Article No. jmbi.1998.2512 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 286, 951–962

JMB

Experimentally Based Orientational Refinement of Membrane Protein Models: A Structure for the *Influenza* A M2 H^+ Channel

Andreas Kukol[1], Paul D. Adams[2], Luke M. Rice[2], Axel T. Brunger[2] and Isaiah T. Arkin[1]*

[1]*Cambridge Centre for Molecular Recognition and Department of Biochemistry University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK*

[2]*The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University New Haven, CT 06529, USA*

**Corresponding author*

The 97-residue M2 protein from *Influenza* A virus forms H^+-selective ion channels which can be attributed solely to the homo-tetrameric α-helical transmembrane domain. Site-directed infrared dichroism spectra were obtained for the transmembrane domain of M2, reconstituted in lipid vesicles. Data analysis yielded the helix tilt angle $\beta = 31.6(\pm 6.2)^\circ$ and the rotational pitch angle about the helix axis for residue Ala29 $\omega_{Ala29} = -59.8(\pm 9.9)^\circ$, whereby ω is defined as zero for a residue located in the direction of the helix tilt. A structure was obtained from an exhaustive molecular dynamics global search protocol in which the orientational data are utilised directly as an unbiased refinement energy term. Orientational refinement not only allowed selection of a unique structure but could also be shown to increase the convergence towards that structure during the molecular dynamics procedure. Encouragingly, the structure obtained is highly consistent with all available mutagenesis and conductivity data and offers a direct chemical insight that relates the altered functionality of the channel to its structure.

Keywords: influenza; M2-protein; ion channel; infrared spectroscopy; molecular dynamics

Introduction

The structure determination of membrane proteins is still a difficult task. Considering the fact that as many as 30 % or more of cellular proteins are associated with the membrane (Arkin *et al.*, 1997a), the paucity of high resolution structures of membrane proteins is in stark contrast to the number of soluble protein structures. Recent successes include cytochrome *c* oxidase (Iwata *et al.*, 1995; Tsukihara *et al.*, 1996), cytochrome bc_1 complex (Akiba *et al.*, 1996) and the transmembrane part of the K^+ channel (Doyle *et al.*, 1998). NMR and diffraction methods encounter numerous experimental obstacles due to the solubility problems of membrane proteins which normally require use of detergents. Other approaches use molecular modelling techniques combined with low resolution experimental data.

Molecular modelling is particularly applicable when analysing structures of symmetric homo-oligomeric transmembrane helical bundles. In such an instance, it is generally feasible to exhaustively sample the conformational space using molecular dynamics protocols leading to different low energy structures (Adams *et al.*, 1995). However, selection of the correct model amongst these local convergence points can be difficult without additional experimental data. Accordingly, mutagenesis data were obtained for the dimerizing transmembrane domain of human glycophorin A (Lemmon *et al.*, 1992, 1994; Treutlein *et al.*, 1992; Adams *et al.*, 1996) and the pentamerising protein phospholamban (Adams *et al.*, 1995; Arkin *et al.*, 1994) which enabled selection of a unique model for these proteins.

The aim of this work is to determine the structure of the transmembrane domain of the M2 protein of the *Influenza* A virus. This protein is the target of the anti-viral drug amantadine (Hay *et al.*, 1985), the only drug against *Influenza* available. As

Abbreviations used: RNP, ribonucleoprotein; FTIR, Fourier transform infrared; ATR, attenuated total reflection; FSD, Fourier self-deconvolution.

E-mail address of the corresponding author: sa232@cam.ac.uk

0022-2836/99/000951–12 $30.00/0

Influenza is the causative agent of the most deadly disease outbreaks of the 19th and 20th century and any insights into the structure of M2 are clearly of biomedical importance.

M2 is a 97 amino acid homo-tetrameric membrane protein (4×97) located in the envelope of the *Influenza* A virus (Holsinger & Lamb, 1991; Lamb *et al*., 1985; Sugrue & Hay, 1991). It is generally believed that M2 functions as an H^+ channel (Lamb & Pinto, 1997) and participates in the virus uncoating process after viral uptake by endocytosis. The passage of H^+ ions from the acidic environment of the endosomal lumen into the virion weakens the interactions between the matrix protein (M1) and the ribonucleoprotein (RNP) core (Bui *et al*., 1996). In the presence of amantadine, a blocker of the M2 H^+ channel, *Influenza* A virus uncoating is incomplete and the RNPs fail to promote infection.

In the native protein the transmembrane domain of M2 corresponds to amino acids 26–43 and has been shown to be α-helical by circular dichroism spectroscopy (Duff *et al*., 1992). A synthetic 25-residue peptide encompassing the transmembrane domain has been shown to form amantadine-sensitive ion channels in lipid bilayers (Duff & Ashley, 1992; Duff *et al*., 1994), implying that the transmembrane domain is responsible for the functional and structural properties of the protein.

Here we apply the new method of site-directed infrared dichroism (Arkin *et al*., 1997b), an extension of previous site-directed UV linear dichroism methods (Hagmar *et al*., 1992; Norden *et al*., 1992) to examine the structure of a 25-residue peptide corresponding to the transmembrane domain of M2 reconstituted in lipid vesicles. Infrared spectroscopy dichroism measurements of single amide-I vibrational modes corresponding to (1-^{13}C)-labelled sites yielded the tilt and rotational pitch angles of the transmembrane domain of M2. Incorporating these data in a molecular modelling process as refinement energy terms, we obtained a structure that is consistent with spectroscopic data and which explains the functional importance of amino acid substitutions in different viral strains.

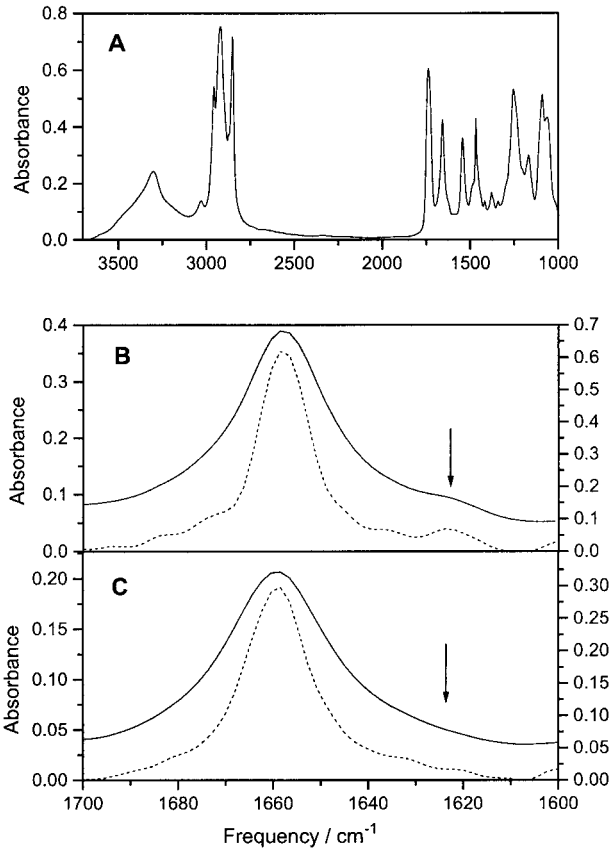

Figure 1. Transmission-FTIR spectra of the [1-^{13}C]Ala29-labelled M2 peptide reconstituted in lipid vesicles and dried on a CaF_2 window. (a) Whole spectrum between 3800 and 1000 cm^{-1}. (b) Amide I region expanded. The dotted line shows the spectrum after Fourier self deconvolution using a band width at half height of 13 cm^{-1} and a resolution enhancement of 2.0. The arrow points to the [1-^{13}C]amide I. (c) Same as (b) but for a peptide containing no (1-^{13}C) label.

