THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse III — Paul Sabatier
Discipline: Biologie Structurale, Biophysique
École Doctorale: Biologie – Santé – Biotechnologies

presentée et soutenue par

Louic S. VERMEER

NMR structure determination and MD simulations of membrane peptides and proteins: a peptide derived from H^+-V-ATPase subunit *a*, and MscL.

---

le 20 Mars 2009

JURY

J.-L. Popot	Directeur de recherche, Université, Paris-7	Président
E.J. Dufourc	Directeur de recherche, Université Bordeaux-1	Rapporteur
A.G. Lee	Professeur, University of Southampton	Rapporteur
M.A. Hemminga	Professeur Associé, Wageningen University	Examinateur
A. Milon	Professeur, Université de Toulouse, UPS	Directeur de thèse

---

Institut de Pharmacologie et de Biologie Structurale, CNRS UMR 5089
205, route de Narbonne, 31077 Toulouse cedex, FRANCE

Acknowledgements

The work presented in this thesis would not have been possible without the help of many other people, which is why a search for "thank" on this section of my thesis reports a word count of 15.

At the IPBS in Toulouse, my colleagues were always prepared to take time and provide advice when asked. For the solid-state NMR work, I would like to thank Valérie Réat. The liquid state NMR in Toulouse was carried out with the help of Pascal Ramos, Olivier Saurel, and Virginie Gervais. I want to thank Pascal Demange for his involvement in the biochemical work at the IPBS, and Jerzy Czaplicki for the discussions about MD simulations, computers, and science in general. I would also like to thank the other Ph.D. students at the IPBS, especially Marie Renault for her occasional help with NMR experiments, and Corinne Hazan, for carrying out a docking experiment involving the peptide studied in this thesis and V-ATPase subunit *c*. Thanks (again) to Valérie Réat, Isabelle Muller, Pascal Demange and Corinne Hazan (whoever happened to be nearest), for their help whenever I had to fill in my French tax forms or other interesting foreign documents.

A special word of thanks is well earned by Alain Milon, who not only gave me a warm welcome in Toulouse, but also was of great help during all the administrative issues that were necessary for me to move to France. Also, regardless of his busy schedule, he always provided me with immediate and accurate answers to questions ranging from biology to organic chemistry and –of course– NMR. Even when I did not ask for it, he was always able to critically look at my analyses and identify any possible problems or inaccuracies. I think this is a very important role for any professor, and I really appreciated his suggestions. Apart from his scientific contributions to my education, Alain has also proven to be a nice and understanding supervisor, as well as a friend.

For the work I carried out in Holland, at good old Wageningen University

where I have passed my M.Sc. thesis several years ago, I owe a word of thanks to Werner Vos and Afonso Duarte, for many useful discussions. I am also grateful to Marcus Hemminga for his critical comments on my publications, and especially for showing me the opportunity to apply for the Marie Curie fellowship that allowed me to start working on this thesis in the first place. I would also like to thank Rainer Wechselberger, who helped record NMR spectra at the large-scale NMR facility in Utrecht, and Tony Lee and Andy Powl, whom I visited in Southampton to learn the expression of MscL in *E. coli*.

Thanks goes out to Bert de Groot as well, whom I collaborated with while writing the paper on lipid order parameters. The collaboration was both efficient and pleasurable, even though we did not meet a single time during the writing of the article. I was positively surprised that he also likes to edit LaTeX sources directly, which made things a lot easier. And fortunately I also had the opportunity to meet him later on during my thesis.

I want to thank the members of the jury, that were prepared to read my thesis and come all the way to Toulouse for the defence despite the fact that most of them were very busy. I appreciated their intelligent questions and the discussions we had after my presentation.

A word of thanks is also appropriate for John Findlay and his entire Marie Curie "BIOMEM" consortium. I had a great time, and really enjoyed the excellent conferences and workshops organised by the members of the consortium. It is my personal opinion that schools and workshops are usually a lot more interesting than the "standard" scientific conferences, and this was certainly the case for the "BIOMEM" courses, where the excellent level of the presentations and modern research methods provided a very solid basis to start a scientific career. Because a large part of the fun I had during these meetings was caused by my fellow Marie Curie students, I want to thank them for being such good company.

Apart from science, I want to thank the many people that supported me during my time in France, Holland and Southampton. This includes (in no particular order) the group members of the labs I visited, the people at gentoo-forum.nl, the (mostly Spanish) gang I met in Toulouse, my housemates in Toulouse, at MDL3000, and at "De Brug", the people at KSV St. Franciscus Xaverius, The Knights of the Round Table and line Quiet, and all my other friends that do not belong to one of these groups. And finally, but importantly, I want to thank my sister, my father and my mother, for their support which has a history much longer than the time it took me to complete this thesis.

Therefore: Thanks!

Contents

List of Abbreviations

ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
CD	circular dichroism
CMC	critical micelle concentration
CP	cross-polarisation
CP-MAS	cross-polarisation magic angle spinning
CSA	chemical shift anisotropy
CSI	chemical shift index
DCl	deuterium chloride
DLS	dynamic light scattering
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DMSO	dimethyl sulfoxide
DNA	desoxyribonucleic acid
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DPC	dodecylphosphocholine
DSS	2,2-dimethyl-2-silapentane-5-sulfonic acid
DSSP	dictionary of protein secondary structure

ESR	electron spin resonance
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	infrared
KcsA	a bacterial potassium channel
LDAO	lauryldimethylamine-oxide
MAS	magic angle spinning
MD	molecular dynamics
MscL	mechanosensitive channel of large conductance
MS	mechanosensitive
MTM7	a peptide derived from H^+-V-ATPase subunit *a* that has been used in previous studies, see also table 4.1 on page 79
NaOD	deuterated sodium hydroxide
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
octylglucoside	n-octyl-β-D-glucopyranoside
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-[phospho-rac(1-glycerol)]
RMSD	root mean square deviation
SDS-d25	uniformly deuterated sodium dodecyl sulfate
SDS	sodium dodecyl sulfate
Tb-MscL	mechanosensitive channel of large conductance (MscL) from *Mycobacterium tuberculosis*
TM1	transmembrane segment 1
TM2	transmembrane segment 2

TM4	transmembrane segment 4
TM7	transmembrane segment 7
TM	transmembrane
TOCSY	total correlation spectroscopy
TPPM	two pulse phase modulation
V-ATPase	vacuolar proton-translocating adenosine triphosphatase
WATERGATE	water suppression through gradient tailored excitation

Summary

The 3D structure of a peptide derived from the putative transmembrane segment 7 (TM7) from subunit *a* of H^+-vacuolar proton-translocating adenosine triphosphatase (V-ATPase) (*Saccharomyces cerevisiae*) has been determined by solution state nuclear magnetic resonance (NMR) in sodium dodecyl sulfate (SDS). A stable helix is formed from L736 up to and including Q745, the lumenal half of the putative TM7. The helical region extends well beyond A738, as was previously suggested based on NMR studies of a similar peptide in dimethyl sulfoxide (DMSO). The pKa of both histidine residues that are important for proton transport was measured in water and in SDS. The differences that were found demonstrate that the histidine residues interact with the SDS polar heads. In detergent, circular dichroism (CD) data indicate that the secondary structure of the peptide depends on the pH and the type of detergent used. Using solid-state NMR, it was shown that the peptide is immobile in the presence of phospholipid bilayers, which means that it is probably not a single transmembrane helix in these samples. The environment is important for the structure of TM7, so in subunit *a* it is probably held in place by the other transmemebrane helices of this subunit.

The reasons behind the experimental observations are studied in atomistic detail by molecular dynamics (MD) simulations. The effect of the pH is simulated by running two simulations: one with uncharged histidine residues and one with charged histidine residues. These simulations are carried out both in SDS and n-octyl-β-D-glucopyranoside (octylglucoside), for 20 ns each. One more simulation was run under the conditions of the NMR experiment, with the nuclear Overhauser effect (NOE) restraints in place. This fifth simulation can be seen as a further refinement of the NMR structure, but it is also used as a reference to compare to the other, unrestrained, simulations.

In SDS, the peptide takes a position in the hydrophobic interior of the micelle if the histidine residues are uncharged, but it aligns itself to the SDS

headgroups on the micelle surface at low pH. In octylglucoside, the peptide does not take a position inside a micelle, regardless of the histidine protonation state. From these data, it is concluded that at high pH, the peptide is protected from aggregation by the SDS micelle, but not in octylglucoside, in agreement with experimental results. Furthermore, it is observed that the structure of the peptide is different in the simulation with the NOE restraints which indicates a bias towards alpha helices by the other simulations. The structural differences are small, but seem responsible for the increased amount of water contacts that are observed in the simulation with the NOEs.

Finally, the residues that are in contact with the water are identified, in order to locate the face of TM7 that lines the lumenal water channel in V-ATPase. Those residues are Y733, S740, W737 and H743. The R735 residue which is essential for proton transport is located on the other, hydrophobic side of the helix. The hydrophobicity of this face of TM7 fits well with the current opinion that it interacts with the hydrophobic subunit *c*.

As a second part of this thesis, around 300 mg of uniformly ^{15}N labelled MscL has been produced. It has been successfully purified and reconstituted in lipid bilayers, in order to prepare oriented samples for solid-state NMR studies. So far, no oriented samples have been obtained, as is shown by ^{31}P solid-state NMR, but suggestions for possible improvements are made.

During this thesis, I also wrote a mini-review about the comparison of order parameters from ^{2}H NMR and MD simulations. Order parameters from deuterium NMRare often used to validate or calibrate molecular dynamics simulations of lipid bilayers. The mini-review is included in this thesis as chapter 3 on page 39, and gives a short overview of the literature in which experimental order parameters from ^{2}H NMRare compared to those calculated from MD simulations. The different ways in which order parameters from experiment are used to calibrate and validate simulations are reviewed. In the second part of this review, a case study of cholesterol in a 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) bilayer is presented. It is concluded that the agreement between experimental data and simulation is favourable in the hydrophobic region of the membrane, for both the phospholipids and cholesterol. In the interfacial region the agreement is less satisfactory, probably because of the high polarity of this region which makes the correct computation of the electrostatics more complex.

1

General Introduction: membranes and membrane proteins

Contents

1.1 Biophysical studies of membrane proteins

The existence of living cells is only possible because they are separated from the world outside. This barrier consists of different types of membranes, composed of mainly lipids and proteins [63]. Many types of these so-called membrane proteins exist, each of them with one or more very specific functions. Membrane proteins can, for example, be pumps or channels that allow specific molecules to enter or leave the cell, which is necessary to maintain the delicate balance between the inside of the cell and its surroundings [63]. For example: molecules that provide the cell with energy (food) must be allowed to enter, but at the same time cells try to keep toxic compounds out. Many ($\sim$ 50%) of known drugs interact with membrane proteins to exert their function.

A better understanding of the structure and function of membrane proteins can contribute to a better fundamental understanding of the workings of a single cell and communication between cells, and it is likely to lead to more insight into the cause of various types of diseases as well as to the development of new drugs and new types of antibiotics.

In the latest decades, scientists have become to understand that the structure and function of membrane proteins can be influenced by the lipids, and that a protein can also influence the properties of the lipids around it [21, 32, 33, 40]. The thickness of a membrane, for example, has been shown to be able to change the shape and orientation of proteins [10, 55] . Examples of a specific type of lipid that is required for a protein to function also exist [39].

