

# **A structure for the trimeric MHC class II-associated invariant chain transmembrane domain**

Running title: Invariant chain transmembrane structure

Andreas Kukol<sup>1\*</sup>, Jaume Torres<sup>2</sup>, Isaiah T. Arkin<sup>3</sup>

1) Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

2) School of Biological Sciences, Nanyang Technological University, 1 Nanyang Walk, Block 5 Level 3, Singapore 637616, Republic of Singapore

3) The Alexander Silberman Institute of Life Sciences, Department of Biological Chemistry, The Hebrew University, Givat-Ram, Jerusalem, 91904, Israel

\*) Corresponding author, Email: [akukol@bio.warwick.ac.uk](mailto:akukol@bio.warwick.ac.uk)

## Abstract

The MHC associated invariant chain (Ii) contains a single transmembrane domain that forms trimers. Ii is involved in the assembly of the MHC complex and antigen presentation and is thus central to the function of the immune system. Here we show by ATR-FTIR-spectroscopy that the transmembrane domain is  $\alpha$ -helical and provide a structural model of the transmembrane domain obtained by a combination of site-specific infrared dichroism and molecular dynamics (MD) simulations. This work resolves the backbone structure of a transmembrane peptide by multiple  $^{13}\text{C}=^{18}\text{O}$  labelling at 10 different residues. A second purely computational approach, based on MD simulations of Ii transmembrane homologous sequences, yields a similar structure that is consistent with our experimental results. The structure presented forms a left handed coiled coil with an average helix tilt of  $13^\circ \pm 6^\circ$ ; 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 bond.

Keywords: associated invariant chain / FTIR spectroscopy / molecular dynamics / site-specific dichroism / transmembrane protein

## 1. Introduction

The MHC class II-associated invariant chain (Ii)\* is an accessory protein involved in the pathway of MHC class II maturation and peptide loading (reviewed in (Busch et al., 2000)). The 279 residue Ii contains a single transmembrane domain and associates after biosynthesis to form a trimer (Lamb & Cresswell, 1992; Marks et al., 1990). This trimeric complex serves as a scaffold for the binding of three MHC class II  $\alpha\beta$  heterodimers (Roche et al., 1991). The assembled  $(\alpha\beta)_3Ii_3$  complex is exported through the endocytic pathway. In the late endocytic compartment Ii is progressively degraded by proteases, which leaves only the residues 81-104, known as CLIP peptides, remaining in the antigen binding groove. The CLIP peptides are finally released by HLA-DM (Wolf et al., 1998), which remains associated with MHC until the loading with antigenic peptides occurs (Denzin & Cresswell, 1995).

Trimerisation of Ii is mediated by two different regions of the polypeptide chain: (1) residues 163-183 (Bertolino et al., 1995; Bijlmakers et al., 1994), and (2) the transmembrane domain, residues 31-54 (Ashman & Miller, 1999). The available data can be interpreted such that the C-terminal domain initiates trimerisation and once a trimer is formed it remains stable even after cleavage of the C-terminal domain. However, a truncated version of the first 80 residues of Ii has been found sufficient for trimerisation and a patch of hydrophilic residues in the transmembrane domain has been shown to be important for trimerisation (Ashman & Miller, 1999).

Taken together it is highly likely that the transmembrane domain on its own assembles into a native like structure resembling the transmembrane part of the whole Ii molecule.

This work uses infrared spectroscopy to confirm for the first time the predicted  $\alpha$ -helical structure and more importantly to obtain a detailed molecular model of the transmembrane domain.

Since structure determination of membrane proteins is still a difficult task for x-ray crystallography and NMR spectroscopy a new, yet already proven successful approach is undertaken here to obtain a structural model for the Ii transmembrane domain trimer. The method of site-specific infrared dichroism (SSID) (Arkin et al., 1997) is applied to an isotopically labelled synthetic transmembrane peptide to yield helix tilt and rotation defining the interaction surface of the helices in the bundle. In case of the transmembrane domain of glycoporphin A (Arkin et al., 1997) and the M2 protein (Kovacs & Cross, 1997) orientational data have been obtained which are in agreement with NMR data. Introducing this data into molecular dynamics

calculations as energy constraints yields a detailed structural model. This method has already been applied to other transmembrane proteins, i.e. the M2 protein of the Influenza A virus (Kukol et al., 1999), the vpu protein of HIV-I virus (Kukol & Arkin, 1999), the CM2 protein of Influenza C virus (Kukol & Arkin, 2000) and phospholamban (Torres et al., 2000). Here an extension of previous applications is described, that makes use of  $^{13}\text{C}=^{18}\text{O}$  labelled aminoacid residues (Torres et al., 2000; Torres et al., 2001b) at each possible position in the transmembrane sequence, thus yielding a more complete picture of the backbone structure compared to previous investigations. Independently a different approach has been employed utilising homologous sequences of the Ii transmembrane domain to select the correct model (Briggs et al., 2001). In this approach molecular dynamics simulations are carried out for each of the homologous sequences. This leads to several possible structures for each sequence. However, only one structure is common to all simulations, and this is selected as the correct model. Encouragingly, both methods result in a similar structure.

## 2. Results and Discussion

### 2.1. Secondary structure and transmembrane orientation

Typical ATR infrared spectra of the Ii transmembrane peptide (Iitp) reconstituted in dimyristoylphosphocholine (DMPC) vesicles are shown in figure 1 at  $0^\circ$  and  $90^\circ$  polarisation. The shape of the amide I absorption band centred at  $1658\text{ cm}^{-1}$  is indicative of a highly  $\alpha$ -helical protein (Byler & Susi, 1986) with the helix oriented perpendicular to the plane of the bilayer membrane. The orientation is confirmed by integration of the amide I absorption bands at  $0^\circ$  and  $90^\circ$  polarisation. Typically the dichroic ratio  $R = A_{\parallel}/A_{\perp}$  of the amide I varies between  $R = 2.6$  and  $R = 4.1$ , which is caused by a variation in sample order. Classical analysis of dichroic ratios (Braiman & Rothschild, 1988) yields order parameters between  $S = 0.36$  and  $S = 4.1$ , which

