structure is mostly conserved in both ROP and Membun. The main differences between the ROP models and the true structure are found at the ends of the helices and in the long side chains at

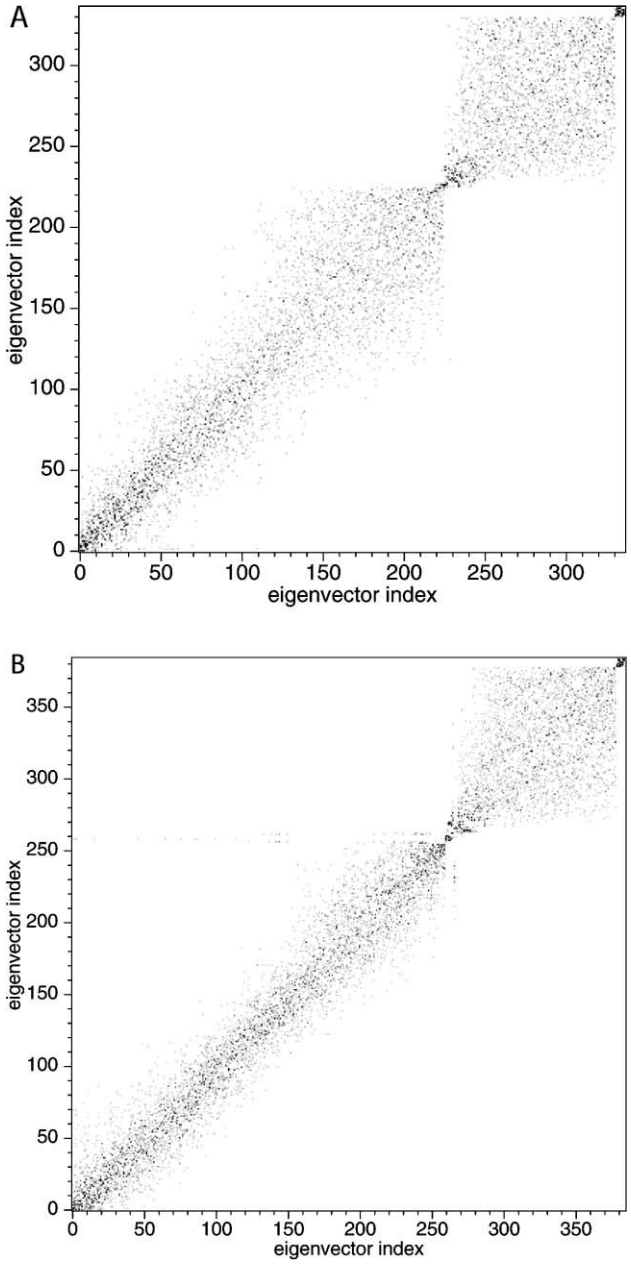
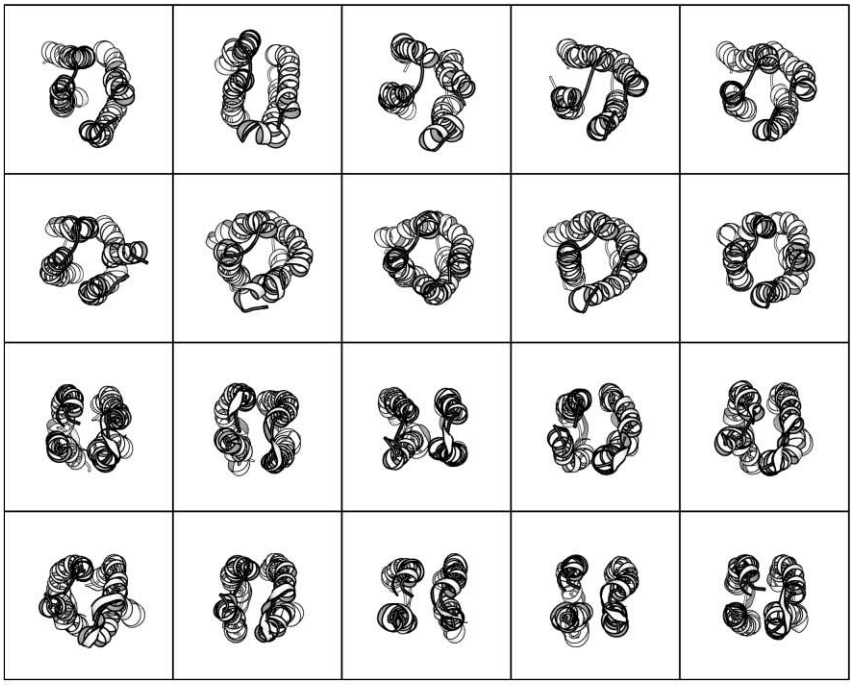