Molecules cannot be observed by the naked eye or by using a regular microscope. Therefore, biophysicists use indirect techniques to learn about the shape of molecules. One such technique is NMR. NMR is a sophisticated and elegant type of spectroscopy, that uses the absorption of radio waves by the nucleus of an atom inside a magnetic field to study the surroundings of that nucleus. By determining the positions of atoms relative to each other, a three-dimensional picture of the molecule emerges: the structure of a protein. NMR can also be used to get information about the internal movements of a molecule, or the interactions it has with its environment.

When the protein is a membrane protein, it is especially interesting to see what its structure looks like while it is inside the membrane, just like in a real cell [48]. But because living cells have very complex membranes that consist of many different types of lipids, proteins, cholesterol, and other molecules, it is hard to acquire high resolution structures in these systems. Therefore, model systems are used that consist of one or two different lipids and a protein of interest that has been purified from bacteria. Usually, these

bacteria are genetically engineered to produce the protein in larger than normal quantities.

Membrane proteins are usually not soluble in water. Although it is relatively straightforward to determine the structure of a water-soluble protein by NMR, a membrane fragment is much bigger than a single molecule, which complicates matters, because the observed NMR signal becomes orientation dependent. This effect occurs because the protein–lipid complex does not rotate completely in the time that the measurement is carried out. There are many proteins and lipids in the sample and they all have a different orientation with respect to the magnetic field. Each orientation gives a slightly different signal, and the recorded signal becomes a superposition of the signals from all the different orientations. This leads to broad lines in the spectrum, that prevent an interpretation in terms of the distances between atoms as is possible for liquid state NMR.

Fortunately, the dependence on orientation can be turned to our advantage if we want to study the way a membrane protein is oriented inside a membrane [2, 37]. To do so, it is possible to prepare membranes that are oriented on a glass surface. This way, the lipids containing the protein do not rotate, and all molecules point in the same direction with respect to the magnetic field. These oriented samples allow measurements to be carried out that reveal the orientation of the protein. Because we know the orientation of the lipids, this experiment can be interpreted in terms of the tilt angle of helices inside a protein, relative to the lipid bilayer. In this thesis, tests are carried out to prepare oriented samples to study the membrane protein called “MscL”.

An other possibility to learn about the structure of membrane proteins by NMR is the use of membrane-mimicking systems. This often involves solubilising the protein in a detergent instead of lipids [61]. The properties of detergents resemble those of lipids in the sense that they also consist of a long apolar tail that interacts with the membrane protein and a polar headgroup that makes the protein-detergent complex (micelle) soluble in water. An important difference with lipids is the shape of the detergent molecules, which causes them to form smaller structures in solution. Because these smaller structures rotate faster, the orientation-dependent interactions are eliminated. This allows measurement of internuclear distances in the protein and the determination of the protein structure. Solubilisation in detergent micelles is used in this thesis to determine the structure of a peptide fragment from the V-ATPase membrane protein.

With the advances in processor speed of computers, the modelling of interactions between atoms has become commonplace in biophysical research [46]. These theoretical models are getting closer and closer to the experi-

mental results, and can provide atomic details that are not available from experiments. It is always important to compare the results of these simulations to experimental data, because that is the only way to know if the simulations correctly reflect the experimental observations. If this is indeed the case, properties of the system that cannot be observed experimentally may be derived from the simulations to construct a model of the atomic interactions. In this thesis, MD simulations are used to study the details of the interactions between the peptide from V-ATPase and the different detergents in which it was solubilised.

In this thesis a peptide is studied using both solid-state NMR to determine its orientation and mobility in lipid vesicles, and liquid state NMR to determine its structure in detergent micelles. Molecular dynamics simulations are used to help in the interpretation of the NMR data, and their results are compared to the experimental data that has been obtained using several different biophysical methods. As a second part of this thesis, an entire membrane protein was reconstituted in lipid bilayers and the system was characterised by solid-state NMR.

1.2 The lipid bilayer and its influence on membrane proteins

Nowadays, it is well known that lipids can influence the structure of membrane proteins, and that membrane proteins can also affect the properties of lipids around them [32, 40]. In fact, lipids have even been shown to regulate the activity of some membrane proteins. There are membrane proteins that require a specific type of lipid for their activity. This lipid usually has an increased affinity for the protein [45]. It has for example been shown that a bacterial potassium channel (KcsA) requires the binding of anionic phospholipids to function, and in this case the lipids were compared to a co-factor [38, 39, 74]. But in most cases, no specific lipids are required for protein function, and the phospholipid bilayer can be seen as the solvent for the membrane proteins. This does not exclude the possibility that certain lipids have a higher affinity for a specific location on the outside of the protein, as has been demonstrated for the MscL [57]. These lipids may also influence the function of the membrane protein [59]. Therefore, a distinction has been made between non-annular lipids (the lipids that act as a co-factor), annular lipids that form the first shell of lipids around the membrane protein, and bulk lipids that confer the general physical properties of the bilayer such as the membrane curvature, elasticity, packing, order and fluidity. It should

be noted that these bulk properties are not independent. More details on the bulk properties of lipid bilayers are discussed in chapter 3 on page 39, references therein, and other review articles [55]. Evidence suggests that membrane peptides and proteins influence the lipids directly around them, but not the properties of the bulk lipids, as is clearly explained in an article by Powl and Lee [55]. An important question in modern scientific research is whether a given membrane protein is regulated by non-annular lipids, annular lipids or the bilayer bulk properties.

An example of a membrane protein that is influenced by the length of the hydrophobic acyl chains is Ca^{+}-ATPase. It has been shown that its activity is highest in lipids with a chain length of 18 carbon atoms, while the activity decreases when it is reconstituted in lipids with shorter or longer fatty acyl chains [55]. More fundamentally, the influence that the hydrophobic thickness has on model peptides has been extensively studied [10, 29, 30, 47].

The lipid headgroups also play an important role in the interaction with membrane proteins. It has been shown that tryptophan residues have a specific affinity for the bilayer interface region [15, 69], near the lipid carbonyl atom. Lysine residues are also known to serve as membrane anchors, but their affinity is higher near the lipid phosphate group, closer to the water phase [11]. It is noteworthy that MscL does not contain any native tryptophan residues, the implications of which will be discussed below (to be precise, in section 1.5.2 on page 12). An overview of several proteins that are regulated by lipid headgroups is given in a review article by Newton [45].

There are other ways in which the bulk properties of a lipid bilayer can influence membrane proteins. The saturation of the acyl chains, bilayer fluidity and curvature have been shown to exert such an influence. The influence of bilayer curvature on membrane protein function was clearly shown in a paper by Perozo et al. [51], where an open form of MscL was stabilised by creating an asymmetric, curved bilayer with lysophosphatidylcholine.

Some properties of the lipid bilayer can be studied by NMR. Especially the combination with Molecular Dynamics simulations is very interesting, because those methods often are complementary. This subject is discussed in greater detail in the mini-review included in this thesis as chapter 3 on page 39.

1.3 Peptides derived from proteins

For the work that will be presented in this thesis, an important question to ask is whether a peptide derived from a (membrane) protein shows the same structural properties as in the native protein. In the protein, the peptide

may be in an environment where its native structure is stabilised by the surrounding residues, which are not present in the single peptide.

Review articles by Popot et al. [53, 54] suggest that membrane protein folding can be described by a two-step process. In the first step, transmembrane helices fold and insert in the membrane, and in the second step they interact with other transmembrane helices. This theory also means that helical fragments of membrane proteins have an intrinsic property to form helices in a hydrophobic environment, such as a lipid bilayer or a detergent micelle. Furthermore, it is very unfavourable for charged atoms to exist in a hydrophobic environment, so transmembrane sections of proteins are expected to form hydrogen bonds and thus form an α-helix or β-barrel inside the membrane. More details about the theories of membrane protein folding are described in a review article by Popot and Engelman [54], and the references therein. Experimental data has also shown that the structure of transmembrane α-helices is very stable in both peptides and membrane proteins [48]. The most important observation for this thesis is that transmembrane sections of integral membrane proteins are expected to fold in a helical conformation without the helix–helix interactions that are present in the completely folded, native protein.

Applications of this two-state folding model exist. It has been shown that peptides derived from Bacteriorhodopsin form α-helices or loops, depending on the structure they have in the entire peptide [24]. The structure of a peptide corresponding to the TM2 domain of $GABA_A$ in oriented bilayers was determined by solid-state NMR [22]. A peptide from TM4 of the Slc11a1 cation transporter was studied by solution NMR in SDS micelles [75], and many other examples of structural studies of short peptides derived from membrane proteins exist.

Peptides derived from soluble membrane proteins have also successfully been studied. The reader should not though, that the two-state folding model is only applicable to membrane proteins. The secondary structure of the soluble protein thymidine kinase has been predicted by CD spectroscopy measurements of many overlapping 15-residue peptides [3] and the "dissection" of Ubiquitin into two peptides has also shown that the amount of secondary structure of the peptides agrees well with the NMR structure of the entire protein [9]. A structural model of elastin has been determined based on a recomposition of peptides derived from its primary structure, whose structures were determined by CD and NMR spectroscopy [68]. Binding properties of ligands or inhibitors to peptides derived from entire proteins have also been studied successfully [28, 71].

1.4 Outline of this thesis

This thesis has two objectives:

- the study of the structure of a peptide derived from V-ATPase TM7, subunit *a* by liquid state NMR, solid state NMR, MD simulations, and complementary biophysical methods. Residues in this peptide play a key role in the proton translocation by V-ATPase.

- the preparation of samples with an entire membrane protein, MscL, in oriented lipid bilayers, for studies by solid-state NMR. If oriented samples can be prepared, important information about the MscL opening mechanism and lipid–protein interactions in general may be obtained.

The first and major part of the research was dedicated to the structural characterisation of a peptide derived from V-ATPase, subunit *a*. This peptide was studied in lipids and different detergents, and at a high and low pH. Although a structure of a similar peptide in DMSO exists, it is important to conduct studies in other membrane mimicking solvents as well, to assess the possible influence of the solvent on the structure of the peptide. The results obtained are interesting, because they show the structure of the peptide that plays a mayor role in the proton translocation by V-ATPase. But when interpreted in a more general way, they add to the knowledge of the interactions between peptides and lipids or membrane mimicking systems. To study the peptide–detergent interactions in atomic detail, molecular dynamics simulations were carried out, and compared with experiments.

As a second part of this thesis, the reconstitution of an entire membrane protein in lipid bilayers was carried out, and the possibility to prepare oriented samples for NMR studies was investigated. The protein–lipid interactions of MscL are especially interesting, because it has been shown that they are directly responsible for the opening of the channel. MscL can be produced in minimal media, in quantities that are suitable for NMR studies. This allows isotopic labelling of the protein for detailed NMR studies. If the protein can be properly oriented in the bilayer, helical tilt angles can be measured with respect to the bilayer normal. A successful reconstitution in oriented lipid bilayers would permit detailed studies of the interactions between MscL and the lipids, such as using phospholipid membranes with different hydrophobic thickness to study their influence on the tilt angle of the MscL transmembrane helices, or studying different mutants of MscL in lipid bilayers to learn more about the interactions of specific residues with the lipids, and possibly the opening mechanism of this membrane channel protein.

During this work, it was found that no review articles seem to exist where the comparison between order parameters from MD simulations and experimental order parameters from ^{2}H NMR is made. We therefore decided to write a mini-review [70], which is included in this thesis as chapter 3 on page 39.

