# Site-specific IR spectroscopy and molecular modelling combined towards solving transmembrane protein structure

Andreas Kukol

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL

Tel.: +44 (0)24 7652 8373, FAX: +44 (0)24 7652 3701, Email:

a.kukol@warwick.ac.uk

Abstract – Membrane protein structures are underrepresented in structural databases despite their abundance and biomedical importance. This review focuses on the novel method of site-specific infrared dichroism (SSID) combined with constraint molecular dynamics simulation, which has recently emerged as a powerful method to obtain structures of transmembrane α-helical bundles. The theory of SSID including its latest developments is reviewed with the aim to encourage widespread application of this method. This is followed by an outline of the conformational search using experimentally constraint molecular dynamics simulations. Finally a critical evaluation of recent applications, namely the Influenza M2 proton channel, the vpu ion channel of HIV-1 and the MHC-class II associated invariant chain, is conducted.

# 1 Introduction

Membrane proteins are estimated to comprise about 30% of most genomes and account for 70% of known pharmaceutical drug targets. Despite their importance, at the time of writing 82 unique membrane protein structures of sufficient resolution to identify secondary structure elements are available [51] in the Protein Structure Database compared to more than 23000 structures of soluble proteins (incl. virus capsids). The discrepancy between the number of known transmembrane protein structures and the number of known soluble protein structures is due to experimental difficulties, i.e. the difficulty to crystallise membrane proteins for x-ray crystallography and the large size of protein/lipid or protein/detergent aggregates, which exceeds the size limit for solution state NMR spectroscopy. There is clearly a need in our post-genomic era for new methods to gain more structural information about membrane proteins. The most promising novel methods, which have emerged during the last decade, are solid-state NMR spectroscopy and site-specific infrared dichroism in combination with molecular dynamics simulation. Solid-state NMR uses highly aligned lipid bilayers on glass slides with peptides, in which a single residue is labelled with $^{15}N$ or a randomly oriented lipid bilayer/peptide sample utilising the technique of magic angle spinning, in which two residues are labelled differently, e.g. with $^{15}N$ and $^{13}C$. Solid state NMR methods have been reviewed elsewhere [11, 32]. This review will focus on the method of site-specific infrared dichroism (SSID), which has been developed by Arkin [2, 25], in combination with a conformational search protocol developed by Adams using molecular dynamics simulations [1]. Infrared (IR) spectroscopy has been used successfully to determine the secondary structure content of proteins as well as solvent accessibility of residues due to hydrogen/deuterium exchange (reviewed in e.g. [18, 19]). Automatic secondary

structure analysis using neural network approaches is emerging as a useful tool for structural proteomics [22, 23]. A unique advantage of IR spectroscopy over x-ray crystallography is that protein dynamics can be studied using difference spectroscopy [4]. In particular for membrane proteins IR spectroscopy is the method of choice as the lipid environment does not perturb the resolution or sensitivity of the spectra. In fact, the lipid molecules can be analysed in addition to the protein embedded. A typical spectrum of an α-helical transmembrane peptide in dimyristoyl-phosphatidylcholine (DMPC) lipids is shown in fig. 1. Various absorption bands can be assigned to molecular vibrations as indicated in fig. 1A. The protein vibrations most commonly analysed are the amide I mode, mainly the C=O stretch vibration, and the amide II mode, mainly an N-H bending mode with a contribution from C-N stretching (fig. 1B). Membrane protein studies often employ the technique of attenuated total reflection (ATR) spectroscopy, in which lipid samples containing proteins are deposited on a planar surface of a diffraction element. Although ATR-IR spectroscopy has been extensively reviewed in the last decade [9, 17, 43], we will give a short introduction into ATR as it is essential to understand site-specific dichroism (SSID). This will be followed by a short treatment of SSID, an overview of molecular dynamics simulation with orientational constraints and finally applications of these methods to biological examples will be reviewed.

## 2 Theory and Methods

### 2.1 Attenuated total reflection IR spectroscopy

Contrary to transmission spectroscopy attenuated total reflection (ATR) IR spectroscopy measures the absorption of material attached to the surface of a

diffraction element, while the beam of radiation is captured in the diffraction element by total internal reflection. Typical experimental setups for membrane proteins use a trapezoidal plate, in which the infrared light beam is captured by multiple total internal reflections (fig. 2). At the points of total internal reflection light penetrates into the adjacent medium by a short distance, the evanescent field, and an absorption spectrum can be measured. The plate is often made of germanium, but also ZnS, ZnSe, KRS-5 or silicon is used. The geometry shown in fig. 2 is the one most commonly used for lipid membrane samples, although a variety of ATR accessories including single reflection devices are available. Using a polarizer an absorption spectrum can be measured at IR light polarised parallel to the plane of incidence and perpendicular to the plane of incidence as defined in fig. 2. For each particular absorption band a dichroic ratio R can be calculated as the ratio of absorption at parallel polarisation $A_{parallel}$ and perpendicular polarisation $A_{perpendicular}$:

$$R = \frac{A_{parallel}}{A_{perpendicular}} \tag{1}$$

The dichroic ratio is related to the order parameter S [16]:

$$S = \frac{E_x^2 - RE_y^2 + E_z^2}{E_x^2 - RE_y^2 - 2E_z^2} \left( \frac{3\cos^2 \alpha - 1}{2} \right)^{-1} \tag{2}$$

whereby $\alpha$ is the angle between the molecular director and the transition dipole moment of the particular vibration mode and $E_x$, $E_y$, $E_z$ are the electric field components of the evanescent field given by Harrick [21] assuming that the thickness of the deposited film is much larger than the penetration depth of the evanescent field (ca. $1\mu m$ for Ge):

$$E_x = \frac{2\cos\varphi\sqrt{\sin^2\varphi - n_{21}^2}}{\sqrt{1 - n_{21}^2}\sqrt{(1 + n_{21}^2)\sin^2\varphi}}$$

$$E_y = \frac{2\cos\varphi}{\sqrt{1 - n_{21}^2}} \tag{3}$$

$$E_z = \frac{2\cos\varphi\sin\varphi}{\sqrt{1 - n_{21}^2}\sqrt{(1 + n_{21}^2)\sin^2\varphi - n_{21}^2}}$$

where $n_{21}$ is the ratio of the refractive indices between the sample ($n_1$ = 1.43 for a lipid

bilayer) and the diffraction element ($n_2$ = 4.0 for Germanium). The angle of incidence

$\varphi$ between the diffraction element and the infrared beam is typically set at 45º.