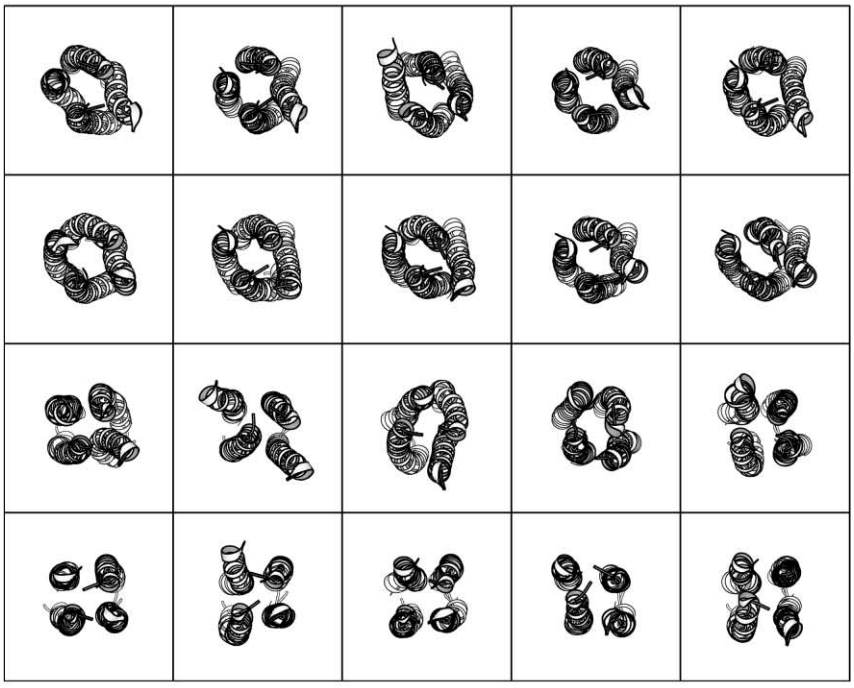
Fig. 7. Squared inner product matrices of the essential dynamics eigenvectors for (A) ROP; and (B) Membun, calculated from the first and the last nanosecond of each trajectory.

positions ‘b’, ‘c’ and ‘f’, which is where one would expect the most structural fluctuation. The average RMSD from the mean structure of the ensemble is 0.19 nm for backbone atoms and 0.25 nm for all heavy atoms. The RMSD for the structures that assumed straight helices is the same, 0.19 nm for the backbone and 0.25 nm for all heavy atoms. Several of these structures form coils.

The RMSD from the mean structure for MembunSA1 (starting from the C-α positions of the Membun design) is considerably lower compared to ROP, 0.13 nm for the backbone atoms and 0.17 nm for all heavy atoms. Interestingly,



(A)



(B)

Fig. 8. (A) Structures generated by simulated annealing starting from the C-α atoms of the ROP crystal structure (ROPSA1, upper 10) and from four straight helices (ROPSA2, lower 10). For each case, 10 out of a total of 25 samples are shown. (B) Structures generated by simulated annealing starting from the C-α atoms of the Membun model (MembunSA1, upper 10) and from four straight helices (MembunSA2, lower 10). For each case, 10 out of a total of 25 samples are shown.