1.5 The proteins studied in this work

1.5.1 The H^+ transporting vacuolar ATPase

It is well known that the pH of cellular compartments is an important factor in many biological processes. Some of these cellular compartments are therefore actively acidified against a proton gradient. The protein that is directly responsible for this acidification is V-ATPase, which uses adenosine triphosphate (ATP) as energy source. The V-ATPase protein has been identified in the intracellular compartments of eukaryotic cells [26]. It has also been found in certain types of specialised cells such as human osteoclasts, where it acidifies the extracellular space called the resorption lacuna. This is a part of the natural process of continuous renewal of the bone tissue: osteoclasts degrade the bone tissue, while osteoblasts create it. If the renewal of the bone tissue is slower than the degradation, as is often the case in older individuals, the bones slowly break down and become weaker, a condition that it is diagnosed as osteoporosis. Specific inhibitors of V-ATPase are therefore an interesting research subject [76], because they can lead to the development of a medicine against osteoporosis. So far, however, the absence of a high-resolution structure of V-ATPase or its transmembrane subunits has hampered this research. In most of the current research and in this thesis, V-ATPase from *Saccharomyces cerevisiae* is used [23, 76].

V-ATPase consists of two domains: the cytoplasmic V_1 domain and the transmembrane V_0 domain (see figure 1.1 on the facing page). Both domains consist of multiple subunits. The V_1 domain is responsible for ATP hydrolysis, which drives the rotation of the A/B-subunit assembly. This rotation is transferred to the c subunit of the V_0 domain by subunit D. This introduction will focus on subunits *a* and c from the V_0 domain.

Subunit *c* is called the rotor, and interacts with subunit *a*. Although the structure of subunit *c* from human and yeast H^+-V-ATPase are unknown, the crystal structures of the related V-type Na^+ ATPase from *Enterococcus hirae* [44] and F-type Na^+ ATPase from *Ilyobacter tartaricus* [42] have been published.

Subunit *a*, also referred to as "the Vph1p subunit", is the stator against

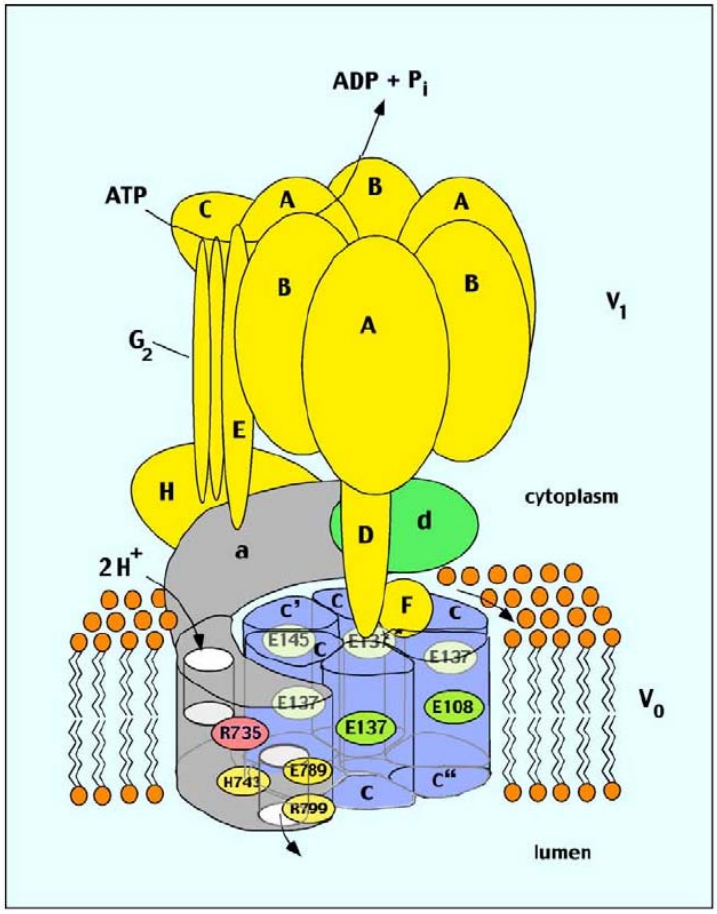

Figure 1.1: *Figure from Kawasaki-Nishi et al. [26], showing the subunit arrangement and some details of the proposed V-ATPase proton translocation mechanism. Note that the position of the residues in subunit a is different from the topological model shown in figure 1.2 on the next page, due to the recent revision of the topological model.*

which subunit *c* rotates. It is believed to contain a separate input and output channel, that are in contact with the cytoplasm and lumen, respectively. This is schematically depicted in figure 1.1 and figure 1.3 on page 11. The structure of subunit *a* is unknown. Even its membrane topology has been the subject of some controversy. It was first believed to consist of 9 transmembrane helices, which is why in the older literature Arg-735 may be referred to as being located in TM8 [36] instead of TM7. This model was revised based on cysteine mutagenesis and maleimide labelling data [26], placing the Arg-735 residue in TM7. Recently, the model was again revised and subunit

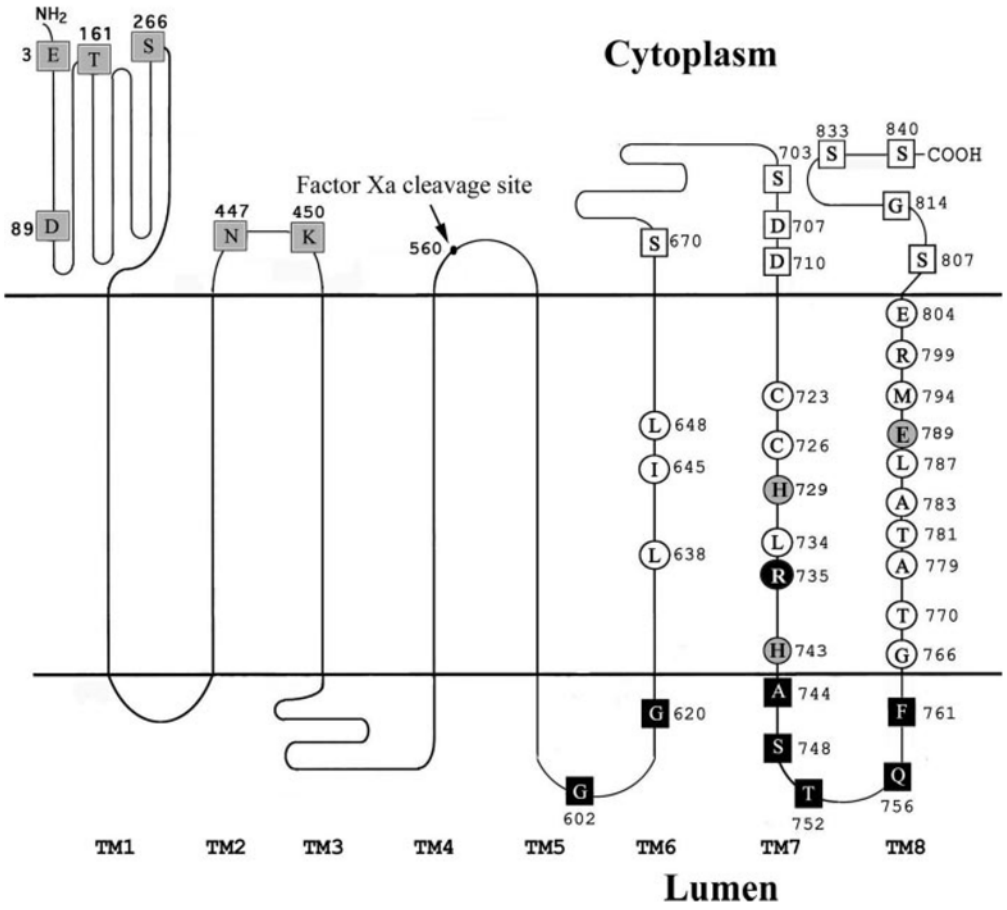

Figure 1.2: *Figure from Wang et al. [73], showing the latest topological model of V-ATPase subunit a. Residues shown in open squares were shown to be located on the cytoplasmic side of the membrane, those in black squares on the lumenal side. The residues in grey circles have been shown to reduce, but not completely inhibit proton translocation [34, 35]. The Arg-735 residue, shown as a black circle, is absolutely essential for proton translocation [25].*

a is now believed to contain eight transmembrane helices [73], as shown in figure 1.2. The histidine residues that have been shown to be important for proton translocation [34, 35] are located in TM7. The Arg-735 residue, which has been shown to be absolutely essential for proton translocation [25] is also located in this helix. The latest topological model places the Arg-735 residue on the lumenal side of the membrane.

Cysteine cross-linking studies have shown that subunit *a* and subunit *c* are in contact. The arginine-735 residue in TM7 from subunit *a* interacts

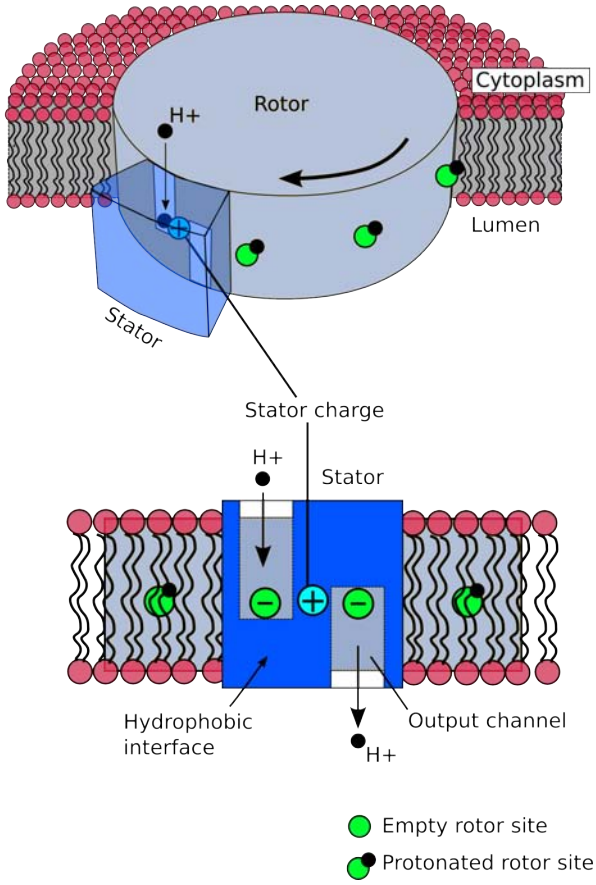

Figure 1.3: *A model for the proton translocation mechanism by V-ATPase, based on a figure by Grabe et al. [17]. Proton translocation takes place at the interface between subunit a and subunit c. The negative charge on subunit c makes it unfavourable for an unprotonated glutamic acid residue to enter the membrane. Once protonated, subunit c can rotate further, exposing the now uncharged glutamic acid residue to the hydrophobic interior of the membrane. After an almost complete rotation, the proton is released in the lumenal hemichannel due to it being destabilised by the charges on the stator, subunit a.*

with the glutamic acid residue in TM4 from subunit c and c′ [27]. In subunit c″, the interacting glutamic acid residue is located in subunit in TM2 [72]. The different cysteine cross-links that were observed indicate mobility [27], suggesting conformational changes or rotation of one or both of these helices during proton translocation.