---

\* Abbreviations used: Ii: MHC associated invariant chain; Iitp: Ii transmembrane peptide; FTIR: Fourier-transform infrared; ATR: attenuated total reflection; SSID: site-specific infrared dichroism

corresponds to helix tilts between  $41^\circ$  and  $12^\circ$  respectively, if a completely ordered sample is assumed. Note that in SSID the sample order is taken into account and samples with any order above zero can be analysed to yield the absolute helix tilt. The small bands centred at  $1618\text{ cm}^{-1}$  and  $1594\text{ cm}^{-1}$  correspond to the  $^{13}\text{C}=\text{}^{16}\text{O}$  and  $^{13}\text{C}=\text{}^{18}\text{O}$  label in the positions indicated in figure 1. In all cases only part of  $^{16}\text{O}$  has exchanged to  $^{18}\text{O}$ , which does not restrict the analysis as long as the  $^{13}\text{C}=\text{}^{18}\text{O}$  amide I band can be integrated. In fact, for nine out of ten labelled peptides the  $^{13}\text{C}=\text{}^{18}\text{O}$  amide band could be integrated without the need for resolution enhancement techniques, like Fourier-self deconvolution, which can introduce artefacts into the results. Furthermore, in all cases the band appears at the same position, indicating that all residues chosen as labels are in an  $\alpha$ -helical environment (Torres et al., 2001a). For Ala45 the  $^{13}\text{C}=\text{}^{16}\text{O}$  amide I band has been integrated after Fourier-self deconvolution.

Quantitative analysis of spectra from adjacent labels according to the theory of SSID, as detailed in Materials and Methods, yields the orientation of each label in terms of rotational pitch angle and local helix tilt shown in table 1. The average helix tilt is  $13^\circ \pm 6^\circ$ , with a tendency to higher helix tilts from Leu44 onwards.

## 2.2. Constraint global molecular dynamics search

To find a detailed molecular structure which represents the experimentally determined orientation and corresponds to an energy minimum of a molecular dynamics simulation, a global search has been carried out (Adams et al., 1995), searching all possible interaction surfaces between helices in a trimeric bundle employing the orientational energy refinement. From the search process, described in Materials and Methods, 10 energy minimised structures have been found. The distribution of all structures as a function of energy and rotational search parameter  $\phi$  is shown in figure 2. The orientation of the labelled sites are compared with the experimental values in table 2. Note, that although the experimental data has been applied as additional energy refinement terms, energy minima at different orientations have been found, because due to CPU time limitations molecular dynamics simulations normally only find local minima and have difficulty overcoming large energy barriers involving the rotation of helices as a whole by angles more than  $\phi = 30^\circ$  to  $40^\circ$  as it can be seen in figure 2. The experimental constraints are fulfilled in all

these structures, because slight distortions from the  $\alpha$ -helical geometry can produce a variation of angles between the C=O bond and the z-axis. However, comparing simulated structures and experiment with regard to the rotational pitch angle, structure 2 emerges as the one closest to the experimental data as well as possessing the lowest energy. The corresponding right handed model is structure 9 as can be seen from figure 2. But due to distortions from helical geometry accurate determination of the rotational pitch angles was impossible. In addition structure 2 has a lower energy than structure 9. Although SSID alone cannot decide between left- and right-handed structures, it can be safely concluded from a comparison between simulation and experiment that Iitp forms a left-handed coiled coil trimer.

### 2.3. Molecular dynamics search of functional mutant sequences

In this approach a similar global molecular dynamics search has been carried out for a number of homologous sequences, but without experimental restraints. Structures resulting from simulations with different sequences are compared to each other with regard to their backbone RMSD. The structure chosen as the best model in this case is the structure, which is common to all simulations performed with different sequences. The reason is, that all sequences form the same native structure, and therefore this structure must be found in all of the simulations. Non native structures found in the simulations, need not persist in all of the sequences insce the sequence variation (a.k.a. silent substitutions) are only silent with respect to the native structure.

Figure 3 shows that there is only one structure (left handed, see open squares), at  $\omega_{42} \approx -10^\circ$ , appearing in all simulations. In this case the structures of different simulations differ from each other by less than 1.0 Å backbone RMSD. Two other sets were found at  $\omega_{42} \approx 0^\circ$  (right handed, grey squares) and  $\omega_{42} \approx 100^\circ$  (right handed, grey squares), but the RMSD is greater than 1.2 Å. Therefore the left handed structure at  $\omega_{42} \approx 0^\circ$  is chosen as the best model of Iitp, resulting from this approach.

### 2.4 Comparison of the two models

Calculation of the backbone RMSD between the two structures obtained independently from experimental and evolutionary data yields 1.1 Å. Since each structure has been obtained from a different approach, the two structures are assumed to be identical at the level of accuracy that can be achieved with the methods employed here. Although SSID enables us to select the native structure from an outcome of a global search, the closeness of the structures provides a further strong argument, that the right structure has been found. For the following discussion, the model obtained from orientational data is chosen. We believe that it is more accurate in detail, because orientational constraints have been applied during the simulation.

## 2.5 Reliability of the model

The viability of the approach using a transmembrane peptide as a model of the transmembrane domain of a 279 residue protein can be questioned. Therefore in this work two alternative approaches have been employed. In addition to orientational data obtained from the transmembrane peptide, homologous transmembrane sequences have been used in simulations. The assumption that the structure adopted by the transmembrane domain is similar to that of the transmembrane part of the whole protein is implicitly contained in these variant sequences, because otherwise they would not be found in other organisms. Since both approaches led to the same model, it can be safely concluded that the structure presented here is a reliable model for the transmembrane domain of the MHC associated invariant chain (Ii). In addition infrared spectroscopic data shows, that the transmembrane peptide exists in an oligomeric form and not as monomer, because the site dichroism  $R_{\text{site}}$  is always different from  $R_{\text{Helix}}$ . Similar dichroisms would indicate free rotation of the helix, since the dichroism is an average over all orientations in the lipid bilayer.