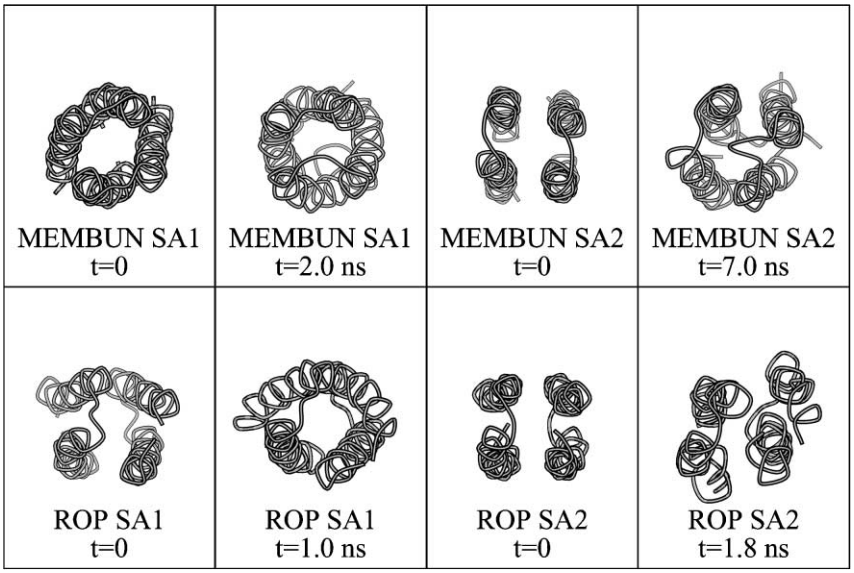


Fig. 9. C-α traces of the starting structure and final structure of runs 4–7. The starting and final structures in runs 1–3 do not significantly differ from each other.

these values are not only much lower than for ROP, but they are also much lower than for the SA2 model (0.18 nm for the backbone, 0.23 nm for the heavy atoms). It appears that packing in Membun is more restricted than in ROP. When starting from the designed backbone, the freedom for side chains to pack in different conformations appears to be less than in ROP. When four straight helices are taken as the initial backbone, the range of side chain conformations that can be assumed is much larger.

Extended MD simulations permit a more accurate assessment of the effect of environment (water or DMPC) on the protein. It was interesting to see if MD simulations could bring the structure generated from the SA procedure closer to that of the "real" structure which, in the case of ROP, is the crystal structure. In Fig. 4C, the RMSD as a function of time for ROP, simulated in water, is plotted for ROPSA1 and ROPSA2. In both cases, the RMSD shows no tendency to approach the crystal structure. In the first case, the RMSD is somewhat higher than for the run starting from the crystal structure. The final structure after 1 ns of simulation in water is still a coiled coil (Fig. 9), and the main differences with the crystal structure are found at the ends of the protein. In the second case, the RMSD with respect to the crystal structure is very high, close to 0.4 nm. The final structure of this run shows that the coiled coil is mostly lost and the helices are no longer packed as tightly (Fig. 9). This emphasises the importance of accurate experimental constraints on model building procedures. The simulation of ROPSA1, which assumed that the C-α positions were known, gives sensible structures for ROP in solution. However, the assumption of just four straight helices facing roughly in the right direction (ROPSA2) is not enough to generate sensible protein structures.

Given these results for ROP, it is doubtful that an extended simulation of an inaccurate starting model (MembunSA2) will improve the prediction of the true structure of Membun. Starting from the model C-α co-ordinates, the MembunSA1 run stays relatively close to the model structure (Fig. 4D). The second run, which is reasonably long by current standards, has a RMSD of 0.45 nm for both backbone and all atoms compared to the Membun model. It shows no trend to move closer toward the designed structure. However, this has limited predictive power because we do not know the true structure of Membun. The final structures after 1 ns for MembunSA1 and 7 ns for MembunSA2 are given in Fig. 9. The final structures of the simulations of the Membun design in DMPC and the crystal structure of ROP in water do not differ significantly from their initial structures.

3.11. Peptide synthesis

The synthesis and initial characterization of the peptide has been described in detail elsewhere [73,74]. Briefly, the 63-residue monomer was synthesized using Fmoc solid phase peptide synthesis. A novel solubilizing tail strategy was employed to facilitate reverse phase HPLC purification of the peptide. The tail, consisting of a repeating Gly–Lys sequence attached via an aqueous base-labile linker to the hydrophobic peptide, was found to make the target peptide water-soluble. After HPLC purification and lyophilization, the hydrophobic peptide was cleaved from the tail by short exposure to mild NaOH, Matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) showed that the correct peptide had been synthesized.

