Biochimica et Biophysica Acta 1828 (2013) 2256–2264

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbamem

Review

Gaining insight into membrane protein structure using isotope-edited FTIR ☆

Joshua Manor, Isaiah T. Arkin *

Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmund J. Safra Campus, Jerusalem, 91904, Israel

ARTICLE INFO

Article history:
Received 9 September 2012
Received in revised form 8 November 2012
Accepted 10 November 2012
Available online 27 November 2012

Keywords:
FTIR
Isotope editing
Membrane protein
Molecular modeling

ABSTRACT

FTIR spectroscopy has long been used as a tool used to gain average structural information on proteins. With the advent of stable isotope editing, FTIR can be used to derive accurate information on isolated amino acids. In particular, in an anisotropic sample such as membrane layers, it is possible to measure the orientation of the peptidic carbonyl groups. Herein, we review the theory that enables one to obtain accurate restraints from FTIR spectroscopy, alongside considerations for sample suitability and general applicability. We also propose approaches that may be used to generate structural models of simple membrane proteins based on FTIR orientational restraints. This article is part of a Special Issue entitled: FTIR in membrane proteins and peptide studies.

Contents

☆ This article is part of a Special Issue entitled: FTIR in membrane proteins and peptide studies.

* Corresponding author. Tel./fax: +972 2 658 4329.
E-mail address: arkin@huji.ac.il (I.T. Arkin).

0005-2736/$ – see front matter © 2012 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.bbamem.2012.11.020

1. Introduction

The structural investigation of integral membrane proteins is of significant importance in structural biology, as this family represents the majority of the entire repertoire of available drug targets [1]. The presence of a lipid membrane, however, imposes significant difficulties when applying techniques relying on water-solubility of the protein under investigation. It is in this realm that Fourier Transform Infra Red (FTIR) spectroscopy may be of significant use.

1.1. FTIR spectroscopy

FTIR spectroscopy presents an IR-absorbance spectrum of amide vibrational modes from peptides separately from the lipid vibrations, thus allowing the study of peptides embedded in their native environment. Moreover, as water absorbs dramatically in the infrared regions in which proteins absorb, it is a technique that is particularly suited for the study of membrane proteins. The reason being, is that bulk water can normally be removed without any deleterious effects when studying membrane proteins, but not from their water-soluble counterparts.

FTIR spectra of proteins have traditionally been used to analyze secondary structure content based on a correlation between the peak frequencies and specific structural elements [2–8]. In this respect, FTIR is similar in terms of information to circular dichroism spectroscopy. However, CD employs short wavelengths (ca. 200 nm) that may result in differential absorption flattening when examining large particles, such as proteins embedded in lipid vesicles [9].

Another feature that FTIR spectra is sensitive to is H^+/D^+ exchange, enabling one to determine the solvent accessibility of the protein. Linewidth information in FTIR has also been used to yield information on the polarity of the surrounding area, mostly by 2D-IR spectroscopy [10–13], and recently also by 1D [14]. The principle of the method is based on the correlation between the inhomogenous linewidth of the vibrational mode and the electrostatic field [11,15]. Finally, one can use linear dichroism spectroscopy in order to obtain spatial information on the subject in question, which is the focus of this review.

1.2. Linear dichroism

At the heart of linear FTIR studies, lies the dichroic absorbance of the anisotropic *Amide I* vibrational mode. As the geometry of the *Amide I* transition dipole moment is known relative to the molecular frame [16], one can translate a measured dichroism into a spatial orientation of the peptide's amide groups. The geometry and theoretical background for analyzing both α-helices and β-strands is indeed well established [17,18], nonetheless in this work we narrow the proposed approach to α-helices.

The wavelength of the absorbance of a particular vibrational mode depends on the masses of its atoms and thus can be changed by replacing the atoms with isotopes of differing masses, as expanded below. In the work presented here, 1-^{13}C = ^{18}O probes are used, shifting significantly the absorbance wavelength of such isotope-edited vibrational modes, allowing one to examine and orient individual transition dipole moments. Every isotopically edited probe therefore produces one spatial restraint: the measured angle between the transition dipole moment vector and the laboratory z axis, which in the context of this work is the lipid membrane's normal. This allows the investigator to extract information on individual transition dipole moments, even when other absorbing vibrations coexist in the sample. The site-specific labeling technique creates a map of structural restraints, or a set of signals interpreted as geometric organizations of atoms, bonds, or in the case presented here – electric dipoles of the molecule. After constructing such a map, one can draw definitive conclusions regarding the structure of the peptide, provided there are sufficient number of restraints.

Generally, every amino acid can reside in a variety of backbone dihedral angles. Therefore, it is difficult to convert the aforementioned restraint map to a unique peptide conformation. However, since secondary structure can also be inferred from FTIR spectra, the addition of limited information may allow one to utilize the restraint map in order to refine secondary structures of peptides. This can be achieved even for peptides that are not strictly confined to a canonical form of secondary structure. Moreover, working within the native environment of the peptide, the lipid bilayer, allows one to investigate changes in the peptide's orientation with the introduction of different environmental parameters, such as pH [13].

The conversion of such a restraint map into an actual structure of a hydrophobic peptide with a known sequence is best achieved using molecular dynamics simulations. Molecular dynamics is a powerful tool that has long been used for structural refinements [19–21]. Assimilating this method in the approach presented here, maximizes the ability to objectively extract the information from site-specific transition dipole moments' orientation, thereby not limiting the approach to canonical secondary structures (*e.g.* [22]). This review integrates isotope-edited FTIR spectroscopy and molecular refinement into an easily applicable approach for the investigation of simple membrane peptides.

2. Isotope edited FTIR

Specific chromophores of a molecule have distinctive absorption capabilities in different wavelengths. In proteins, the IR absorption spectra are dominated by vibrational modes arising from the peptide bond (for review see [2–8]). For example, the *Amide I* vibrational mode has a transition dipole moment close to the carbonyl group stretching coordinate and absorbs at 1600–1700 cm^{-1} ($\lambda \approx 6$ μm). The *Amide II* arises mostly from the deformation of the peptidic C – N bond, and has a distinctive absorbance band around 1550 cm^{-1}. Henceforth, all discussion in this review we will focus only on the *Amide I* vibrational mode.

As the transition dipole moment is highly dependent on the spatial arrangement of the molecule, the frequency of *Amide I* vibrational mode is sensitive to the secondary structure. For example, in a β-sheet structure the main *Amide I* absorbance is around 1630 cm^{-1}, while the α-helical peak is normally around 1655 cm^{-1}. Furthermore, there is a slight detectable shift between *Amide I* vibration mode of parallel and antiparallel β-sheets. Taken together, Fourier-transformed IR absorbance spectroscopy is sensitive enough to distinguish between known secondary structure motifs of the molecule.

