Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 23 (2004) 99–110
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Structure prediction of small transmembrane helix bundles

Kay-Eberhard Gottschalk *

Department of Biological Chemistry, Weizmann Institute of Science, Herzl St 1, 76100 Rehovot, Israel

Received 22 September 2003; received in revised form 4 January 2004; accepted 19 February 2004
Available online 10 April 2004

Abstract

In this work, we will introduce a novel computational approach to predict the structures of small helical hetero-oligomeric transmembrane bundles. The approach is based on the generation and evaluation of a large library of randomly generated helix bundle conformations. This library will be evaluated by energy-dependent distributions of the structural parameters of the conformations. The approach enables us to model a subunit of cytochrome *c* oxidase (occ), consisting of four TM helices, to an accuracy of 1.7 Å and the transducer protein of the sensory Rhodopsin II–transducer complex to an accuracy of 2.3 Å when including two transducer-contacting Rhodopsin helices. As the approach does not afford a unique solution for each protein, experimental data would be needed to discriminate the possible models. In addition to predicting the structure of helix bundles, one can also gain insight into possible higher-energy conformations or flexible regions of the protein.

Keywords: Molecular modelling; Structure prediction; Membrane protein; Conformational search; Folding emulation

1. Introduction

Transmembrane proteins are pivotal for life. They transmit signals through the cell membrane and regulate the chemical composition of the cell lumen. Despite their importance, structural studies on TM proteins are rare. Experimental difficulties in crystallizing proteins in a membraneous environment or their stabilization in conditions suitable for NMR studies hamper structural investigations on TM proteins. Therefore, computational approaches are desirable. These can provide low-resolution structural information guiding further experimental studies as well as lead to a deeper understanding of TM protein function [1].

While the prediction of membrane-spanning helices at the sequence level as well as the prediction of the topology of these helices is well established [2–7], structure prediction of membrane proteins is hindered by the higher complexity of the problem.

Nevertheless, the low-dielectric environment of the membrane facilitates computational modeling as it restricts the secondary structure of the membrane buried parts of the proteins to either all-helical or beta-barrel structures. The anisotropy of the environment reduces further the degrees of freedom for the arrangement of the secondary structure elements [8].

Since computational modeling of proteins in principle constitutes a folding emulation, considerations of possible folding pathways might simplify the computational approach. A conceptual advantage of membrane protein modeling is the simplicity of the two-stage model of TM protein folding. The model states that individually stable, membrane-spanning helices are formed in the first step of folding and then aggregate to the final structure in the second step of folding. For computational studies, only the second step has to be emulated; i.e., the computational problem is reduced to optimally pack helices against each other [9–11]. In case of multimers, the monomers can hence be modeled independently from each other and put together in a second step.

It has been shown previously that a global search of the conformational space in combination with simulated annealing/molecular dynamics calculations of helix dimers resulted in a near-native structure for the homo-dimeric TM protein glycophorin A (GpA) [12]. The same approach was extended to model the homo-pentameric TM protein phospholamban [8]. It was demonstrated that a search of the

* Tel.: +972-8-934-3199; fax: +972-8-934-4118.
E-mail address: kay.gottschalk@weizmann.ac.il.(K.-E. Gottschalk)

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2004.02.002

conformational space of helix–pair interactions of phospholamban results in the same model as a symmetric search of the whole pentamer, indicating that in homo-oligomeric proteins the many-body problem of TM helix packing can potentially be reduced to a number of two-body problems. A related approach for modeling of homo-oligomeric proteins by a two-body search with subsequent clustering has been introduced recently [13]. Normally, experimental data are used after the initial modeling step to distinguish between possible models. An approach, which incorporates the biochemical data already into the modeling step, resulted in a slightly different model of phospholamban [14]. Also other homo-oligomeric bundles of TM helices have been modeled extensively with simulated annealing techniques, using a high number of experimental constraints to guide the modeling procedure [13,15–23].

The gating mechanism of the mechanosensitive channel MscI has been elegantly modeled using the X-ray structure and different biochemical data as input [24,25]. Another field, which has been extensively explored by different modeling techniques, is the structure as well as gating and translocation mechanisms of transmembrane channel proteins [26–38].

Molecular dynamic studies with and without membrane have furthermore explored the different structural possibilities as well as ion permeation capabilities of pore forming peptides like alamethicin or gramicidin [39–43].

Ab initio modeling without experimental data is still in its infancy for transmembrane proteins. A potential smoothing algorithm could successfully and unambiguously predict the structure of GpA [44], while recently a low-resolution prediction of the structure and topology of small transmembrane proteins using a modified version of a program originally designed for soluble proteins was successful by including membrane specific energy terms [45].

Four-helix bundles are an interesting target for membrane protein prediction for a variety of reasons. An important class of membrane proteins, the small multidrug resistance (SMR) family of multidrug transporters, consists of four tightly packed helices. Other important protein families containing four TM helices include the transmembrane Helices 2–5 from rotary ATP synthase, the proton channel of the energy-transducing nicotineamide transhydrogenase of *Escherichia coli* and the transmembrane domain M2 from influenza A virus. From a practical point of view, four-helix bundles have the advantage that the packing of the helices is much less ambiguous than for five-helix proteins: while for five helices, a variety of packing arrangements like a symmetrical ring-like structure (for example phospholamban) or a distorted rings or even a central helix surrounded by four outer helices can be imagined, four helices most probably pack into a compact bundle structure.

Nevertheless, the structural complexity of hetero-oligomeric helix bundles is still a major challenge in the field of transmembrane protein structure prediction [1,46]. Here a novel approach to optimize the packing of helices in hetero-oligomeric helix bundles is described and tested for two proteins: a subunit of the cytochrome *c* oxidase (occ) [47] and the transducer protein of the Rhodopsin–transducer complex [48] (Fig. 1). The described approach is based on a library of random conformations coupled with simple energy considerations as *the* discriminatory function. While the approach does find near-native structures for these bundles, the solution is not unambiguous. Therefore, experimental data are needed to discriminate between the possible outcomes. The work is conceptually related to a previous method for calculating structural changes of transmembrane proteins [49] by treating the helices as rigid bodies and performing a search of the available conformational space.

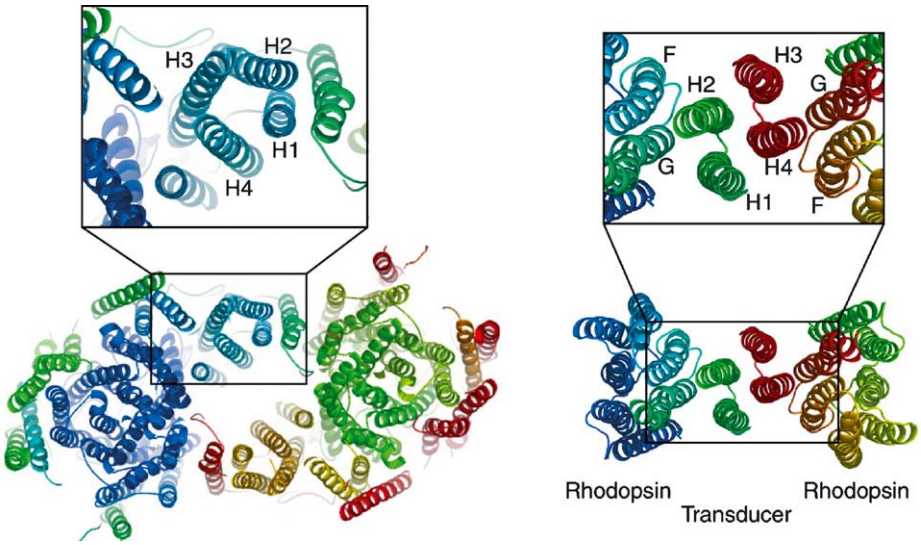

Fig. 1. Model systems. The approach was validated on these two model system: first on a subunit of cytochrome *c* oxidase (left) and second on the transducer protein of the Rhodopsin–transducer complex (right). Two calculations, one with Helix F and Helix G and one without, have been performed for the transducer protein.