4. Discussion and conclusions

Although membrane proteins are designed by nature to function in an environment very different from that inhabited by water-soluble proteins, membrane proteins appear to be surprisingly similar to water-soluble proteins in many respects. The overall hydrophobicity of interior residues, and the close-packing of side chains in the interiors, are similar in both classes of proteins [33]. Furthermore, at present there is no evidence that water- and membrane-soluble proteins have different preferred side chain rotamers. In addition, the overall topology of helix bundle proteins (helix orientation and coiling) found in both aqueous and membrane environments also seems to be similar, although a wide variation in the tilt of membrane helices with respect to the lipid bilayer has been observed in membrane peptides [75,76]. Of course, only a few structures of membrane proteins are known at high resolution, and as more structures become available, differences between the two classes of proteins may become apparent. In light of our present knowledge, however, it appears that the design of a membrane protein bundle can be successfully undertaken using much the same approach and following similar rules to those applied successfully to the design of water-soluble bundles. Our intent was to design a simple membrane protein (Membun) using the available structural and database information, and test the design using computational approaches.

The structural and dynamic properties obtained from a simulation of the Membun design in a lipid environment give no indication that Membun is unstable. The essential dynamics analysis demonstrated that there are internal constraints on the structure characteristic of native proteins. It also showed that as in other proteins, most of Membun's motions could be described by just a few collective degrees of freedom. Structural properties such as the RMSD and secondary structure analysis indicate a stable structure, comparable to the behavior of the high-resolution crystal structure of ROP in water. In contrast, ROP in DMPC continues to diverge from its crystal structure. This is encouraging, as it shows that the simulations can distinguish, at least in this obvious case, between a favorable and an unfavorable situation. This was also found in a simulation of two models of an alamethicin hexameric bundle in a lipid bilayer. In this case, one model that seemed reasonable turned out to be unstable [35]; this instability was verified by additional calculations. We can reasonably expect that major flaws in the design of Membun would show up during its simulation in a lipid bilayer. A second difference between ROP and Membun is the more limited freedom of the side chains in Membun, indicated by the small difference between the backbone and all-atom RMSD. This is consistent with the design of Membun, in which the side chain packing was optimized.

The simulated annealing/restrained molecular dynamics procedure yielded a number of structures for both ROP and Membun. When a very simple template consisting of four parallel helices is used as the starting point, the difference between the resulting structures and the known high-resolution structure of ROP is large. When the C-α positions of the known ROP structure are used as initial information, the resulting structures are quite reasonable, although not as accurate as the structure of a two-helix leucine zipper that was predicted by a similar method [29]. In the future, a fruitful approach might be the combination of the physics in the SA/MD and molecular dynamics methods with the statistical and rational design knowledge used for the initial design of Membun.

It was necessary to make a number of assumptions in the course of the simulations, the most important being that Membun will actually insert as a tetrameric helix bundle into a bilayer. However, it seems that as long as there is a continuous lengthwise hydrophobic face covering over 50% of the helix, helix formation and transversal in membranes is spontaneous and rather indifferent to sequence [77,78].

5. Outlook

In spite of these assumptions and limitations, simulations and modeling with an empirical force field are likely to be the only method that can give detailed information on the accuracy of a model before the protein is actually synthesized or expressed. If helices unfold in a simulation, if the structure drifts far away from the model, or if the radius of gyration changes significantly, then there is likely a major problem with the design and the design can be altered before the protein is synthesized. If there are small problems, the models and simulations will give a range of possible structures that might indicate the tightness of the design. It is becoming increasingly common that model building or protein structure prediction methods yield a series of possible structures of membrane proteins. In the absence of a high-resolution crystal structure, this is the best that modeling and simulation can achieve [79–82].

MD cannot be used to distinguish between a large set of different structures to find the globally most favorable structure. However, when given a set of possible structures that are relatively close, it might be able to select among those. An intriguing possibility is to use relatively coarse protein folding potentials to reduce the number of folds for a given sequence to a relatively small set, and then to use simulations to further reduce this set. Such protein folding potentials currently can predict the structures of small water-soluble proteins such as ROP to within 3–6 Å. This is currently not possible for membrane proteins as there is insufficient experimental structural data to parameterize the protein folding potentials. As the number of membrane protein crystal structures increases, it should be possible to predict the fold of a designed membrane protein to intermediate resolution based only on the sequence, and then refine the structure to atomic resolution using simulation.

Simulations of small membrane proteins such as Membun can be performed on desktop workstations in a few

weeks. In the case of Membun, this is a relatively small effort compared to the design, synthesis and characterization of the protein. If such simulations can dismiss a model before a synthesis is attempted, they are easily worth the effort. In the present study, the additional simulations of ROP were useful in order to provide a reference, but as more simulations of membrane proteins become available, those will serve as more appropriate references.