However, the peak width of a single *Amide I* vibrational mode is significantly larger than the differences between different peptide groups. Therefore, spectral overlap entails that the data that one derives from FTIR spectroscopy normally pertain only to the average properties of the protein. For example, while it is possible to use an FTIR spectrum to state that the protein is mostly helical, it is not possible to state which specific residues in the proteins are helical and which are not.

The advent of isotope-edited FTIR has lifted this limitation due to the fact that stable isotopes can appreciably shift the vibrational mode at the labeled site [23]. Initially, 1-^{13}C = ^{16}O labels were introduced [24] resulting in a downward shift in the vibrational frequency of the *Amide I* mode of approximately 36 cm^{-1}. However, due to the width of the natural abundance (*i.e.* unlabeled protein) of the *Amide I* mode, the 1-^{13}C = ^{16}O isotope edited peak is still not baseline resolved and is present as a shoulder on the main peak. Hence, any detailed analysis that relied on integration necessitated peak deconvolution, and/or peak fitting procedures.

The use of a newer isotope label, 1-^{13}C = ^{18}O alleviated the spectral overlap problem almost entirely, due to the increased frequency shift of -63 cm^{-1} [25,26]. See Fig. 1 for a comparison of spectra from an unlabeled peptide, a peptide labeled with a single 1-^{13}C = ^{16}O label

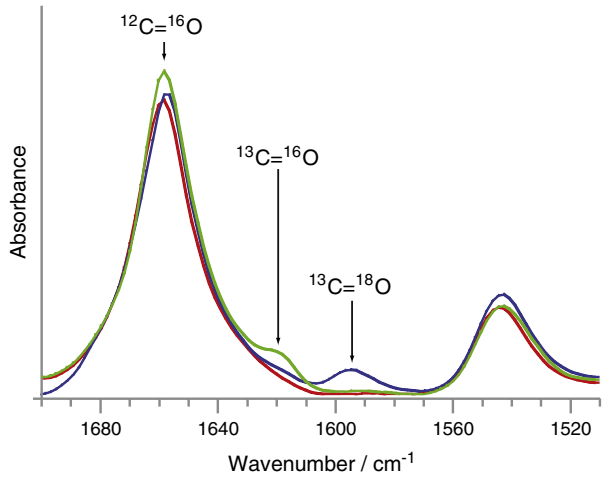

Fig. 1. Effect of labeling on the *Amide I* vibrational mode. FTIR spectra in the region of the *Amide I* and *Amide II* vibrational modes of the influenza A M2 transmembrane domain reconstituted in lipid bilayers [40]. The spectra shown are of samples containing no label (red), a single 1-$^{13}C=^{16}O$ label (green) or a single 1-$^{13}C=^{18}O$ label (blue). The location of the labeled peaks are indicated in the figure.

and a peptide labeled with a single 1-$^{13}C=^{18}O$ label. This probe is now the label of choice when analyzing structures of proteins using 1 or 2D FTIR [10–13,19–21,27–37].

Finally, one obvious limitation in the above strategy is that incorporation of isotopic labels, is only straightforward when studying synthetic peptides. While techniques do exist to incorporate isotopic labels at specific residues in larger proteins, it is still an arduous task. However, synthetic approaches to synthesize larger and larger peptides are continuously improving, including routine techniques for peptide ligation [38,39].

3. Orientation from FTIR spectra

The purpose of the next section is to detail the procedure in which one can derive accurate orientational information from isotope-edited linear dichroism FTIR spectroscopy. Specifically, the aim is to obtain the tilt angle between the *Amide I* transition dipole moment and the *z* axis.

3.1. Dichroic ratio

Absorption is obtained by using parallel ($A_{\parallel}$) and perpendicular ($A_{\perp}$) polarized light. In each polarization, the area of the peak created by the isotope-edited *Amide I* absorption is then measured. The quotient of the above areas yields the dichroic ratio $\mathcal{R}$, of a particular vibrational mode:

$$\mathcal{R} \equiv \frac{A_{\parallel}}{A_{\perp}} \tag{1}$$

Absorption is proportional to the squared scalar product of the light electric field vector $\mathcal{E}$, and the transition dipole moment d, of a particular vibrational mode:

$$A \propto (\mathcal{E} \cdot d)^2 \tag{2}$$

3.2. ATR FTIR

A common geometric configuration that is used in analyzing membrane proteins by FTIR is Attenuated Total internal Reflection (ATR). In this geometry, light is passed through a medium with a high refractive index, such as Germanium, at an angle that leads to total internal reflection (*i.e.* an angle that is greater than the critical angle). At the point of total internal reflection, an evanescent wave is created, that may lead to absorption in the rarer medium (see Fig. 2).

The main advantage of ATR versus transmission spectroscopy in orientational analyses is the dynamic range difference between parallel and polarized absorbances. In brief, a lipid bilayer is a two-dimensional fluid with uniaxial averaging (*i.e.* free rotation) in the *xy* plane. Consequently, the orientational data are obtained with respect to the normal of the lipid bilayer, designated as the *z* axis. Therefore, the biggest range in dichroic ratios would be obtained when the sample is analyzed with light that is parallel to the *z* axis and light that resides in the *xy* plane (the perpendicular beam). In transmission mode this is not possible since the parallel beam is only obtained when the sample is at an angle of 90° to the incident beam. The reason being that the lipid bilayers are stacked in the plane of transmission window. As a result, tilt angles up to 45° are normally used and not 90° (see [41] for a detailed analysis of isotope edited FTIR in transmission geometry). In contrast, in ATR the geometry of the evanescent wave enables a full range of polarizations to be measured, from 0° to 90°.

Therefore, in ATR geometry, decomposing the electric field vector and the transition dipole moment into their axial components, one obtains the dichroic ratio, as follows [42]:

$$\mathcal{R}^{ATR} = \frac{\mathcal{E}_z^2 \mathcal{K}_z^2 + \mathcal{E}_x^2 \mathcal{K}_x^2}{\mathcal{E}_y^2 \mathcal{K}_y^2} \tag{3}$$

whereby $\mathcal{K}_x$, $\mathcal{K}_y$ and $\mathcal{K}_z$ are the corresponding integrated, dimensionless, absorption coefficients for each of the three axes, and $\mathcal{E}_x$, $\mathcal{E}_y$ and $\mathcal{E}_z$ are the axial electric field components.

3.3. Evanescent wave penetration

In order to obtain the values of the axial electric field components in ATR geometry, it is essential to know if the sample is thicker or thinner than the evanescent wave penetration depth d_p, which is the distance in the rarer medium (*i.e.* sample) where the electric field amplitude falls to $1/e$:

$$d_p = \frac{\lambda}{2\pi\sqrt{sin^2\phi - n_{21}^2}} \tag{4}$$

where ϕ is the angle of incidence (normally 90°), and n_{21} is the ratio between the refractive indices of the sample and the internal reflection element. Since the refractive index of a hydrated lipid bilayer is $n_2 = 1.43$, the penetration depth using a Germanium internal reflection element ($n_1 = 4.0$) is approximately 1 μm. See Fig. 3a for a correlation between the evanescent wave penetration depth and the refractive index of the internal reflection element. See [41] for a discussion on using isotope edited IR in transmission geometry.