## 2.6 Structural implications

Figure 4 shows the structure of Ii<sub>tp</sub>, with the residues Tyr33, Gln47, Thr49 and Thr50 highlighted. A triple mutant of Ii at the position Gln47, Thr49 and Thr50 has been shown to fail to assemble with MHC class II because of a failure of Ii trimerisation (Ashman & Miller, 1999). From the structure shown in fig. 4 it is clear that Gln47 and Thr50 interact with each other and may stabilise the trimer structure. Thr49 points to the lipid phase and forms as would be expected intrahelical hydrogen bonds to main chain carbonyl-oxygens at position i-4 in the sequence. It can be predicted that mutation of Gln47 and Tyr50 only will be sufficient to disrupt trimerisation of Ii. It has been shown recently that polar residues, capable of serving simultaneously as hydrogen bond donor and acceptor (Asn, Asp, Gln, Glu, His), may stabilise transmembrane helix association (Zhou et al., 2001). The role of Gln47 in our model can be understood in the light of the above study, whereas Thr has been shown not to facilitate helix-helix association. Contrary to that, our model provides an interesting example of the involvement of Thr in helix-helix association.

Other residues involved in stabilising the trimer structure are Tyr33 and Leu43, which are highlighted in figure 2a. Minor interactions exist between Val36 and Val40 (not shown).

## 4 Materials and Methods

### 4.1 Peptide purification

Synthetic peptides corresponding to the predicted transmembrane domain of Ii (mouse) and a few adjacent aminoacids to reduce the hydrophobic character were made by solid-phase Fmoc (N-(9-fluorenyl)methoxycarbonyl) chemistry, cleaved from the resin with trifluoroacetic acid, and lyophilised. Ten peptides corresponding to residues 29-60 of Ii with the following sequence RGALYTGVSVLVALLLAGQATTAYFLYQQQGR where synthesised each containing one 1-<sup>13</sup>C=<sup>18</sup>O labelled aminoacid residue in positions 38, 39, 40, 41, 42, 43, 44, 45 and 48. The <sup>13</sup>C=<sup>18</sup>O labelled aminoacids have been prepared from 1-<sup>13</sup>C labelled aminoacids (Cambridge Isotopes Laboratories, Andover, MA) by acid catalysed oxygen exchange in H<sub>2</sub><sup>18</sup>O as described elsewhere (Torres et al., 2001a; Torres et al., 2001b). The peptides were further purified as described elsewhere (Kukul et al., 1999) for analogous transmembrane peptides. Briefly, the



peptide was dissolved in trifluoroacetic acid and purified by reversed phase chromatography (Jupiter 5C4-300 Å column, Phenomenex, Cheshire, UK). Peptide elution was achieved with a linear gradient to a final solvent composition of 5 % H<sub>2</sub>O, 38% acetonitrile and 57% 2-propanol. All solvents contained 0.1% (v/v) trifluoroacetic acid. After lyophilisation of the pooled fractions the peptide were dissolved in 1,1,1,3,3,3-hexafluoro-2-propanol, containing dimyristoylphosphocholine (Sigma), to a peptide to lipid ration of 1:25 (w/w).

## 4.2 Infrared spectroscopy

Fourier transform infrared spectra (FTIR) were recorded on a Nicolet Magna 560 spectrometer (Nicolet, Madison, WI) equipped with a high sensitivity liquid nitrogen cooled MCT/A detector. Attenuated total reflection (ATR) – FTIR spectra were measured with a 25 reflections ATR accessory from Graseby Specac (Kent, UK) and a wire grid polariser (0.25 µm, Graseby Specac). 100 µL of the peptide/lipid solution in hexafluoroisopropanol were dried onto a germanium trapezoidal internal reflection element (50 x 2 x 20 mm) under a stream of nitrogen. After the solvent was removed, the lipid/protein film was hydrated for 12h with a water saturated stream of nitrogen. After removal of bulk water an FTIR spectrum was obtained from 1000 averaged interferograms and 1 point zero filling Happ-Genzel apodisation. The dichroic ratio,  $R$ , was calculated as the ratio between the integrated absorption of parallel and perpendicular polarised light of the absorption bands (between 1670 and 1645 cm<sup>-1</sup> for the <sup>12</sup>C=<sup>16</sup>O amide I, and between 1600 and 1583 cm<sup>-1</sup> for the <sup>13</sup>C=<sup>18</sup>O amide I mode).

## 4.3 Data Analysis

The data were analysed according to the theory of site-specific dichroism presented in detail elsewhere (Arkin et al., 1997). Briefly, the measured dichroic ratio, the absorption between parallel and perpendicular polarised light  $R = A_{\parallel}/A_{\perp}$  of a particular transition dipole moment is a function of its spatial orientation. For the amide I mode (mainly the C=O bond vibration) of an  $\alpha$ -helical protein the geometric relation between the transition dipole moment and the helix is known from fibre diffraction studies (Tsuboi, 1962). Therefore, by measuring the orientation of the amide I transition dipole moment one can determine the helix tilt angle  $\beta$  and the rotational

pitch angle  $\omega$  of the specific dipole moment about the helix axis. The rotational pitch angle  $\omega$  is arbitrarily defined as  $0^\circ$  when the transition dipole moment, the helix director, and the z-axis all reside in a single plane. Thus, measuring the site-specific dichroic ratio  $R_{\text{site}}$  of the  $^{13}\text{C}=^{18}\text{O}$  amide I mode from a particular label and the helix dichroic ratio  $R_{\text{helix}}$  allows calculation of the helix tilt  $\beta$  and the rotational pitch angle  $\omega$  of a particular label if measurements from two samples with labels at different positions are analysed together. The pair of labels used in the analysis has been chosen such, that they are closest in the sequence; i.e. V38/L39, L39/V40, V40/A41, A41/L42, L42/L43, L43/L44, L44/L45, L45/A48, A48/A51. For each pair of labels  $X_n/Y_{n+1}$  the calculation yields the rotational pitch angle and local helix tilt for the label in position  $n$ . The difference of  $\omega$  between two consecutive residues is assumed to be  $100^\circ$  as in a canonical  $\alpha$ -helix.