The angle $\alpha$ between the molecular director and the transition dipole moment is 90º

for the symmetric and antisymmetric stretch vibrations of aliphatic lipid chains and

also for the C=O stretch vibrations of the lipid ester group [15]. The amide I transition

dipole moment is oriented at $\alpha$ = 38º to 39º [35, 49], although lower values have been

reported.

In case of transmembrane helical peptides the order parameter S is defined as follows:

$$S = \frac{3 < \cos^2\beta > -1}{2} \tag{4}$$

whereby $\beta$ is the angle between the helix and the z-axis, which coincides with the

membrane normal. In conventional analysis of ATR spectra equation (4) is used to

calculate the helix tilt angle against the membrane normal. However, it should be

stressed, as pointed out by Arkin et al. [2], that the helix tilt obtained in this way is

only a maximum value, assuming that the protein sample is completely ordered. The

actual helix tilt can range from 0º to $\beta$ dependent on the sample order.


2.2 Site-specific infrared dichroism

Site-specific infrared dichroism (SSID) uses the fact that the frequency of molecular vibrations is sensitive to the mass of the participating atoms. Thus introducing an aminoacid residue where the carbonyl group is isotopically labelled, e.g. $^{13}C=^{16}O$, results in a shift of the corresponding absorption frequency for that residue. In case of a peptide composed of the transmembrane region of glycophorin A, a shift of 44 cm$^{-1}$ towards lower wave numbers from the carbon-12 amide I absorption band located at 1657 cm$^{-1}$ has been reported [2]. Recently a number of other labels have been developed namely the $^{13}C=^{18}O$ carbonyl group, which results in a shift of 64 cm$^{-1}$ towards lower wave numbers [46], the double deuterated glycine [45] and the deuterated methyl group –CD$_3$ of alanine [47]. The deuterated labels have the advantage, that the symmetric and antisymmetric stretch vibrations can be analysed separately thus requiring only one label to deduce helix tilt and orientation in the presence of sample disorder. Another advantage is that the vibrations of the deuterated labels are not coupled to nonlabelled groups, while the vibration of carbonyl group labels might couple to the unlabelled peptide backbone carbonyl groups thus reducing the accuracy of the derived geometrical parameters. However, the $^{13}C=^{16}O$ and $^{13}C=^{18}O$ labels have clearly visible bands in the absorption spectrum, while the absorption bands of deuterated labels are difficult to detect. That is the reason, why most applications to date have used the carbonyl group labels. Especially the oxygen-18 containing carbonyl bond has an absorption band in a transparent region of the spectrum, while the $^{13}C=^{16}O$ appears as a shoulder of the unlabelled amide I peak (fig. 1B) and mathematical band sharpening techniques have to be used to resolve this absorption band. The theory of SSID has been developed by Arkin et. al. [2] originally modelling the sample order as a fraction f of the sample that is completely ordered, while the rest of the sample (1-f) is at random orientation.

Recently this theory has been developed further with regards to sample disorder assuming a more realistic Gaussian distribution of a peptide α-helix around a particular tilt angle [25].

The theory of SSID establishes a relationship between the measured dichroic ratio R defined in eq. (1), the geometry of a polypeptide α-helix as defined in fig. 3 and the orientation of a particular transition dipole moment vector $\vec{P}$. The α-helix is described by a helix tilt angle β and the distribution around this tilt angle characterised by the standard deviation σ. The orientation of the particular transition dipole moment of the labelled site is described in addition by the rotational pitch angle ω (fig.3), which is arbitrarily defined as zero, when the transition dipole moment lies in the same plane with the helix director and the z-axis. The measured dichroic ratio of the unlabelled helix $R_{Helix}$ for a particular sample 1 can then be expressed as (for detailed equations see appendix):

$$R_{Helix1} = function(\beta, \sigma_1) \tag{5a}$$

while the dichroic ratio of the site $R_{site1}$ is additionally dependent on the rotational pitch angle ω:

$$R_{Site1} = function(\omega, \beta, \sigma_1) \tag{6a}$$

If the adjacent aminoacid residue in the sequence is labelled, the rotational pitch angle of this residue is given by ω + 100° for a standard α-helix [39], thus the dichroic ratios for a second peptide with a label in an adjacent position are given by:

$$R_{Helix2} = function(\beta, \sigma_2) \tag{5b}$$

$$R_{site2} = function(\omega + 100°, \beta, \sigma_2) \tag{6b}$$

The equations 5a-6b form a nonlinear system of four equations with four unknowns, $\beta$, $\omega$, $\sigma_1$ and $\sigma_2$ which can be solved with standard mathematical techniques. The equations 5 and 6 correspond to eq. (A5) and (A6) in the appendix.

This analysis taking into account a Gaussian distribution of the helix tilt, described by $\sigma$, around a mean tilt angle presents a significant advantage over the earlier method, which assumed that a fraction of the sample was ordered, while the rest was completely disordered. The standard deviation $\sigma$ yields a new experimental parameter, which can be related to the mosaicity of the membrane preparation and thermal fluctuations of the transmembrane helices [25]. However, the experimental applications so far have made use of the fractional sample order model.

## 2.2 Conformational search and constraint molecular dynamics simulations

Because of the necessity to satisfy hydrogen bonding requirements in the apolar lipid environment, the transmembrane part of proteins has a well defined secondary structure. Transmembrane protein structures can be classified in beta-barrel and helical bundle structures [51]. So far SSID has been applied to helical bundle structures, which occur frequently in animals as well as in lipid enveloped animal viruses. Small bitopic transmembrane proteins forming oligmerising $\alpha$-helical bundles are conceptionally well defined structures and can be described by three structural parameters, the helix tilt $\beta$, the helix crossing angle $\Omega$ and the helix rotation $\phi$ (fig.4), which defines the sides of the helices interacting with each other [3]. This leads to a relatively straightforward conformational search algorithm [1] in which a symmetrical helical bundle (as shown in fig. 4 for a dimer) is generated and a rotation to all helices is applied simultaneously between $\phi = 0^\circ$ and $\phi = 360^\circ$. The crossing angle is set at $\Omega$

= 25º for left-handed helical bundles and at $\Omega$ = -25º for right-handed bundles. The helix rotation is typically varied in steps of 10º giving raise to 36 x 2 = 72 structures. In molecular dynamics simulations the initial atom velocities can determine the outcome of the simulation; for that reason each structure should be simulated several times with different atom velocities assigned randomly. Four simulations with different random velocities are commonly used giving raise to 72 x 4 = 288 structures, which are subjected to a simulated annealing molecular dynamics simulation and energy minimisation protocol *in vacuo*. Clusters of similar structures are identified on the basis of the root mean square deviation of the peptide backbone coordinates not being larger than a value in the range of 1.0 to 2.0 Å. For each cluster an average structure is calculated, energy minimised and subjected to the same simulated annealing/energy minimisation protocol used in the conformational search.