A detailed model of the proton translocation mechanism has been suggested by Grabe et al. [17], and is summarised in figure 1.3. In short, the

model supposes that the glutamic acid residue on the rotating subunit *c* becomes protonated when it comes in contact with the cytoplasmic hemichannel. The protonation of glutamic acid neutralises its negative charge, allowing it to rotate further, where it comes in contact with the hydrophobic acyl chains of the lipid bilayer. After a rotation of *almost* 360°, the proton is released in a second hemichannel that is in contact with the lumenal side of the membrane. The deprotonation of the glutamic acid residue is induced because its pKa is lowered by the proximity of positively charged residues on the stator (subunit *a*). The absence of a structure of subunit *a* has made it difficult to choose between a one-channel and a two-channel model for proton translocation, although the current consensus is that the two-channel model is the more likely because it does not expose hydrophobic parts of subunit *c* to water[13, 17]. Structural information about especially subunit *a* has therefore long been a hot topic in V-ATPase and membrane protein research. The structure of helix 7 from subunit *a* that is presented in this work is a modest contribution that may help to better understand the proton translocation mechanism.

1.5.2 The mechanosensitive channel of large conductance

Around 20 years ago, it was shown that the membrane of *E.coli* bacteria contained a channel that opened or closed depending on the suction applied to the capillary of a patch-clamping pipette [41]. A couple of years later, several different classes of mechanosensitive proteins were identified. The most studied ones, named according to their conductive properties, are MscS for "mechanosensitive channel of small conductance" [1] and MscL for "mechanosensitive channel of large conductance" [66, 67]. Review articles about the two families of mechanosensitive (MS) channels are available [8, 52], and more information can also be found in the references therein.

Mechanosensitivity is an important property that allows organisms to respond to physical stimuli from their environment. Examples involving mechanosensitivity are touch, hearing, sensing pressure and gravity etc. Multiple different mechanisms of mechanosensitation exist in nature [16], but in this thesis our interest goes out to MscL, which senses the tension in the lipid bilayer in which it is located.

The bacterial channel MscL, has often been compared to an "escape valve", because it serves to release pressure when before it becomes too high inside the cell (see figure 1.4 on the facing page). Such high pressures may arise when water enters the cell due to osmosis, which occurs when the medium

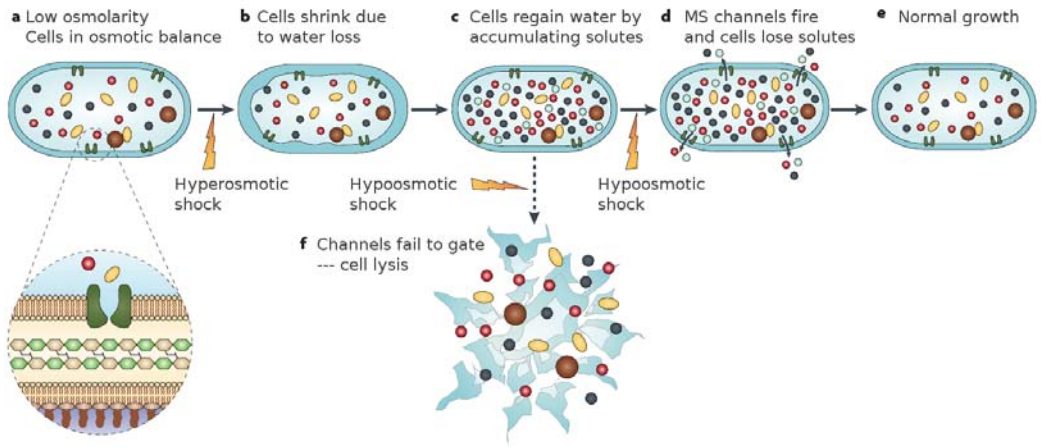

Figure 1.4: *Schematic explanation of the MscL function as an escape valve in the case of osmotic downshock. Slightly adapted from Booth et al. [5], a review paper on the function on mechanosensitive channels.*

where the bacteria live is suddenly diluted with respect to the bacterial interior. Such a difference can for example be caused by rain. If the cell's environment changes quickly from a high osmotic value to a low one, the osmotic pressure between the bacterial interior and exterior will cause water to enter the cell, leading to an increase of the pressure on the membrane. If this pressure becomes too high and is not released quickly, the cell will lyse. The bacterium therefore opens its mechanosensitive channels, allowing solutes and water to be released to the external medium, releasing the pressure and equilibrating the osmotic value of the inside of the cell with the outside [5]. This is of course a very unfavourable situation for a bacterium. Under normal circumstances, it is of vital importance that the channels remain closed, for the bacterium to remain separated from its environment. For example, many cellular processes depend on the impermeability of the membrane for ions such as H^+, Na^+ and K^+. It can therefore be expected that the mechanosensitive channels remain fully closed until a critical pressure is reached, at which point they would have to open fast enough to prevent lysis of the cell.

It was found that an increase in the tension in the lipid bilayer triggered the opening of MscL, and it has even been shown that MscL is functional when reconstituted in synthetic lipids [20]. This important finding means that it does not need any other protein or cell-signalling pathway to function. Therefore, MscL is a very interesting protein to use in studies of protein–lipid interactions. There are also remaining questions on how exactly MscL itself

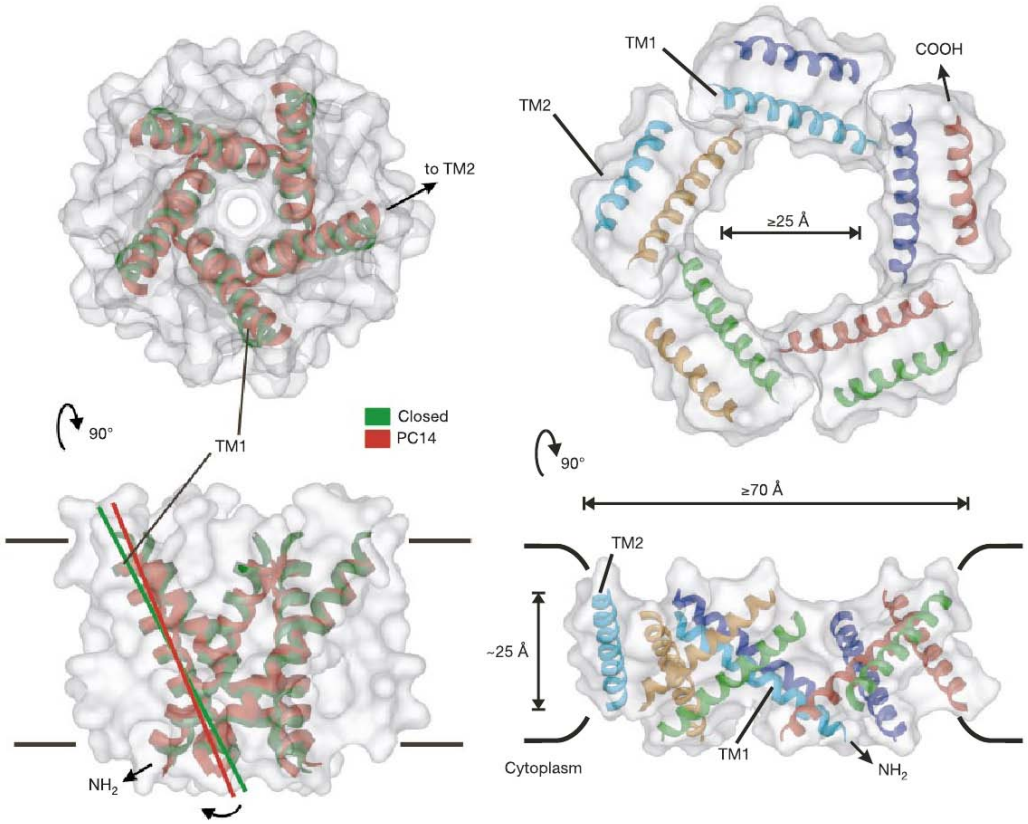

Figure 1.5: *A figure from Perozo et al. [51], showing a structural model of the closed state (left) and open state (right) of MscL as determined from ESR constraints.*

"senses" the changes in pressure that are transmitted to it by means of the phospholipid bilayer, and how its opening mechanism exactly works.

In 1998, the crystal structure of the closed form of MscL from *Mycobacterium tuberculosis* was solved by X-ray crystallography, with a resolution of 3.5 Å [6]. The crystal structure of MscS from *Escherichia coli* was elucidated in 2002, at a 3.9 Åresolution [1]. MscL was shown to be a homopentamer, with each subunit consisting of two transmembrane α-helices and a cytoplasmic α-helix. A structural model of the open form of MscL was derived from ESR distance constraints in 2002 [51], showing a pore size of at least 25 Å (see figure 1.5). The available structural information quickly led to the development of structure-based theoretical models of the channel gating mechanism [4, 31, 64, 65].

In the closed state of MscL, the five inner helices (TM1) from each of the

five homopentameric subunits form a small ($\sim 2\,\text{Å}$) hydrophobic pore that is impermeable to ions and water. Five outer helices (TM2) are in contact with the lipids. Multiple studies have addressed the opening mechanism [50], and not surprisingly many of them involve computer simulations [12, 14, 18, 43] because no experimental methods can directly follow the opening of the channel in atomic detail. A clear consensus on a model with a quantitative description of the channel gating and its intermediate states has so far not been established. Many of the currently existing theories on the opening mechanism of MscL and MscS are summarised in a recent review [8]. A recent publication by Powl et al. [60] suggests a direct interaction between MscL and its annular lipids to be responsible for the channel opening.

A study of the separate transmembrane helices of MscL revealed that TM1 creates ion channels when reconstituted in a membrane, even without the presence of TM2. Reconstitution of both helices together led to functional mechanosensitive channels with a conductance similar to native MscL, although it opened at lower pressure [49]. The outer helices alone did not form any pores or channels. These results also show that a single transmembrane helix may have intrinsic properties that reflect its function in the entire protein, which proves again that single transmembrane helices may be used as a model for a certain section of the entire protein, a method that is used in this thesis on a peptide derived from V-ATPase.

It is interesting to note that MscL from *Mycobacterium tuberculosis* does not contain any tryptophan or cysteine residues. This greatly facilitated cysteine scanning mutagenesis [62], site directed spin labelling [51] and tryptophan fluorescence studies [56]. From a functional point of view, the absence of the tryptophan residues in the headgroup region (where they are usually located in membrane proteins), permits the channel to exhibit a much larger helical tilt than would be possible if the protein were anchored to the headgroup region by Trp residues [7, 51].

The hydrophobic matching between MscL and the lipid bilayer is very efficient: lipids with a chain length of 14 or 24 carbon atoms did not lead to differences in the fluorescence of the residues in the interface region [58]. The binding constants of the lipids with different chain lengths were different; the most efficient binding was found for C16 lipids [56] with a hydrophobic thickness of about 24 Å, but because the differences were small, it is concluded that MscL distorts to match the changes in hydrophobic thickness of the bilayer. On the cytoplasmic side of the protein, the effect of different lipid chain lengths on lipid binding constants was bigger than on the periplasmic side, indicating that asymmetry in the membrane could be important for the MscL opening mechanism.

The interactions between MscL and its surrounding lipids have also been

studied by the quenching of fluorescence by brominated lipids. It was shown that a cluster of positively charged residues interacts with the negatively charged phospholipid headgroups. The residues involved in this interaction were identified as Arg-98, Lys-99 and Lys-100 [57]. This cluster occurs on the cytoplasmic side of the membrane, and it was found that the specific interaction with anionic phospholipids did not occur on the periplasmic side. Considering this, is interesting to note that MD simulations of channel opening have shown that the channel gated most easily when a force applied predominantly to the cytoplasmic side of MscL [18, 19].