#### 4.4 Global molecular dynamics search

A global search with respect to rotation about the helix axis, assuming trimeric symmetry, was carried out as described in detail elsewhere (Adams et al., 1995; Kukol et al., 1999). In brief, all calculations were performed with the parallel processing version of the Crystallography and NMR System Version 0.40 (Brunger et al., 1998) and CHI, using the OPLS parameter set with a united atom topology that explicitly represents the polar hydrogen and aromatic side chain atoms (Jorgensen & Tirado-Rives, 1988). All calculations were carried out in vacuo with the initial coordinates of a canonical  $\alpha$ -helix (3.6 residues/turn). Symmetric trimers were generated from the sequence ALYTGVSVLVALLAGQATTAYFL, acetylated at the N-terminus and methylaminated at the C-terminus, by replicating the helix and rotating it by  $360^\circ/3$  around the centre of the trimer. An initial crossing angle of  $25^\circ$  for left-handed and  $-25^\circ$  for right-handed structures was introduced by rotating the long helix axis with respect to the long bundle axis. The symmetric search was carried out by applying a rotation to all helices simultaneously between  $\phi = 0^\circ$  and  $\phi = 360^\circ$  in  $10^\circ$  steps. Four trials were carried out for each starting structure, using different random initial atom velocities in each case at both right- and left-handed crossing angles yielding  $36 \times 4 \times 2$  equalling 288 structures. Each structure was subjected to a simulated annealing and energy minimisation protocol. Clusters of similar structures were defined such that the root mean square deviation of the co-ordinates between all structures within a cluster was not larger than  $1 \text{ \AA}$ ; a cluster was formed by a minimum of 10 structures. For each cluster an average

structure was calculated, energy minimised and subjected to the same simulated annealing protocol used in the systematic search.

During all simulations orientation refinement energy terms have been incorporated as described elsewhere (Kukol et al., 1999). In this case for each of the 10 aminoacid residues an orientational restraint has been applied by setting the angles between the  $^{13}\text{C}=\text{}^{18}\text{O}$  bond and the z-axis to those obtained from the experiment. This angle is dependent on the rotational pitch angle and the helix tilt. The overall weight constant  $k_{\text{dichro}}$  for these orientational constraints was chosen to be  $k_{\text{dichro}} = 1000 \text{ kcal/rad}^2$ , determined empirically by selecting the minimum value necessary to obtain an identical outcome from MD protocols with different constants.

#### 4.5 Molecular dynamics search of mutant sequences

The same global search protocol, as described above (without orientational restraints), was applied to sequences from different organisms. In this case no orientational refinement energy terms have been applied. The following sequences have been used:

```
Q9XRE4  LKITGLTVLA  CLLLAGQALT  AYMV (zebra fish)
k01144  ALYTGFSILV  TLLLAGQATT  AYFL (human)
jc4796  ALYTGFSVLV  ALLLAGQATT  AYFL (horse_pig_bovine)
p04441  ALYTGVSVLV  ALLLAGQATT  AYFL (mouse)
x13044  VLYTSVSVLV  ALLLAGQATT  AYFL (rat)
```

These sequences have been chosen as homologous but not identical to the human sequence and represent all which could be found in the sequence database OWL (Bleasby & Wooton, 1990). The results from the global searching molecular dynamics simulations were represented graphically by plotting each cluster representative (candidate models) as a function of two parameters, the helix tilt  $\beta$ , and the rotational orientation  $\omega$ , as described previously (Briggs et al., 2001). The tilt angle of the model,  $\beta$ , was taken as the average of the angles between each helix axis and the z-axis, i.e., coincident with the normal to the bilayer and was calculated by CHI (Adams et al., 1995; Adams et al., 1996). The helical axis is a vector with starting and end point above and below the defined residue, these points corresponding to the mean of the co-ordinates of the five  $\alpha$  carbons N-terminal and the five  $\alpha$  carbons C-terminal to the defined residue.

The rotational orientation  $\omega$  is defined relative to an arbitrarily chosen, specified residue. The angle  $\omega$  is defined by the angle between a vector perpendicular to the helix axis, oriented towards the middle of the peptide C=O bond of the residue, and a plane that contains both the helical axis and the normal to the bilayer. This angle is  $0^\circ$  when the residue is located in the direction of the tilt (Arkin et al., 1997). The structures identified were plotted against the  $\omega$  angle of residue 42. Precise comparisons between similar clusters obtained from different variants were made by calculating the RMSD between its  $\alpha$  carbon backbones. In the simulations, the handedness of the bundle is indicated by the helix tilt sign, positive or negative, which corresponds to left and right handed bundles, respectively. Simulations were performed without any restraint.

## 5 Acknowledgements

This work was supported by grants from the Wellcome Trust and BBSRC to ITA.

## 6 References

- Adams, P. D., Arkin, I. T., Engelman, D. M. & Brünger, A. T. (1995). Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Structural Biology* **2**(2), 154-162.
- Adams, P. D., Engelman, D. M. & Brünger, A. T. (1996). Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins* **26**(3), 257-261.
- Arkin, I. T., MacKenzie, K. R. & Brünger, A. T. (1997). Site-directed dichroism as a method for obtaining rotational and orientational constraints for orientated polymers. *Journal of the American Chemical Society* **119**(38), 8973-8980.
- Ashman, J. B. & Miller, J. (1999). A role for the transmembrane domain in the trimerization of the MHC class II-associated invariant chain. *Journal of Immunology* **163**(5), 2704-2712.
- Bertolino, P., Staschewski, M., Trescol-Biemont, M. C., Freisewinkel, I. M., Schenck, K., Chretien, I., Forquet, F., Gerlier, D., Rabourdin-Combe, C. & Koch, N. (1995). Deletion of a C-terminal sequence of the class II-associated invariant chain abrogates invariant chains oligomer formation and class II antigen presentation. *Journal of Immunology* **154**(11), 5620-5629.