Orientational data from SSID is incorporated into the forcefield for molecular dynamics (MD) simulations and energy minimisation as additional energy terms in the form of:

$$E = k_{dichro}\left(\delta - \delta_0\right)^2 \tag{7}$$

where $\delta$ represents the actual angle and $\delta_0$ the target angle. The helix tilt angle and the angle between the C=O bond and the z-axis for all labelled residues are constraint in this way. Constraints between pairs of atoms and external axes are not standard in most MD simulation programs, but energy terms similar to eq. (7) have been incorporated into the simulation software GROMACS from version 3.1 onwards [7, 31].

**3 Biological applications**

So far six different applications of SSID combined with MD simulations have been reported leading to transmembrane structures of several biomedical important systems, namely a structure for the M2 proton channel from the Influenza A virus [26], a structure for the transmembrane domain of vpu from HIV-I virus [27], the structure of the CM2 transmembrane domain from Influenza C virus [28], a structural model of human phospholamban [44], a structure of the transmembrane domain of human CD3-$\zeta$ [48] and a structure of the human MHC class II-associated invariant chain [29]. In the following paragraphs three of the applications will be reviewed in greater detail, because each introduces a novel aspect of SSID application.

3.1 The M2 proton channel from Influenza A virus

M2 is a 97 aminoacid bitopic transmembrane protein located in the lipid envelope of the Influenza A virus. M2 forms a tetrameric proton channel and participates in the virus uncoating process after virus uptake by endocytosis as well as in the budding of newly made virus particles in later stages (reviewed in [30]). The method of SSID has been applied to a synthetic peptide containing the transmembrane domain of the sequence SSDPLVV_AA_SIIGILHLILWILDRL. Two different peptides have been synthesised with one $^{13}C=^{16}O$ labelled alanine residue each in the positions shown underlined. The SSID analysis gives a helix tilt of $\beta = 32°$ and the rotational pitch angle about the helix axis for Ala29 is $\omega = -60°$. A global conformational search has been carried out without orientational energy refinement terms and with energy refinement terms. The results of this conformational search are striking with respect to the energy distribution of individual structures. In absence of orientational energy refinement terms all structures of the conformational search show little differences in

energy, while with energy refinement terms the structures separate in higher and lower energy structures, while one structure has a significantly lower energy than all others; its structure is also generated from the largest number of substructures forming a cluster of similar structures. The structure is shown in fig. 5A with the residues Ala29, Ser31 and His 37 highlighted. It is instructive to compare this model with the M2 structure determined by solid-state NMR shown in fig. 5B. In this work a variety of sites have been labelled with $^{15}$N at their amide group and orientational data has been obtained by solid-state NMR of highly aligned samples [50]. In addition the distance between $^{15}$N$_\pi$ labelled His37 and 13C$_\gamma$ labelled Trp41 has been determined by magic angle spinning of a randomly oriented sample [37]. This structure is fully experimentally defined including the orientations of the side chains His37 and Trp41, while other side chain orientations have been taken from a rotamer library. It can be seen that the orientation of the highlighted residues in fig. 5 is similar defining the interacting residues of the helices and the channel lining residues. However, the SSID model (fig. 5A) shows a stronger coiled coil structure implicating strong interaction of the helices throughout their length. This is clearly an artefact of the *in vacuo* simulation procedure, in which a constant interhelical distance restraint of 10.5 Å has been applied to prevent the bundle from falling apart. In conclusion the similarity of the SSID model based on labelling of only two sites with the NMR structure is convincing, yet the *in vacuo* modelling does not adequately approximate the structure in a native lipid bilayer. A conformational search in a lipid bilayer/water environment would improve the model.

3.2 The vpu protein from HIV-1 virus

The 81-residue vpu protein belongs to the auxiliary proteins of the human immunodeficiency virus type 1 (HIV-1) [10, 42]. The bitopic transmembrane protein contains a single membrane spanning segment and forms homooligomers of at least four subunits according to SDS gel electrophoresis studies [33]. The 5 residues short N-terminal domain is responsible for virus particle release and the C-terminal cytoplasmic domain causes the degradation of one of the HIV-1 coreceptor molecules, CD4 [40]. The transmembrane domain has been studied independently and ion channel activity for monovalent cations has been observed in *Xenopus* oocytes and in planar lipid bilayers [13, 41]. The SSID analysis has been applied to a peptide of the sequence MQPIQIAIVALV<u>VA</u>IIIAI<u>VV</u>WSIVIIEYRK, using two peptides each containing a $^{13}C=^{16}O$ double label at positions Val13, Val20 and Ala14, Val21 (labels shown underlined). The rationale for double labelling is that in a standard $\alpha$-helix two residues 7 positions apart have approximately the same rotational pitch angle, i.e. $\omega_i = \omega_{i+7} - 20°$, thus increasing the signal of the $^{13}C=^{16}O$ absorption band with respect to the background with some sacrifice of accuracy. SSID analysis yields a helix tilt of $\beta = 6.5°$ and the rotational pitch angle for the label pair Val30/Val20 is $\omega = 283°$. The conformational search applying orientational constraints given in eq. (7) does not lead to a distinction of structures in terms of energy. By inserting eq. (A13) into eq. (7) it can be shown that the energetic discrimination between different rotational pitch angles is weak at low tilt angles:

$$E = k_{dichro} \arccos(\cos \alpha \cos \beta - \sin \alpha \sin \beta \cos(\omega + 17°)) - \theta_0)^2 \qquad (8)$$

Plotting the energy against the rotational pitch angle $\omega$ (fig. 6) shows that for low helix tilt angles the energetic discrimination between different rotational pitch angles is very weak. However, conducting the conformational search and analysing the

resulting structures for the rotational pitch angle, yields a unique structure with ω

close to the experimental data. As the oligomerisation number of vpu was unknown,

simulations for tetrameric, pentameric and hexameric bundles have been performed

and only the pentameric assembly gave a structure whose rotational pitch angle was

close to the experimental data. This structure shown fig. 7A reveals a slightly tilted

left-handed coiled coil. Interestingly the potential ion channel pore is occluded by

Trp22, while the only hydrophilic residue Ser23 points to the lipid phase and the

hydroxyl group is hydrogen-bonded to the carbonyl backbone atoms.