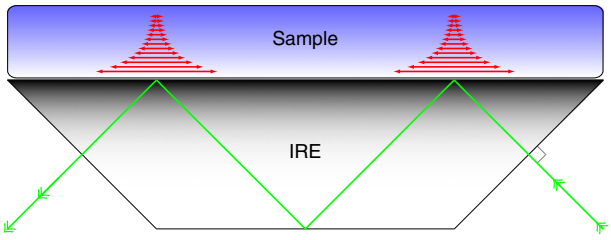

Fig. 2. Schematic digram of the geometric configuration of attenuated total internal reflection spectroscopy containing three reflections. The internal reflection element (IRE) is in gray, also known as the denser medium, while the sample is the rarer medium shown in blue. The incident beam is shown in green, while the evanescent wave is in red.

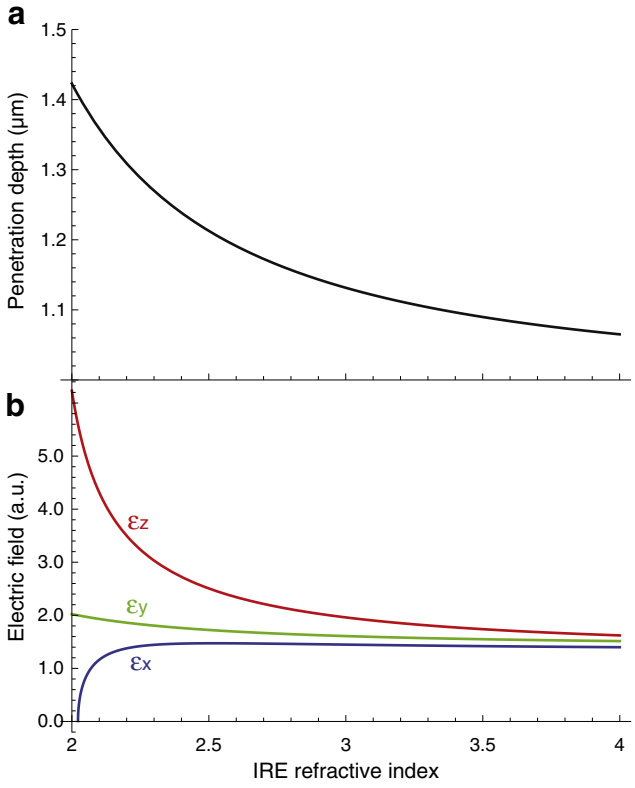

Fig. 3. a. Dependence of the electric field penetration depth in the rarer medium (*i.e.* sample) as a function of the refractive index of the internal reflection element. The wavelength in question is 6.25 μm, which is equivalent to 1600 cm^{-1}, which is around the position of the isotope edited *Amide I* vibrational mode. b. Strength of the axial electric field components as a function of the refractive index of the internal reflection element.

3.4. Electric field components

Based on the above, when the sample thickness is sufficiently larger than the evanescent wave penetration depth (ca. 1 μm, in the case of a Ge internal reflection element), we can use the thick film approximation [43]. See Section 4 for the experimental details. In such an instance the electric field components are given by Harrick [42] as:

$$\mathcal{E}_x = \frac{2\cos\varphi\left(\sin^2\varphi - n_{21}^2\right)^{1/2}}{\left(1-n_{21}^2\right)^{1/2}\left[\left(1+n_{21}^2\right)\sin^2\varphi - n_{21}^2\right]^{1/2}} \tag{5}$$

$$\mathcal{E}_y = \frac{2\cos\varphi}{\left(1-n_{21}^2\right)^{1/2}} \tag{6}$$

$$\mathcal{E}_z = \frac{2\cos\varphi\sin\varphi}{\left(1-n_{21}^2\right)^{1/2}\left[\left(1+n_{21}^2\right)\sin^2\varphi - n_{21}^2\right]^{1/2}} \tag{7}$$

where φ is the angle of incidence between the infrared beam and the internal reflection element (normally 45°). In the case of a Ge internal reflection element ($n_1 = 4.0$), the values are $\mathcal{E}_x = 1.398$, $\mathcal{E}_y = 1.516$ and $\mathcal{E}_z = 1.625$, as shown in Fig. 3b.

3.5. Absorption coefficient

For any particular axis *i*, the integrated absorption coefficient is the axial projection of the transition dipole moment on the axis multiplied by its spatial distribution:

$$\mathcal{K}_i = \int F(\alpha)\mathcal{K}\cdot\hat{i}\, d\alpha \tag{8}$$

where α is the angle between the axis and the transition dipole moment. When the axis in question is the *z* axis, the angle is commonly referred to as θ. $F(\theta)$ is normally modeled as a Gaussian distribution [44], with the following probability of finding a transition dipole moment with a tilt angle θ:

$$F(\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} \tag{9}$$

where μ and σ are the mean tilt angle and the standard deviation, respectively. However, we have previously shown that when sample mosaicity is lower than a given level ($\sigma < 5^\circ$), it may be neglected altogether [22]. In such systems, in a uniaxial symmetric system, such as a lipid bilayer, the integrated absorption coefficients are readily given by:

$$\mathcal{K}_z^2 = \cos^2\theta \tag{10}$$

$$\mathcal{K}_x^2 = \mathcal{K}_y^2 = \frac{1}{2}\sin^2\theta \tag{11}$$

where θ is the angle between the *z* axis and the transition dipole moment of the vibrational mode in question.

3.6. Tilt angle derivation

Assimilating the electric field components, alongside the axial integrated absorption coefficients, we obtain the dichroic ratio, $\mathcal{R}^{ATR}$ as a function of the aforementioned tilt angle, using the following relationship:

$$\mathcal{R}^{ATR} = \frac{\mathcal{E}_z^2\ \cos^2\theta + \mathcal{E}_x^2\ \frac{1}{2}\sin^2\theta}{\mathcal{E}_y^2\ \frac{1}{2}\sin^2\theta}. \tag{12}$$

and the tilt angle of the transition dipole moment from the *z* axis may be obtained from the measured dichroic ratio using the following simple relationship (shown in Fig. 4):

$$\theta = \arctan\sqrt{\frac{2.14153}{\mathcal{R}^{ATR} - 0.923832}} \tag{13}$$

It is important to note that in such experiments the calculated angle is invariably in the first quadrant ($0 \le \theta \le \pi/2$), despite the fact that the actual tilt angle might be one of the following four equal possibilities: θ, $-\theta$, $\pi+\theta$ or $\pi-\theta$.