The present paper represents a first attempt to design a membrane protein with native-like sequence complexity from first principles. As discussed above, numerous assumptions were unavoidable in the design. The principal unknowns in the present design are: (i) do membrane bundles coil to the same extent as water-soluble proteins? (ii) How is the association state of multi-helix membrane proteins controlled? (iii) Are there essential differences in the packing of membrane proteins compared to water-soluble proteins [83]? The Membun peptide has been synthesized and the next step is the protein's structural characterization, which is currently underway. The results from the characterization, together with the rapidly growing body of information on membrane proteins, will form the basis for the next generation of de novo designed membrane proteins.

Acknowledgements

DPT is a Scholar of the Alberta Heritage Foundation, and thanks Mark Sansom, Oxford University, for his interest in this work.

References

[1] Y. Fezoui, E.H. Braswell, W. Xian, J.J. Osterhout, Dissection of the de novo designed peptide ata: stability and properties of the intact molecule and its constituent helices, Biochemistry 38 (1999) 2796–2804.

[2] J.W. Bryson, J.R. Desjarlais, T.M. Handel, W.F. DeGrado, From coiled coils to small globular proteins: design of a native-like three-helix bundle, Protein Sci. 7 (1998) 1404–1414.

[3] S.P. Ho, W.F. DeGrado, Design of a four-helix bundle protein: synthesis of peptides which self-associate into a helical protein, J. Am. Chem. Soc. 109 (1987) 6751–6758.

[4] S.F. Betz, W.F. DeGrado, Controlling topology and native-like behavior of de novo-designed peptides: design and characterization of antiparallel four-stranded coiled coils, Biochemistry 35 (1996) 6955–6962.

[5] R.B. Hill, W.F. DeGrado, Solution structure of α_2D, a native-like de novo-designed protein, J. Am. Chem. Soc. 120 (1998) 1138–1145.

[6] C. Das, S. Raghothama, P. Balaram, A designed three-stranded β-sheet peptide as a multiple β-hairpin model, J. Am. Chem. Soc. 120 (1998) 5812–5813.

[7] K. Goraj, A. Renard, J.A. Martial, Synthesis, purification and initial structural characterization of octarellin, a de novo polypeptide modelled on the α/β barrel proteins, Protein Eng. 3 (1990) 259–266.

[8] C.T. Choma, J. Lear, M. Nelson, D. Robertson, P.L. Dutton, W.F. DeGrado, Design of a heme-binding four-helix bundle, J. Am. Chem. Soc. 116 (1994) 856–865.

[9] B.R. Gibney, F. Rabanal, K.S. Reddy, P.L. Dutton, Effect of four-helix bundle topology on heme-binding and redox properties, Biochemistry 37 (1998) 4635–4643.

[10] K.S. Broo, L. Brive, P. Ahlerg, L. Baltzer, Catalysis of hydrolysis and transesterification reactions of *p*-nitrophenyl esters by a designed helix–loop–helix dimer, J. Am. Chem. Soc. 119 (1997) 11362–11372.

[11] M.M. Javadpour, M.D. Barkley, Self-assembly of designed antimicrobial peptides in solution and micelles, Biochemistry 36 (1997) 9540–9549.

[12] C.T. Choma, H. Gratkowski, J.D. Lear, W.F. DeGrado, Asparagine-mediated self-association of a model transmembrane helix, Nature Struct. Biol. 7 (2000) 161–166.

[13] L.-P. Li, S.C. Li, N.K. Goto, C.M. Deber, Threshold hydrophobicity dictates helical conformations of peptides in membrane environments, Biopolymers 39 (1996) 465–470.

[14] W.C. Wimley, S.H. White, Experimentally determined hydrophobicity scale for proteins at membrane interfaces, Nature Struct. Biol. 3 (1996) 842–848.

[15] S.-C. Li, C.M. Deber, A measure of helical propensity for amino acids in membrane environments, Nature Struct. Biol. 1 (1994) 368–373.

[16] Y. Zhou, J. Wen, J.U. Bowie, A passive transmembrane helix, Nature Struct. Biol. 4 (1997) 986–990.