The pentameric oligomerisation has been confirmed recently by an intriguing

experiment chemically linking vpu transmembrane peptides to a carrier template

assembling a tetramer or pentamer of transmembrane helices [6]. Ion channel

measurements in planar lipid bilayers led to the conclusion that the native state

corresponds to a pentameric assembly. However, a recent study by solid state NMR

utilising uniformly $^{15}$N-labeled peptides of residues 2-30 with a C-terminal GGKKKK

attachment to aid solubility reports a helix tilt angle of 12º for residues 8-16 and 15º

for residues 17-25 with a slight kink located at Ile17 [38]. Furthermore the helix

rotation is markedly different to the model proposed on the basis of SSID data; the

residue Trp22 pointing inside the pore in the SSID model points into the lipid phase in

the NMR structure (fig. 7B). Several reasons could account for this discrepancy,

namely the difference in the sequences studied, i.e. the NMR structure contains a C-

terminal attachment not found in naturally occurring viral strains as well as a Tyr29-

Gly mutation, the different lipid systems used, i.e. dimyristoylphosphatidylcholine

(DMPC) for the SSID study and a mixture of dioleoylphosphatidylcholine (DOPC) /

dioleoylphosphatidylglycerol (DOPG) for the NMR study. Unsaturated chains adopt

different conformations than the saturated aliphatic chains in DMPC and could thus

favour a different structure of the embedded peptide. Furthermore strong electrostatic interactions between the negatively charged DOPG lipids and the C-terminal GGKKKK attachment can be expected. The small kink in the helix at Ile17, which is actually in between two $^{13}$C=O labels, should not obscure the SSID analysis as to give a structure of a completely opposite helix orientation and different tilt angle.

## 3.3 The MHC class II-associated invariant chain

The class II major histocompatibility complex (MHC) plays an important role in the human immune defence system presenting peptide fragments of intruding organisms at the surface of B-cells. The MHC class II-associated invariant chain (Ii) is involved in the pathway of MHC maturation and peptide loading [8]. The 279 residue Ii contains a single transmembrane domain and associates to trimers after biosythesis [34]. The trimeric complex serves as a scaffold for the binding of three MHC class II $\alpha\beta$ heterodimers forming an $(\alpha\beta)_3$Ii$_3$ complex, which is exported through the endocytic pathway. In later stages Ii is progressively degraded and remaining fragments are finally released from MHC until the loading with antigenic peptides occurs [12]. With a truncated version of Ii containing the first 80 residues it has been shown that a patch of hydrophilic residues in the transmembrane domain is important for trimerisation [5].

The transmembrane domain has been modelled with a peptide containing residues 29-60 of Ii with the sequence RGALYTGVS<u>VLVALLLA</u>GQ<u>A</u>TT<u>A</u>YFLYQQQGR, each underlined residue containing a 1-$^{13}$C=$^{18}$O label amounting to 10 different peptides. This represents the first example, where the peptide has been labelled over the whole transmembrane region [29]. The infrared spectroscopic data confirmed the predicted

α-helical structure, in particular the position of the $^{13}C=^{18}O$ absorption bands indicated that each labelled residue is in an α-helical environment. The results of the SSID analysis are presented in table 1. There is considerable variation in the local helix tilt β ranging from 6º to 25º with an average of 13º; this indicates that the helix is not a straight cylinder, but shows local deviations from the ideal α-helix geometry. The structure as a result of constraint MD simulations (fig. 7) forms a left-handed coiled coil; the residue Gln47 implicated in trimer formation forms strong interhelical contacts. Thr50 points to the inside of the trimeric coil and forms a network of hydrogen bonds, while Tyr33 and Leu43 are also involved in stabilising the trimer structure.

**4 Conclusions and Outlook**

It has been shown that SSID can make a significant contribution to the structural elucidation of transmembrane proteins. The applications so far cover bitopic oligomerising transmembrane proteins, rather than larger proteins with multiple membrane spanning domains. Given that peptides have been used successfully to dissect the structural domains of voltage gated ion channels e.g. [20], application to larger integral transmembrane proteins are expected. The site specific labels used so far have been aminoacids labelled at their carbonyl groups with $^{13}C=^{16}O$ or $^{13}C=^{18}O$. The $^{13}C=^{16}O$ label is limited in its application, because of the 1% natural abundance of $^{13}C$; in a protein of 100 aminoacid residues half of the signal would be expected to originate from naturally occurring $^{13}C$ carbonyl groups at random positions. The $^{13}C=^{18}O$ therefore enables us to study proteins of any size. The same applies to the $CD_2$ labelled glycine and the $-CD_3$ labelled alanine, which are also predicted to find

increased use in SSID applications. Recently the theory of SSID has been developed towards the analysis of β-sheet proteins [36] and we expect applications to expand in the field of transmembrane β-sheet proteins.

Another field of improvement are the constraint molecular dynamics simulations. It has been discussed above that the *in vacuo* simulations provide significant limitations for the outcome of the final structure in the way that helix distance constraints as well as constraints to maintain helix geometry have to be imposed. Furthermore the delicate interactions between the protein and lipid molecules with hydrophilic headgroups and hydrophobic acyl chains cannot be modelled in the framework of an *in vacuo* approach. It is common practice to simulate transmembrane peptides in lipid bilayers e.g. [14] and we predict that with increasing computer power and resolution of methodological difficulties the global conformational search outlined above will be conducted in a lipid bilayer system with water and ions added.

**Acknowledgements**

**Appendix**

This section gives the relevant equations for the application of SSID; for a detailed mathematical treatment the reader is referred to [2, 25].