- Bijlmakers, M. J., Benaroch, P. & Ploegh, H. L. (1994). Mapping functional regions in the luminal domain of the class II- associated invariant chain. *Journal of Experimental Medicine* **180**(2), 623-629.
- Bleasby, A. J. & Wooton, J. C. (1990). Construction of validated, non-redundant composite protein sequence databases. *Protein Engineering* **3**(3), 153-191.
- Braiman, M. S. & Rothschild, K. J. (1988). Fourier transform infrared techniques for probing membrane protein structure. *Annual Review of Biophysics and Biomolecular Structure* **17**, 541-570.
- Briggs, J. A., Torres, J. & Arkin, I. T. (2001). A new method to model membrane protein structure based on silent amino acid substitutions. *Proteins* **44**(3), 370-375.
- Brunger, A. T., Adams, P. D., Clore, G., Gros, W., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L. M., Simonson, T. & Warren, G. (1998). Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica D Biological Crystallography* **54**, 905-921.
- Busch, R., Doebele, R., Patil, N. S., Pashine, A. & Mellins, E. D. (2000). Accessory molecules of MHC class II peptide loading. *Current Opinion in Immunology* **12**, 99-106.
- Byler, D. M. & Susi, H. (1986). Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **25**, 469-487.
- Denzin, L. K. & Cresswell, P. (1995). HLA-DM induces clip dissociation from MHC class II alpha-beta dimers and facilitates peptide loading. *Cell* **82**(1), 155-165.
- Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modelling. *Electrophoresis* **18**, 2714-2723.
- Jorgensen, W. & Tirado-Rives, J. (1988). The OPLS potential function for proteins, energy minimization for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **110**, 1657-1666.
- Kovacs, F. A. & Cross, T. A. (1997). Transmembrane four-helix bundle of Influenza A M2 protein channel: Structural implications from helix tilt and orientation. *Biophysical Journal* **73**(11), 2511-2517.
- Kukul, A., Adams, P. D., Rice, L. M., Brunger, A. T. & Arkin, I. T. (1999). Experimentally based orientational refinement of membrane protein models: A structure for the influenza A M2 H<sup>+</sup> channel. *Journal of Molecular Biology* **286**(3), 951-962.
- Kukul, A. & Arkin, I. T. (1999). vpu Transmembrane peptide structure obtained by site-specific fourier transform infrared dichroism and global molecular dynamics searching. *Biophysical Journal* **77**(3), 1594-1601.
- Kukul, A. & Arkin, I. T. (2000). Structure of the Influenza C virus CM2 protein transmembrane domain obtained by site-specific infrared dichroism and global molecular dynamics searching. *Journal of Biological Chemistry* **275**(6), 4225-4229.
- Lamb, C. A. & Cresswell, P. (1992). Assembly and transport properties of invariant chain trimers and HLA-DR- invariant chain complexes. *Journal of Immunology* **148**(11), 3478-3482.
- Marks, M. S., Blum, J. S. & Cresswell, P. (1990). Invariant chain trimers are sequestered in the rough endoplasmic reticulum in the absence of association with HLA class II antigens. *Journal of Cell Biology* **111**, 839-845.

- Roche, P. A., Marks, M. S. & Cresswell, P. (1991). Formation of a nine-subunit complex by HLA class II glycoproteins and the invariant chain. *Nature* **354**, 392-394.
- Torres, J., Adams, P. D. & Arkin, I. T. (2000). Use of a new label,  $^{13}\text{C} = 18\text{O}$ , in the determination of a structural model of phospholamban in a lipid bilayer. Spatial restraints resolve the ambiguity arising from interpretations of mutagenesis data. *Journal of Molecular Biology* **300**(4), 677-685.
- Torres, J., Kukol, A. & Arkin, I. T. (2001a). Site specific examination of secondary structure and orientation determination in membrane proteins: The peptidic  $^{13}\text{C}=^{18}\text{O}$  group as a novel infrared probe. *Biopolymers* **59**, 396-401.
- Torres, J., Kukol, A., Goodman, J. M. & Arkin, I. T. (2001b). Site-specific examination of secondary structure and orientation determination in membrane proteins: The peptidic  $^{13}\text{C}=^{18}\text{O}$  group as a novel infrared probe. *Biopolymers* **in press**.
- Tsuboi, M. (1962). Infrared dichroism and molecular conformation of a-form poly-gamma-benzyl-L-glutamate. *Journal of Polymer Science* **59**, 139-153.
- Wolf, P. R., Tourne, S., Miyazaki, T., Benoist, C., Mathis, D. & Ploegh, H. L. (1998). The phenotype of H-2M-deficient mice is dependent on the MHC class II molecules expressed. *European Journal of Immunology* **28**(9), 2605-2618.
- Zhou, F. X., Merianos, H. J., Brunger, A. T. & Engelman, D. M. (2001). Polar residues drive association of polyleucine transmembrane helices. *Proceedings of the National Academy of Sciences of the United States of America* **98**(5), 2250-2255.