Results

FTIR measurements

Transmission FTIR

The transmission FTIR-spectrum in Figure 1(a) was obtained from the M2 transmembrane domain reconstituted in lipid vesicles. The expanded amide I region (Figure 1(b)), shows virtually no intensity arising from β-sheet structures (Byler & Susi, 1986), consistent with the results obtained by CD measurements (Duff *et al*., 1992). The peak at 1623 cm^{-1} arises from the (1-^{13}C)-labelled amino acid Ala29 and the 1% naturally occurring (1-^{13}C) (Tadesse *et al*., 1991; Anderson *et al*., 1992; Ludlam *et al*., 1996) as seen by comparing the spectra to that obtained for a peptide which contains no ^{13}C label (Figure 1(c)). The intensity ratio $^{12}C/^{13}C$ of 0.06 corresponds roughly to one (1-^{13}C) label per 20 residues long transmembrane region, as expected.

ATR-FTIR

The ATR-FTIR spectra of the amide I region for [1-^{13}C]Ala29 (Figure 2(a) and (b)) and [1-^{13}C]Ala30 (Figure 2(c) and (c)) labelled peptides obtained with either parallel or perpendicular polarised infrared light, respectively, show that the absorbance obtained with parallel polarised light is more intense. This reflects the preferential orientation of the helix perpendicular to the membrane plane. The results of five separate measurements from each of the peptides resulted in the dichroic ratios depicted in Table 1.

The order parameters calculated for the helices and the lipid bilayer are indicative of a well ordered lipid bilayer and of helices possessing a net transbilayer orientation (Braiman & Rothschild,

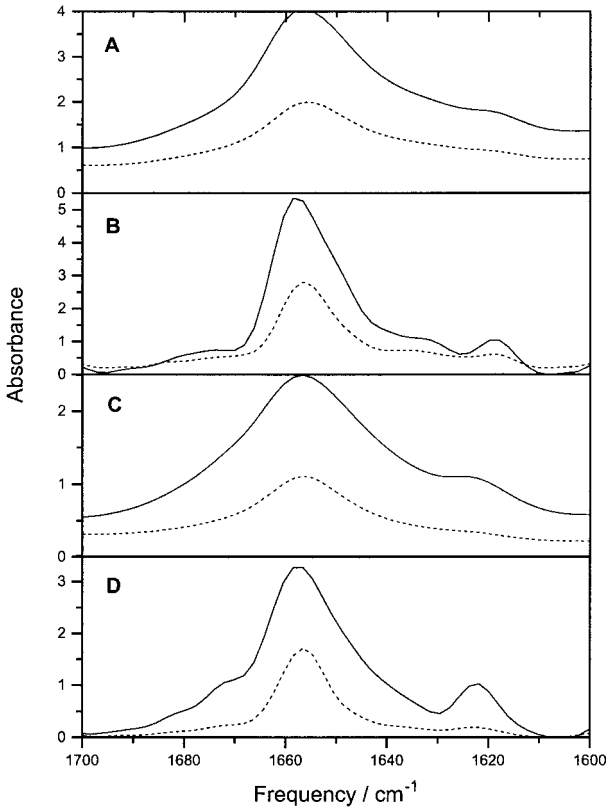

Figure 2. ATR-FTIR spectra of the amide I region obtained with parallel (continuous line) and perpendicular (dotted line) polarised light. Samples are from the lipid-vesicle reconstituted transmembrane domain of M2 dried on a KRS internal reflection element. (a) and (c) The spectra of (1-^{13}C) labelled Ala29 and Ala30 peptides, respectively. (b) and (d) The Fourier self-deconvoluted spectra of (a) and (c).

1988). Table 1 also presents the results from the analysis of the dichroic ratios according to the theory for site-specific dichroism as detailed in Materials and Methods.

The variation in the dichroic ratios reflects the variability of sample fractional order between the different samples. The larger variation in the values of $\mathcal{R}_{\mathrm{Site}}$ of Ala30 are expected, since the variation is proportional to the extent of orientation (i.e. the higher the averaged dichroism the higher the variation). As an example, a randomly ordered sample will exhibit a low dichroism with virtually no variation (Arkin *et al.*, 1997b). Note that the calculations of the helix tilt and rotational pitch angles take the variation of sample disorder into account.

Molecular modelling of the M2 transmembrane domain

Global search protocols were employed in which each symmetrical starting structure was subjected to a molecular dynamics and energy minimisation protocol. The search protocol consisted of changing the helix rotational angle ϕ, defined in Figure 6, in increments of 10 degrees (36 structures), each with two different crossing angles ($\pm 25^\circ$) and four trials with different initial random velocities yielding $36 \times 2 \times 4 = 288$ structures. The root-mean-square deviation (RMSD) between each pair of structures was then calculated, leading to the delineation of clusters containing similar structures (RMSD < 1 Å). Each of the clusters was then averaged and the average structure was subjected to energy minimisation protocols.

The results of two such protocols are presented in Figure 3 and Table 2. As a control, in the first protocol only empirical energy terms were used in the calculations. In the second protocol orientational energy refinement was undertaken.

It can be seen that employing orientation energy refinement resulted in a larger number of clusters and that each cluster contains, on average, more structures. This results from the increased convergence seen in Figure 3(b), i.e. greater variation between starting and ending structures. It can also be seen that the clusters obtained when employing orientation energy refinement are of lower energy, and that cluster 2 has both the lowest energy of $E = -182$ kcal/mol and is calculated from the largest number of structures. The second lowest energy structure is represented by cluster 4 with $E = -175$ kcal/mol, but only 17 structures have converged to form this cluster.

Discussion

The aim of this study was to obtain a reliable structure for the *Influenza* A M2 H^+ channel. To that end a novel method based on site-directed dichroism to obtain experimental orientational restraints was employed. These restraints were then implemented in the molecular dynamics global search protocol as an unbiased refinement

Table 1. Helix dichroism $\mathcal{R}_{\mathrm{Helix}}$, site-specific dichroism $\mathcal{R}_{\mathrm{Site}}$ and lipid CH_2 asymmetric stretching modes $\mathcal{R}_{\mathrm{Lipid}}$ for the two M2 transmembrane peptides as noted

Peptide	$\mathcal{R}_{\mathrm{Helix}}$	S_{Helix}	$\mathcal{R}_{\mathrm{Site}}$	$\mathcal{R}_{\mathrm{Lipid}}$	S_{Lipid}	β (deg.)	ω (deg.)	θ (deg.)
Ala29	2.81 ± 0.56	0.34 ± 0.22	3.94 ± 0.73	1.04 ± 0.20	0.53 ± 0.14	31.6 ± 6.2	-59.8 ± 9.9	17.1 ± 7.6
Ala30	2.76 ± 0.36	0.33 ± 0.14	7.22 ± 2.3	1.14 ± 0.28	0.46 ± 0.18	31.6 ± 6.2	40.2 ± 9.9	7.0 ± 5.2

Also shown are the helix and lipid order parameters S calculated as detailed in Materials and Methods. Variation of the data reflects the different sample order achieved by drying the lipid vesicles on the ATR internal reflection element. The calculated helix tilt angle β, and the rotational pitch angle about the helix axis ω, are obtained from the analysis of the site-specific dichroism as detailed in Materials and Methods. θ is the angle between the C = O bond and the z-axis, as shown in Figure 5.