2. Methods

2.1. General procedure

Our general procedure to model four-helix bundles consists of three steps: as a first step, a large library of random structures is generated by rigid-body displacement of the four helices in the bundle. In a second step, a brief energy minimization to resolve major clashes is applied. In a third step, the energy serves as criterion for a frequency-based analysis of low-energy conformations.

The new method is tested on two different four-helix bundles. The two test cases are a four-helix bundle forming subunit of cytochrome *c* oxidase and the transducer protein of the sensory Rhodopsin II–transducer complex.

An initial test has been performed on occ, starting from the X-ray structure and changing just the rotation angle of each helix. In this initial test the rotations around the long helix axis were performed randomly, and 10,000 structures with random rotational orientations were obtained. Since the complex membranous environment is not included in the calculation, it is not clear whether or not helix–helix interactions are sufficient to identify near-native conformations. This test shall establish whether in general simple energy criteria are sufficient to distinguish between near-native and non-native conformations. Since the start conformation is the X-ray structure, the sidechains are correctly positioned. Therefore, the test cannot show whether the minimization scheme is sufficient for cases where the sidechain position is not known. This is tested later. Equally, it is not meant to provide a complete description of the energy landscape of a four-helix bundle.

This initial test is further extended to a more realistic challenge: not the X-ray structure, but a random bundle of four canonical α-helices served as input structure. All degrees of freedom were changed. This realistic test was performed on occ and on the transducer complex. Here, the sidechain conformation was consistent with a helical secondary structure, but no information of the tertiary structure was used for sidechain positioning. While the initial test (starting from the X-ray structure) is meant to test whether the interaction energy can in principle distinguish between correct and wrong structures, here it is tested whether the minimization scheme is sufficient for cases in which the sidechain conformation is not known.

Since the transducer protein is structurally coupled to the Rhodopsin, two calculations have been performed for the transducer complex: one including Rhodopsin helices, the other excluding Rhodopsin helices.

In a helix bundle, each helix has four structural parameters (Fig. 2). These structural parameters have to be optimized simultaneously for the structure prediction of such a bundle. To this end, we generated a library of conformations by varying these four structural parameters in a random fashion. For each helix, a random value for vertical and lateral translation, rotation around the long axis and tilt was assigned (Fig. 2).

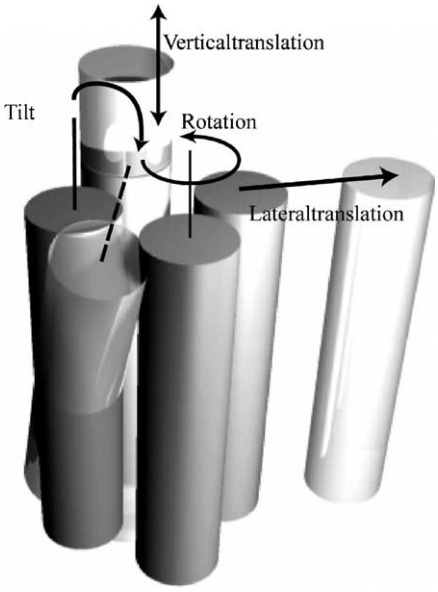

Fig. 2. Structural determinants of a four-helix bundle. The degrees of freedom in a helical bundle are the tilt- and rotation-angle as well as vertical and lateral shift. These four degrees of freedom have to be optimized simultaneously.

In our approach, the interaction energy between the helices serves to discriminate the generated conformations. The interaction energy can be calculated either as total energy, regarding all helices, or as interaction of a single helix with the rest of the protein. Distributions of structural parameters of low energy structures were evaluated. For this evaluation, energy-dependent distributions of the structural parameters were calculated. Peaks in the low-energy range of these distributions were then regarded as probable values for the specific structural parameter.

All the calculations have been performed with the molecular modeling and manipulation program CNS, version 1.1 [50], applying the OPLS forcefield parameters [51]. The energy-dependent distributions have been generated using Origin v. 6.1.

2.2. Helix bundle generation

As input for the library generation served a bundle of four caconical α-helices. The helices were built one by one as follows.

First, the backbone was built so that the Cα-atoms rotated by 98.99° per residues and had a rise of 1.5 Å per residue. The generated helix was initially minimized with 3000 steps of Powell minimization with fixed Cα-atoms. The electrostatic and the van der Waals energy term were cut off at 13 Å. A shift function was applied starting at 10 Å to ensure a smooth energy transition to zero. Then 500 steps of a molecular dynamics simulation in Cartesian space with NOE restraints of 2.8 Å between N_i and O_{i+4}

and of 1.8 Å between H_i and O_{i+4} with the temperature bath set to 300 K, and a timestep of 0.5 fs followed. This was followed by 3000 steps of Powell minimization with the van der Waals repel term set to 1.5, followed by 3000 steps of Powell minimization using the standard van der Waals term, followed by another 3000 steps of Cartesian dynamics at a temperature of 300 K with 3000 steps and a timestep of 0.5 fs. During all these steps the Cα-atoms were kept fixed at their position. The helices were then minimized with 1000 steps of Powell minimization releasing the Cα-atoms, using NOE restraints of 2.8 Å between N_i and O_{i+4} and of 1.8 Å between H_i and O_{i+4}, followed by 1000 steps of a Cartesian dynamics calculation coupled to a temperature bath at 300 K with a stepsize of 0.5 fs, followed by another 1000 steps of Powell minimization. The applied building scheme assures that the $\chi 1$ angles are in accordance with an α-helical secondary structure. This restricts the rotamers of β-branched sidechains like I or V to a single value, but does not determine the rotamers of all the sidechains unambiguously. The single helices were than put together to form a bundle by separating the helix centers to 10.4 Å at the appropriate angle (90° between three neighboring helix centers for a four-helix bundle). No further minimization was applied for the start structure. Only the random conformations generated in the next step were further minimized.

3. Library generation

3.1. Starting from X-ray structure

In a first test, the crystal structure of a subunit of cytochrome *c* oxidase (1occ.pdb) was taken as input structure. Ten thousand structures were calculated with randomly assigned rotations around the central helix axis. The central helix axis was calculated as the vector between the center of mass of the backbone atoms of the first and last seven residues (two helical turns) of each helix. After assignment of the random values, a short Powell minimization with 50 steps and a harmonic restraint on the C, Cα and N atoms of the protein was performed. The electrostatic energy terms were cut off at 13 Å. The energy was smoothly reduced to zero starting at 10 Å *using a shift function*.