In this thesis, the possibility to prepare oriented samples of MscL in lipid bilayers is explored, in order to study the interactions between MscL and lipids in a bilayer.

Bibliography

[1] R. B. Bass, P. Strop, M. Barclay, and D. C. Rees. Crystal structure of escherichia coli MscS, a voltage-modulated and mechanosensitive channel. *Science*, 298(5598):1582–1587, Nov 2002.

[2] B. Bechinger, C. Aisenbrey, and P. Bertani. The alignment, structure and dynamics of membrane-associated polypeptides by solid-state nmr spectroscopy. *Biochim Biophys Acta*, 1666(1-2):190–204, Nov 2004.

[3] H. W. Behrends, A. G. Beck-Sickinger, and G. Folkers. Evaluation of the secondary structure of vaccinia-virus thymidine kinase by circular-dichroism spectroscopy of overlapping synthetic peptides. *Eur J Biochem*, 241(1):126–132, Oct 1996.

[4] M. Betanzos, C.-S. Chiang, H. R. Guy, and S. Sukharev. A large iris-like expansion of a mechanosensitive channel protein induced by membrane tension. *Nat Struct Biol*, 9(9):704–710, Sep 2002.

[5] I. R. Booth, M. D. Edwards, S. Black, U. Schumann, and S. Miller. Mechanosensitive channels in bacteria: signs of closure? *Nat Rev Microbiol*, 5(6):431–440, Jun 2007.

[6] G. Chang, R. H. Spencer, A. T. Lee, M. T. Barclay, and D. C. Rees. Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. *Science*, 282(5397):2220–2226, Dec 1998.

[7] C.-S. Chiang, L. Shirinian, and S. Sukharev. Capping transmembrane helices of MscL with aromatic residues changes channel response to membrane stretch. *Biochemistry*, 44(37):12589–12597, Sep 2005.

[8] B. Corry and B. Martinac. Bacterial mechanosensitive channels: experiment and theory. *Biochim Biophys Acta*, 1778(9):1859–1870, Sep 2008.

[9] J. P. Cox, P. A. Evans, L. C. Packman, D. H. Williams, and D. N. Woolfson. Dissecting the structure of a partially folded protein. circular dichroism and nuclear magnetic resonance studies of peptides from ubiquitin. *J Mol Biol*, 234(2):483–492, Nov 1993.

[10] M. de Planque and J. Killian. Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and

interfacial anchoring (review). *Molecular Membrane Biology*, 20(4):271–284, 2003.

[11] M. R. de Planque, J. A. Kruijtzer, R. M. Liskamp, D. Marsh, D. V. Greathouse, R. E. Koeppe, B. de Kruijff, and J. A. Killian. Different membrane anchoring positions of tryptophan and lysine in synthetic transmembrane alpha-helical peptides. *J Biol Chem*, 274(30):20839–20846, Jul 1999.

[12] G. Debret, H. ValadiÃľ, A. M. Stadler, and C. Etchebest. New insights of membrane environment effects on MscL channel mechanics from theoretical approaches. *Proteins*, 71(3):1183–1196, May 2008.

[13] P. Dimroth, H. Wang, M. Grabe, and G. Oster. Energy transduction in the sodium F-ATPase of Propionigenium modestum. *Proc Natl Acad Sci U S A*, 96(9):4924–4929, Apr 1999.

[14] D. E. Elmore and D. A. Dougherty. Investigating lipid composition effects on the mechanosensitive channel of large conductance (MscL) using molecular dynamics simulations. *Biophys J*, 85(3):1512–1524, Sep 2003.

[15] D. N. Ganchev, D. T. Rijkers, M. M. Snel, J. A. Killian, and B. de Kruijff. Strength of intergration of transmembrane α-helical peptides in lipid bilayers as determined by atomic force spectroscopy. *Biochemistry*, 43: 14987–14993, 2004.

[16] P. G. Gillespie and R. G. Walker. Molecular basis of mechanosensory transduction. *Nature*, 413(6852):194–202, Sep 2001.

[17] M. Grabe, H. Wang, and G. Oster. The mechanochemistry of V-ATPase proton pumps. *Biophys J*, 78(6):2798–2813, Jun 2000.

[18] J. Gullingsrud and K. Schulten. Gating of MscL studied by steered molecular dynamics. *Biophys J*, 85(4):2087–2099, Oct 2003.

[19] J. Gullingsrud and K. Schulten. Lipid bilayer pressure profiles and mechanosensitive channel gating. *Biophys J*, 86(6):3496–3509, Jun 2004.

[20] C. C. Häse, A. C. L. Dain, and B. Martinac. Purification and functional reconstitution of the recombinant large mechanosensitive ion channel (MscL) of *escherichia coli*. *J Biol Chem*, 270(31):18329–18334, Aug 1995.

[21] M. Ø. Jensen and O. G. Mouritsen. Lipids do influence protein function — the hydrophobic matching hypothesis revisited. *Biochim Biophys Acta*, 1666(1-2):205–226, Nov 2004.

[22] S. K. Kandasamy, D.-K. Lee, R. P. R. Nanga, J. Xu, J. S. Santos, R. G. Larson, and A. Ramamoorthy. Solid-state nmr and molecular dynamics simulations reveal the oligomeric ion-channels of tm2-gaba(a) stabilized by intermolecular hydrogen bonding. *Biochim Biophys Acta*, 1788(3): 686–695, Mar 2009.

[23] P. M. Kane. The where, when, and how of organelle acidification by the yeast vacuolar H^+-ATPase. *Microbiol Mol Biol Rev*, 70(1):177–191, Mar 2006.

[24] M. Katragadda, J. L. Alderfer, and P. L. Yeagle. Assembly of a polytopic membrane protein structure from the solution structure of overlapping peptide fragments of bacteriorhodopsin. *Biophysical Journal*, 81:1029–1036, 2001.

[25] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Arg-735 of the 100-kDa subunit a of the yeast V-ATPase is essential for proton translocation. *Proc. Natl. Acad. Sci. U.S.A.*, 98(22):12397–12402, Oct 2001.

[26] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Proton translocation driven by ATP hydrolysis in V-ATPases. *FEBS Lett*, 545(1):76–85, Jun 2003.

[27] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Interacting helical surfaces of the transmembrane segments of subunits a and c′ of the yeast V-ATPase defined by disulfide-mediated cross-linking. *J Biol Chem*, 278 (43):41908–41913, Oct 2003.

[28] T. Kikkou, O. Matsumoto, T. Ohkubo, Y. Kobayashi, and G. Tsujimoto. NMR structure of an intracellular third loop peptide of human GABA(B) receptor. *Biochem Biophys Res Commun*, 366(3):681–684, Feb 2008.

[29] J. A. Killian. Synthetic peptides as models for intrinsic membrane proteins. *FEBS Lett*, 555(1):134–138, Nov 2003.

[30] J. A. Killian and T. K. M. Nyholm. Peptides in lipid bilayers: the power of simple models. *Curr Opin Struct Biol*, 16(4):473–479, Aug 2006.

[31] Y. Kong, Y. Shen, T. E. Warth, and J. Ma. Conformational pathways in the gating of Escherichia coli mechanosensitive channel. *Proc Natl Acad Sci U S A*, 99(9):5999–6004, Apr 2002.

[32] A. G. Lee. How lipids affect the activities of integral membrane proteins. *Biochim Biophys Acta*, 1666(1-2):62–87, Nov 2004.

[33] B. W. Lee, R. Faller, A. K. Sum, I. Vattulainen, M. Patra, and M. Karttunen. Sructural effects of small molecules on phospholipid bilayers investigated by molecular simulations. *Fluid Phase Equilib*, 228–229: 135–140, 2005.

[34] X. H. Leng, M. F. Manolson, Q. Liu, and M. Forgac. Site-directed mutagenesis of the 100-kDa subunit (Vph1p) of the yeast vacuolar (H^+)-ATPase. *J Biol Chem*, 271(37):22487–22493, Sep 1996.

[35] X. H. Leng, M. F. Manolson, and M. Forgac. Function of the COOH-terminal domain of vph1p in activity and assembly of the yeast V-ATPase. *J Biol Chem*, 273(12):6717–6723, Mar 1998.

[36] X.-H. Leng, T. Nishi, and M. Forgac. Transmembrane topography of the 100-kDa a subunit (Vph1p) of the yeast vacuolar proton-translocating atpase. *The Journal of Biological Chemistry*, 274:14655–14661, 1999.

[37] F. Marassi. NMR of peptides and proteins in oriented membranes. *Concepts in magnetic resonance*, 14(3):212–224, 2002.

[38] P. Marius, S. J. Alvis, J. M. East, and A. G. Lee. The interfacial lipid binding site on the potassium channel KcsA is specific for anionic phospholipids. *Biophys J*, 89(6):4081–4089, Dec 2005.

[39] P. Marius, M. Zagnoni, M. E. Sandison, J. M. East, H. Morgan, and A. G. Lee. Binding of anionic lipids to at least three nonannular sites on the potassium channel KcsA is required for channel opening. *Biophys J*, 94(5):1689–1698, Mar 2008.

[40] D. Marsh. Protein modulation of lipids, and vice-versa, in membranes. *Biochim Biophys Acta*, 1778(7-8):1545–1575, 2008.

[41] B. Martinac, M. Buechner, A. H. Delcour, J. Adler, and C. Kung. Pressure-sensitive ion channel in Escherichia coli. *Proc Natl Acad Sci U S A*, 84(8):2297–2301, Apr 1987.

[42] T. Meier, P. Polzer, K. Diederichs, W. Welte, and P. Dimroth. Structure of the rotor ring of F-type Na+-ATPase from Ilyobacter tartaricus. *Science*, 308(5722):659–662, Apr 2005.

[43] G. R. Meyer, J. Gullingsrud, K. Schulten, and B. Martinac. Molecular dynamics study of MscL interactions with a curved lipid bilayer. *Biophys J*, 91(5):1630–1637, Sep 2006.

[44] T. Murata, I. Yamato, Y. Kakinuma, A. G. W. Leslie, and J. E. Walker. Structure of the rotor of the V-type Na+-ATPase from Enterococcus hirae. *Science*, 308(5722):654–659, Apr 2005.

[45] A. C. Newton. Interaction of proteins with lipid headgroups: lessons from protein kinase C. *Annu Rev Biophys Biomol Struct*, 22:1–25, 1993.

[46] J. Norberg and L. Nilsson. Advances in biomolecular simulations: methodology and recent applications. *Q Rev Biophys*, 36(3):257–306, Aug 2003.

[47] T. K. M. Nyholm, S. Ozdirekcan, and J. A. Killian. How protein transmembrane segments sense the lipid environment. *Biochemistry*, 46(6): 1457–1465, Feb 2007.

[48] R. C. Page, C. Li, J. Hu, F. P. Gao, and T. A. Cross. Lipid bilayers: an essential environment for the understanding of membrane proteins. *Magn Reson Chem*, 45(S1):S2–S11, Dec 2007.

[49] K.-H. Park, C. Berrier, B. Martinac, and A. Ghazi. Purification and functional reconstitution of N- and C-halves of the MscL channel. *Biophys J*, 86(4):2129–2136, Apr 2004.

[50] E. Perozo. Gating prokaryotic mechanosensitive channels. *Nat Rev Mol Cell Biol*, 7(2):109–119, Feb 2006.

[51] E. Perozo, D. M. Cortes, P. Sompornpisut, A. Kloda, and B. Martinac. Open channel structure of MscL and the gating mechanism of mechanosensitive channels. *Nature*, 418(6901):942–948, Aug 2002.