[17] W.C. Wimley, S.H. White, Designing transmembrane α-helices that insert spontaneously, J. Am. Chem. Soc. 39 (2000) 4432–4442.

[18] J.D. Lear, Z.R. Wasserman, W.F. DeGrado, Synthetic amphiphilic peptide models for protein ion channels, Science 240 (1988) 1177–1181.

[19] S. Seth, P. Balaram, M.K. Mathew, Characterization of a 22-residue peptide derived from a designed ion channel, Biochim. Biophys. Acta 1328 (1997) 177–184.

[20] A.V. Starostin, R. Butan, V. Borisenko, D.A. James, H. Wenschuh, M.S. Sansom, G.A. Woolley, An anion-selective analogue of the channel-forming peptide alamethicin, Biochemistry 38 (1999) 6144–6150.

[21] L.H. Pinto, G.R. Dieckmann, C.S. Gandhi, C.G. Papworth, J. Braman, M.A. Shaughnessy, J.D. Lear, R.A. Lamb, W.F. DeGrado, A functionally defined model for the M2 proton channel of influenza A virus suggests a mechanism for its ion selectivity, Proc. Natl. Acad. Sci U.S.A. 94 (1997) 11301–11306.

[22] P. Whitley, I. Nilsson, G. von Heijne, De novo design of integral membrane proteins, Nature Struct. Biol. 1 (1994) 858–862.

[23] G. Vriend, WhatIf: a molecular modeling and drug design program, J. Mol. Graph. 8 (1990) 52–56.

[24] J.W. Ponder, F.M. Richards, Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes, J. Mol. Biol. 193 (1987) 775–791.

[25] P. Tuffery, C. Etchebest, S. Hazout, R. Lavery, A new approach to the rapid determination of protein side chain conformations, J. Biomol. Struct. Dyn. 8 (1991) 1267–1289.

[26] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kolman, AMBER: a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules, Comput. Phys. Commun. 91 (1995) 1–41.

[27] A.T. Brunger, X-PLOR Version 3.1: A System for X-ray Crystallography and NMR, Yale University Press, New Haven, CT, 1992.

[28] M. Nilges, A.T. Brunger, Automated modeling of coiled coils: application to the GCN4 dimerization region, Protein Eng. 4 (1991) 649–659.

[29] M. Nilges, A.T. Brunger, Successful prediction of the coiled-coil geometry of the GCN4 leucine zipper domain by simulated annealing: comparison to the X-ray structure, PROTEINS: Struct. Funct. Genet. 15 (1993) 133–146.

[30] I.D. Kerr, R. Sankaramakrishnan, O.S. Smart, M.S.P. Sansom, Parallel helix bundles and ion channels: molecular modelling via