3.2. Starting from random structure

In a second test, not a crystal structure, but a bundle of four canonical α-helices with the sequence of cytochrome *c* oxidase or the transducer protein was the input structure. The bundle had a random rotational orientation, with an initial tilt of the helix axis of 0° relative to the symmetry axis, with a helix–helix distance of 10.5 Å and with the centers of the helices positioned at the same height. Starting from this structure, 200,000 structures were generated by rigid-body displacements of the helices. For these rigid-body displacements, different tilt angles, rotation angles, vertical and horizontal shifts to each helix were randomly assigned. Two different ranges of tilt angles were tested for occ. First, the assigned tilt angle had a range of $20 \pm 20^\circ$, and second it had a range from $0 \pm 20^\circ$. For the transducer complex, only a tilt angle range of $0 \pm 20^\circ$ was tested. The rotation angle had a range of 360°, the horizontal shift of 3 Å towards the center of the helix bundle and 5 Å away from center and the vertical shift 0 ± 4 Å. The center of the helix bundle was defined as the center of mass of all backbone atoms.

The sidechains were initially positioned so that they are in accordance with a helical secondary structure. This restraints the rotamers of β-branched residues to just a single value. The rigid-body displacements were followed by a brief Powell minimization with 50 steps and a harmonic constraint with a force constant of 10 kcal/mol on the C, Cα and N atoms of the protein was performed and the interaction energies between the helices calculated. This minimization does not include a thorough sidechain positioning. In initial tests we tried to position the sidechains either using the dead end elimination theorem or a mean field approach. Both methods proved to be too time consuming.

Although the transducer protein is a homo-dimeric protein with two helices per monomer, no symmetry was introduced in the generation of the conformational library and therefore the individual conformations were not symmetric. Two calculations have been performed for the transducer protein: one with only the protein, another with two contacting Rhodopsin helices per monomer. In the calculation including these two helices, their X-ray structure was taken as input and they were not changed during the library generation step.

For occ, the modeled helices ranged from residues C129–C148 (Helix 1), C161–C180 (Helix 2), C200–C219 (Helix 3) and C234–C253 (Helix 4). For the homo-dimeric transducer protein, the modeled helices ranged from residues 24–42 (Helix 1 and Helix 3) and from residues 60–78 (Helix 2 and Helix 4). The two transducer-contacting Rhodopsin helices range from residue 165–180 (Helix F) and 190–204 (Helix G).

4. Library evaluation

4.1. Starting from X-ray structure

For the evaluation of the structure library, the total non-bonded energy between the helices as well as the non-bonded interaction of each helix with the other three helices was calculated for the first calculation of the cytochrome *c* oxidase. Electrostatic energy term was cut of at 13 Å, using a shift function starting at 10 Å to smoothly reduce the energy to zero. The backbone atoms were restrained with a harmonical potential with a force constant of 10 kcal/mol.

4.2. Starting from random bundle

Only the interaction between each single helix with the remaining of the bundle was calculated. The 1000 lowest energy structures were evaluated for each helix. A three-dimensional histogram for each of the four structural parameters (tilt angle, rotation angle, vertical shift and horizontal shift) was generated using Origin v. 6.1. In these histograms, the respective structural parameter is on the *x*-axis, the energy on the *y*-axis and the frequency on the *z*-axis. Peaks in each distribution of the structural parameter were determined by visual inspection. The input structure was then changed according to these peak-values (as in Fig. 3). Afterwards a Powell minimization with 1000 steps including NOE-restraints for helicity and helix–helix axis restraints was performed. For maintaining helicity, the distance between N_i and O_{i+4} was restrained to 3.1 Å, while the centers of the first and last seven residues of neighboring helices were restrained to be less than 14 Å apart, respectively.

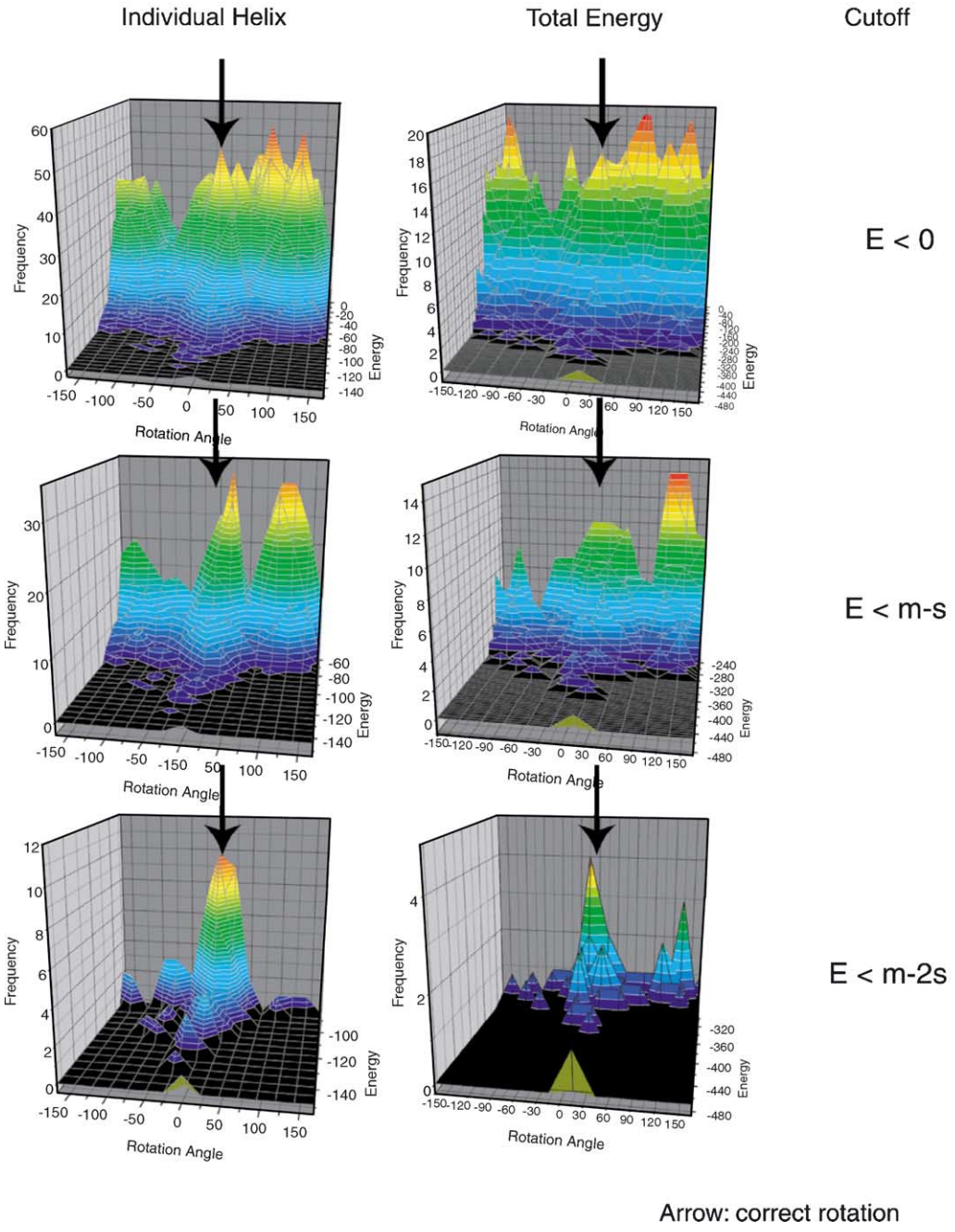

Fig. 3. Energy dependency of rotation angle distribution. As the crystal structure served as input structure for this calculation, a narrow peak around 0° is expected as a good result. The lower the energy cutoff for the conformations taken into consideration, the better is the result. The agreement is better for the interaction energy calculated for Helix 1 with the other helices (left) than for the total non-bonded energy (right).