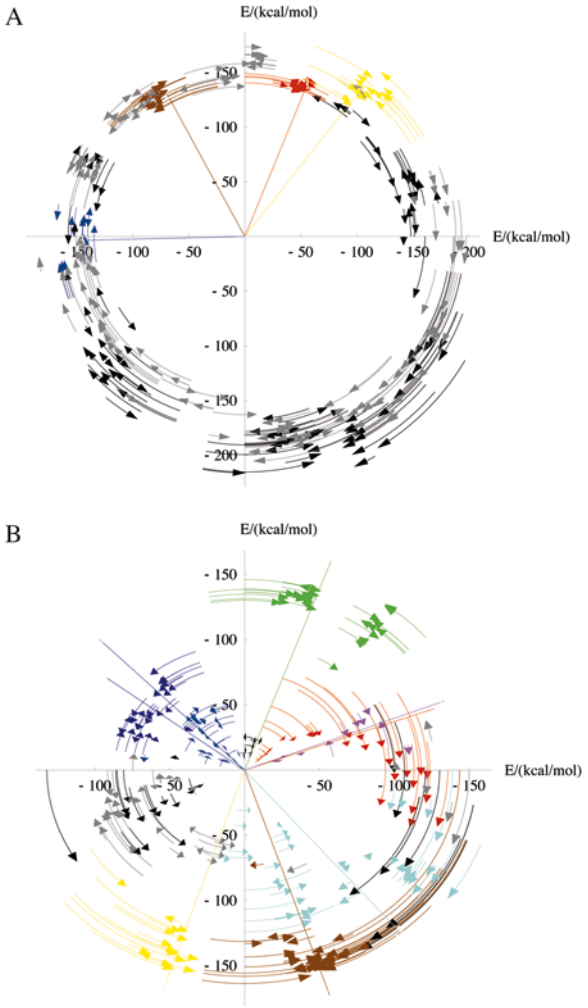

Figure 3. Energy plots (in polar format) of all the structures obtained from the global search protocol as a function of the helix rotation parameter ϕ (defined in Figure 6). The arcs represent each of the individual structures whereby the arrows designate the direction in which the starting structure moved during the simulation. The energy is measured as the distance from the origin. Each of the structures are coloured according to their cluster affiliation. Left-handed unclustered structures are depicted in black and right-handed unclustered structures in grey. The cluster averages are represented as azimuthal lines ending with a circle at the energy level of the average. (a) Structures obtained without orientational refinement. (b) Structures obtained with orientational energy refinement.

energy term. Initially, the experimental results which lead to a unique structure are discussed, followed by a description of the structure and how it explains all the experimental data, such as conductivity data and viral strain variation to amantadine, and anti-*Influenza* drug in clinical use. Finally, a comparison between the approach taken herein and previous modelling exercises is undertaken which demonstrates its advantages over approaches.

Table 2. Parameters of averaged structures obtained from the global search protocol with and without orientational refinement

Cluster	E/(kcal mol^{-1})	N	Ω (deg.)	β (deg.)	ϕ (deg.)	θ_{29} (deg.)	θ_{30} (deg.)
A. *Without orientational refinement*							
1	−150	16	7.44	9.76	68.2	24.4	23.8
2	−149	20	12.9	14.9	118.0	27.0	24.1
3	−146	13	16.0	14.0	181.4	31.3	6.1
4	−172	18	−9.1	11.7	52.3	20.1	13.4
B. *With Orientational refinement*							
1	−135	28	29.4	31.1	20.4	17.2	7.7
2	−182	36	31.4	31.2	288	17.2	7.3
3	−139	32	28.5	31.0	134	17.7	7.9
4	−175	17	31.2	31.2	252	17.4	7.1
5	−161	40	−27.9	31.0	310	17.8	7.5
6	−143	11	−29.2	30.7	21.7	17.5	7.1
7	−170	28	−31.1	31.2	70.2	17.2	7.4
8	−118	28	−30.8	31.2	143	17.1	7.5

These include the energy E, of each structure, the number of structures considered for the average N, the helix crossing angle Ω, the helix tilt angle β and the helix rotation parameter ϕ as well as the angles between the z axis and the C═O bond of residues 29 (θ_{29}) and 30 (θ_{30}). The angles are averages over the four helices of a structure. Note that the energy depicted in the Table does not take into account the refinement energy terms, only empirical energy terms.

Secondary structure and orientation

As expected, FTIR transmission spectra have clearly indicated that the peptide is mostly helical. Both the narrow width of the amide I mode and the lack of any components aside from the helical signature are highly consistent with previous CD spectroscopic results (Duff *et al.*, 1992). Both transmission FTIR and CD spectroscopy yield information about the average secondary structure of a protein. However, by utilising (1-^{13}C) labels it is possible to determine the secondary structure at the site of labelling, since the magnitude of the shift assigned to the isotope-edited residue is known to be ca. − 40 cm^{-1} (Tadesse *et al.*, 1991; Anderson *et al.*, 1992; Ludlam *et al.*, 1996). Upon comparison of the spectra corresponding to labelled and unlabelled peptides (Figure 1(b) and (c), respectively) one can observe a peak around 1622 cm^{-1} that is present only in the labelled peptide spectrum. This peak results from a single (1-^{13}C) label at position Ala29 and is shifted roughly by −40 cm^{-1} from the main ^{12}C-labelled helical amide I mode (Tadesse *et al.*, 1991; Anderson *et al.*, 1992; Ludlam *et al.*, 1996). Transmission FTIR spectra of the peptide labelled at Ala30 yielded identical results (data not shown), showing that not only is the protein highly helical but that the region between Ala29 and Gly34 (Gly34 is the hydrogen bonding partner) is definitively helical.

ATR-FTIR spectra of oriented peptides in a lipid environment, utilising conventional analysis tools, can yield an order parameter for the peptide in question (Braiman & Rothschild, 1988). As evident