simulated annealing and restrained molecular dynamics, Biophys. J. 67 (1994) 1501–1515.
[31] D.W. Banner, M. Kokkinidis, D. Tsernoglou, Structure of the ColE1 ROP protein at 1.7 Å resolution, J. Mol. Biol. 196 (1987) 657–675.
[32] S.J. Marrink, O. Berger, D.P. Tieleman, F. Jaehning, Adhesion forces of lipids in a phospholipid membrane studied by molecular dynamics simulations, Biophys. J. 74 (1998) 931–943.
[33] H.J.C. Berendsen, D. Van der Spoel, R. Van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56.
[34] O. Berger, O. Edholm, F. Jahnig, Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature, Biophys. J. 72 (1997) 2002–2013.
[35] D.P. Tieleman, H.J.C. Berendsen, M.S.P. Sansom, An alamethicin channel in a bilayer: molecular dynamics simulations, Biophys. J. 76 (1999) 1757–1769.
[36] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, in: B. Pullman (Ed.) Intermolecular Forces, Reidel, Dordrecht, The Netherlands, 1981, pp 331–342.
[37] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472.
[38] H.J.C. Berendsen, J.P.B. Postma, W.F. Van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[39] J.-L. Popot, D.M. Engelman, Membrane protein folding and oligomerization: the two-state model, Biochemistry 29 (1990) 4031–4037.
[40] D.C. Rees, L. DeAntonio, D. Eisenberg, Hydrophobic organization of membrane proteins, Science 245 (1989) 510–513.
[41] W.F. DeGrado, C.M. Summa, V. Pavone, F. Nastri, A. Lombardi, De novo design and structural characterization of proteins and metalloproteins, Annu. Rev. Biochem. 68 (1999) 779–819.
[42] A.D. McLachlan, M. Stewart, Tropomyosin coiled-coil interactions: evidence for an unstaggered structure, J. Mol. Biol. 98 (1975) 293–304.
[43] P.B. Harbury, T. Zhang, P.S. Kim, T. Alber, A switch between two-, three- and four-stranded coiled coils in GCN4 leucine zipper mutants, Science 262 (1993) 1401–1407.
[44] M. Lemmon, J.M. Flanagan, J.F. Hunt, B.D. Adair, B.J. Bormann, C.E. Dempsey, D.M. Engelman, Glycophorin A: dimerization is driven by specific interactions between transmembrane α-helices, J. Biol. Chem. 267 (1992) 7683–7689.
[45] F.X. Zhou, M.J. Cocco, W.P. Russ, A.T. Brunger, D.M. Engelman, Interhelical hydrogen bonding drives strong interactions in membrane proteins, Nature Struct. Biol. 7 (2000) 154–159.
[46] H. Gratkowski, J.D. Lear, W.F. DeGrado, Polar side chains drive the association of model, transmembrane peptides, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 880–885.
[47] J.U. Bowie, Helix packing in membrane proteins, J. Mol. Biol. 272 (1997) 780–789.
[48] T. Haltia, E. Freire, Forces and factors that contribute to the structural stability of membrane proteins, Biochem. Biophys. Acta 1228 (1995) 1–27.
[49] C. Landolt-Marticorena, K.A. Williams, C.M. Deber, R.A.F. Reithmeier, Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins, J. Mol. Biol. 229 (1993) 602–608.
[50] W.F. DeGrado, Z.R. Wasserman, J.D. Lear, Protein design, a minimalist approach, Science 243 (1989) 622–628.
[51] G. von Heijne, Membrane proteins: from sequence to structure, A. Rev. Biophys. Biomol. Struct. 23 (1994) 167–192.
[52] M. Monne, I.M. Nilsson, M. Johansson, N. Elmhed, G. von Heijne, Positively and negatively charged residues have different effects on the position in the membrane of a model transmembrane helix, J. Mol. Biol. 284 (1998) 1177–1183.
[53] R.A.F. Reithmeier, Characterization and modeling of membrane proteins using sequence analysis, Curr. Opin. Struct. Biol. 5 (1995) 491–500.
[54] W.M. Yau, W.C. Wimley, K. Gawrisch, S.H. White, The preference of tryptophan for membrane interfaces, Biochemistry 37 (1998) 14713–14718.
[55] A. Ridder, S. Morein, J.G. Stam, A. Kuhn, B. de Kruijff, J.A. Killian, Analysis of the role of interfacial tryptophan residues in controlling the topology of membrane proteins, Biochemistry 39 (2000) 6521–6528.
[56] S. Mall, R. Broadbridge, R.P. Sharma, A.G. Lee, J.M. East, Effect of aromatic residues at the ends of transmembrane α-helices on helix interactions with lipid bilayers, Biochemistry 39 (2000) 2071–2078.
[57] J.W. Seale, R. Srinivansan, G.D. Rose, Sequence determinants of the capping box, a stabilizing motif at the N-termini of α-helices, Protein Sci. 3 (1994) 1741–1745.
[58] E.T. Harper, G.D. Rose, Helix stop signals in proteins and peptides: the capping box, Biochemistry 32 (1993) 7605–7609.
[59] V. Muñoz, J. Blanco, L. Serrano, The hydrophobic-staple motif and a role for loop-residues in α-helix stability and protein folding, Nature Struct. Biol. 2 (1995) 380–385.
[60] M.A. Jimenez, V. Muñoz, M. Rico, L. Serrano, Helix stop and start signals in peptides and proteins: the capping box does not necessarily prevent helix elongation, J. Mol. Biol. 242 (1994) 487–496.
[61] H.X. Zhou, P. Lyu, D.E. Wemmer, N.R. Kallenbach, Alpha helix capping in synthetic model peptides by reciprocal side chain–side chain interactions: evidence for an N-terminal capping box, PROTEINS: Struct. Funct. Genet. 18 (1994) 1–7.
[62] R. Aurora, G.C. Rose, Helix capping, Protein Sci. 7 (1998) 21–38.
[63] P.F. Predki, V. Agrawal, A.T. Bruger, L. Regan, Amino acid substitutions in a surface turn modulate protein stability, Nature Struct. Biol. 3 (1996) 54–58.
[64] D.P. Tieleman, S.J. Marrink, H.J.C. Berendsen, A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems, Biochim. Biophys. Acta 1331 (1997) 235–270.
[65] L.R. Forrest, M.S.P. Sansom, Membrane simulations: bigger and better, Curr. Opin. Struct. Biol. 10 (2000) 174–181.
[66] L.R. Forrest, A. Kukol, I.T. Arkin, D.P. Tieleman, M.S.P. Sansom, Exploring models of the Influenza M2 channel — molecular dynamics simulations in a lipid bilayer, Biophys. J. 78 (2000) 55–69.
[67] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[68] H.J.C. Berendsen, S. Hayward, Collective protein dynamics in relation to function, Curr. Opin. Struct. Biol. 10 (2000) 165–169.
[69] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Essential dynamics of proteins, PROTEINS: Struct. Funct. Genet. 17 (1993) 412–425.
[70] S. Hayward, H.J.C. Berendsen, Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme, PROTEINS: Struct. Funct. Genet. 30 (1998) 144–154.
[71] B.L. De Groot, D.M.F. Van Aalten, A. Amadei, H.J.C. Berendsen, The consistency of large concerted motions in proteins in molecular dynamics simulations, Biophys. J. 71 (1996) 1707–1713.
[72] B.L. De Groot, D.M.F. Van Aalten, R.M. Scheek, A. Amadei, G. Vriend, H.J.C. Berendsen, Prediction of protein conformational freedom from distance constraints, PROTEINS: Struct. Funct. Genet. 29 (1997) 240–251.
[73] C.T. Choma, G.T. Robillard, D.R. Englebretsen, Synthesis of hydrophobic peptides: an Fmoc solubilizing tail method, Tetrahedron Lett. 39 (1998) 2417–2420.
[74] D.R. Englebretsen, C.T. Choma, G.T. Robillard, Synthesis of a designed transmembrane protein by thioether ligation of solubilized segments, Tetrahedron Lett. 39 (1998) 4929–4932.
[75] F.A. Kovacs, J.K. Denny, Z. Song, J.R. Quine, T.A. Cross, Helix tilt of the M2 transmembrane peptide from influenza A virus: an intrinsic property, J. Mol. Biol. 295 (2000) 117–125.