In the ATR configuration the dichroic ratio R can be expressed with the electric field components $E_x$, $E_y$, $E_z$ and the integrated absorption coefficients K:

$$R = \frac{E_z^2 K_z + E_x^2 K_x}{E_y^2 K_y} \tag{A1}$$

For the site-specific dichroic ratio the integrated absorption coefficients depend on the rotational pitch angle $\omega$ of the labelled site:

$$R_{site} = \frac{E_z^2 \langle K_z(\omega) \rangle + E_x^2 \langle K_x(\omega) \rangle}{E_y^2 \langle K_y(\omega) \rangle} \tag{A2}$$

The angular brackets $\langle\rangle$ denote averaging over all possible tilt angles using a probability distribution function $F(\beta)$:

$$\langle K(\omega) \rangle = \int_{-\infty}^{+\infty} F(\beta) K(\omega) d\beta \tag{A3}$$

The probability $F(\beta)$ of finding a helix with tilt angle $\beta$ is given by a Gaussian distribution with mean tilt angle $\mu$ and standard deviation $\sigma$.

$$F(\beta) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(\beta-\mu)^2}{2\sigma^2}} \tag{A4}$$

In the $\alpha$-helix the non-labelled transition dipole moments are distributed around the helix, therefore the integrated absorption coefficients are not dependent on the rotational pitch angle $\omega$. The dichroic ratio is then given by:

$$R_{Helix} = \frac{E_z^2 \langle K_z^{Helix} \rangle E_x^2 \langle K_x^{Helix} \rangle}{E_y^2 \langle K_y^{Helix} \rangle} \tag{A5}$$

The averaged absorption coefficients are given again by integrating over all possible tilt angles:

$$\left\langle K^{Helix} \right\rangle = \int\limits_{-\infty}^{+\infty} F(\beta) K^{Helix} d\beta \tag{A6}$$

In particular the site-specific averaged integrated absorption coefficients are obtained by taking into account uniaxial symmetry, thus integrating over all rotation angles around the z-axis:

$$\left\langle K_x(\omega) \right\rangle = \frac{\cos^2\alpha \left( \frac{1}{2} - \frac{\cos(2\mu)}{2e^{2\sigma^2}} \right)}{2} + \frac{\cos^2\omega \sin^2\alpha \left( \frac{1}{2} + \frac{\cos(2\mu)}{2e^{2\sigma^2}} \right)}{2}$$
$$+ \frac{\cos\alpha \cos\mu \cos\omega \sin\alpha \sin\mu}{e^{\sigma^2}} + \frac{\sin^2\alpha \sin^2\omega}{2} \tag{A7}$$

$$\left\langle K_y(\omega) \right\rangle = \left\langle K_x(\omega) \right\rangle \tag{A8}$$

$$\left\langle K_z(\omega) \right\rangle = \cos^2\alpha \left( \frac{1}{2} + \frac{\cos(2\mu)}{2e^{2\sigma^2}} \right) + \cos^2\omega \sin^2\alpha \left( \frac{1}{2} - \frac{\cos(2\mu)}{2e^{2\sigma^2}} \right)$$
$$- \frac{2\cos\alpha \cos\mu \cos\omega \sin\alpha \sin\mu}{2} \tag{A9}$$

The averaged integrated absorption coefficients for the helix require apart from uniaxial symmetry integration over all rotational pitch angles ω and are thus independet of ω:

$$\left\langle K_x^{Helix} \right\rangle = \frac{1}{4} \left( 2\cos^2\alpha \left( \frac{1}{2} - \frac{\cos(2\mu)}{2e^{2\sigma^2}} \right) + \sin^2\alpha + \sin^2\alpha \left( \frac{1}{2} + \frac{\cos(2\mu)}{2e^{2\sigma^2}} \right) \right) \tag{A10}$$

$$\left\langle K_y^{Helix} \right\rangle = \left\langle K_x^{Helix} \right\rangle \tag{A11}$$

$$\left\langle K_z^{Helix} \right\rangle = \cos^2 \alpha \left( \frac{1}{2} + \frac{\cos(2\mu)}{2e^{2\sigma^2}} \right) + \frac{1}{2} \left( \sin^2 \alpha \left( \frac{1}{2} - \frac{\cos(2\mu)}{2e^{2\sigma^2}} \right) \right) \tag{A12}$$

Please note that contrary to the standard definition of the angle $\alpha$ between the transition dipole moment and the molecular director the angle $\alpha = 180^\circ - 38^\circ = 142^\circ$ has been used in the equations above [2].

The angle between the C=O bond and the z-axis $\theta_z$ used in restraint molecular dynamics simulations is a function of the rotational pitch angle $\omega$ and the mean helix tilt angle $\mu$.

$$\cos \theta_z = \cos \alpha \cos \mu - \sin \alpha \sin \mu \cos(\omega + 17^\circ) \tag{A13}$$

Please note, that the equation given by Marsh [36] is reported here, as the relative orientations of the transition dipole moment and the peptide carbonyl bond are treated exactly in this work. The angle $\alpha$ in eq. (A13) is used according to its standard definition, i.e. $\alpha = 38^\circ$.

## References

[1] Adams, P.D., Arkin, I.T., Engelman, D.M. and Brünger, A.T., Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban, *Structural Biology* **2** (1995), 154-162.

[2] Arkin, I.T., MacKenzie, K.R. and Brünger, A.T., Site-directed dichroism as a method for obtaining rotational and orientational constraints for orientated polymers, *Journal of the American Chemical Society* **119** (1997), 8973-8980.

[3] Arkin, I.T., Structural aspects of oligomerization taking place between transmembrane alpha-helices of bitopic membrane proteins, *Biochimica et Biophysica Acta* **1565** (2002), 347-363.

[4] Arrondo, J.L.R. and Goni, F.M., Structure and dynamics of membrane proteins as studied by infrared spectroscopy, *Progress in Biophysics & Molecular Biology* **72** (1999), 367-405.

[5] Ashman, J.B. and Miller, J., The role for the transmembrane domain in the trimerization of the MHC Class II-associated invariant chain, *The Journal of Immunology* (1999), 2704-2712.

[6] Becker, C.F., Oblatt-Montal, M., Kochendorfer, G.G. and Montal, M., Chemical synthesis and single channel properties of tetrameric and pentameric TASPs (template-assembled synthetic proteins) derived from the transmembrane domain of HIV virus protein u (Vpu), *J. Biol. Chem.* **279** (2004), 17483-17489.