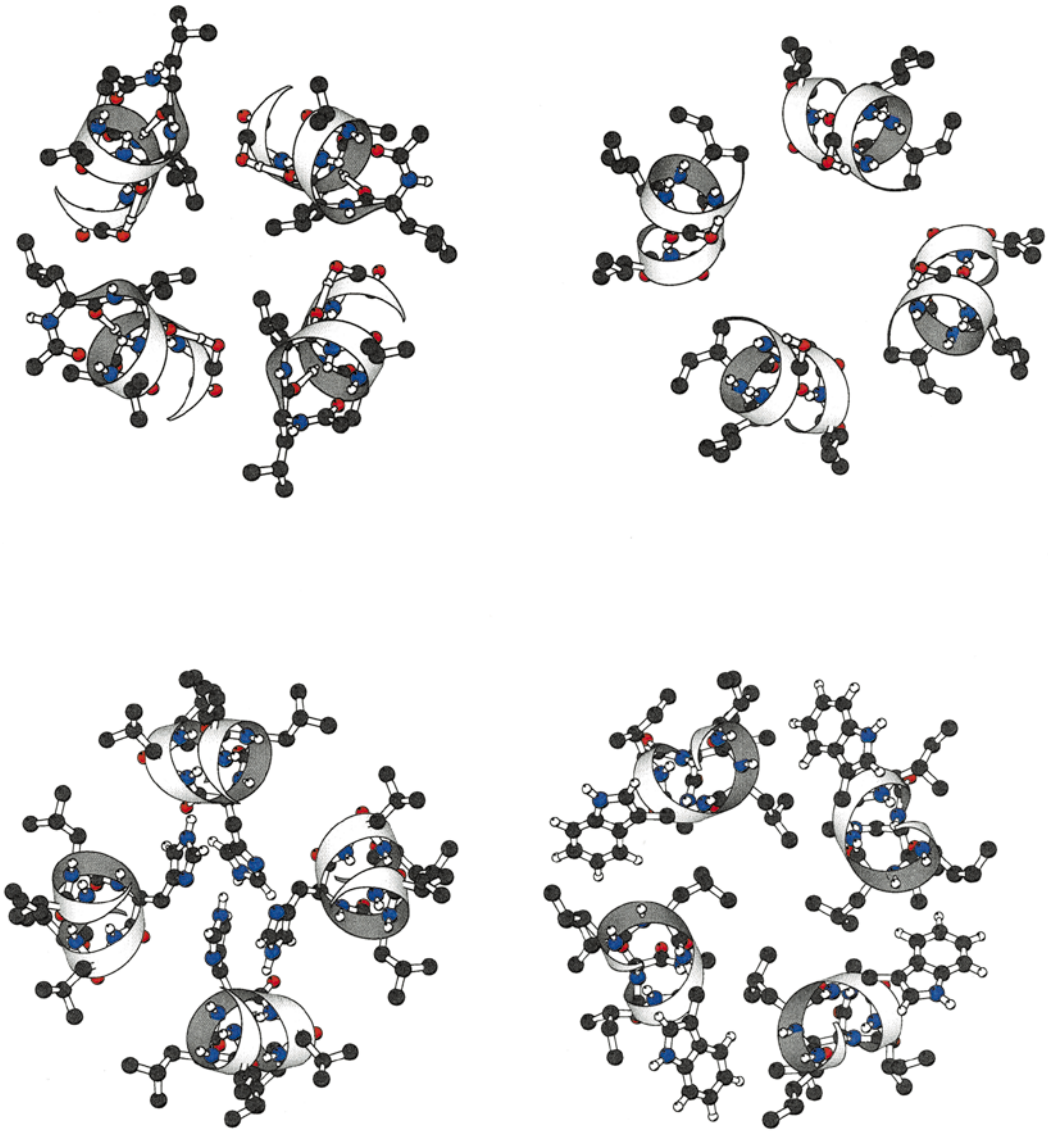

Figure 4. Consecutive slices corrsponding to structure 2 showing the helix as a ribbon diagram and the conformation of the side-chains in a ball and stick representation. Top left, Leu26-Ser31; top right, Ser31-Ile35. Bottom left, Ile35-Ile39; bottom right, Ile39-Leu43. The Figure was created with MOLSCRIPT (Kraulis, 1991).

from spectra depicted in Figure 2 and Table 1, the protein is highly oriented whereby the helix dichroic ratio ($\mathcal{R}_{\text{Helix}} = 2.8$) yields an order parameter of $s = 3\langle\cos^2\theta\rangle - 1/2 = 0.36$ (Braiman & Rothschild, 1988; Arkin *et al*., 1997b). Due to the fact that an order parameter is a convolution of both orientation and sample order, these results only indicate a maximum angle (41°) between the helix and the bilayer normal. Since the tilts of most transmembrane helices observed so far are between 0°-40° (e.g. see Iwata *et al*., 1995; Tsukihara *et al*., 1996; Akiba *et al*., 1996; Doyle *et al*., 1998), the results obtained from conventional order parameter analysis can be perceived as not being dramatically useful. This underscores the need to deconvolute the sample fractional order from the order parameter which would yield a direct estimation of the helix tilt. Furthermore, as the range of the rotational pitch angle about the helix axis can vary anywhere from 0°-359°, it is clear that both these parameters must be obtained in order to model the helical bundle properly. Analysis of site-directed dichroism spectroscopy can yield both these parameters (Arkin *et al*., 1997b).

Calculation of tilt and pitch angles

Comparison of the spectra in Figure 2(b) and (d) yields two important results (see Table 1). First, the dichroism of the isotope-shifted mode ($\mathcal{R}_{\text{Site}}$) is significantly larger than that of the helix ($\mathcal{R}_{\text{Helix}}$). Second, there is significant difference between the site-specific dichroism obtained from a (1-^{13}C) label at Ala29 and that obtained from a label at Ala30. Both these results indicate that the peptide is tilted from the bilayer normal and does not rotate about its axis. If either of these statements were untrue then the three measured dichroisms would have

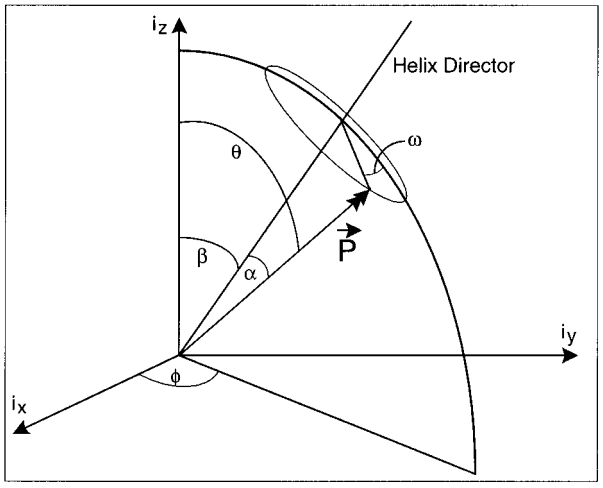

Figure 5. Schematic describing the mathematical dependence of the measured dichroism of a particular transition dipole moment. The helix is tilted from the *z*-axis by the angle β, containing a vibrating bond, P, related to the helix director by the angle α. The bond is positioned with a rotational pitch angle ω around the helix director. The angle θ is the absolute angle between the transition dipole moment and the *z*-axis. Uniaxial symmetry results in rotational averaging about the angle ϕ.

been identical. As this is not the case, one can analytically solve for both the helix tilt angle β, and rotational pitch angle ω, as well as the deconvoluted sample fractional order. Error estimates are obtained by solving for any pair-wise combination of the data and calculating the standard deviation over the spread of the results (see Materials and Method).

The values obtained were $\beta = 31.6(\pm 6.2)^\circ$ and $\omega_{Ala29} = -59.8(\pm 9.9)^\circ$, whereby ω is defined as zero for a residue located in the direction of the helix tilt. Note that ω_{Ala30} is simply $\omega_{Ala29} + 100^\circ$. It is highly encouraging to note that recent application of solid state NMR spectroscopy has resulted in similar tilt angles (Kovacs & Cross, 1997).

Structure calculations and refinement

Simulation of transmembrane α-helical bundles has traditionally been done utilising two approaches. In the first approach, the helices are positioned according to some arbitrary assumptions which may include mutagenesis data and/or hydrophobicity considerations. The resulting bundle is then energy minimised to alleviate any steric clashes (Sansom & Kerr, 1993; Sansom *et al*., 1997). In our approach, a global search protocol is employed on a series of bundles whereby each bundle contains helices which are rotated with respect to each other at a different incremental level. In this way no assumptions are made with respect to preconditions (Adams *et al*., 1996). Therefore, the main difference between the two approaches is that the first approach assumes a particular rotational pitch angle before any simulation takes place.

How does one select the correct model amongst the clusters of structures that are obtained by the global search protocols? Without any experimental data, two criteria can be used: lowest energy and cluster abundance (i.e. how many structures have converged to that local minima). Previous studies on the pentamerising transmembrane helices of human phospholamban have compared each cluster average with the mutagenesis results, noting that the structure most consistent with the experimental data is not the one with the lowest energy (Arkin *et al*., 1994; Adams *et al*., 1995).

The approach taken herein is fundamentally distinct. Not only are orientational restraints far more interpretable than mutagenesis results, but they also lend themselves to implementation as refinement tools taking place during the simulation. Thus, data from the site-directed dichroism were implemented to restrain the tilts of the two C═O bonds of residues Ala29 and Ala30 with respect to the *z*-axis, which coincides with the bundle symmetry axis. Since the structure is tetrameric this results in a torque about the helix axis. Furthermore, knowledge of the helix tilt enabled refinement of that parameter by constraining the angle between the *z* axis and the vectors made by connecting every C^α of residue *i* and C^α of residue $i + 7$ to a target angle obtained experimentally (the angle β, see Figure 7).