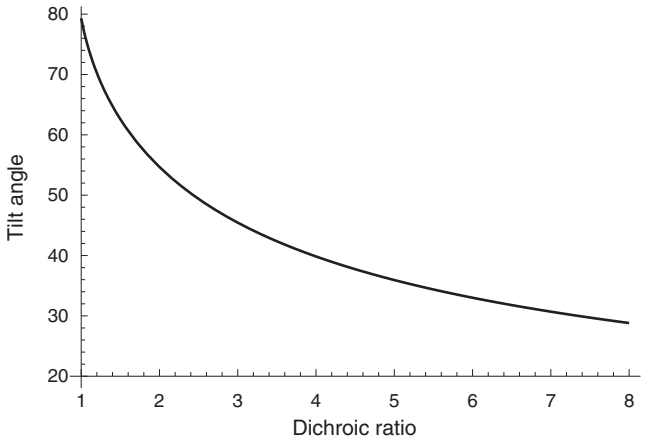

Fig. 4. Tilt angle between the *z* axis and the transition dipole moment of the *Amide I* mode as a function of the dichroic ratio.

4. Experimental details

Below we describe in brief, the experimental procedures one uses in order to obtain isotope edited FTIR spectra of transmembrane peptides.

4.1. Sample preparation

4.1.1. Isotopic labeling

The biggest limitation of isotope edited-FTIR analysis is the need to incorporate isotopic labels in the protein. In synthetic peptides, the procedure is straightforward, whereby the 1-$^{13}C=^{18}O$ labeled amino acid simply replaces an unlabeled residue at the appropriate position. In terms of length, synthetic peptides are normally limited to a few dozen residues, and ligation may be used to double the size of the final product [45]. Incorporation of isotopic labels into biologically derived proteins is more of a challenge, requiring techniques such as tRNA Mediated Protein Engineering (Ambergene, Watertown, MA). Finally, the synthesis of the 1-$^{13}C=^{18}O$ amino acid precursors is once more straightforward, and has been described in [25,46].

4.1.2. Peptide synthesis, purification and membrane reconstitution

Peptide synthesis is now a routine procedure, even for hydrophobic transmembrane peptides, normally employing Fmoc chemistry. The purification of such peptides is normally achieved by using reverse phase HPLC employing less hydrophobic columns than those that are used to purify water soluble peptides. For example, C4 or cyano stationary phases are common choices for purification, using 2-propanol or 2,2,2-Trifluoroethanol as eluants [47]. The fractions that contain the peptides, which can be confirmed by mass spectrometry, are then lyophilized under high vacuum in systems that are geared to capture organic solvents with low melting points. Finally reconstitution can be achieved by using two common routes:

1 *Organic solvent co-solubilization:* The factions that contain the peptides are dissolved by an organic solvent, normally 1,1,1,3,3,3-Hexafluoro-2-propanol. To this solution lipids are added at the appropriate ratio. Finally, reconstitution is achieved after the evaporation of the organic solvent and subsequent hydration.
2 *Detergent dialysis*: The factions that contain the peptides are dissolved in an aqueous solution of a detergent with a high critical micellar concentration (CMC), such as n-β-octyl glucoside. Lipids are then added to this solution leading to the formation of mixed micelles that contain detergent, peptide and lipid. Finally, reconstitution is achieved by dialysis of the detergent into the external solution leaving the peptide and lipid together to form bilayers.

4.2. FTIR spectra collection

As stated above, ATR-FTIR is the preferred geometry for orientational analysis. Therefore, the first step in spectra collection is the deposition of the peptide-containing lipid bilayers onto the internal reflection element. As an example, ca. 400 µl of the vesicle solution are pipetted on a $50\times20\times2$ mm Ge crystal, after which the bulk solvent is removed by blowing dry air on the sample. If the vesicle solution contains 5 mg of lipid then it is possible to calculate that roughly 1500 bilayers will be formed. If the thickness of every bilayer including the intervening hydration layer is around 50 Å (easily measured by using X-ray reflectivity [22]) then the total thickness of the sample is ca. 7.5 µm. Note that this is much larger than the penetration depth of the evanescent wave, which in Ge at 1600 cm^{-1} is approximately 1 µm (see Fig. 3). This justifies the use of the thick film approximation since the evanescent wave decays entirely within the sample [43].

The final consideration that merits attention is the integration of the isotope-edited peak. Since the peak at times is not completely baseline resolved from the unlabeled *Amide I*, one needs to estimate the baseline. In our experience the best results are obtained by averaging the two extreme cases: (i) The peak sits entirely on the margins of the natural-abundance peak, (ii) versus a complete separation between the two peaks and the baseline is a straight line from one end of the peak to the other. Subsequently a Voigt profile is used to integrate the peak according to the selected baseline. The reader is referred to [14] for a detailed description of the peak fitting procedure.

5. Structural modeling

We have previously shown that FTIR spectroscopy can yield highly accurate orientational restraints. For example, in the case of the influenza M2 channel, the average tilt angle difference between the results from FTIR spectroscopy and those calculated from a structure solved by X-ray crystallography was only 6° [13]. This poses the following challenge: can the experimental angles that are derived from isotope edited FTIR spectroscopy be used in structural modeling of simple membrane proteins? Below we show the approaches that we have taken to meet the above challenge. We focus on two routes with different levels of complexity: (i) rigid body modeling that may be suited for simple helical bundles; and (ii) molecular dynamics based refinement that is suited for more complex structures.

5.1. Rigid body modeling

In the above study on the M2 channel [13], we were able to derive an accurate backbone model of the M2 channel based on the orientational restraints using rigid body modeling. In this instance the protein is constructed as a canonical helix with its director aligned along the *z* axis. The peptide is then tilted in 1° increments until it resides in the *xy* plane (*i.e.* tilt angle of 90°), and at each tilt angle the peptide is rotated about its director by 1° increments until an entire revolution was obtained. Taken together, $32{,}400=90\times360$ tilt and rotation combinations are created. Subsequently, at each tilt and rotation combination, the angles between the l *Amide I* transition dipole moments and the *z* axis are calculated and compared to the experimentally derived values.