5. Results

5.1. Interaction energy as discriminatory criterion

It has been shown previously that in a pool of possible conformations of transmembrane proteins, the correct conformation resides in an energy-minimum, even if the membrane is not included in the calculations [52]. In order to check the ability of the interaction energy to discriminate between different conformations, only the rotation angle of each of the helices of the four-helix bundle of the subunit of cytochrome *c* oxidase was randomly changed. Since the X-ray structure was the starting point in these calculations, an ideal result would be if the interaction energy leads to a narrow distribution of rotation angles centered at 0°. The following discussion is restricted to Helix 1 of the bundle as an example.

If all structures are considered, the rotational angles are equally often populated (data not shown). As shown exemplarily for Helix 1, the rotation angle distribution is strongly dependent on the interaction energy of the considered conformations (Fig. 3): if all conformations with an interaction energy smaller than zero are considered, no preferential angle can be deduced. The lower the energy cutoff of the considered conformations, the better is the agreement between rotation angle distribution and experimental structure: at a cutoff of the interaction energy of $E \leq m-s$, with m being the mean and s the standard deviation of a Gaussian approximation of the interaction energy distribution, two preferred angle can be observed, accompanied by a less preferred rotation angles using the non-bonded interaction energy of Helix 1 as criterion. On the other hand, using the total non-bonded energy as criterion, one preferred angle at a wrong value and a less preferred angle at the correct orientation can be observed.

At a cutoff of $E \leq m-2s$, one angle at 0° is clearly preferred. This preference is much more pronounced when using just the non-bonded interaction energy of Helix 1 instead of the total non-bonded interaction energy as discriminative criterion. Therefore, the interaction-energy of a single helix with the rest of the protein favors native-like values for low-energy conformations of this helix and suppresses wrong conformations, while the suppression of wrong values is less efficient when regarding the total non-bonded interaction energy. Probably, the total energy can be low even when the regarded helix is misplaced, provided that the other helices are in a correct conformation. The individual interaction energy on the other hand requires the regarded helix to be placed correctly. Nevertheless, the lowest peak is slightly off the correct rotation, though only 25°. This offset is within the error of the method. Furthermore, we do not exclude the possibility that certain conformations exist, which display a very low energy although they have the wrong orientation. Nevertheless, it is shown here that in the ensemble of low-energy structures the near-native values are predominant.

The very distinct peak at 0° stems partially from a bias of the calculation: since the correct sidechain conformation was used, no clashes occurred at the correct orientation. It is not clear from this calculation whether the minimization scheme is sufficient when the sidechain positions are not correct.

'Real' helices are not ideal and it cannot be predicted how the helices deviate from a canonical α-helix. Therefore, in the following calculations the pool of random conformations was generated starting from canonical α-helices in a random start orientation with random sidechain orientations, which are compatible with a helical secondary structure. While the first test checked whether or not the interaction energy can in principle distinguish between near-native and non-native conformations, here it is tested whether the limited minimization applied after rigid-body displacement is sufficient to detect near-native conformations.

5.2. The library of random conformations

As the starting structure served a bundle of four canonical α-helices. Although different stereo-isomers of a four-helix bundle exist, only the correct orientation was calculated in order to save computer time. Starting from this structure, a library of random conformations with 2×10^5 members was generated as described.

The lowest energy structure of these 2×10^5 conformations has a RMSD of the Cα-atoms to the crystal structure of 3.2 Å, demonstrating that the total energy of a single structure is not a good indicator of structural correctness.

Therefore, distributions of structural parameters of the 1000 lowest-energy conformations were further evaluated, assuming that in low-energy structures the predominant structural parameters are native-like. The energy was calculated for each helix individually, so that for two different helices different sets of conformations were evaluated. The energy-dependent distributions of each of the parameters of the 1000 lowest-energy conformations have been visualized for each helix (Fig. 4).

For some structural parameters, no unique solution was found. Helix 2 and Helix 3 have two peaks in the rotation angle distribution. Also the longitudinal shift of Helix 2 has two maxima. This indicates that either interactions to other helices are important for the correct orientation of Helix 2 and Helix 3, or that other influences apart from helix–helix interactions, like helix–loop or helix–membrane interactions, are structure-determining. It might furthermore indicate certain flexibility in the structure, hinting to possible structural sub-states of the system. In general it seems as if the rotation angles are rather well defined by this approach, while the other degrees of freedom, especially the longitudinal shift, have broader distributions. Broader distributions might result in greater errors.

The appearance of multiple peaks per structural degree of freedom leads to multiple possible models. Eight different combinations are possible here. When adjusting the