[7] Berendsen, H.J.C., van der Spoel, D. and van Drunen, R., GROMACS: A message passing parallel molecular dynamics implementation, *Comp. Phys. Comm.* **91** (1995), 43-56.

[8] Busch, R., Doebele, R., Patil, N.S., Pashine, A. and Mellins, E.D., Accessory molecules of MHC class II peptide loading, *Current Opinion in Immunology* **12** (2000), 99-106.

[9] Chittur, K.K., FTIR/ATR for protein absorption to biomaterial surfaces, *Biomaterials* **19** (1998), 357-69.

[10] Cohen, E.A., Terwillinger, E.F., Sodroski, J.G. and Haseltine, W.A., Identification of a protein encoded by the vpu gene of HIV-1, *Nature* **334** (1988), 532-534.

[11] Cross, T.A. and Fu, R., Solid-state nuclear magnetic resonance investigation of protein and polypeptide structure, *Annual Review of Biophysics and Biomolecular Structure* **28** (1999), 235-268.

[12] Denzin, L.K. and Cresswell, P., HLA-DM induces CLIP dissociation from MHC class II alpha-beta dimers and facilitates peptide loading, *Cell* **82** (1995), 155-165.

[13] Ewart, G.D., Sutherland, T., Gage, P.W. and Cox, G.B., The vpu protein of human immunodeficiency virus type 1 forms cation selective ion channels, *Journal of Virology* **70** (1996), 7108-7115.

[14] Forest, L.R., Kukol, A., Arkin, I.T., Tieleman, D.P. and Sansom, M.S.P., Exploring models of the influenza A M2 channel - MD simulations in a phospholipid bilayer, *Biophysical Journal* **78** (1999), 55-69.

[15] Frey, S. and Tamm, L.K., Orientation of melittin in phospholipid bilayers. A polarized attenuated total reflection study., *Biophysical Journal* **60** (1991), 922-930.

[16] Fringeli, U.P. and Gunthard, H.H., *Infrared membrane spectroscopy*, in *Membrane Spectroscopy*, E. Grell, Editor. 1981, Springer Verlag: Berlin. p. 270-332.

[17] Goormaghtigh, E., Raussens, V. and Ruysschaert, J.-M., Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes, *Biochimica et Biophysica Acta* **1422** (1999), 105-185.

[18] Haris, P.I. and Chapman, D., Does Fourier-transform infrared spectroscopy provide useful information on protein structures?, *Trends in Biochemical Sciences* **17** (1992), 328-333.

[19] Haris, P.I. and Chapman, D., The conformational analysis of peptides using Fourier transform IR spectroscopy, *Biopolymers* **37** (1995), 251-263.

[20] Haris, P.I., Structural model of a voltage-gated potassium channel based on spectroscopic data, *Biochemical Society Transactions* **29** (2001), 589-593.

[21] Harrick, N.J., *Internal reflection spectroscopy*. 2nd ed. 1979, Ossining: Harrick Scientific Corporation. 13-63.

[22] Hering, J.A., Innocent, P.R. and Haris, P.I., Automatic amide I frequency selection for rapid quantification of protein secondary structure from fourier transform infrared spectra of proteins, *Proteomics* **2** (2002), 839-849.

[23] Hering, J.A., Innocent, P.R. and Haris, P.I., Neuro-fuzzy structural classification of proteins for improved protein secondary structure prediction, *Proteomics* **3** (2003), 1464-1475.

[24] Humphrey, W., Dalke, A. and Schulten, K., VMD - Visual molecular dynamics, *J. Mol. Graphics* **14** (1996), 33-38.

[25] Kass, I., Arbely, E. and Arkin, I.T., Modeling sample disorder in site-specific dichroism studies of uniaxial systems, *Biophysical Journal* **86** (2004), 2502-2507.

[26] Kukol, A., Adams, P.D., Rice, L.M., Brunger, A.T. and Arkin, I.T., Experimentally based orientational refinement of membrane protein models: A structure for the Influenza A M2 H+ channel, *Journal of Molecular Biology* **286** (1999), 951-962.

[27] Kukol, A. and Arkin, I.T., vpu transmembrane peptide structure obtained by site-specific fourier transform infrared dichroism and global molecular dynamics searching, *Biophysical Journal* **77** (1999), 1594-1601.

[28] Kukol, A. and Arkin, I.T., Structure of the Influenza C CM2 protein transmembrane domain obtained by site-specific infrared dichroism and global molecular dynamics searching., *J. Biol. Chem.* **275** (2000), 4225-4229.

[29] Kukol, A., Torres, J. and Arkin, I.T., A structure for the trimeric MHC class II-associated invariant chain transmembrane domain., *Journal of Molecular Biology* **320** (2002), 1109-1117.

[30] Lamb, R.A. and Pinto, L.H., 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** (1997), 1-11.

[31] Lindahl, E., Hess, B. and van der Spoel, D., Gromacs 3.0: A package for molecular dynamics simulation and trajectory analysis, *J. Mol. Mod.* **7** (2001), 306-317.

[32] Luca, S., Heise, H. and Baldus, M., High-resolution solid-state NMR applied to polypeptides and membrane proteins, *Accounts of Chemical Research* **36** (2003), 858-865.

[33] Maldarelli, F., Chen, M.-Y., Willey, R.L. and Strebel, K., Human immunodeficiency virus type 1 vpu protein is an oligomeric type 1 integral membrane protein, *Journal of Virology* **67** (1993), 5056-5061.

[34] Marks, M.S., Blum, J.S. and Cresswell, P., Invarian chain trimers are sequestered in the rough endoplasmatic reticulum in the abscence of association with HLA class II antigens, *J. Cell Biol.* **111** (1990), 839-845.

[35] Marsh, D., Muller, M. and Schmitt, F.-J., Orientation of the infrared transition moments for an alpha-helix, *Biophysical Journal* **78** (2000), 2499-2510.

[36] Marsh, D., Infrared dichroism of isotope-edited alpha-Helices and beta-sheets, *Journal of Molecular Biology* **338** (2004), 353-367.