[52] C. D. Pivetti, M.-R. Yen, S. Miller, W. Busch, Y.-H. Tseng, I. R. Booth, and M. H. Saier. Two families of mechanosensitive channel proteins. *Microbiol Mol Biol Rev*, 67(1):66–85, table of contents, Mar 2003.

[53] J. L. Popot and D. M. Engelman. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*, 29(17):4031–4037, May 1990.

[54] J. L. Popot and D. M. Engelman. Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem*, 69:881–922, 2000.

[55] A. M. Powl and A. G. Lee. Lipid effects on mechanosensitive channels. *Current Topics in Membranes*, 58:151–178, 2007.

[56] A. M. Powl, J. M. East, and A. G. Lee. Lipid-protein interactions studied by introduction of a tryptophan residue: the mechanosensitive channel MscL. *Biochemistry*, 42(48):14306–14317, Dec 2003.

[57] A. M. Powl, J. M. East, and A. G. Lee. Heterogeneity in the binding of lipid molecules to the surface of a membrane protein: hot spots for anionic lipids on the mechanosensitive channel of large conductance MscL and effects on conformation. *Biochemistry*, 44(15):5873–5883, Apr 2005.

[58] A. M. Powl, J. N. Wright, J. M. East, and A. G. Lee. Identification of the hydrophobic thickness of a membrane protein using fluorescence spectroscopy: studies with the mechanosensitive channel MscL. *Biochemistry*, 44(15):5713–5721, Apr 2005.

[59] A. M. Powl, J. M. East, and A. G. Lee. Anionic phospholipids affect the rate and extent of flux through the mechanosensitive channel of large conductance MscL. *Biochemistry*, 47(14):4317–4328, Apr 2008.

[60] A. M. Powl, J. M. East, and A. G. Lee. Importance of direct interactions with lipids for the function of the mechanosensitive channel MscL. *Biochemistry*, 47(46):12175–12184, Nov 2008.

[61] M. Renault, O. Saurel, J. Czaplicki, P. Demange, V. Gervais, F. LÃűhr, V. RÃľat, M. Piotto, and A. Milon. Solution state nmr structure and dynamics of kpompa, a 210 residue transmembrane domain possessing a high potential for immunological applications. *J Mol Biol*, 385(1): 117–130, Jan 2009.

[62] G. Shapovalov, R. Bass, D. C. Rees, and H. A. Lester. Open-state disulfide crosslinking between mycobacterium tuberculosis mechanosensitive channel subunits. *Biophys J*, 84(4):2357–2365, Apr 2003.

[63] L. Stryer. *Biochemistry*. W.H. Freeman and company, New York, 1998.

[64] S. Sukharev, M. Betanzos, C. S. Chiang, and H. R. Guy. The gating mechanism of the large mechanosensitive channel MscL. *Nature*, 409 (6821):720–724, Feb 2001.

[65] S. Sukharev, S. R. Durell, and H. R. Guy. Structural models of the MscL gating mechanism. *Biophys J*, 81(2):917–936, Aug 2001.

[66] S. I. Sukharev, P. Blount, B. Martinac, F. R. Blattner, and C. Kung. A large-conductance mechanosensitive channel in E. coli encoded by mscL alone. *Nature*, 368(6468):265–268, Mar 1994.

[67] S. I. Sukharev, P. Blount, B. Martinac, and C. Kung. Mechanosensitive channels of Escherichia coli: the MscL gene, protein, and activities. *Annu Rev Physiol*, 59:633–657, 1997.

[68] A. M. Tamburro, A. Pepe, and B. Bochicchio. Localizing alpha-helices in human tropoelastin: assembly of the elastin "puzzle". *Biochemistry*, 45(31):9518–9530, Aug 2006.

[69] P. C. van der Wel, N. D. Reed, D. V. Greathouse, and R. E. Koeppe. Orientation and motion of tryptophan interfacial anchors in membrane-spanning peptides. *Biochemistry*, 46(25):7514–7524, 2007.

[70] L. S. Vermeer, B. L. de Groot, V. Réat, A. Milon, and J. Czaplicki. Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with ^{2}H NMR experiments. *Eur Biophys J*, 36(8):919–931, Nov 2007.

[71] B. Vincent, L. Mouledous, B. Bes, H. Mazarguil, J.-C. Meunier, A. Milon, and P. Demange. Description of the low-affinity interaction between nociceptin and the second extracellular loop of its receptor by fluorescence and NMR spectroscopies. *J Pept Sci*, 14(11):1183–1194, Nov 2008.

[72] Y. Wang, T. Inoue, and M. Forgac. TM2 but not TM4 of subunit c″ interacts with TM7 of subunit a of the yeast V-ATPase as defined by disulfide-mediated cross-linking. *J Biol Chem*, 279(43):44628–44638, Oct 2004.

[73] Y. Wang, M. Toei, and M. Forgac. Analysis of the membrane topology of transmembrane segments in the c-terminal hydrophobic domain of the yeast vacuolar ATPase subunit a (vph1p) by chemical modification. *J Biol Chem*, 283(30):20696–20702, Jul 2008.

[74] I. M. Williamson, S. J. Alvis, J. M. East, and A. G. Lee. The potassium channel KcsA and its interaction with the lipid bilayer. *Cell Mol Life Sci*, 60(8):1581–1590, Aug 2003.

[75] R. Xue, S. Wang, H. Qi, Y. Song, S. Xiao, C. Wang, and F. Li. Structure and topology of slc11a1(164-191) with g169d mutation in membrane-mimetic environments. *J Struct Biol*, 165(1):27–36, Jan 2009.

[76] G. Yao, H. Feng, Y. Cai, W. Qi, and K. Kong. Characterization of vacuolar-ATPase and selective inhibition of vacuolar-H(+)-ATPase in osteoclasts. *Biochem Biophys Res Commun*, 357(4):821–827, Jun 2007.

2

Theory of the biophysical methods used

Contents

2.1 Tryptophan fluorescence of membrane proteins

The fluorescence emission of a tryptophan residue is very sensitive to the polarity of its environment. The emission maximum varies from 350 nm in a polar environment to 307 nm in an extremely apolar environment. In other words, a blue-shift of the tryptophan fluorescence is detected when the residue moves from a polar to an apolar environment.

In a lipid bilayer, a polarity profile exists, going from polar in the interface region to apolar in the bilayer centre (see Koehorst et al. [11] and references therein). This has been proven to be extremely useful to estimate the position of a tryptophan residue with respect to the middle of a bilayer, provided this residue is positioned on the outside of the membrane protein (exposed to the environment).

When a tryptophan residue is excited (usually with light at a wavelength of 280–295 nm), it shows an emission maximum around 350 nm when in contact with free water molecules. A blue-shift to 340 nm indicates that the tryptophan residue is in contact with bound water or other polar groups. An emission maximum at 330 nm is usually interpreted as a tryptophan in contact with polar groups inside a protein. A combination (superposition) between this spectrum and a spectrum with a maximum at 316-317 nm is considered a residue in an apolar environment. In the case of a bilayer, it indicates a tryptophan residue that is exposed to the apolar acyl chains of the lipids. More details may be found in publications that discuss in detail how to measure and analyse tryptophan fluorescence in membranes [12, 13].

The reader should note that other methods to estimate the depth of a tryptophan residue in a bilayer exist. Examples of such methods are the use of water soluble and membrane soluble fluorescence quenchers, or lipids that are brominated at specific positions. The quenchers have been applied previously to the KMTM7 peptide [7]. The interactions between different lipids and MscL has been studied extensively by the use of brominated lipids [18]. In this thesis, the blue-shift of the fluorescence was used to check the proper reconstitution of KMTM7 and MscL in lipid bilayers.

The fluorescence spectra can be analysed by fitting them to equation 2.1, adapted from Ladokhin et al. [13], to extract the spectral parameters $\lambda_{\max}$ and the width at half height.

$$I(\lambda) = \exp\left(-\frac{\ln 2}{\ln^2 \rho} \ln^2 \frac{a-\lambda}{a-\lambda_{\max}}\right) \tag{2.1}$$

Where $I_{\max}$ is the intensity at position $\lambda_{\max}$, and the spectral asymmetry ρ

is given by:

$$\rho = \frac{\lambda_{\max} - \lambda^-}{\lambda^+ - \lambda_{\max}} \tag{2.2}$$

The value of a is calculated as follows:

$$\mathrm{a} = \lambda_{\max} + \frac{(\lambda^+ - \lambda^-)\rho}{\rho^2 - 1} \tag{2.3}$$

where λ^- and λ^+ are the wavelengths where the intensity equals half of $\mathrm{I}_{\max}$. Fitting of four parameters ($\mathrm{I}_{\max}$, $\lambda_{\max}$, λ^-, λ^+) is required to obtain a unique solution.

2.2 Circular dichroism for secondary structure determination of proteins

Circular dichroism (CD) is a form of absorption spectroscopy that is often used for the determination of the secondary structure of proteins. The method is based on the fact that chiral (optically active) molecules absorb left-handed and right-handed circularly polarised light in different amounts, which contributes to an ellipticity of the measured signal. In the case of proteins, the ellipticity is sensitive to the relative orientation of two consecutive peptide bonds, and can thus be used to determine the secondary structure of a protein in solution. The measured ellipticity is usually expressed in molar ellipticity (degrees·M^{-1}·cm^{-1}), a unit that is independent of the path length through the sample and the sample concentration. To permit the comparison to different proteins, the signal is divided by the number of residues in the protein under study, giving the "per-residue molar ellipticity".

In figure 2.1 on the next page, spectra of proteins with 100% α-helical structure, 100% β-sheet structure and 100% random coil structure are shown. Spectra of α-helices have two minima: at 208 and 222 nm, and a maximum at 190 nm. A β-sheet has a minimum at 217 nm, the maximum is shifted to a slightly higher wavelength than in an α-helix and the spectrum is lower in intensity. A random coil structure displays only a very low intensity spectrum in the 210–250 wavelength range, and therefore mainly influences the intensity of the minima and maxima observed, which explains the sensitivity of the deconvolution methods to the sample concentration. It has a minimum at around 198 nm

The interpretation of CD spectra is usually done by deconvoluting the spectra with signals that represent a 100% α-helical, a 100% β-sheet and a 100% random coil structure. Databases with CD spectra from proteins of