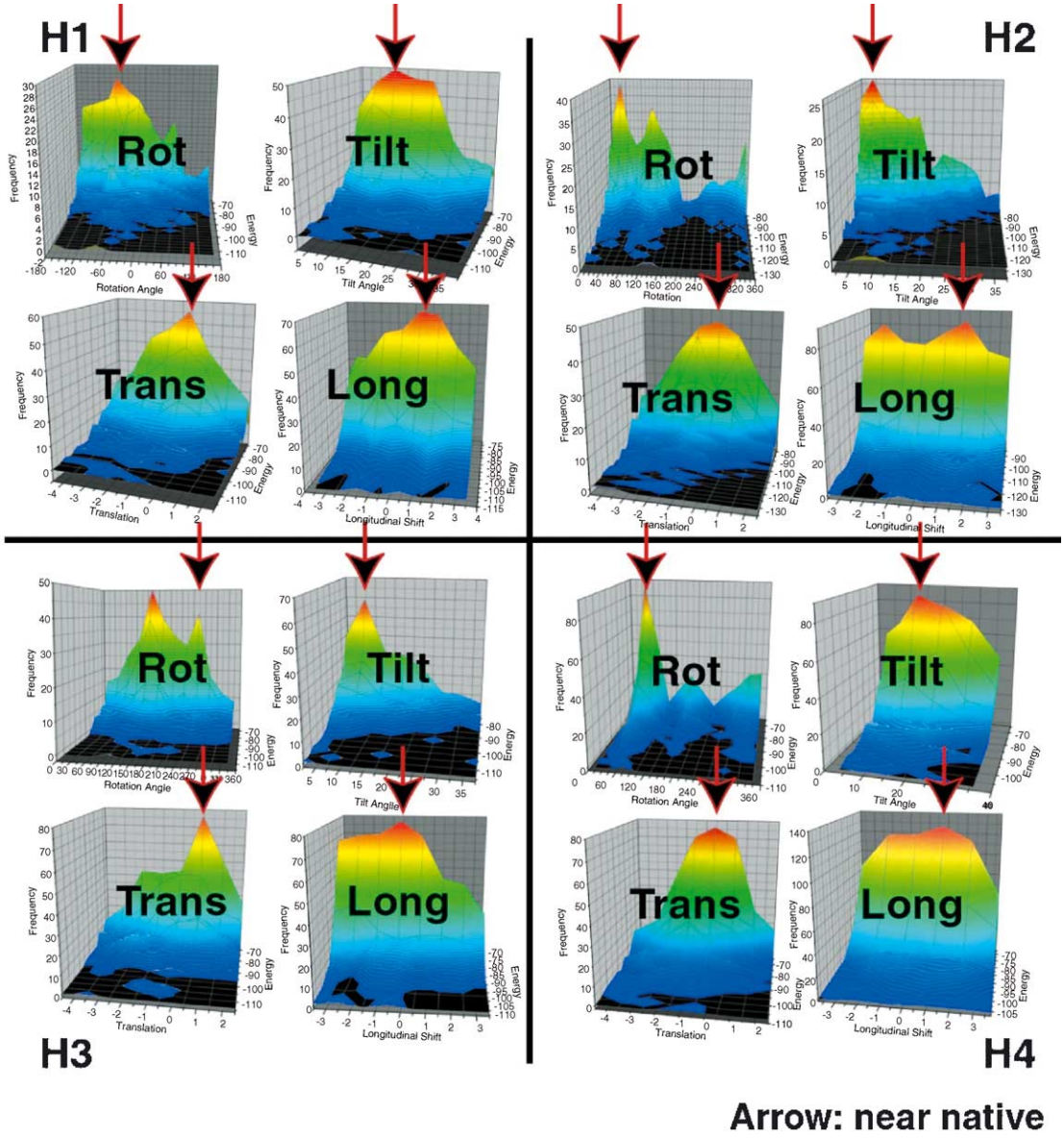

Fig. 4. Energy-dependent distributions of structural parameters. The peaks in the energy-dependent distributions of all four structural parameters relate to highly populated low-energy conformations. The values of these maxima were taken to obtain the models for occ. The near-native values of the 1.7 Å model are highlighted with an arrow.

structural parameters of the input structure according the eight different combinations possible, the range of Cα-RMSDs between model and X-ray structure is 1.7–3.1 Å. That means that even the worst case is slightly better than the lowest energy structure (RMSD of 3.2 Å). The initial structure has a RMSD of 3.6 Å. The best RMSD in our ensemble (1.7 Å) demonstrates that the modeling approach is powerful and finds a near-native structure (Fig. 5A). The structural parameters of this native-like structure are pointed out by an arrow in the parameter distributions shown in Fig. 4. It can be observed that the peaks are much less distinct in this real case scenario than the peak at the first test. This stems most probably from mis-placed sidechains, which potentially might disfavor native-like conformations. And yet, despite the very limited minimization, peaks corresponding to near-native conformations are found (Fig. 4). This indicates that the limited minimization is sufficient for the identification of correct conformations.

In a real modeling case, experimental data are mandatory to distinguish between the obtained conformations. This might be performed by mutational or spectroscopic techniques [12,14,17].

5.3. *Influence of tilt angle*

In the previous *section*, the tilt angle was varied from 0 to 40°, assuming a left-handed helix bundle. Although the handedness can be deduced from mutational experiments or conservation patterns of the amino-acid sequence of the helices [53–55], it is interesting to test how the assigned tilt

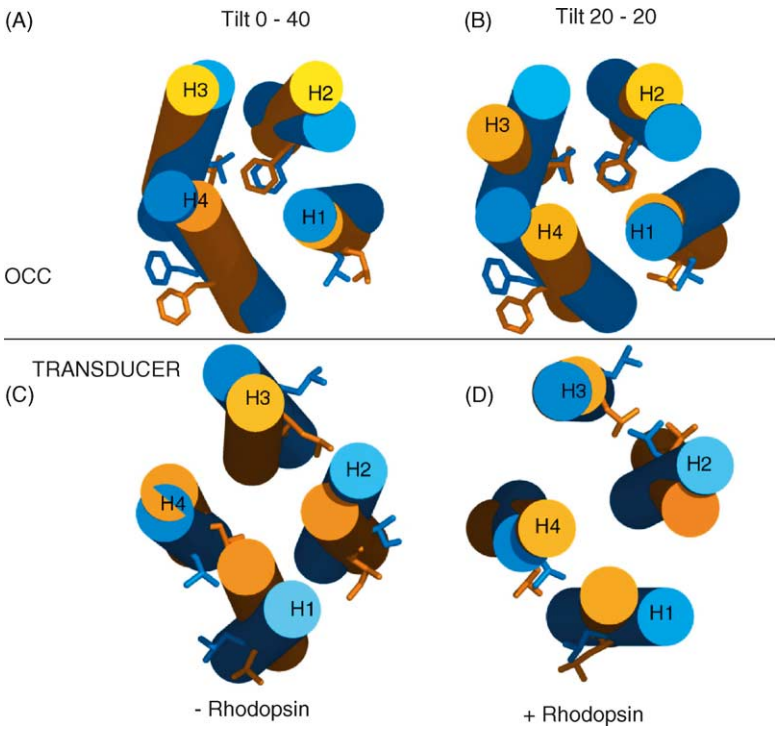

Fig. 5. Superimposition of best models on X-ray structures. The models are shown in orange, the experimental structure in blue. To indicate the rotation angle, one sidechain per helix is depicted: (A) occ-calculation with tilt angle range from 0 to 40°; RMSD: 1.7 Å; (B) occ-calculation with tilt angle range from −20 to 20°; RMSD: 2.5 Å; (C) transducer-calculation without complexing Rhodopsin helices; RMSD: 2.9 Å; (D) transducer-calculation with complexing Rhodopsin helices; RMSD: 2.3 Å.

angle values influence the overall result. Preliminary tests showed that a variation of more than 40° would need an even larger structural library in order to obtain good statistics. This is not possible within a reasonable timeframe. Hence, we assigned values ranging from −20 to 20° for the tilt angle. This still entails the correct value of ∼10°, but also includes right-handed tilt angles as a possible source of error.

The results are not quite as good as those of the previous section. Although the correct value for the rotation angle has been found for each helix, the peaks in the tilt angle distribution of Helix 3 and Helix 4 are at a negative (right-handed) value. The best model obtained has an RMSD of 2.5 Å (Fig. 5B). The main error in the model is the tilt angle of Helix 3 and Helix 4. 2.5 Å is still a satisfactory result, especially bearing in mind that the rotational orientation of the helices has been modeled correctly. This test nevertheless demonstrates that the procedure is robust concerning the rotation angle, but less so for the tilt angle. The procedure is therefore especially well suited for systems for which one has indications about the average tilt of the helices relative to the membrane normal.

5.4. The free transducer protein of the Rhodopsin–transducer complex

Recently, the structure of the sensory Rhodopsin II–transducer complex has been reported [48]. In the X-ray structure, the four-helix bundle of the transducer protein is sandwiched between two Rhodopsin molecules (Fig. 1, right). The transducer protein is assumed to undergo structural re-arrangements upon signaling, initiated by structural changes of the Rhodopsin molecule.

The energy-dependent parameter distributions are cleft. While for occ the helices display one to two peaks in the rotation angle distribution, they display more peaks for the transducer (Fig. 6, right panel). Also the areas between the peaks are more populated, especially for Helices 2 and 4, which contact Rhodopsin. This indicates that the protein has a rather shallow energy landscape and can switch between different conformational states.

The lowest RMSD between a minimized model and the X-ray structure is 2.9 Å (Fig. 5C). One has to bear in mind the special function of the transducer protein when evaluating this result: the protein is designed to transmit signals received from Rhodopsin. This means that the conformational state of the transducer protein should be tightly coupled to the conformational state of Rhodopsin, otherwise the signal transduction would be error prone. Here we tried to model the transducer protein without Rhodopsin while comparing the results to a crystal structure with Rhodopsin. It is not necessarily so that a free transducer protein has the same structure as a transducer protein in complexes with Rhodopsin. We therefore tried in a second step to model the transducer protein in the context of the Rhodopsin structure.