[37] Nishimura, H., Kim, S., Zhang, L. and Cross, T.A., The closed state of a H+ channel helical bundle combining precise orientational and distance restraints from solid state NMR, *Biochemistry* **41** (2002), 13170-13177.

[38] Park, S.H., Mrse, A.A., Nevzorov, A.A., Mesleh, M.F., Oblatt-Montal, M., Montal, M. and Opella, S.J., Three dimensional structure of the channel-forming trans-membrane domain of virus protein "u" (vpu) from HIV-1, *Journal of Molecular Biology* **333** (2003), 409-424.

[39] Pauling, L.R., Corey, R. and Branson, H., The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain, *Proceedings of the National Academy of Sciences USA* **37** (1951), 205-211.

[40] Schubert, U., Bour, S., Ferrer-Montiel, A.V., Montal, M. and Maldarelli, F., The two biological activities of human immunodeficiency virus type 1 vpu protein involve two seperable structural domains, *Journal of Virology* **70** (1996), 809-819.

[41] Schubert, U., Ferrer-Montiel, A.V., Oblatt-Montal, M., Henklein, P., Strebel, K. and Montal, M., Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells, *FEBS Letters* **398** (1996), 12-18.

[42] Strebel, K., Klimkait, T., Maldarelli, F. and Martin, M.A., Molecular and biochemical analyses of human immunodeficiency virus type 1 *vpu* protein, *Journal of Virology* **63** (1989), 3784-3791.

[43] Tatulian, S.A., Attenuated total reflection fourier transform infrared spectroscopy: A method of choice for studying membrane proteins and lipids, *Biochemistry* **42** (2003), 11898-11907.

[44] Torres, J., Adams, P.D. and Arkin, I.T., Use of a New Label, 13C=18O, in the Determination of a Structural Model of Phospholamban in a Lipid Bilayer. Spatial Restraints Resolve the Ambiguity Arising from Interpretations of Mutagenesis Data, *Journal of Molecular Biology* **300** (2000), 677-685.

[45] Torres, J., Kukol, A. and Arkin, I.T., Use of a single glycine residue to determine tilt and orientation of a transmembrane helix. A new structural lable for infrared spectroscopy, *Biophysical Journal* **79** (2000), 3139-3143.

[46] Torres, J., Kukol, A., Goodman, J.M. and Arkin, I.T., Site-specific examination of secondary structure and orientation determination in membrane proteins: the peptidic 13C=18O group as a novel infrared probe., *Biopolymers* **59** (2001), 396-401.

[47] Torres, J. and Arkin, I.T., C-deuterated alanine: a new label to study membrane protein structure using site-specific infrared dichroism, *Biophysical Journal* **82** (2002), 1068-1075.

[48] Torres, J., Briggs, J.A.G. and Arkin, I.T., Multiple site-specific infrared dichroism of CD3-zeta, a transmembrane helix bundle., *Journal of Molecular Biology* **316** (2002), 365-374.

[49] Tsuboi, M., Infrared dichroism and molecular conformation of α-form poly-gamma-benzyl-L-glutamate, *Journal of Polymer Science* **59** (1962), 139-153.

[50] Wang, J., Kim, S., Kovacs, F.A. and Cross, T.A., Structure of the transmembrane region of the M2 protein H+ channel, *Protein Science* **10** (2001), 2241-2250.

[51] White, S., Membrane proteins of known 3D structure, http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html. 2004, University of California, Irvine.

Table 1: Rotational pitch angle ω and local helix tilt β, for each $^{13}C=^{18}O$ labeled Ii

transmembrane peptide.

| Label | ω | β |
|---|---|---|
| V38 | 290° ± 10° | 10° ± 2° |
| L39 | 359° ± 10° | 10° ± 2° |
| V40 | 108° ± 12° | 6° ± 2° |
| A41 | 216° ± 10° | 8° ± 2° |
| L42 | 316° ± 14° | 14° ± 4° |
| L43 | 359° ± 12° | 7° ± 2° |
| L44 | 83° ± 13° | 21° ± 5° |
| A45 | 133° ± 14° | 25° ± 8° |
| A48 | 299° ± 20° | 15° ± 7° |
| A51[a] | 239° | 15° |

a) Values are calculated from A48 assuming Δω = 300°.

**Figure captions**

Fig. 1: A) Infrared spectrum of a transmembrane peptide derived from the Influenza M2 protein reconstituted in dimyristoyl-phosphatidylcholine lipids, with characteristic absorption bands indicated. B) The amide I and amide II region expanded. The amide I peak exhibits a small shoulder due to $^{13}$C labelling.

Fig. 2: Schematic drawing showing a trapezoidal ATR plate for multiple reflection ATR. The electric field vectors of the incident infrared beam are shown for parallel polarisation and perpendicular polarisation. The grey pattern denotes the sample deposited on top of the ATR plate.

Fig. 3: Definition of orientational parameters of a transition dipole moment $\vec{P}$ in an $\alpha$-helix. The helix is tilted against the z-axis by the angle $\beta$, the transition dipole moment's rotational orientation with respect to the helix axis is given by the rotational pitch angle $\omega$, while the angle between the helix axis and the transition dipole moment is given by $\alpha$. The angle $\theta$ between the transition dipole moment and the z-axis is a function of $\omega$ and $\beta$ at constant $\alpha$, this parameter is used as orientational constraint in molecular dynamics simulations.

Fig. 4: Structural parameters of a transmembrane helical bundle: the helix tilt angle $\beta$, helix rotation $\phi$ and the helix crossing angle $\Omega$. Please note that only for a dimer as shown here the helix crossing angle is twice the helix tilt angle.

Fig. 5: Structure of the M2 transmembrane domain. Ala29, Ser 31 and His37 are highlighted. A) The structure of the model generated by constraint MD simulation

based on SSID data [26]. B) The structure determined by NMR [37]. The figure was created with VMD [24]

Fig. 6: The refinement energy E for the angle between the carbonyl bond and the z-axis in dependence of the rotational pitch angle $\omega$ for different helix tilt angles $\beta$. Calculated from eq. (8), setting $k_{dichro} = 800$ kcal/grad$^2$ and $\theta_0 = 21°$.

Fig. 7: A) Structural model of the vpu transmembrane domain generated by constraint MD simulations based on SSID data [27], showing residues Ile6 to Ile27. Residues Trp22 and Ser23 are highlighted. B) The structure based on NMR orientational data [38]. The figure was created with VMD [24].