As seen from Figure 3 and Table 2, the use of the orientation refinement term has resulted in a

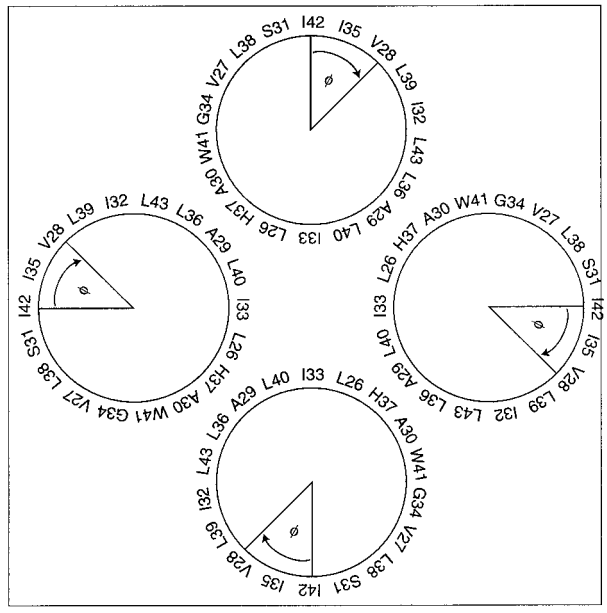

Figure 6. A sketch of the initial structure in a helical wheel representation for the global search protocol with respect to the helix rotation ϕ, which is applied simultaneously to all helices.

cluster that not only has the lowest energy but also the largest number of structures. Satisfaction of both criteria by a single cluster is indicative that it is the most consistent with the experimental results. In comparison, results obtained from unrefined simulations indicate that nearly all of the clusters were indistinguishable in terms of energies and cluster population. It is also clear that the use of orientation refinement can be seen to increase the convergence to the various clusters. In conclusion, this analysis points to structure 2 as the current structure.

Structure description and implications

The details of the structure obtained from the analysis described above are presented in Figure 4, which depicts the orientation of the side-chains in slices of four to five amino acids residues throughout the bundle axis.

The M2 H^+ channel has been studied extensively, since it has been implicated as the target of the only anti-*Influenza* drug, amantadine. It is therefore interesting to compare this structure with functional data of different virus strains (Grambas *et al*., 1992; Hay *et al*., 1985) and site-directed mutagenesis data (Holsinger *et al*., 1994). Mutations at positions Val27, Ala30, Ser31 and Gly34 have been found in amantadine-resistant viruses (Grambas *et al*., 1992; Wang *et al*., 1993). Inspection of Figure 4 shows that Val27, Ala30 and Gly34 point into the channel lumen, whereas Ser31 is located at the helix-helix interface with its hydroxyl pointing into the channel lumen. Mutation of the His37 residue into Ala has resulted in channel activity that is not pH-activated (Wang *et al*., 1995). Accordingly, the His37 residue in our structure points into channel lumen. The orientation of His37 strengthens a model for channel activation by pH, which was presented earlier (Wang *et al*., 1995). Figure 4 shows that the pore is occluded by the His37 residues. Protonation of one histidine nitrogen atom from the virus exterior may result in deprotonation of the other nitrogen atoms because the positive charge is distributed over the π-electron ring system. Tautomerisation or a ring flip could restore the starting position. At low pH multiple protonation may result in a conformational change which allows the observed passage of cations (Shimbo *et al*., 1996). Taking into account different mutagenesis data, a group of amino acid residues lying on one face of the α-helix has been presented (Holsinger *et al*., 1994) which is in concurrence with our model. It places residues Val27, Ala30, Ser31, Gly34, His37, Leu38 and Trp41 on one face of the helix. Taken together it is highly encouraging that our structure is consistent with all the available data, noting that we have made no such assumptions during the refinement and simulation process.

The side-chain conformations of the residues facing the lipid phase remain uncertain, as they could not be deduced by our experimental approach. Their conformations may be obtained by two different approaches: extending the molecular model of the system by including lipid and water molecules, or by using additional geometric constraints.

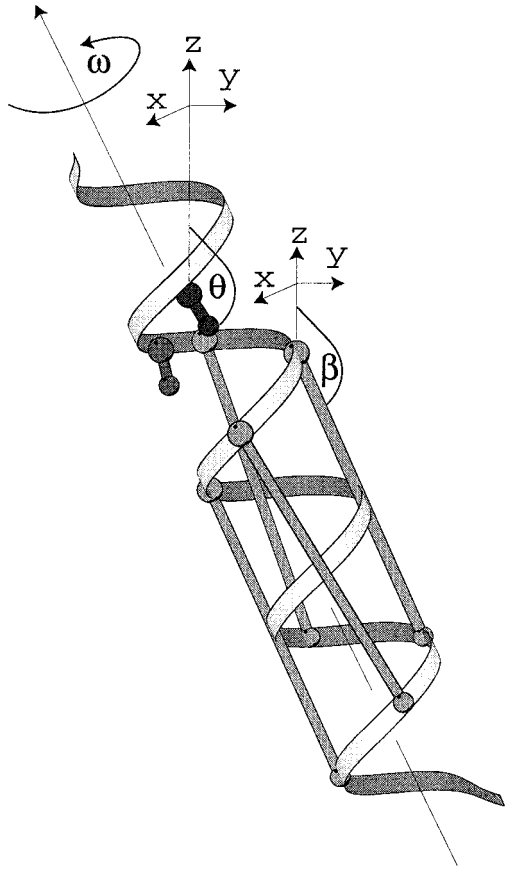

Figure 7. Schematic of the orientational restraints employed during the molecular dynamics and energy minimisation calculations. The helix tilt was restrained to the angle β by defining a vector between every C^α of residue i and C^α of residue $i+7$. For clarity only four different vectors are shown. The angles θ_{Ala29} and θ_{Ala30} formed between the C═O bonds and the z-axis were restrained to the values calculated from β and the rotational pitch angle ω. The rotational pitch angle ω is also shown in the diagram.

Comparison with previous modelling efforts

The *Influenza* A M2 H^+ channel has been a popular choice for molecular modelling exercises (Pinto *et al*., 1997; Sansom & Kerr, 1993; Sansom *et al*., 1997; Forrest *et al*., 1998). It is therefore imperative to outline the fundamental differences that distinguish our methods from previous efforts.

Firstly, an exhaustive global searching of rotational space was employed with no assumptions placing particular residues at specific locations. This global searching protocol has previously been shown to yield a structure of the dimerising human glycophorin A (Lemmon *et al*.,

1992; Adams *et al*., 1996), which was in remarkable agreement with that derived from high-resolution NMR studies (MacKenzie *et al*., 1997). In contrast, in previous efforts, a single structure based on mutagenesis data (Sansom *et al*., 1997), or a couple of distinct structures based on Cys scanning mutagenesis (Pinto *et al*., 1997) were used as an initial structure for molecular dynamics simulations.

Secondly, a novel method has been employed which yields long range orientational restraints between transmembrane α-helices. The method is applied to peptides reconstituted in lipid bilayers, the native environment of the protein. Furthermore, in contrast with NMR methods, due to the high sensitivity of infrared spectroscopy, there is no need to employ high protein to lipid ratios as is a requirement in solid state NMR methods, which may lead to sample inhomogeneity due to non-specific aggregation.

Finally, the experimental orientational restraints are employed in an unbiased fashion utilising a new orientational refinement energy term during the molecular dynamics global searching protocol. The implementation of experimental data in this fashion is fundamentally different from the situation in which an agreement between the experimental data and structures that arise from simulations is sought *a posteriori*.

Based on the above three distinctions we believe that this approach offers significant advantages to that of previous studies in that global searching is undertaken with no previous assumption, and the use of experimental data is made as a refinement tool during simulations.