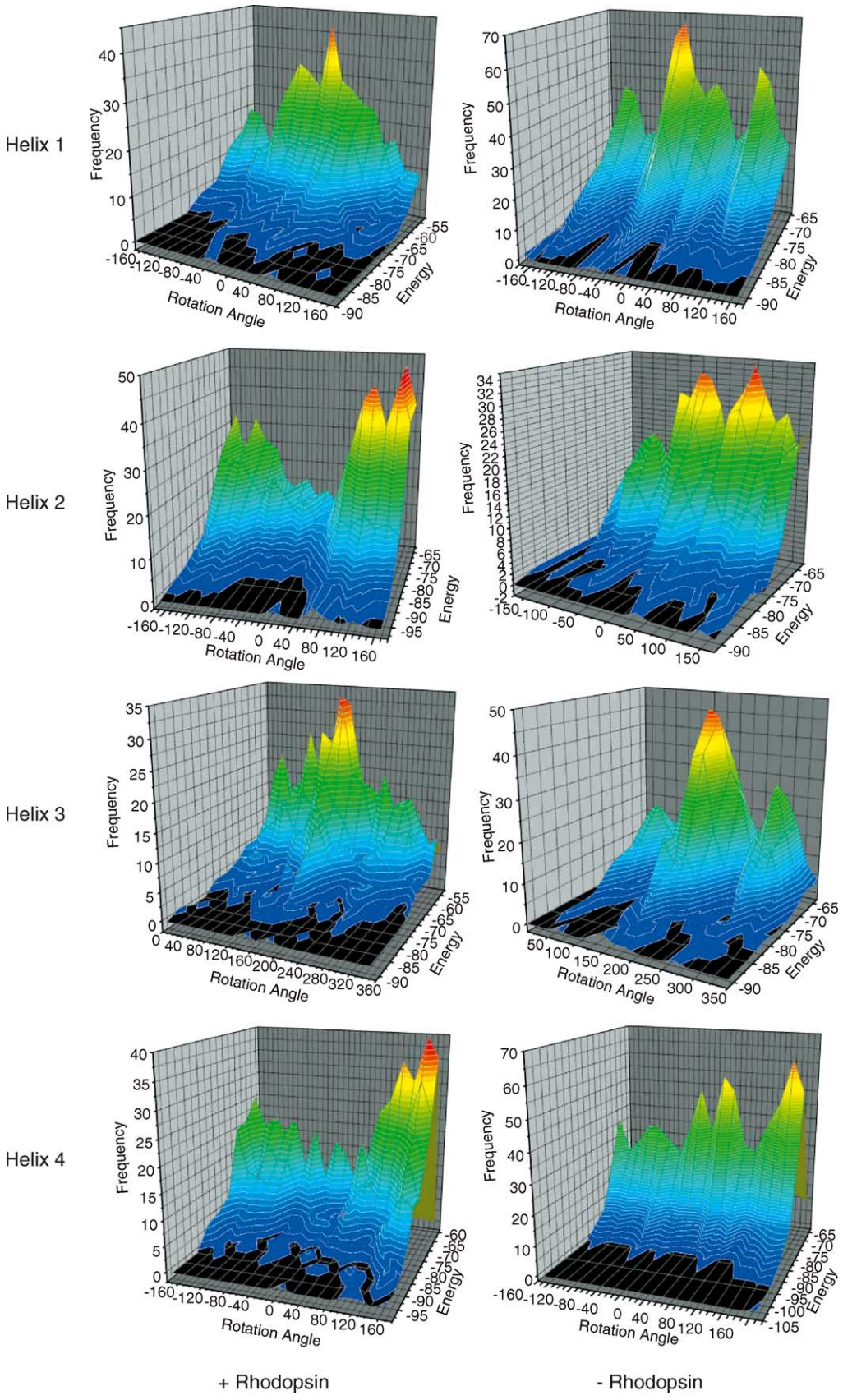

Fig. 6. Energy-dependent rotation angle distribution of the transducer protein. The rotation angle distribution for the uncomplexed transducer (right) shows especially for Helix 2 and Helix 4 (the Rhodopsin-contacting helices) many peaks and hardly any minima. The whole distribution changes dramatically when adding two complexing helices of Rhodopsin (left).

5.5. The transducer protein in complex with Rhodopsin

The transducer-contacting helices Helix F and Helix G from Rhodopsin were added to the input structure (see Fig. 1, right). While the experimentally determined conformation of the Rhodopsin-helices was kept unchanged, a random structure library of the four-helix bundle was generated as described above. Apart from the two Rhodopsin helices, the initial structure of the four-helix bundle was identical to the input structure of the previous section.

The energy-dependent distributions of the four-helix bundle change significantly from the previous calculation (Fig. 6). Helix 2 and Helix 4 of the transducer, the two helices that contact Rhodopsin, display now pronounced minima at rotation angles where peaks were observed before. Additional maxima (which turn out to be at the correct rotation angle) appear. Similar changes are true for Helix 1 and Helix 3, though to a lesser extent.

The lowest RMSD between the different minimized models and the X-ray structure is now 2.3 Å (Fig. 5D). The main error is (as in the case of occ) in the tilt angle, while the rotational orientation is modeled nearly correctly.

Apparently the energy landscape of the four-helix bundle changes significantly upon complexation with Rhodopsin, indicating that the transducer protein (a) is capable of major structural re-arrangements, and (b) probably adopts a different structure when being uncomplexed than when being in complex with Rhodopsin. This is biologically reasonable: the interior of the cell has to know whether the transducer protein is poised to transmit signals or is in an uncomplexed resting state. This corroborates the assumption that the conformation of the transducer protein is indeed tightly coupled to the conformation of Rhodopsin.

6. Discussion

This study describes the development and validation of a novel computational tool to model bundles of TM helices. The generation of a large library of possible conformations allows the sampling of nearly the complete conformational space of each helix in the bundle. As shown for the subunit of cytochrome *c* oxidase, simple energy criteria can serve to identify low energy conformations, which correspond to the true conformation of the bundle. Analysis of the structural parameters of each helix dependent on the interaction energy of each individual helix with the rest of the protein yielded a structural model of a four-helix bundle that is in good agreement with the crystal structure. To our knowledge, this is the first time that a hetero-oligomeric four-helix bundle was modeled in this accuracy with no experimental data during the modeling procedure. The ab initio prediction of transmembrane protein structures reported recently has an RMSD range of above 3 Å for helix dimers [45]. A direct comparison of the results is not possible, though, as Pellegrini-Calace et al. tackled the more complex problem of an ab initio prediction, while here a knowledge of the secondary structure and topology of the protein is assumed.

For the transducer complex, the results are not as straightforward as for occ. Without the complexing Rhodopsin helices, larger errors between the best model and the X-ray structure are found. Despite these errors, the structure still resembles the correct structure. Main differences are the transversal shift of Helix 1 and Helix 3 as well as the tilt angles (Fig. 5C). The obtained structure, although displaying a rather large RMSD, could nevertheless be useful for the design of new experiments, as the peaks in the distributions provide testable hypothesizes.

When adding the complexing Rhodopsin helices, a structure close to the reported X-ray structure is found. The differences between the results including respectively excluding the Rhodopsin helices can be rationalized in functional terms: the transducer conformation has to be coupled to the Rhodopsin conformation in order to reliably transmit signals through the cell membrane. This indicates that in addition to structural insight, one might gain insight into structural changes of the proteins. In this sense, the described approach is similar to a previously published method for describing conformational changes in transmembrane proteins [49].