The success of the above procedure (see Fig. 5), yielding a structure that is within 2.5 Å backbone RMSD from the X-ray structure, most likely stems from the fact that the influenza M2 channel adopts a canonical helix structure [48]. Therefore, a procedure that is based

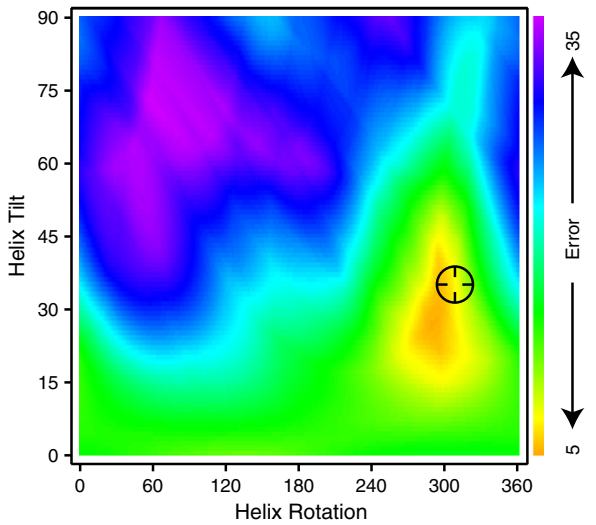

Fig. 5. Rigid body modeling of the Influenza M2 channel based on FTIR derived orientational restraints [13]. The average difference per residue between the transition dipole moment tilt angle, as a function of the helix tilt and rotation angles is depicted in color (see scale on right). The crosshairs represent the tilt and rotation of the helices of the X-ray structure [49].

on rigid body modeling has a high success rate. However, the same approach will probably fail when applied to structures that are not strictly canonical.

Below we propose a procedure to employ orientational restraints to produce a backbone model for simple transmembrane structures that need not be strictly canonical, using MD based refinement. We test the procedure on two proteins with known structures and demonstrate its applicability.

5.2. MD based refinement

In order to incorporate orientational restraints in a model building exercise we made use of molecular dynamics simulations. The reason being is that it allows the model to conform to the experimental results (*i.e.* orientational restraints), while maintaining correct protein geometry [20]. Such an approach may prove useful to model structures that differ from canonical helices, for which rigid body modeling may suffice [13].

5.2.1. Error calculation

In any molecular dynamics simulation, the aim is to lower the energy of the system. Since the total energy of the system is a sum of the normal force field components (*i.e.* bonded and non-bonded terms) and the refinement energy (a.k.a. penalty function), it is important to know exactly what one measures experimentally. Since the absorption of light is proportional to the squared scalar product of the electric field vector and the integrated absorption coefficient, angular ambiguity arises. In other words, while the calculated angle from a dichroism experiment is invariably in the first quadrant ($0 \le \theta \le \pi/2$), the actual angle might be one of the following four equal options: θ, $-\theta$, $\pi+\theta$ or $\pi-\theta$. Therefore, the difference Δ, between the experimental tilts θ, and the tilt angles for any model obtained ϑ, should be calculated as the minimum of the following four options:

$$\Delta = min\{|\vartheta-\theta|, |\vartheta+\theta|, |\vartheta-(\pi-\theta)|, |\vartheta-(\pi+\theta)|\} \quad (14)$$

5.2.2. Molecular dynamics details

5.2.2.1. Set up. The molecular dynamics (MD) simulations were carried out using CNS [50,51], employing torsion angle dynamics [52–54], with an annealing protocol for both the temperature and the Van der Waals force constant.

The peptide was constructed as an ideal helix using PyMOL (Schrödinger, NY) and included two dummy atoms in every residue. The exact geometry of the atoms within the peptide group plane, was such that the angle between the dummy atoms coincided exactly with the transition dipole moment of the *Amide I* vibrational mode (see green atoms in Fig. 6a). The two dummy atoms were constrained to their position using bond, angle and dihedral terms. However, they did not participate in any non-bonded interactions. All other force field parameters employed default values using the OPLS parameter set with a united atom topology [55]. Calculations were carried out *in vacuo* with a dielectric constant of 1, using a non-bonded cutoff of 13 Å, and a switching function was applied to Van der Waals interactions between 10 Å and 12 Å.

5.2.2.2. Orientational refinement. At each simulation step the angle between the vector connecting the two dummy atoms and the z axis was calculated (ϑ) and restrained to the target angles (θ) for the corresponding residue according to the following energy penalty function:

$$E = k\Delta^2 \quad (15)$$

where k is the harmonic force content equal to 400 kcal/rad^2. Δ is the difference between the target angles and those measured currently, as detailed in Section 5.2.1.

5.2.2.3. Enforcing helical geometry. The information content in a single orientational restraint per residue is insufficient to produce a complete structure due to the large number of degrees of freedom that exist. However, one can make use of the fact that the frequency of the isotope edited *Amide I* peak is also indicative of the secondary structure of the labeled site [23–25]. Therefore, as indicated, different segments of the proteins were restrained to a helical geometry by an NOE restraint between the carbonyl oxygen of residue i and the amide proton of residue $i+4$. The NOE distance was set to 3.15 Å with a force constant of 200 kcal/Å^2 with a ceiling of 1000 kcal/Å^2.

5.2.2.4. Angular ambiguity in a continuous helix. All transition dipole moment tilt angles in a mildly inclined, continuous helix will either be in quadrants I and IV (right hemisphere, $\pm\theta$) or II and III (left hemisphere, $\pi\pm\theta$). The reason for this geometric fact is that in an α-helix, the angle between the transition dipole moment and the helix axis is relatively small, 35° [16]. Therefore, in regions that were restrained to a helical geometry, the orientational refinement was calculated as follows: The differences between the current angle ϑ, and the target angle θ, was the minimum of the following two terms:

$$\Delta = min\{|\vartheta-\theta|, |\vartheta+\theta|\} \quad (16)$$

whenever the helix that is tilted from the z axis by an angle smaller than $^{\pi}/_{4}$. When the helix tilt angle was larger than $^{\pi}/_{4}$, the difference

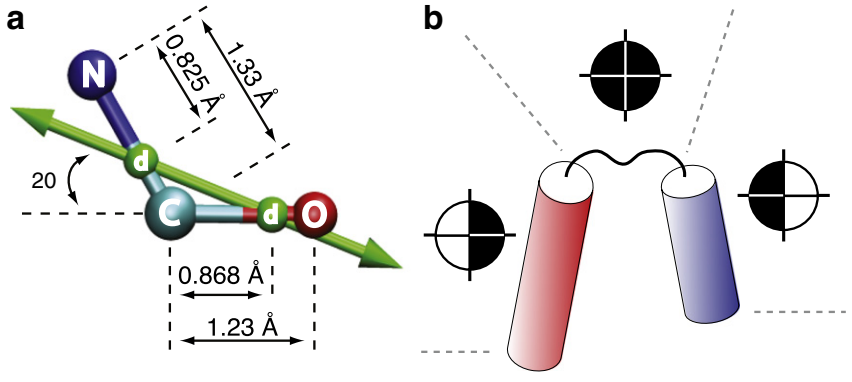

Fig. 6. a. Geometry of the *Amide I* vibrational mode's transition dipole moment (green arrow) according to [16]. The two dummy atoms used in the molecular dynamics refinement process are in green. b. Schematic of the refinement process showing the quadrants of the angular restraints that were imposed during the simulation. In the first helix (red) the angles were restrained to $\pm\theta$ while in the second helix (blue) they were restrained to $\pi\pm\theta$. In the region between the two helices the angles were refined to any of the four possible quadrants (*i.e.* θ, $-\theta$, $\pi+\theta$ or $\pi-\theta$).