One can compare the results of previous modelling studies (Sansom & Kerr, 1993; Pinto *et al*., 1997) with the structure obtained herein, focusing on the three major parameters describing a symmetrical helical bundle. As both of these previous models are very similar to one another C^α RMSD = 1.7 Å (Forrest *et al*., 1998), a comparison will be made to their averaged parameters.

(1) Helix tilt angle. The helix tilt angle obtained in this study 31.2(±1.2)°, is markedly different from that obtained previously 15.6(±0.6)° (Forrest *et al*., 1998). However, lower helix tilt angles were obtained when orientation refinement was not undertaken in the energy calculation protocols (Table 2). We therefore reason that this low helix tilt angle results from the fact that in the absence of lipid (all studies were *in vacuo* simulations) the helices tend to maximise van der Waals interactions by orienting themselves perpendicular to the membrane plane.

(2) Helix crossing angle: Similarly to the helix tilt angles, the crossing angle obtained in this study 30(±2)°, is larger than that obtained previously 23(±2)° (Forrest *et al*., 1998). The reasoning is identical to that stated above.

(3) Helix rotational pitch angle. The helix rotational pitch angles are not listed in both previous publications, but are similar to those obtained herein, because both previous modelling studies have positioned the helices in the bundle so as to place the residues implicated in altered functionality (see above) in the bundle core. In the current study no assumptions were made as to the pre-defined rotational pitch angle, but rather a global search was employed. Encouragingly, the resultant structure has placed the residues implicated in altered functionality in the bundle core.

Conclusion

The method of site-directed infrared dichroism has been shown to produce data about the orientation of helical peptides within a membrane, which are in accordance with recently published solid-state NMR results (Kovacs & Cross, 1997). By introducing the orientational restraints into molecular dynamics calculations and energy minimisation as a refinement procedure, we derived a detailed structure of the *Influenza* A virus M2-protein transmembrane domain. This structure has been shown to be in accordance with evolutionary and mutagenesis data on the function of single amino acid residues. Therefore, it is possible with our experimentally based molecular modelling approach to obtain a highly reliable structure for a transmembrane α-helical bundle, a goal which has presented significant difficulties to other approaches.

Materials and Methods

Peptide purification

Synthetic peptides corresponding to *Influenza* A M2 protein sequence Ser22-Leu46 were made by standard solid-phase F-moc chemistry, cleaved from the resin with hydrofluoric acid and lyophilised. Each of the two peptides contains a carbonyl [1-^{13}C]amino acid (Cambridge Isotopes Laboratories, Andover, MA) at positions Ala29 and Ala30, respectively:

SSDPLVV($^{1-13C}$A)ASIIGILHLILWILDRL

SSDPLVVA($^{1-13C}$A)SIIGILHLILWILDRL

As any two labels within the transmembrane segment would enable the particular analysis detailed below, we have chosen Ala due to its lower cost. The lyophilised peptides were dissolved in 2 ml of trifluoroacetic acid (final concentration ca. 5 mg/ml) and immediately injected onto a 20 ml Jupiter 5C4-300 Å column (Phenomex, Cheshire, UK) equilibrated with 95% H_2O, 2% (w/v) acetonitrile and 3% (v/v) 2-propanol. Peptide elution was achieved with a linear gradient to a final solvent composition of 5% H_2O, 38% acetonitrile and 57% 2-propanol (Biocad Sprint, Perceptive Biosystems, Cambridge, USA). All solvents contained 0.1% (v/v) trifluoroacetic acid. Pooled fractions were lyophilised and dissolved in a solution of 5% (w/v) β-octylglycopyranoside, 12.5 mg/ml dimyristoylphosphocholine to a final volume of 2 ml.

Peptide reconstitution

Peptide reconstitution into lipid vesicles was achieved by dialysis (Spectra/Por 4 MWCO:12-14,000, Spectrum laboratories, Laguna Hills, CA, USA) for 24 hours in water, followed by six hours dialysis against 10 mM phosphate $K_2HPO_4 \cdot KH_2PO_4$ buffer (pH 7.0). Successful reconstitution and α-helical conformation was checked qualitatively by both transmission FTIR and ATR-FTIR order parameter calculations (see below).

Infrared spectroscopy

FTIR spectra were recorded on a Nicolet Magna 560 spectrometer (Madison, WI, USA). The spectrometer was purged with N_2 and equipped with a liquid nitrogen cooled high sensitivity MCT/A detector. Attenuated total reflection (ATR) spectra were measured with a 25 reflection ATR accessory from Grasbey Specac (Kent, UK) and a wire grid polariser (0.25 μM, Graseby Specac). A 200 μl sample (ca. 2.5 mg/ml protein and 12.5 mg/ml lipid) was dried onto a KRS-5 or Ge trapezoidal internal reflection element (50 mm × 2 mm × 20 mm). One thousand interferograms were averaged for every sample and processed with one point zero filling Happ-Genzel apodisation.

Fourier self-deconvolution (FSD; Kauppinen *et al*., 1982) was applied to the spectra in the amide I region in order to separate the overlapping ^{12}C and isotope shifted (1-^{13}C) (Tadesse *et al*., 1991) amide I peaks. The enhancement factor used in FSD was 2.0, lower than that previously reported (Byler & Susi, 1986) in order to minimise Fourier ringing.

Peak integration was performed on Fourier self-deconvoluted spectra in the regions 1670-1645 cm^{-1} and 1610-1630 cm^{-1} for the helix (^{12}C) and the site-specific label (^{13}C), respectively. The dichroic ratio was calculated as the ratio between the integrated absorptions of parallel and perpendicular polarised light and corrected for natural abundance (1-^{13}C) as described (Arkin *et al*., 1997b).

Data analysis

The data were analysed according to the theory of site-specific dichroism presented in detail elsewhere (Arkin *et al*., 1997b). Briefly, the measured dichroism $\mathcal{R}$ of a particular transition dipole moment is defined as the ratio between the absorption of parallel and perpendicular polarised light. Hence the dichroism is a function of the sample fractional order f and the orientation of the transition dipole moment, which is described by the following parameters shown schematically in Figure 5: the helix-tilt angle β, the angle α, relating the transition dipole moment to the helix director and the rotational pitch angle ω. The rotational pitch angle ω is arbitrarily defined as 0° when the transition dipole moment, the helix director and the z-axis all reside in a single plane. In this study the transition dipole moment utilised is that of the amide I vibrational mode (mostly the peptide C═O stretch). Since the angle α is known from fibre diffraction studies (Tsuboi, 1962), the measured dichroism $\mathcal{R}_{\mathrm{Site}_i}$ of the ^{13}C amide I mode at a particular label is a function of three unknowns:

$$\mathcal{R}_{Site_i} = F(\beta, \omega, f_i) \tag{1}$$

whereby f_i represents the fractional sample order. In case of the helix dichroism, multiple amide I modes exist (at 100° intervals of ω) that eliminate the dependency upon the rotational pitch angle ω. Thus, the measured helix dichroism is a function of two unknowns:

$$\mathcal{R}_{Helix_i} = F(\beta, f_i) \tag{2}$$

Thus, each measurement i yields two different equations, one for the helix dichroism and the other for the site-specific dichroism. By obtaining a spectrum for a different peptide j, containing a (1-^{13}C) label at an adjacent residue, two additional equations are obtained:

$$\mathcal{R}_{Site_j} = F(\beta, \omega + 100^\circ, f_j) \quad \text{and} \tag{3}$$

$$\mathcal{R}_{Helix_j} = F(\beta, f_j) \tag{4}$$

With the above data it is possible to solve analytically the four equations for all the unknowns: β, ω, f_i and f_j (Arkin *et al*., 1997b). Note that the difference of ω between two consecutive residues is assumed to be 100° as in a standard α-helix. As each preparation yields a slightly different fractional sample order the dichroic ratios are not generally identical for different samples. Therefore different measurements will lead to equations differing only in their fractional sample orders, $f_i, f_j \ldots f_n$, whereby n represents the number of different spectra obtained from different samples. By extending the formerly presented method (Arkin *et al*., 1997b), all samples with a fractional sample order greater than zero could be utilised. Results from five different measurements of a peptide (1-^{13}C)-labelled at Ala29 and five different measurements with a peptide labelled at Ala30 are combined yielding 25 different input sets. Solving the four equations for each i and j pair, yields β_{ij}, ω_{ij}, f_i and f_j, whereby β_{ij} and ω_{ij} are the results obtained from the combinations of sample i and sample j. Thus, β and ω are obtained from the average of all of the results:

$$\beta = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \tag{5}$$

$$\omega = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{ij} \tag{6}$$

Solving the four non-linear equations was done with Newton's method as implemented in the FindRoot function in Mathematica 3.0 (Wolfram Research, Champaign, USA).