The applied modeling scheme suffers from the fact that the correct sidechain configuration is not known. Preliminary attempts to more rigorously refine the sidechain conformation for each of the 200,000 members of the structure library turned out to be too time consuming to be practical. Therefore, it cannot be ruled out that the initial sidechain conformation influences the final result. But it has been demonstrated that β-branched or small residues are likely to be at the helix–helix interface [56,57]. For β-branched or small residues, the sidechain conformation is fixed due to possible sidechain-backbone clashes (for the β-branched residues like V or I) or due to the existence of just a single rotamer (as for A and G). Also for Ser and Thr, it is rather likely that hydrogen bonds with the backbone of the same helix are formed, reducing the rotameric space to just one rotamer. Large unbranched sidechains like Met or to a lesser extent Leu are nevertheless a possible source of error. Yet, the applied minimization should reduce artificially high energies if these sidechains overlap. Backbone overlaps, on the other hand, cannot be efficiently removed by the applied minimization scheme. Conformations with overlapping backbones will therefore be discarded by the evaluation procedure. The success of the modeling procedure in identifying peaks corresponding to near-native conformations indeed indicates that the limited minimization is sufficient.

Furthermore, it has been suggested that TM helices pack in a knob-into-hole fashion [58,59]. This way of packing will be favored by the applied scheme, since knobs-into-holes packing effectively reduces the danger of overlapping sidechains. Therefore, one might think that the approximation

used here, namely to allow the sidechains to adapt to the conformation during a brief minimization, but not to perform a complete search of the rotameric space, is a valid approximation.

In all cases, experimental data would be needed in order to distinguish between the possible models. While the rotation angle can be modeled rather accurately, the tilt angle is more problematic. Therefore it is desirable to have experimental information about the tilt angle, which can be incorporated into the modeling scheme. Also the vertical and lateral shifts have broader distributions. Thus, it appears as if the latter degrees of freedom are not solely determined by helix–helix interactions, but also by helix–loop or helix–membrane interactions. As a consequence of this behavior, these degrees of freedom will be modeled with less accuracy than the rotation angle.

The occurrence of more than one peak in the distributions also reflects the influence of other parameters apart from helix–helix interactions and might furthermore be a consequence of wrongly positioned sidechains. More than one maximum in the distributions or very broad distributions might also reflect a certain flexibility of the structure, possibly indicating structural sub-states of the system. Thus, in addition to gain insight into preferred orientations, one may also gain insight into possible dynamical regions as well as into the nature of helix–helix interactions. A further advantage of the method is that also non-equilibrium, higher-energy states of the system might be detectable. These conformational sub-states are usually not amenable to crystallization.

It appears advantageous to treat the interactions of each helix individually as opposed to regarding the total interaction energy. This is probably caused by the short ranging nature of the non-bonded interactions, especially the van der Waals interactions. Due to the short range of the van der Waals interactions, not the whole protein has to be in the correct conformation to generate a low energy conformation for the helix under consideration, but just that helix and its direct neighbors have to be in a conformation that is close to the correct one. So the many-body problem is effectively reduced to a three-body problem. Therefore, the method may be well suited to model even larger systems.

Experimental distance restraints can be easily incorporated by modifying the energy landscape with Nuclear Overhauser Enhancement (NOE)-like restraints. A technical advantage of the procedure is the ease with which it can be parallelized. One can either calculate 200,000 random conformations on one processor or 50,000 on each of four processors with no need to change the code of the procedure.

A disadvantage of the approach is that all helices have to be treated as rigid bodies. A simulated annealing/molecular dynamics treatment of each individual conformation is prohibited by the large number of structures in combination with the time needed for the MD calculations.

Acknowledgements

I want to thank Horst Kessler for many helpful discussions and suggestions and the Minerva Foundation for financial support.

References

[1] M.S. Sansom, L. Davison, Modeling transmembrane helix bundles by restrained MD simulations, Methods Mol. Biol. 143 (2000) 325–347.
[2] G. von Heijne, C. Manoil, Membrane proteins: from sequence to structure, Protein Eng. 4 (1990) 109–112.
[3] G. von Heijne, Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule, J. Mol. Biol. 225 (1992) 487–494.
[4] G. von Heijne, Membrane protein assembly: rules of the game, Bioessays 17 (1995) 25–30.
[5] M. Punta, A. Maritan, A knowledge-based scale for amino acid membrane propensity, Proteins 50 (2003) 114–121.
[6] C.M. Ott, V.R. Lingappa, Integral membrane protein biosynthesis: why topology is hard to predict, J. Cell Sci. 115 (2002) 2003–2009.
[7] S. Jayasinghe, K. Hristova, S.H. White, Energetics, stability, and prediction of transmembrane helices, J. Mol. Biol. 312 (2001) 927–934.
[8] P.D. Adams, I.T. Arkin, D.M. Engelman, A.T. Brunger, Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban, Nat. Struct. Biol. 2 (1995) 154–162.
[9] J.L. Popot, D.M. Engelman, Membrane protein folding and oligomerization: the two-stage model, Biochemistry 29 (1990) 4031–4037.
[10] J.L. Popot, D.M. Engelman, Helical membrane protein folding, stability, and evolution, Annu. Rev. Biochem. 69 (2000) 881–922.
[11] S.H. White, W.C. Wimley, Membrane protein folding and stability: physical principles, Annu. Rev. Biophys. Biomol. Struct. 28 (1999) 319–365.
[12] P.D. Adams, D.M. Engelman, A.T. Brunger, Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching, Proteins 26 (1996) 257–261.
[13] S. Kim, A.K. Chamberlain, J.U. Bowie, A simple method for modeling transmembrane helix oligomers, J. Mol. Biol. 329 (2003) 831–840.
[14] P. Herzyk, R.E. Hubbard, Using experimental information to produce a model of the transmembrane domain of the ion channel phospholamban, Biophys. J. 74 (1998) 1203–1214.
[15] F.S. Cordes, A. Kukol, L.R. Forrest, I.T. Arkin, M.S. Sansom, W.B. Fischer, The structure of the HIV-1 Vpu ion channel: modelling and simulation studies, Biochim. Biophys. Acta 1512 (2001) 291–298.
[16] F.S. Cordes, A.D. Tustian, M.S. Sansom, A. Watts, W.B. Fischer, Bundles consisting of extended transmembrane segments of Vpu from HIV-1: computer simulations and conductance measurements, Biochemistry 41 (2002) 7359–7365.
[17] K.G. Fleming, D.M. Engelman, Computation and mutagenesis suggest a right-handed structure for the synaptobrevin transmembrane dimer, Proteins 45 (2001) 313–317.
[18] L.R. Forrest, D.P. Tieleman, M.S. Sansom, Defining the transmembrane helix of M2 protein from influenza A by molecular dynamics simulations in a lipid bilayer, Biophys. J. 76 (1999) 1886–1896.
[19] A.L. Grice, I.D. Kerr, M.S. Sansom, Ion channels formed by HIV-1 Vpu: a modelling and simulation study, FEBS Lett. 405 (1997) 299–304.