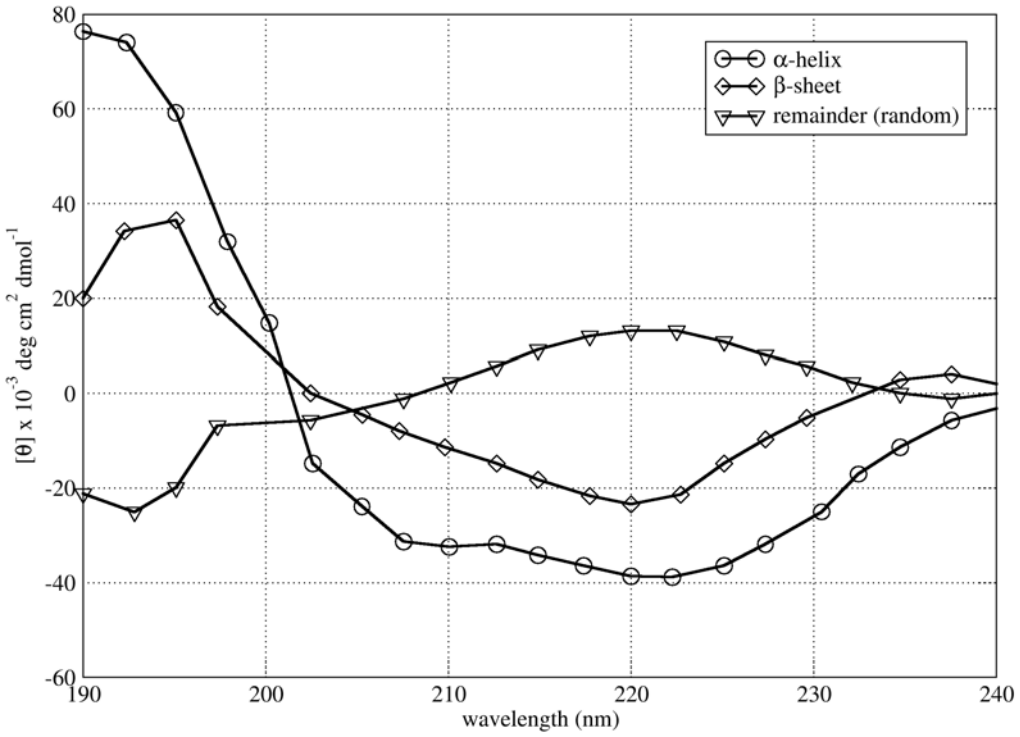

Figure 2.1: *CD spectra of 100% α-helix, 100% β sheet and 100% random coil structures. Data from Saxena and Wetlaufer [20], who used X-ray diffraction data in combination with experimental CD spectra to calculate the "pure" theoretical spectra for different secondary structures.*

known structure exist for this purpose. The well-known software package CDPro [23, 24] includes several different methods to carry out this deconvolution. It should be noted, however, that only a rough estimate of the secondary structure is given. Especially in the case of peptides the accuracy of these methods has been questioned, because peptides lack both long-range interactions and multiple twisted and packed secondary structure fragments [17]. The databases used for deconvolution of the spectra include only bigger proteins, where those interactions are present. Furthermore, the accuracy of the secondary structure quantification is very sensitive to the (accuracy of) the protein concentration in the sample, which may be difficult to assess precisely in the case of hydrophobic peptides due to micro-precipitation and/or adsorption to surfaces.

2.3 Structure determination of membrane proteins by (2D) liquid state ^{1}H NMR

In this work, nuclear Overhauser effect spectroscopy (NOESY) and total correlation spectroscopy (TOCSY) spectra are used to determine the structure of a peptide in detergent. Note that the theory described here is oversimplified, but a detailed discussion of NMR theory is beyond the scope of this thesis. For a thorough discussion of the principles of NMR, please refer to the literature. A particularly well-written book that covers both the basics and interesting details of modern liquid state NMR spectroscopy is "Spin dynamics" by Malcolm Levitt [14].

In NMR, the environment of atomic nuclei is studied by looking at their spin, a property of certain nuclei. These spins precess around the magnetic field vector, and the frequency of this precession can be measured. Because the atomic nuclei are present in different positions in a molecule, they experience different interactions with the surrounding atoms. Especially important is the presence of the electron clouds, because they shield the nuclei from the magnetic field. This shielding reduces the effective field at the site of the nucleus, which in turn reduces the precession frequency of the nuclear spin. This effect is called chemical shift, and can be used to distinguish most of the protons (or other nuclei such as ^{13}C or ^{15}N) in the different amino acids.

A second important influence on the spins is the so-called NOE. The NOE is a cross-relaxation between spins, caused by the dipolar coupling. Because the dipolar coupling is a through-space coupling, it does not depend on the number of bonds between atoms, only on their distance. As a consequence, it scales with r^{-6}, where r is the distance between the spins involved in the coupling[1] This effect makes it possible to measure internuclear distances, which are extremely valuable for use in a structure calculation.

A 25 residue peptide, such as the one that is studied in this thesis typically displays several hundreds of NOEs, and the only reasonable way to take these all into account in a structure calculation is by using a computer program that fits a model for the peptide structure to the distance restraints that are imposed by the NOESY data. Typically, a software package starts with an extended structure of the peptide, and uses a restrained molecular dynamics simulation to fold it into the secondary structure that best satisfies the experimental distance restraints. In a final step, the quality of the structure is checked, by calculating the number of violated NOE restraints, the root mean square deviation (RMSD) of the structures to the average

[1]Note that the NOE also depends on the internal dynamics of the spins in interaction. Molecular tumbling or exchange phenomena may influence the signal.

structure, the deviation of the dihedral angles from their expected values, and more.

Because several different programs are used from the NOE peak assignment to the structure calculation and validation, different formats are used, and some scripts are necessary to convert them from one format to the other. For future reference, a short tutorial on structure calculation "from Sparky to ARIA and procheck_nmr" was added to this thesis as appendix C on page 171.

2.4 Solid-state NMR

In an isotropic liquid, a molecule quickly samples all possible orientations due to its fast isotropic motion, which leads to identical nuclei in different molecules experiencing the same magnetic field. This is why in liquid-state NMR, the chemical shift does not depend on the orientation of the molecule with respect to the magnetic field. However, in (liquid-)crystalline samples or solids, each molecule has a certain orientation with respect to B_0, and its Larmor precession frequency therefore becomes dependent on its orientation. This means that molecules with a different orientation give a different signal, and the observed spectrum is a sum of the spectra from all the possible orientations. This effect is called the chemical shift anisotropy (CSA). The same complexity arises with the other anisotropic interactions, such as dipolar couplings and quadrupolar interactions.

If the sample is a powder, which means that each molecule has a different, random orientation, more molecules are oriented perpendicular to B_0 than parallel to B_0. This can easily be understood by imagining a sphere with a vertical z-axis through its middle and the origin in its centre of mass, and noting that the radius of the circle that lies in the xy-plane is highest at $z = 0$ and decreases to zero for increasing $|z|$. The number of molecules with a certain orientation with respect to the z-axis is proportional to this radius. The NMR spectrum of such a distribution of orientations is called a powder spectrum. Both the frequency and intensity of the signal depend on the angle between the molecule and the magnetic field (details in chapter 3 on page 39). A simulation of such a spectrum is shown in figure 2.2 on the facing page.

If the molecules in a sample are all oriented in the same direction, only one sharp peak is visible in the spectrum at the frequency corresponding to that particular orientation. Examples for a perfect orientation at 0°and 90°are also visible in figure 2.2 on the next page. In practice, the orientation

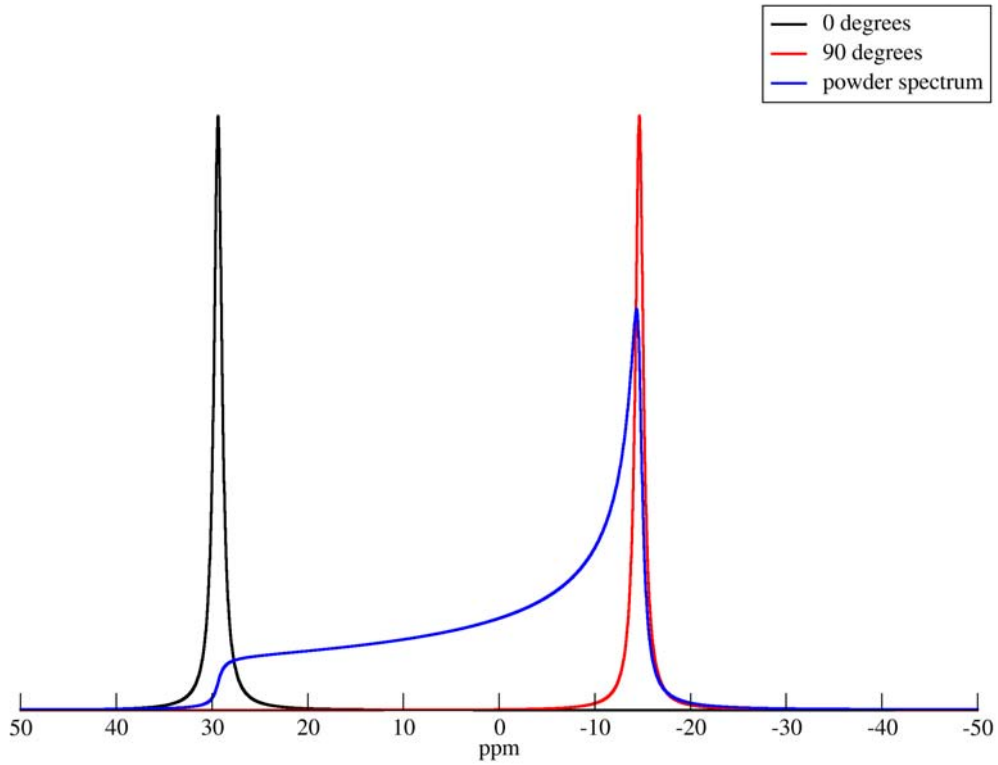

Figure 2.2: *Simulated ^{31}P spectra at 500 MHz of DOPC lipids (assuming a CSA of 44 ppm [8] and a line broadening of 500 Hz). The spectra at orientations of 0° and 90° are scaled down 5x to fit on the scale. In reality, the integral under these spectra is equal to that under the powder spectrum (assuming the number of molecules in the samples is the same). The isotropic chemical shift was set to 0. The spectra were simulated using the SIMPSON software [2].*

is never perfect: even in well-oriented samples, the mobility of lipids[2] leads to the simultaneous observation of multiple orientations, causing spectral broadening. Because the orientation and mobility both contribute to the final spectrum, it is useful to define a parameter that describes this "orientational mobility", the order parameter. The details of its definition are given in chapter 3 on page 39.

[2]The timescale of the mobility is important here, and determines the influence (if any) on the NMR spectrum.

2.4.1 Solid state NMR of quadrupolar nuclei

A spin 1/2 particle has a spherical charge distribution in the nucleus, and is therefore not affected by an electric field gradient. Nuclei with $I > 1/2$ have a non-spherical charge distribution, which causes an asymmetric electric field inside the nucleus, a quadrupole moment. These nuclei are called "quadrupolar". The nuclear spin energy levels of such an asymmetric nucleus are affected by electric field gradients. Such a gradient is mostly caused by the surrounding electrons. Because a nucleus is fixed in a molecule due to molecular bonding, and the shape of the electron cloud is a molecular property, the quadrupolar splitting is mainly caused by intramolecular interactions. It is sensitive to the magnetic field strength, and therefore depends on the orientation and mobility of the molecule with respect to B_0.

In isotropic liquid samples, the quadrupolar splitting is averaged to zero due to molecular motions. In solid-state samples of lipid bilayers however, the quadrupolar splitting becomes dependent on the orientation and mobility of the carbon–deuterium bond vector with respect to a reference axis. Because in general the mobility increases down the lipid acyl chain, the order parameter along the acyl chain decreases towards the bilayer centre. The lipid order parameters for the acyl chain carbon atoms can be calculated directly from the measured ^{2}H quadrupolar splitting, as is discussed in detail in chapter 3 on page 39.

2.5 Molecular dynamics simulations

MD simulations are computer simulations of interactions between molecules, calculated by solving the Newtonian equations of motion $\mathrm{F} = \mathrm{m} \cdot \mathrm{a}$ for all particles in the simulated system. The forces and masses are defined in a force field, and used to calculate the new acceleration vector of a molecule or atom, typically for intervals of 2 picoseconds. Although this means that many simplifications are made, it has often been shown that results from MD simulations are in good agreement with experimental data [6, 27]. For more information about the equations used and the approximations made, the GROMACS manual [25] contains an interesting introduction. Some good advice on how to setup and run MD simulations is given in a recent review article [10].