Order parameter calculation

Order parameters for the transmembrane helices and the lipids, were calculated as described (Arkin *et al*., 1997b). In brief, the order parameter $\mathcal{S}$ is defined as follows:

$$\mathcal{S} = \frac{3\langle \cos^2 \beta \rangle - 1}{2} \tag{7}$$

whereby β represents the angle between the helix director and the z-axis which is coincident with the membrane normal. The order parameter can be determined experimentally from the dichroic ratio $\mathcal{R}^{\mathrm{ATR}}$ by the following equation:

$$\mathcal{S} = \frac{\varepsilon_x^2 - R^{ATR}\varepsilon_y^2 + \varepsilon_z^2}{\varepsilon_x^2 - R^{ATR}\varepsilon_y^2 - 2\varepsilon_z^2} \left(\frac{3\cos^2 \alpha - 1}{2} \right)^{-1} \tag{8}$$

whereby the electric field components of the evanescent

wave, $\mathcal{E}_x = 1.398$, $\mathcal{E}_y = 1.516$ and $\mathcal{E}_z = 1.625$ are given by Harrick (1967) (for a Ge internal reflection element), and $\alpha = 39^\circ$ is the angle between the helix director and the transition dipole moment of the amide I vibrational mode, (Tsuboi, 1962). When calculating the order parameter for the lipid CH_2, the angle α is set to 0. These equations are based on the reasonable assumption that the thickness of the deposited film (>20 μm) is much larger than the penetration depth (ca. 1 μm) of the evanescent wave (Harrick, 1967).

Molecular modelling

A global search with respect to rotation about the helix direction, assuming tetrameric symmetry, was carried out as described (Adams *et al.*, 1995). In brief, all calculations were performed with the parallel processing version of the crystallography and NMR system (CNS version 0.3; Brunger *et al.*, 1998). The OPLS parameter set with a united atom topology was used, explicitly representing all polar hydrogen and aromatic side-chain atoms (Jorgensen & Tirado-Rives, 1988). All calculations were carried out *in vacuo* with the initial co-ordinates of a canonical α-helix (3.6 residues per turn). Symmetric tetramers were generated by duplicating the helix and rotating it by 90° about the helix axis (inter-helix distance of 10 Å). An initial crossing angle of 25° for left-handed and -25° for right-handed structures was introduced by rotating the long helix axis with respect to the long bundle axis. The values for the crossing angle were found to be of little influence on the outcome of the structures as one can deduce from the results presented in Table 2 in which the crossing angle varied between 7° and 31°.

The symmetric tetramer search was carried out by applying a rotation to all helices simultaneously between $\phi = 0^\circ$ and $\phi = 360^\circ$ in 10 degree steps, as defined in Figure 6. Each starting structure was then subjected to a simulated annealing and energy minimisation protocol as follows: 600 steps of Powel energy minimisation followed by Cartesian molecular dynamics at 600 K for 5 ps and at 300 K for 10 ps. Finally an additional 250 steps of Powell energy minimisation were undertaken. Throughout all of these energy calculations, harmonic restraints were employed to maintain helical geometry and distance restraints were employed to maintain the helix centres within 10.5 Å of one another (Adams *et al.*, 1995). For each starting structure four trials were carried out, using different initial random velocities at both right and left-handed crossing angle, yielding $36 \times 4 \times 2 = 288$ structures to be searched.

After the tetramer search ended, clusters of similar structures were chosen, in which the RMSD between the structures is not larger than 1.0 Å. A cluster was defined by a minimum of ten structures. For each cluster an average structure was calculated, energy minimised and then subjected to the same simulated annealing protocol used in the systematic search.

Orientation refinement

In order to take into account the results obtained from the site-directed dichroism analysis, we have incorporated a new orientation refinement energy term in all molecular dynamics and energy minimisation calculations:

$$E = k_{dichro}(\theta - \theta_0)^2 \qquad (9)$$

whereby θ represents the actual angle and θ_0 the target angle. The overall weight for these orientational constraints $k_{\mathrm{dichro}} = 800$ cal/rad^2, was determined empirically by comparing the results from several MD protocols differing in the orientational constraints weight. The value selected was the minimal weight necessary to obtain an identical outcome. According to the experimental data, two different kinds of restraints were applied (see Figure 7).

(1) A helix tilt restraint, by setting the angle between the vectors connecting every C^α of residue i and C^α of residue $i + 7$ and the z-axis to the value obtained for the helix tilt (the angle β).

(2) The two site-specific dichroisms restraints, by setting the target angles between the corresponding ($1\text{-}^{13}C$)═O bonds of Ala29 and Ala30 and the z-axis to that obtained from the analysis (Arkin *et al.*, 1997b). It was the application of this restraints that resulted in a torque about the helix axis.

Acknowledgements

This work was supported by a grant from the Wellcome Trust. L.M.R. is supported by a predoctoral fellowship from the Howard Hughes Medical Institute.

References

Adams, P., Arkin, I., Engelman, D. & Brunger, A. (1995). Computational searching and mutagenesis suggests a structure for the petameric transmembrane domain of phospholamban. *Nature Struct. Biol.* **2**, **(2)**, 154-162.

Adams, P., Engelman, D. & Brunger, A. (1996). Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins: Stuct. Funct. Genet.* **26**, **(3)**, 257-261.

Akiba, T., Toyoshima, C., Matsunaga, T., Kawamoto, M., T, K., Fukuyama, K., Namba, K. & Matsubara, H. (1996). Three-dimensional structure of bovine cytochrome *bc*1 complex by electron cryomicroscopy and helical image reconstruction. *Nature Struct. Biol.* **3**, 553-561.

Anderson, T., Hellgeth, J. & Lansbury, P. (1992). Isotope edited infrared linear dichroism-determination of amide orientational relationships. *J. Am. Chem. Soc.* **31**, **(51)**, 6540-6546.

Arkin, I., Adams, P., MacKenzie, K., Lemmon, M., Brunger, A. & Engelman, D. (1994). Structural organization of the pentameric transmembrane alpha-helices of phospholamban, a cardiac ion channel. *EMBO J.* **13**, **(20)**, 4757-4764.

Arkin, I., Brunger, A. & Engelman, D. (1997a). Are there dominant membrane protein families with a given number of helices?. *Proteins: Struct. Funct. Genet.* **28**, **(4)**, 465-466.

Arkin, I., MacKenzie, K. & Brunger, A. (1997b). Site-directed dichroism as a method for obtaining rotational and orientational constraints for oriented polymers. *J. Am. Chem. Soc.* **119**, 8973-8190.