Figure 1

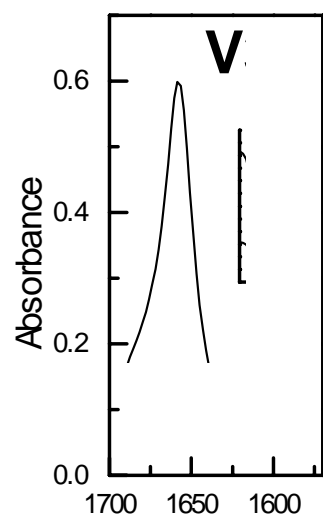


Figure 2

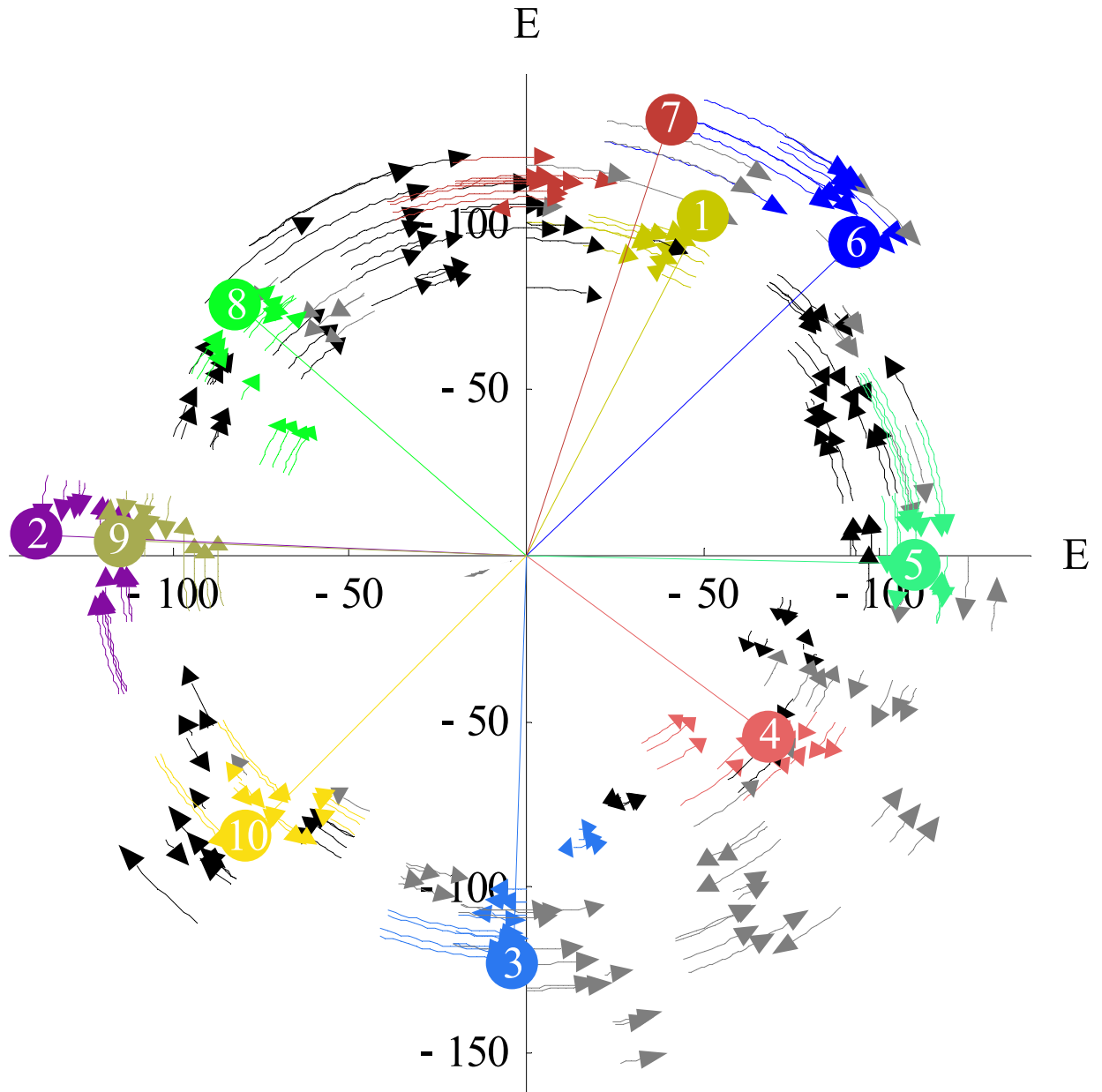
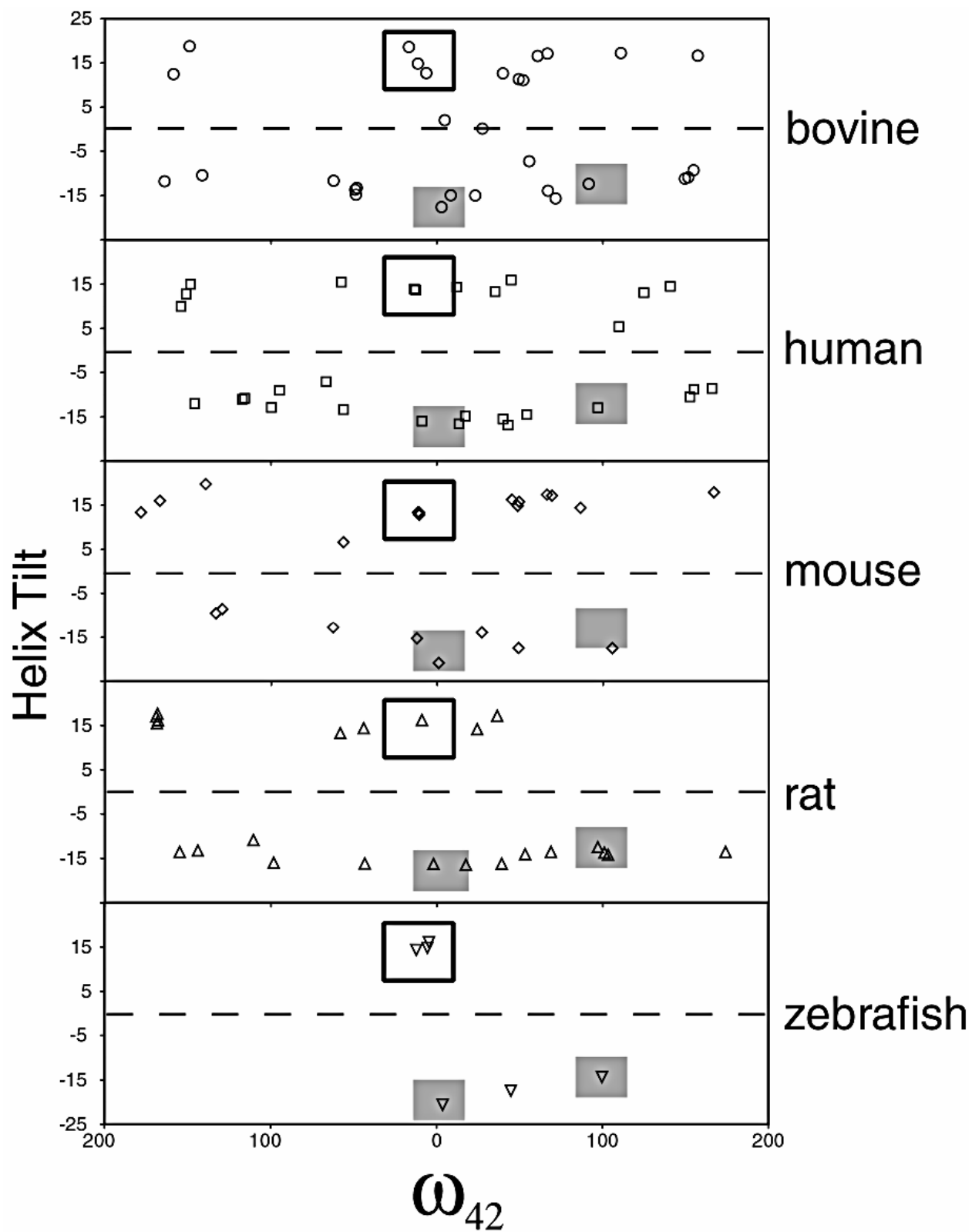