[76] J. Torres, P.D. Adams, I.T. Arkin, Use of a new label C-13 = O-18 in the determination of a structural model of phospholamban in a lipid bilayer: spatial restraints resolve the ambiguity arising from interpretations of mutagenesis data, J. Mol. Biol. 300 (2000) 677–685.
[77] M. Dathe, M. Schumann, T. Wieprecht, A. Winkler, M. Beyermann, E. Krause, K. Matsuzaki, O. Murase, M. Bienert, Peptide helicity and membrane surface charge modulate the balance of electrostatic and hydrophobic interactions with lipid bilayers and biological membranes, Biochemistry 35 (1996) 12612–12622.
[78] T. Kiyota, S. Lee, G. Sugihara, Design and synthesis of amphiphilic alpha-helical model peptides with systematically varied hydrophobic-hydrophilic balance and their interaction with lipid- and bio-membranes, Biochemistry 35 (1996) 13196–13204.
[79] P.D. Adams, A.T. Brunger, Towards prediction of membrane protein structure, in: G. von Heijne (Ed.), Membrane Protein Assembly, Chapman and Hall, New York, 1997, pp 251–267.
[80] C. Etchebest, J.-L. Popot, Packing transmembrane alpha-helices into bundles: computational versus experimental approaches, in: G. von Heijne (Ed.), Membrane Protein Assembly, Chapman and Hall, New York, 1997, pp 22–251.
[81] G.R. Dieckmann, W.F. DeGrado, Modeling transmembrane helical oligomers, Curr. Opin. Struct. Biol. 7 (1997) 486–494.
[82] M.S.P. Sansom, Models and simulations of ion channels and related membrane proteins, Curr. Opin. Struct. Biol. 8 (1998) 237–244.
[83] A. Killian, G. von Heijne, How proteins adapt to a membrane–water interface, TIBS 25 (2000) 429–452.