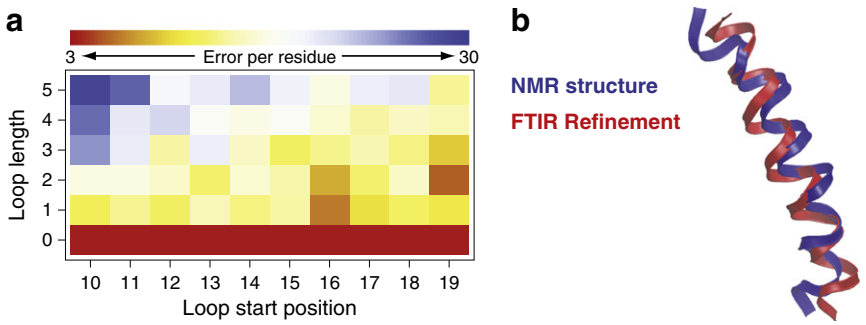

Fig. 7. a. Refinement results of Neuropeptide Y peptide showing the average angular error per residue as a function of the loop length and position. The color coding represents the error as shown in the scale bar. The loop starting position reflects the number of amino acids from the start of the protein and not the actual position in the sequence of the original protein. The bottom row of the plot depicts the refinement of a system without a loop (loop size 0). b. Superpositioning of the NMR structure of Neuropeptide Y (blue) with the results of the orientational refinement (red).

between the current tilt angle and the experimental results were taken as the minimum of:

$$\Delta = min\{|\vartheta-(\pi-\theta)|, |\vartheta-(\pi+\theta)|\} \quad (17)$$

See Fig. 6b for a scheme of the two different calculation options in case of a helical geometry. When the region was not restrained to an α-helical geometry, all four quadrants were used to calculate the difference from the experiential data (Eq. (14)).

5.2.2.5. Loop searching protocol. Simulations were set up to seek a short, non-helical/loop segment flanked by two helical regions and to test whether such a motif properly approximates the target angles, rather than a straight canonical helix. A schematic representation of the loop search algorithm is shown in Fig. 6b. A canonical helix is built according to the peptide sequence. A loop segment is sought throughout the helix, in various positions and sizes. Flanking helices integrity is preserved by maintaining proper hydrogen bonding using NOE restraints. Each flanking helix's transition dipole moment is forced to align with the target FTIR angles in either the first quadrant (preceding helix) or in the second quadrant (succeeding helix). Loop area orientational restraints are free to align to either quadrant and helical integrity is not enforced by using NOE restraints. Therefore it is free to break the α-helix conformation.

5.2.2.6. Annealing protocol. Numerous protocols were tested in order to reach a consistent answer that yields the smallest difference to the target data (*i.e.* maximal refinement) while maintaining correct geometry. The temperature in the simulations was reduced linearly from 598°K to 273°K in 15°K increments (22 cycles), while the Van der Waals force was multiplied by an exponentially increasing constant from 0.03 to 1 in each cycle. The number of simulation steps in each cycle was determined by the total simulation length. Simulation steps were 0.1 fs and the total length of the simulation was 20 ps, after which additional convergence was not obtained.

5.3. Refinement examples

In order to test the suitability of the approach, two different transmembrane segments were tested. Each was taken from an experimentally determined structure of a protein, with an idea of extracting the orientational restraints and using them to regenerate the structure in a refinement process. FTIR target tilt angles were derived from direct calculations of the PDB structure, serving as an input for the molecular dynamics simulation. The initial structure of the two peptides was an ideal α-helix residing in the membrane plane (*xy* plane) and the target angles were added as running restraints. As stated above, a variety of simulations was conducted, with varying parameters, such as: the length of simulations, the convergence criteria, and the alignment of the individual residues to the measured angles (strength and distance).

5.3.1. Neuropeptide Y

The first non-canonical helix that we attempted to model using the above procedure was a fragment from Neuropeptide Y (PDB code 1RON [56]). The peptide contained 24 amino acids (residues 12 to 36) with a wide bend.

The refinement process produces a clear result. Specifically, the criterion we used to select the best outcome is the smallest average angular error per residue, which reflects how closely the resultant structure abides by the orientational restraints. The results show that a single helix ($\Delta = 3.3^{\circ}$) is much better than the best helix–loop–helix structure ($\Delta = 5.2^{\circ}$), as shown in Fig. 7. Furthermore, comparing the structure obtained from the orientational refinement to

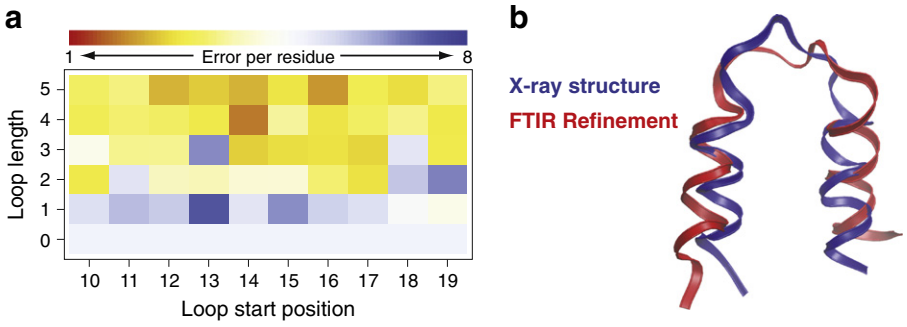

Fig. 8. a. Refinement results of the bacteriorhodopsin helix–loop–helix fragment showing the average angular error per residue as a function of the loop length and position. The color coding represents the error as shown in the scale bar. The loop starting position reflects the number of amino acids from the start of the protein and not the actual position in the sequence of the original protein. The bottom row of the plot depicts the refinement of a system without a loop (loop size 0). b. Superpositioning of the X-ray structure of bacteriorhodopsin (blue) with the results of the orientational refinement (red).

the NMR structure [56], from which the angular information was taken, is highly encouraging, yielding a backbone RMSD of 2.3 Å.

It is important to note that in the case of neuropeptide Y, the refinement points to it being a helix rather than a helix–loop–helix despite the additional degrees of freedom that exists in a loop. In other words, in a single helix the transition dipole tilt angles are all forced to be in one hemisphere (see discussion above). However, in a loop all four quadrants are possible. Hence, all else being equal, a loop is more permissive in a refinement process than a continuous helix. Therefore, the fact the procedure was able to rule out a loop containing structure is encouraging.