Fig. 8: Structural model of the MHC class II-associated transmembrane domain generated by constraint MD simulations based on SSID data [29]. Residues Tyr33, Gln47, Thr49, Thr50 and Tyr52 are highlighted. The figure was created with VMD [24].
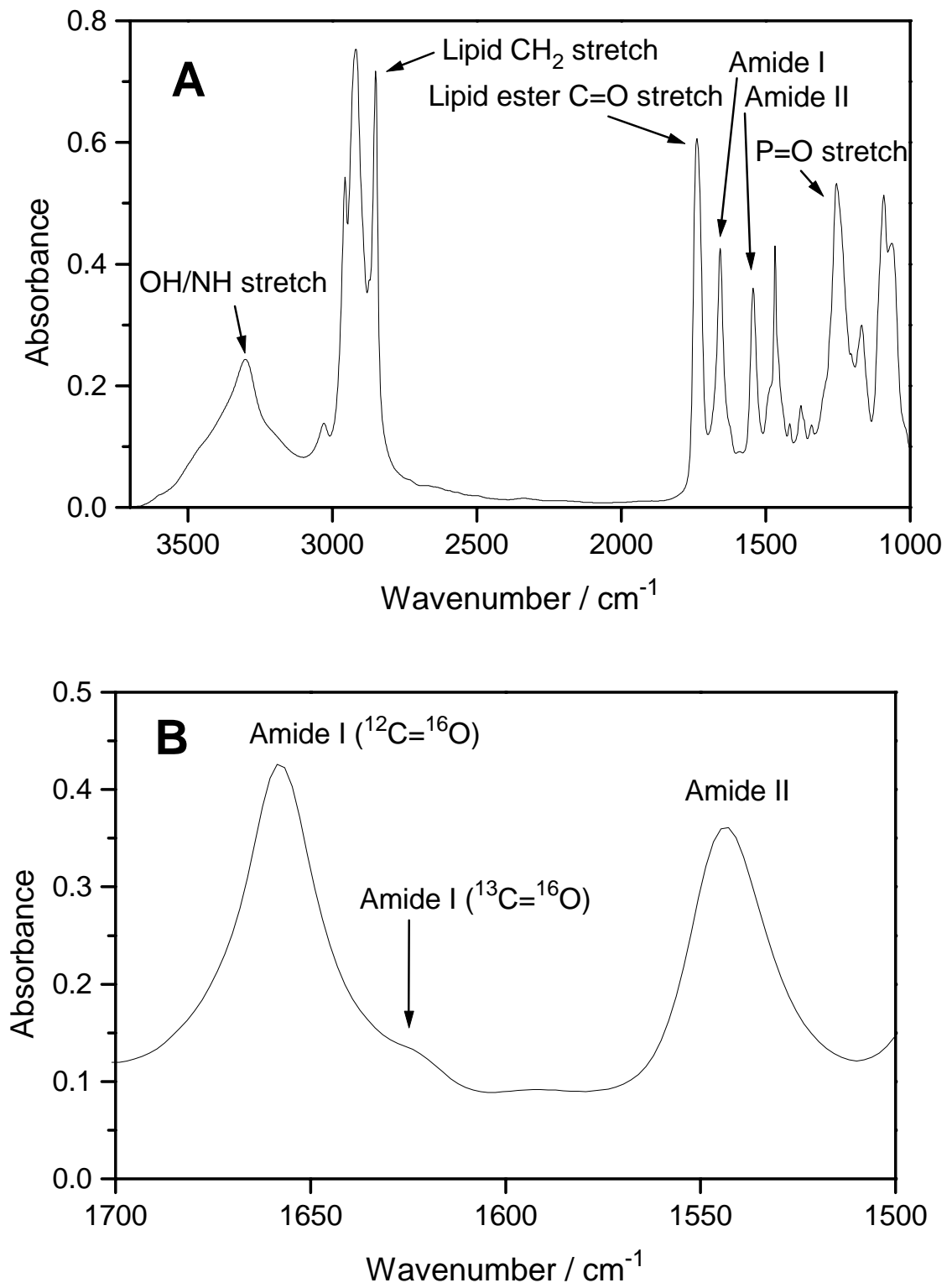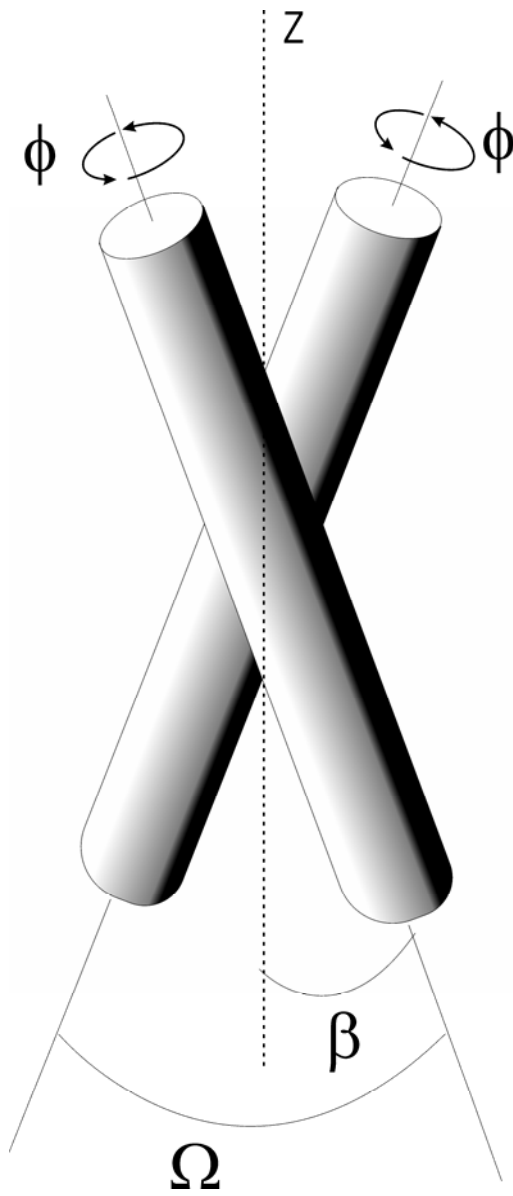
Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig. 5

Fig. 6

Fig. 7



Fig. 8