Figure 3:



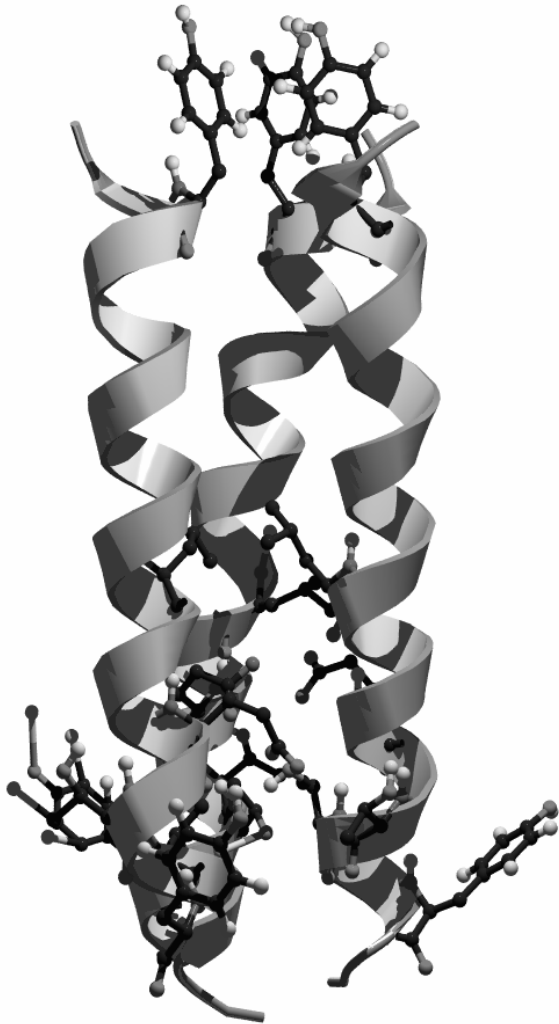
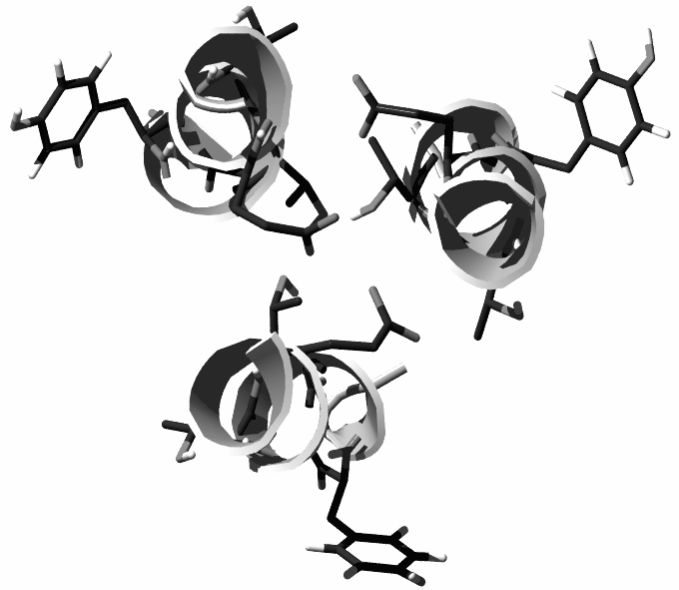
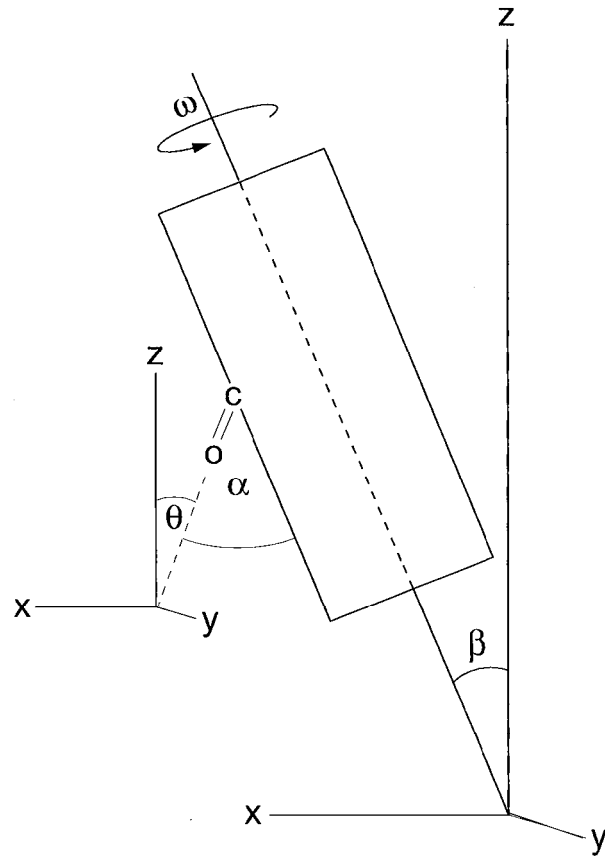
**Figure 4:****A****B**

Figure 5



### Legends to the figures:

**Figure 1:** Representative ATR FTIR-spectra of the lipid reconstituted Ii transmembrane peptide for all 10 different labels, at  $0^\circ$  (solid line) and  $90^\circ$  (dashed line) polarisation. The inset shows the region with the  $^{13}\text{C}=\text{O}$  amide I absorption band vertically expanded.