5.3.2. Helix–loop–helix from bacteriorhodopsin

The second refinement target that we chose was a symmetric 31 amino acids long helix–loop–helix from bacteriorhodopsin (PDB code 3HAO, residues 87 to 117 [57]). The loop segment contains five amino acids from residues 15 to 19 (numbered from the start of the segment).

The refinement results show that the structure with the lowest error relative to the experiential data was one with a loop of four residues starting at residue 14 ($\Delta = 1.8^{\circ}$). The average error is much lower when compared to that obtained from a structure of a single helix ($\Delta = 5.1^{\circ}$). In other words the procedure was readily able to delineate a helix–loop–helix from a continuous helix based on the orientational restraints. A superposition of the X-ray structure with the refinement results (shown in Fig. 8) exemplifies the successes of the approach, yielding a backbone RMSD of 4.3 Å between the two.

6. Concluding remarks

The aim of this work was to provide an overview of how isotope edited FTIR spectroscopy may be used to gain insight into the structure of small membrane proteins. The theory and the data analysis needed to derive accurate orientational restraints from FTIR are both straightforward and robust. Finally, a refinement procedure to incorporate the orientational data in an objective manner is presented and tested on two known structures.

Acknowledgments

This work was supported in part by grants from the National Institutes of Health (R21AI064797), the Israeli Science Foundation (784/01, 1249/05, 1581/08) and the Horowitz Foundation. ITA is the Arthur Lejwa Professor of Structural Biochemistry at the Hebrew University of Jerusalem.

References

[1] C. Zheng, L. Han, C.W. Yap, B. Xie, Y. Chen, Progress and problems in the exploration of therapeutic targets, Drug Discov. Today 11 (2006) 412–420.
[2] W.K. Surewicz, H.H. Mantsch, D. Chapman, Determination of protein secondary structure by Fourier transform infrared spectroscopy: a critical assessment, Biochemistry 32 (1993) 389–394.
[3] P.I. Haris, D. Chapman, The conformational analysis of peptides using Fourier transform ir spectroscopy, Biopolymers 37 (1995) 251–263.
[4] C. Vigano, L. Manciu, F. Buyse, E. Goormaghtigh, J.M. Ruysschaert, Attenuated total reflection ir spectroscopy as a tool to investigate the structure, orientation and tertiary structure changes in peptides and membrane proteins, Biopolymers 55 (2000) 373–380.
[5] I. Martin, E. Goormaghtigh, J.-M. Ruysschaert, Attenuated total reflection ir spectroscopy as a tool to investigate the orientation and tertiary structure changes in fusion proteins, Biochim. Biophys. Acta 1614 (2003) 97–103.
[6] M.C. Manning, Use of infrared spectroscopy to monitor protein structure and stability, Expert Rev. Proteomics 2 (2005) 731–743.
[7] A. Barth, Infrared spectroscopy of proteins, Biochim. Biophys. Acta 1767 (2007) 1073–1101.
[8] M.S. Braiman, K.J. Rothschild, Fourier transform infrared techniques for probing membrane protein structure, Annu. Rev. Biophys. Biophys. Chem. 17 (1988) 541–570.
[9] B.A. Wallace, C.L. Teeters, Differential absorption flattening optical effects are significant in the circular dichroism spectra of large membrane fragments, Biochemistry 26 (1987) 65–70.
[10] P. Mukherjee, I. Kass, I.T. Arkin, M.T. Zanni, Structural disorder of the CD3zeta transmembrane domain studied with 2D IR spectroscopy and molecular dynamics simulations, J. Phys. Chem. B 110 (2006) 24740–24749.
[11] P. Mukherjee, I. Kass, I.T. Arkin, I. Arkin, M.T. Zanni, Picosecond dynamics of a membrane protein revealed by 2D IR, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 3528–3533.
[12] P. Mukherjee, A.T. Krummel, E.C. Fulmer, I. Kass, I.T. Arkin, M.T. Zanni, Site-specific vibrational dynamics of the CD3zeta membrane peptide using heterodyned two-dimensional infrared photon echo spectroscopy, J. Chem. Phys. 120 (2004) 10215–10224.
[13] J. Manor, P. Mukherjee, Y.-S. Lin, H. Leonov, J.L. Skinner, M.T. Zanni, I.T. Arkin, Gating mechanism of the influenza A M2 channel revealed by 1D and 2D IR spectroscopies, Structure 17 (2009) 247–254.
[14] J. Manor, E.S. Feldblum, M.T. Zanni, I.T. Arkin, Environment polarity in proteins mapped noninvasively by FTIR spectroscopy, J. Phys. Chem. Lett. 3 (2012) 939–944.
[15] L. Wang, C.T. Middleton, M.T. Zanni, J.L. Skinner, Development and validation of transferable amide i vibrational frequency maps for peptides, J. Phys. Chem. B 115 (2011) 3713–3724.
[16] H. Torii, M. Tasumi, Model-calculations on the amide-i infrared bands of globular-proteins, J. Chem. Phys. 96 (1992) 3379–3387.
[17] I. Arkin, K. MacKenzie, A. Brunger, Site-directed dichroism as a method for obtaining rotational and orientational constraints for oriented polymers, J. Am. Chem. Soc. 119 (1997) 8973–8980.
[18] D. Marsh, Dichroic ratios in polarized Fourier transform infrared for nonaxial symmetry of beta-sheet structures, Biophys. J. 72 (1997) 2710–2718.
[19] M. Nilges, A.M. Gronenborn, A.T. Brünger, G.M. Clore, Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2, Protein Eng. 2 (1988) 27–38.
[20] A.T. Brünger, Crystallographic refinement by simulated annealing. Application to a 2.8 a resolution structure of aspartate aminotransferase, J. Mol. Biol. 203 (1988) 803–816.
[21] A.T. Brünger, J. Kuriyan, M. Karplus, Crystallographic r factor refinement by molecular dynamics, Science 235 (1987) 458–460.
[22] J. Manor, Z. Khattari, T. Salditt, I.T. Arkin, Disorder influence on linear dichroism analyses of smectic phases, Biophys. J. 89 (2005) 563–571.
[23] I.T. Arkin, Isotope-edited IR spectroscopy for the study of membrane proteins, Curr. Opin. Chem. Biol. 10 (2006) 394–401.
[24] L. Tadesse, R. Nazarbaghi, L. Walters, Isotopically enhanced infrared-spectroscopy — a novel method for examining secondary structure at specific sites in conformationally heterogeneous peptides, J. Am. Chem. Soc. 113 (1991) 7036–7037.
[25] J. Torres, A. Kukol, J.M. Goodman, I.T. Arkin, Site-specific examination of secondary structure and orientation determination in membrane proteins: the peptidic (13)c=(18)o group as a novel infrared probe, Biopolymers 59 (2001) 396–401.
[26] J. Torres, P.D. Adams, I.T. Arkin, Use of a new label, (13)==(18)o, in the determination of a structural model of phospholamban in a lipid bilayer. Spatial restraints resolve the ambiguity arising from interpretations of mutagenesis data, J. Mol. Biol. 300 (2000) 677–685.
[27] E.B. Dunkelberger, L.E. Buchanan, P. Marek, P. Cao, D.P. Raleigh, M.T. Zanni, Deamidation accelerates amyloid formation and alters amylin fiber structure, J. Am. Chem. Soc. 134 (30) (2012) 12658–12667.
[28] A. Remorino, R.M. Hochstrasser, Three-dimensional structures by two-dimensional vibrational spectroscopy, Acc. Chem. Res. 45 (11) (2012) 1896–1905.
[29] C. Falvo, W. Zhuang, Y.S. Kim, P.H. Axelsen, R.M. Hochstrasser, S. Mukamel, Frequency distribution of the amide-i vibration sorted by residues in amyloid fibrils revealed by 2D-IR measurements and simulations, J. Phys. Chem. B 116 (2012) 3322–3330.
[30] A. Remorino, I.V. Korendovych, Y. Wu, W.F. DeGrado, R.M. Hochstrasser, Residue-specific vibrational echoes yield 3D structures of a transmembrane helix dimer, Science 332 (2011) 1206–1209.
[31] D.C. Urbanek, D.Y. Vorobyev, A.L. Serrano, F. Gai, R.M. Hochstrasser, The two dimensional vibrational echo of a nitrile probe of the villin hp35 protein, J. Phys. Chem. Lett. 1 (2010) 3311–3315.
[32] J. Torres, J.A.G. Briggs, I.T. Arkin, Multiple site-specific infrared dichroism of CD3-zeta, a transmembrane helix bundle, J. Mol. Biol. 316 (2002) 365–374.
[33] J. Torres, J.A.G. Briggs, I.T. Arkin, Convergence of experimental, computational and evolutionary approaches predicts the presence of a tetrameric form for CD3-zeta, J. Mol. Biol. 316 (2002) 375–384.
[34] J. Torres, J. Wang, K. Parthasarathy, D.X. Liu, The transmembrane oligomers of coronavirus protein e, Biophys. J. 88 (2005) 1283–1290.
[35] J. Torres, K. Parthasarathy, X. Lin, R. Saravanan, A. Kukol, D.X. Liu, Model of a putative pore: the pentameric alpha-helical bundle of SARS coronavirus e protein in lipid bilayers, Biophys. J. 91 (2006) 938–947.
[36] K. Pervushin, E. Tan, K. Parthasarathy, X. Lin, F.L. Jiang, D. Yu, A. Vararattanavech, T.W. Soong, D.X. Liu, J. Torres, Structure and inhibition of the SARS coronavirus envelope protein ion channel, PLoS Pathog. 5 (2009) e1000511.
[37] A. Kukol, J. Torres, I.T. Arkin, A structure for the trimeric MHC class II-associated invariant chain transmembrane domain, J. Mol. Biol. 320 (2002) 1109–1117.
[38] I. Ghosh, N. Considine, E. Maunus, L. Sun, A. Zhang, J. Buswell, T.C. Evans Jr., M.-Q. Xu, Site-specific protein labeling by intein-mediated protein ligation, Methods Mol. Biol. 705 (2011) 87–107.
[39] P.J. Focke, F.I. Valiyaveetil, Studies of ion channels using expressed protein ligation, Curr. Opin. Chem. Biol. 14 (2010) 797–802.
[40] 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.
[41] E. Arbely, I. Kass, I.T. Arkin, Site-specific dichroism analysis utilizing transmission FTIR, Biophys. J. 85 (2003) 2476–2483.