Braiman, M. & Rothschild, K. (1988). Fourier transform infrared techniques for probing membrane protein structure. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 541-570.

Brunger, A., Adams, P., Clore, G., Gros, W., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T. & Warren, G. (1999). Crystallography and NMR system: A new software system for macromolecular structure determination. *Acta Crystallog. sect. D*, in the press.

Bui, M., Whittaker, G. & Helenius, A. (1996). Effect of m1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *J. Virol.* **70**, 8391-8401.

Byler, D. & Susi, H. (1986). Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers*, **25**, **(3)**, 469-487.

Doyle, D., Cabral, J., Pfuetzner, R., Kuo, A., Gulbis, J., Cohen, S., Chait, B. & MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science*, **280**, **(3)**, 69-77.

Duff, K. & Ashley, R. (1992). The transmembrane domain of influenza A M2 protein forms amantadine-sensitive proton channels in planar lipid bilayers. *Virology*, **190**, **(1)**, 485-489.

Duff, K., Kelly, S., Price, N. & Bradshaw, J. (1992). The secondary structure of influenza A M2 transmembrane domain. A circular dichroism study. *FEBS Letters*, **311**, **(3)**, 256-258.

Duff, K., Gilchrist, P., Saxena, A. & Bradshaw, J. (1994). Neutron diffraction reveals the site of amantadine blockade in the influenza A M2 ion channel. *Virology*, **202**, **(1)**, 287-293.

Forrest, L., DeGrado, W., Dieckmann, G. & Sansom, M. (1998). Two models of the influenza A M2 channel domain: verification by comparison. *Fold. Design*, **3**, 443-448.

Grambas, S., Bennett, M. & Hay, A. (1992). Influence of amantadine resistance mutations on the pH regulatory function of the M2 protein of influenza A viruses. *Virology*, **191**, **(2)**, 541-549.

Hagmar, P., Norden, B., Baty, D., Chartier, M. & Takahashi, M. (1992). Structure of DNA-RecA complexes studied by residue differential linear dichroism and fluorescence spectroscopy for a genetically engineered RecA protein. *J. Mol. Biol.* **226**, **(4)**, 1193-1205.

Harrick, N. (1967). *Internal Reflection Spectroscopy*, 1st edit., Interscience Publishers, New York.

Hay, A., Wolstenholme, A., Skehel, J. & Smith, M. (1985). The molecular basis of the specific anti-influenza action of amantadine. *EMBO J.* **4**, **(11)**, 3021-3024.

Holsinger, L. & Lamb, R. (1991). Influenza virus M2 integral membrane protein is a homotetramer stabilized by formation of disulfide bonds. *Virology*, **183**, **(1)**, 32-43.

Holsinger, L., Nichani, D., Pinto, L. & Lamb, R. (1994). Influenza A virus M2 ion channel protein: a structure-function analysis. *J. Virol.* **68**, **(3)**, 1551-1563.

Iwata, S., Ostermeier, C. & Michel, H. (1995). Structure at 2.8 Å resolution of cytochrome c oxidase from paraoccus denitrificans. *Nature*, **376**, 660-669.

Jorgensen, W. & Tirado-Rives, J. (1988). The OPLS potential function for proteins, energy minimization for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, **(3)**, 1657-1666.

Kauppinen, J., Moffatt, D., Mantsch, H. & Cameron, D. (1982). Fourier self-deconvolution: a method for resolving intrinsically overlapped bands. *Appl. Spectrosc.* **35**, 271-276.

Kovacs, F. & Cross, T. (1997). Transmembrane four-helix bundle of influenza A M2 protein channel: structural implications from helix tilt and orientation. *Biophys. J.* **73**, **(5)**, 2511-2517.

Kraulis, P. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystalog.* **24**, 946-950.

Lamb, R. & Pinto, L. (1997). Do vpu and vpr of human immunodeficiency virus type 1 and nb of influenza b virus have ion channel activities in the viral life cycles? *Virology*, **229**, 1-11.

Lamb, R., Zebedee, S. & Richardson, C. (1985). Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. *Cell*, **40**, **(3)**, 627-633.

Lemmon, M., Flanagan, J., Treutlein, H., Zhang, J. & Engelman, D. (1992). Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry*, **31**, **(51)**, 12719-12725.

Lemmon, M., Treutlein, H., Adams, P., Brunger, A. & Engelman, D. (1994). A dimerization motif for transmembrane alpha-helices. *Nature Struct. Biol.* **1**, **(3)**, 157-163.

Ludlam, C., Arkin, I., Liu, X., Rothman, M., Rath, P., Aimoto, S., Engelman, D. & Rothschild, K. (1996). FTIR spectroscopy and site-directed isotope labelling as a probe of the local secondary structure of in the transmembrane domain of phospholamban. *Biophys. J.* **70**, 1728-1736.

MacKenzie, K., Prestegard, J. & Engelman, D. (1997). A transmembrane helix dimer: structure and implications. *Science*, **276**, **(5309)**, 131-133.

Norden, B., Elvingson, C., Kubista, M., Sjoberg, B., Ryberg, H., Ryberg, M., Mortensen, K. & Takahashi, M. (1992). Structure of RecA-DNA complexes studied by combination of linear dichroism and small-angle neutron scattering measurements on flow-oriented samples. *J. Mol. Biol.* **226**, **(4)**, 1175-1191.

Pinto, L., Dieckmann, G., Gandhi, C., Papworth, C., Braman, J., Shaughnessy, M., Lear, J., Lamb, R. & DeGrado, W. (1997). A functionally defined model for the M2 proton channel of influenza A virus suggests a mechanism for its ion selectivity. *Proc. Natl Acad. Sci. USA*, **94**, **(21)**, 11301-11306.

Sansom, M. & Kerr, I. (1993). Influenza virus M2 protein: a molecular modelling study of the ion channel. *Protein Eng.* **6**, **(1)**, 65-74.

Sansom, M., Kerr, I., Smith, G. & Son, H. (1997). The influenza A virus M2 channel: a molecular modeling and simulation study. *Virology*, **233**, **(1)**, 163-173.

Shimbo, K., Brassard, D., Lamb, R. & Pinto, L. (1996). Ion selectivity and activation of the M2 ion channel of influenza virus. *Biophys. J.* **70**, **(3)**, 1335-1346.

Sugrue, R. & Hay, A. (1991). Structural characteristics of the M2 protein of influenza A viruses: evidence that it forms a tetrameric channel. *Virology*, **180**, **(2)**, 617-624.

Tadesse, L., Nazarbaghi, R. & Walters, L. (1991). Isotopically enhanced infrared spectroscopy: a novel method for examining the secondary structure at specific sites in conformationally heterogenous peptides. *J. Am. Chem. Soc.* **113**, 7036-7037.

Treutlein, H., Lemmon, M., Engelman, D. & Brunger, A. (1992). The glycophorin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. *Biochemistry*, **31**, **(51)**, 12726-12732.

Tsuboi, M. (1962). Infrared dichroism and molecular conformation of a-form poy-g-benzyl-l-glutamat. *J. Polym. Sci.* **59**, 139-153.

Tsukihara, T., Aoyama, H., Yamashita, T., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., R, Y. & Yoshikawa, S. (1996). The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å resolution. *Science*, **272**, 1136-1144.

Wang, C., Takeuchi, K., Pinto, L. & Lamb, R. (1993). Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block. *J. Virol.* **67**, **(9)**, 5585-5594.

Wang, C., Lamb, R. & Pinto, L. (1995). Activation of the M2 ion channel of influenza virus: a role for the transmembrane domain histidine residue. *Biophys. J.* **69**, **(4)**, 1363-1371.

Edited by G. von Heijne

(Received 4 December 1998; accepted 22 December 1998)