**Figure 2:** Energy and helix rotation parameter  $\phi$  for the 288 structures of a global molecular dynamics search shown in a polar plot. The arcs represent each individual structures, where the position on the circle indicates  $\phi$  and the distance from the centre the negative energy. Arrows indicate how the structure  $\phi$  changed during simulation. Structures are coloured according to their cluster affiliation, while the cluster averages are shown as circles connected to the origin by azimuthal lines.

**Figure 3:** Structures obtained from the simulation of homologous sequences. The obtained cluster averages are shown as a function of helix tilt and rotational pitch angle corresponding to residue 42,  $\omega_{42}$ . The horizontal broken line in each case separates left handed bundles (above the line) from right handed ones (below the line). Open squares indicate the closest model that persists in all simulations. Grey squares indicate models that also persist, but where structures are more disperse (see text).

**Figure 4:** Model of the Ii transmembrane peptide. A: The whole structure is shown with the residues Tyr33, Gln47, Thr49, Thr50 and Tyr52 in ball-and-stick representation. B: View of the of the patch of hydrophilic residues implicated in trimerisation of the helices. Gln47, Thr49, Thr50 and Tyr52 are shown in ball-and-stick representation. The figure has been created with Swiss-PDB viewer (Guex & Peitsch, 1997).

**Figure 5:** Schematic diagram of a helix shown as rectangular shape, defining the helix tilt angle  $\beta$  and the rotational pitch angle  $\omega$  in relation to the z-axis and to the transition dipole moment of the amide I mode (shown as C=O). The angle  $\alpha = 141^\circ$  is fixed in  $\alpha$ -helix geometry and known from fibre diffraction studies. The angle  $\theta$  is the absolute angle between the C=O bond and the z-axis; it is used as a constraint in MD simulations.

**Tables:**

**Table 1:** Rotational pitch angle  $\omega$  and local helix tilt  $\beta$ , determined for all  $^{13}\text{C}=^{18}\text{O}$  labelled residues; n gives the number of experiments for each label. From these values the angle between the C=O bond the z-axis,  $\theta$  has been calculated. The peptide sequence with labelled residues in boldface is shown. Note that each peptide contains only one label.

RGALY <b>TGVS</b> <b>VLVALL</b> LAGQATTAYFLYQQQGR				
Label	$\omega$	$\beta$	n	$\theta$
V38	$290^\circ \pm 10^\circ$	$10^\circ \pm 2^\circ$	6	$19.5^\circ$
L39	$359^\circ \pm 10^\circ$	$10^\circ \pm 2^\circ$	3	$12.1^\circ$
V40	$108^\circ \pm 12^\circ$	$6^\circ \pm 2^\circ$	4	$24.2^\circ$
A41	$216^\circ \pm 10^\circ$	$8^\circ \pm 2^\circ$	3	$28.3^\circ$
L42	$316^\circ \pm 14^\circ$	$14^\circ \pm 4^\circ$	7	$13.2^\circ$
L43	$359^\circ \pm 12^\circ$	$7^\circ \pm 2^\circ$	3	$15.1^\circ$
L44	$83^\circ \pm 13^\circ$	$21^\circ \pm 5^\circ$	5	$24.4^\circ$
A45	$133^\circ \pm 14^\circ$	$25^\circ \pm 8^\circ$	10	$39.3^\circ$
A48	$299^\circ \pm 20^\circ$	$15^\circ \pm 7^\circ$	12	$16.9^\circ$
(A51	$239^\circ$	$15^\circ$		$34.1^\circ)^1$

1) Calculated from A48 assuming  $\Delta\omega = 300^\circ$

**Table 2:** Rotational pitch angles  $\omega$  for all structures resulting from the constrained molecular dynamics search.

Structure:	1	2	3	4	5	6	7	8	9	10	Exp.
<b>Dimer search</b>											
V38	19	73	324	8.6	35	104	211				290
L39	110	111	78	113	141	285	344				359
V40	207	308	148	210	230	353	34				108
A41	280	2.7	239	298	322	46	78				216
L42	15	75	333	45	85	96	347				316
L43	132	281	98	137	177	297	4.1				359
L44	246	331	190	252	278	349	48				83
A45	318	36.5	305	334	8.0	71	269				133
A48	272	5.6	239	279	310	25	73				299
A51	209	308	182	204	258	325	29				239
Deviation <sup>1</sup> :	338	330	260	345	386	359	304				0
<b>Trimer search</b>											
V38	203	290	30	90	339	9.1	47	92	292	176	290
L39	300	31	151	216	66	102	143	213	60	323	359
V40	52	169	227	294	137	203	233	321	-	26	108
A41	147	251	307	4.6	243	301	330	28	-	88	216
L42	260	331	81	109	342	47	92	97	-	235	316
L43	344	77	183	267	81	136	187	225	-	336	359
L44	83	142	272	336	191	250	282	350	-	47	83
A45	198	291	7.2	44	300	340	16	50	-	117	133
A48	121	246	308	2	222	297	306	349	-	74	299
A51	69	197	240	299	199	214	216	287	-	16	239
Deviation <sup>1</sup> :	295	213	384	397	251	335	389	412	- <sup>2</sup>	288	0
<b>Tetramer search</b>											
V38	18	39	119	214	289	191	27	94	195		290
L39	113	168	292	320	32	322	130	189	339		359
V40	208	250	3.4	50	151	32	228	306	36		108
A41	285	329	54	147	246	94	320	28	91		216
L42	20	93	292	282	333	311	69	110	272		316
L43	128	212	41	5.5	61	355	163	230	13		359
L44	255	301	12	78	157	50	274	343	50		83
A45	329	21	106	216	285	121	8.1	51	113		133
A48	280	332	56	117	228	71	299	13	72		299
A51	220	274	23	75	154	41	248	290	37		239
Deviation <sup>1</sup> :	353	441	336	287	226	321	339	420	277		0

1) The deviation between the simulated  $\omega_s$  and experimental values  $\omega_e$  is calculated according to

$$D = \sqrt{\sum_{i=1}^{10} (\omega_{s,i} - \omega_{e,i})^2}$$

2) The rotational pitch angles of structure 9 could not be determined, because of distortions.