[42] N.J. Harrick, Internal Reflection Spectroscopy, Interscience Publishers, New York, 1967.
[43] P.H. Axelsen, M.J. Citra, Orientational order determination by internal reflection infrared spectroscopy, Prog. Biophys. Mol. Biol. 66 (1996) 227–253.
[44] I. Kass, E. Arbely, I.T. Arkin, Modeling sample disorder in site-specific dichroism studies of uniaxial systems, Biophys. J. 86 (2004) 2502–2507.
[45] P.E. Dawson, S.B. Kent, Synthesis of native proteins by chemical ligation, Annu. Rev. Biochem. 69 (2000) 923–960.
[46] J. Marecek, B. Song, S. Brewer, J. Belyea, R.B. Dyer, D.P. Raleigh, A simple and economical method for the production of 13C,18O-labeled Fmoc-amino acids with high levels of enrichment: applications to isotope-edited IR studies of proteins, Org. Lett. 9 (2007) 4935–4937.
[47] L.E. Fisher, D.M. Engelman, High-yield synthesis and purification of an alpha-helical transmembrane domain, Anal. Biochem. 293 (2001) 102–108.
[48] F.A. Kovacs, J.K. Denny, Z. Song, J.R. Quine, T.A. Cross, Helix tilt of the M2 transmembrane peptide from influenza a virus: an intrinsic property, J. Mol. Biol. 295 (2000) 117–125.
[49] A.L. Stouffer, R. Acharya, D. Salom, A.S. Levine, L. Di Costanzo, C.S. Soto, V. Tereshko, V. Nanda, S. Stayrook, W.F. DeGrado, Structural basis for the function and inhibition of an influenza virus proton channel, Nature 451 (2008) 596–599.
[50] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR system: a new software suite for macromolecular structure determination, Acta Crystallogr. D Biol. Crystallogr. 54 (1998) 905–921.
[51] A.T. Brunger, Version 1.2 of the crystallography and NMR system, Nat. Protoc. 2 (2007) 2728–2733.
[52] A.T. Brünger, P.D. Adams, L.M. Rice, Recent developments for the efficient crystallographic refinement of macromolecular structures, Curr. Opin. Struct. Biol. 8 (1998) 606–611.
[53] E.G. Stein, L.M. Rice, A.T. Brünger, Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation, J. Magn. Reson. 124 (1997) 154–164.
[54] L.M. Rice, A.T. Brünger, Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement, Proteins 19 (1994) 277–290.
[55] W.L. Jorgensen, J. Tirado-Rives, The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, J. Am. Chem. Soc. 110 (1988) 1657–1666.
[56] S.A. Monks, G. Karagianis, G.J. Howlett, R.S. Norton, Solution structure of human neuropeptide y, J. Biomol. NMR 8 (1996) 379–390.
[57] N.H. Joh, A. Oberai, D. Yang, J.P. Whitelegge, J.U. Bowie, Similar energetic contributions of packing in the core of membrane and water-soluble proteins, J. Am. Chem. Soc. 131 (2009) 10846–10847.