[20] A. Kukol, P.D. Adams, L.M. Rice, A.T. Brunger, T.I. Arkin, Experimentally based orientational refinement of membrane protein models: a structure for the Influenza A M2 H^+ channel, J. Mol. Biol. 286 (1999) 951–962.
[21] R.J. Law, L.R. Forrest, K.M. Ranatunga, P. La Rocca, D.P. Tieleman, M.S. Sansom, Structure and dynamics of the pore-lining helix of the nicotinic receptor: MD simulations in water, lipid bilayers, and transbilayer bundles, Proteins 39 (2000) 47–55.
[22] M.S. Sansom, G.R. Smith, O.S. Smart, S.O. Smith, Channels formed by the transmembrane helix of phospholamban: a simulation study, Biophys. Chem. 69 (1997) 269–281.
[23] M.S. Sansom, I.D. Kerr, G.R. Smith, H.S. Son, The influenza A virus M2 channel: a molecular modeling and simulation study, Virology 233 (1997) 163–173.
[24] S. Sukharev, S.R. Durell, H.R. Guy, Structural models of the MscL gating mechanism, Biophys. J. 81 (2001) 917–936.
[25] S. Sukharev, M. Betanzos, C.S. Chiang, H.R. Guy, The gating mechanism of the large mechanosensitive channel MscL, Nature 409 (2001) 720–724.
[26] C.E. Capener, H.J. Kim, Y. Arinaminpathy, M.S. Sansom, Ion channels: structural bioinformatics and modelling, Hum. Mol. Genet. 11 (2002) 2425–2433.
[27] S.R. Durell, E.P. Bakker, H.R. Guy, Does the KdpA subunit from the high affinity K(+)-translocating P-type KDP-ATPase have a structure similar to that of K(+) channels? Biophys. J. 78 (2000) 188–199.
[28] S.R. Durell, H.R. Guy, Atomic scale structure and functional models of voltage-gated potassium channels, Biophys. J. 62 (1992) 238–247, discussion 247–250.
[29] S.R. Durell, H.R. Guy, Structural models of the KtrB, TrkH, and Trk1, 2 symporters based on the structure of the KcsA K(+) channel, Biophys. J. 77 (1999) 789–807.
[30] S.R. Durell, Y. Hao, H.R. Guy, Structural models of the transmembrane region of voltage-gated and other K^+ channels in open, closed, and inactivated conformations, J. Struct. Biol. 121 (1998) 263–284.
[31] S.R. Durell, Y. Hao, T. Nakamura, E.P. Bakker, H.R. Guy, Evolutionary relationship between K(+) channels and symporters, Biophys. J. 77 (1999) 775–788.
[32] S.R. Durell, G. Raghunathan, H.R. Guy, Modeling the ion channel structure of cecropin, Biophys. J. 63 (1992) 1623–1631.
[33] A. Giorgetti, P. Carloni, Molecular modeling of ion channels: structural predictions, Curr. Opin. Chem. Biol. 7 (2003) 150–156.
[34] H.R. Guy, P. Seetharamulu, Molecular model of the action potential sodium channel, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 508–512.
[35] E. Jakobsson, Using theory and simulation to understand permeation and selectivity in ion channels, Methods 14 (1998) 342–351.
[36] M.S. Sansom, P. Bond, O. Beckstein, P.C. Biggin, J. Faraldo-Gomez, R.J. Law, G. Patargias, D.P. Tieleman, Water in ion channels and pores—simulation studies, Novartis Found Symp. 245 (2002) 66–78, discussion 79–83, 165–168.
[37] Y. Fujiyoshi, K. Mitsuoka, B.L. de Groot, A. Philippsen, H. Grubmuller, P. Agre, A. Engel, Structure and function of water channels, Curr. Opin. Struct. Biol. 12 (2002) 509–515.
[38] B.L. de Groot, H. Grubmuller, Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF, Science 294 (2001) 2353–2357.
[39] B. Roux, Computational studies of the gramicidin channel, Acc. Chem. Res. 35 (2002) 366–375.
[40] B. Roux, B. Prod'hom, M. Karplus, Ion transport in the gramicidin channel: molecular dynamics study of single and double occupancy, Biophys. J. 68 (1995) 876–892.
[41] M.S. Sansom, D.P. Tieleman, H.J. Berendsen, The mechanism of channel formation by alamethicin as viewed by molecular dynamics simulations, Novartis Found Symp. 225 (1999) 128–141, discussion 141–125.
[42] K. Siva, R. Elber, Ion permeation through the gramicidin channel: atomically detailed modeling by the Stochastic Difference Equation, Proteins 50 (2003) 63–80.
[43] D.P. Tieleman, M.S. Sansom, H.J. Berendsen, Alamethicin helices in a bilayer and in solution: molecular dynamics simulations, Biophys. J. 76 (1999) 40–49.
[44] R.V. Pappu, G.R. Marshall, J.W. Ponder, A potential smoothing algorithm accurately predicts transmembrane helix packing, Nat. Struct. Biol. 6 (1999) 50–55.
[45] M. Pellegrini-Calace, A. Carotti, D.T. Jones, Folding in lipid membranes (FILM): a novel method for the prediction of small membrane protein 3D structures, Proteins 50 (2003) 537–545.
[46] M.S. Sansom, I.D. Kerr, R. Law, L. Davison, D.P. Tieleman, Modelling the packing of transmembrane helices: application to aquaporin-1, Biochem. Soc. Trans. 26 (1998) 509–515.
[47] T. Tsukihara, H. Aoyama, E. Yamashita, T. Tomizaki, H. Yamaguchi, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, S. Yoshikawa, The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 A, Science 272 (1996) 1136–1144.
[48] V.I. Gordeliy, J. Labahn, R. Moukhametzianov, R. Efremov, J. Granzin, R. Schlesinger, G. Buldt, T. Savopol, A.J. Scheidig, J.P. Klare, M. Engelhard, Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex, Nature 419 (2002) 484–487.
[49] P. Sompornpisut, Y.S. Liu, E. Perozo, Calculation of rigid-body conformational changes using restraint-driven Cartesian transformations, Biophys. J. 81 (2001) 2530–2546.
[50] P.D. Adams, A.S. Lee, A.T. Brunger, D.M. Engelman, Models for the transmembrane region of the phospholamban pentamer: which is correct? Ann. N. Y. Acad. Sci. 853 (1998) 178–185.
[51] W. Jorgensen, J. Tirado-Rives, The OPLS potential function for proteins, energy minimization for crystals of cyclic peptides and crambin, J. Am. Chem. Soc. 110 (1988) 1657–1666.
[52] J. Torres, J.A. Briggs, I.T. Arkin, Contribution of energy values to the analysis of global searching molecular dynamics simulations of transmembrane helical bundles, Biophys. J. 82 (2002) 3063–3071.
[53] I.T. Arkin, P.D. Adams, K.R. MacKenzie, M.A. Lemmon, A.T. Brunger, D.M. Engelman, Structural organization of the pentameric transmembrane alpha-helices of phospholamban, a cardiac ion channel, EMBO J. 13 (1994) 4757–4764.
[54] H.R. Treutlein, M.A. Lemmon, D.M. Engelman, A.T. Brunger, The glycophorin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices, Biochemistry 31 (1992) 12726–12732.
[55] J. Torres, I.T. Arkin, Recursive use of evolutionary conservation data in molecular modeling of membrane proteins A model of the multidrug H^+ antiporter emrE, Eur. J. Biochem. 267 (2000) 3422–3431.
[56] K.R. MacKenzie, D.M. Engelman, Structure-based prediction of the stability of transmembrane helix–helix interactions: the sequence dependence of glycophorin A dimerization, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 3583–3590.
[57] K.R. MacKenzie, J.H. Prestegard, D.M. Engelman, A transmembrane helix dimer: structure and implications, Science 276 (1997) 131–133.
[58] A.K. Dunker, T.C. Jones, Proposed knobs-into-holes packing for several membrane proteins, Membr. Biochem. 2 (1978) 1–16.
[59] D. Langosch, J. Heringa, Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils, Proteins 31 (1998) 150–159.