Many of the biological processes we want to study take place on timescales ranging from nanoseconds to seconds. For biological molecules such as desoxyribonucleic acid (DNA), proteins, peptides or lipid bilayers, solving the Schrödinger equation is out of the question, because with the currently avail-

able processor speeds it would take far too long for molecules that consist of more than several atoms. Although these so called "ab initio" calculations are very accurate and allow the simulation of breaking and forming chemical bonds, induced dipoles, movement of electrons, tunnelling, etc., a compromise has to be made between accuracy and processor time. Because the biological systems we want to study consist of many thousands of atoms, that often have to be simulated for tens of nanoseconds, a different, simplified method has to be used. MD simulations have proven to be very useful for this kind of work, including many examples of simulations of entire membrane proteins in lipids [15].

A great advantage of using MD simulations to interpret experimental data, is that the simulations can provide atomic detail that is not available from experiments. Another advantage of MD simulations is the fact that dynamics can be studied on picosecond to microsecond timescales, which is often difficult or impossible in laboratory experiments. But care should be taken that the results are not over-interpreted. It is tempting to believe that the beautiful pictures that MD simulations produce are the reality, but they may suffer from uncertainties in the force field and the relatively short simulation times. The only way to validate the results is by comparing the simulated results to experimental data, a fact that seems to be neglected too often, even in published papers in peer-reviewed journals. Unfortunately, this validation is not as easy as it seems. Simulations and experiments may have different timescales, and data obtained by these two methods can be fundamentally different. There is also a significant amount of uncertainty in some experimental data. The accuracy of MD simulations for predicting lipid order parameters is reviewed in chapter 3 on page 39.

2.5.1 Topology files

An important consideration when preparing for MD simulations may be the availability of topology files for the molecule of interest in the chosen force field. Topology files contain all the necessary information to describe a molecule, such as bond lengths and angles. Particularly important are the partial charges, because they can be rather difficult to predict for new molecules. It is vital to use the best possible topology files to do meaningful simulations. If topology files are not available from previously published simulations, they have to be designed in a way that is consistent with the rest of the force field [1]. This usually means creating the files with the same methods that were used for the original topology files. Different packages for MD simulations often use different force fields, and therefore different ways the topology files should be generated. Converting them from one force

field to another is usually not recommended, although examples exist in the literature.

In GROMACS [26], topology files are generally derived in a modular fashion, from existing building blocks. Ideally, the topology files are verified against experimental data such as density, evaporation enthalpy and/or critical micelle concentration (CMC) in the case of micelles. However, in the recent literature a detailed verification is often omitted, because it is argued that a combination of two or more building blocks that behave correctly separately, will also yield reliable results when combined to form a new molecule. In the case of detergents, the correct self-assembly of micelles and an aggregation number that agrees with experimental data are often used to validate the results. To give an example: in a paper by Löw et al. [16] the prodrg2 server [21] was used to generate the molecular structure of SDS, but the partial charges were copied from the CHARMM force field.

The antechamber package allows the calculation of partial charges and the generation of topology files that can be used in conjunction with the GAFF force field and the AMBER software package. Different methods to calculate the partial charges are available in antechamber, such as ab-initio calculations and AM1-BCC [9]. AM1-BCC has been shown to approximate the partial charges from ab-initio calculations, with a significantly reduced computation time. The use of topology files calculated with AM1-BCC is well established (there are more than 100 citations of the AM1-BCC papers). The antechamber package was specifically designed to generate topology files for use in AMBER simulations.

2.5.2 MD simulations of peptides and micelles

It has been shown in several papers that self-assembly from a random distribution can be used to construct lipid bilayers [5, 22] and detergent micelles [3, 19]. The advantages of this method are that no artificial starting conformation is imposed, but the obvious disadvantage is the longer time needed to equilibrate the system. It has been shown that the final result after auto-assembly of micelles around membrane proteins gives a result that is very similar to simulations with preformed micelles [3]. But in a paper by Löw et al. [16] different starting conformations were used to investigate the position of a peptide with respect to an SDS micelle, because they did not obtain equilibrated micelles by auto-assembly, even after 40 ns. In their paper, peptides that were put inside the micelle at the beginning of the simulation adopted a position on the micelle surface after 50 ns, in agreement with their experimental data.

Examples exist of peptides that interact differently with different micelles.

A difference between the location of a peptide in SDS and DPC micelles has for example been shown [4]. In this thesis, we use auto-assembly of detergent around a peptide to study the solvation of this peptide by SDS and octylglucoside micelles.

Bibliography

[1] GROMACS wiki: Parameterization. `http://wiki.gromacs.org/index.php/Parameterization`.

[2] M. Bak, J. T. Rasmussen, and N. C. Nielsen. SIMPSON: a general simulation program for solid-state NMR spectroscopy. *J Magn Reson*, 147(2):296–330, Dec 2000.

[3] P. J. Bond, J. M. Cuthbertson, S. S. Deol, and M. S. P. Sansom. Md simulations of spontaneous membrane protein/detergent micelle formation. *J Am Chem Soc*, 126(49):15948–15949, Dec 2004.

[4] S. Bourbigot, E. Dodd, C. Horwood, N. Cumby, L. Fardy, W. H. Welch, Z. Ramjan, S. Sharma, A. J. Waring, M. R. Yeaman, and V. Booth. Antimicrobial peptide RP-1 structure and interactions with anionic versus zwitterionic micelles. *Biopolymers*, 91(1):1–13, Jan 2009.

[5] S. Esteban-Martín and J. Salgado. Self-assembling of peptide/membrane complexes by atomistic molecular dynamics simulations. *Biophys J*, 92 (3):903–912, Feb 2007.

[6] S. Esteban-MartÃŋn and J. Salgado. The dynamic orientation of membrane-bound peptides: bridging simulations and experiments. *Biophys J*, 93(12):4278–4288, Dec 2007.

[7] R. W. Hesselink, R. B. Koehorst, P. V. Nazarov, and M. A. Hemminga. Membrane-bound peptides mimicking transmembrane Vph1p helix 7 of yeast V-ATPase: a spectroscopic and polarity mismatch study. *Biochimica et Biophysica Acta*, 1716(2):137–145, 2005.

[8] G. P. Holland, S. K. McIntyre, and T. M. Alam. Distinguishing individual lipid headgroup mobility and phase transitions in raft-forming lipid mixtures with ^{31}p MAS NMR. *Biophys J*, 90(11):4248–4260, Jun 2006.

[9] A. Jakalian, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J Comput Chem*, 23(16):1623–1641, Dec 2002.

[10] C. Kandt, W. L. Ash, and D. P. Tieleman. Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, 41(4): 475–488, Apr 2007.

[11] R. B. M. Koehorst, R. B. Spruijt, F. J. Vergeldt, and M. A. Hemminga. Lipid bilayer topology of the transmembrane alpha-helix of M13 major coat protein and bilayer polarity profile by site-directed fluorescence spectroscopy. *Biophys J*, 87(3):1445–1455, Sep 2004.

[12] A. S. Ladokhin. *Fluorescence Spectroscopy in Peptide and Protein Analysis*, pages 5762–5779. John Wiley & Sons Ltd, Chichester, 2000.

[13] A. S. Ladokhin, S. Jayasinghe, and S. H. White. How to measure and analyze tryptophan fluorescence in membranes properly, and why bother? *Anal Biochem*, 285(2):235–245, Oct 2000.

[14] M. H. Levitt. *Spin dynamics*. John Wiley & Sons, Ltd, 2001.

[15] E. Lindahl and M. S. P. Sansom. Membrane proteins: molecular dynamics simulations. *Curr Opin Struct Biol*, 18(4):425–431, Aug 2008.

[16] C. Löw, U. Weininger, H. Lee, K. Schweimer, I. Neundorf, A. G. Beck-Sickinger, R. W. Pastor, and J. Balbach. Structure and dynamics of helix-0 of the N-BAR domain in lipid micelles and bilayers. *Biophys J*, 95(9):4315–4323, Nov 2008.

[17] B. C. Poschner, J. Reed, D. Langosch, and M. W. Hofmann. An automated application for deconvolution of circular dichroism spectra of small peptides. *Anal Biochem*, 363(2):306–308, Apr 2007.

[18] A. M. Powl, J. N. Wright, J. M. East, and A. G. Lee. Identification of the hydrophobic thickness of a membrane protein using fluorescence spectroscopy: studies with the mechanosensitive channel MscL. *Biochemistry*, 44(15):5713–5721, Apr 2005.

[19] M. Sammalkorpi, M. Karttunen, and M. Haataja. Structural properties of ionic detergent aggregates: a large-scale molecular dynamics study of sodium dodecyl sulfate. *J Phys Chem B*, 111(40):11722–11733, Oct 2007.

[20] V. P. Saxena and D. B. Wetlaufer. A new basis for interpreting the circular dichroic spectra of proteins. *Proc Natl Acad Sci U S A*, 68(5): 969–972, May 1971.

[21] A. W. Schüttelkopf and D. M. F. van Aalten. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 8):1355–1363, Aug 2004.

[22] K. A. Scott, P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid, and M. S. P. Sansom. Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure*, 16(4):621–630, Apr 2008.

[23] N. Sreerama and R. W. Woody. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem*, 287(2):252–260, Dec 2000.

[24] N. Sreerama and R. W. Woody. On the analysis of membrane protein circular dichroism spectra. *Protein Science*, 13(1):100–112, Jan 2004.

[25] D. van der Spoel, E. Lindahl, and B. Hess. *GROMACS User Manual Version 3.3*. Department of Biophysical Chemistry, University of Groningen, The Netherlands.

[26] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible and free. *J Comput Chem*, 26:701–1719, 2005.

[27] L. S. Vermeer, B. L. de Groot, V. Réat, A. Milon, and J. Czaplicki. Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with ^{2}H NMR experiments. *Eur Biophys J*, 36(8):919–931, Nov 2007.

3

Mini-review: Acyl chain order parameter profiles in lipid bilayers: computation from molecular dynamics simulations and comparison with ^{2}H NMR experiments

Contents

Summary

Order parameters from deuterium NMR are often used to validate or calibrate molecular dynamics simulations. This paper gives a short overview of the literature in which experimental order parameters from ^{2}H NMR are compared to those calculated from MD simulations. The different ways in which order parameters from experiment are used to calibrate and validate simulations are reviewed. In the second part of this review, a case study of cholesterol in a DMPC bilayer is presented. It is concluded that the agreement between experimental data and simulation is favourable in the hydrophobic region of the membrane, for both the phospholipids and cholesterol. In the interfacial region the agreement is less satisfactory, probably because of the high polarity of this region which makes the correct computation of the electrostatics more complex.

This chapter (with minor changes) has been published as:
L.S. Vermeer, B.L. de Groot, V. Réat, A. Milon and J. Czaplicki, Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with ^{2}H NMR experiments. *European Biophysics Journal*, 36:919–931, November 2007.

3.1 Introduction

In the latest decades, research on membranes and membrane proteins has been growing steadily [59, 61, 103]. Lipid bilayers in the liquid crystalline phase, also known as the L_α or liquid disordered (L_d) phase, are generally considered to be relevant for models of biological membranes. The liquid ordered (L_o) phase is described as an intermediate state between the L_d and the L_β (gel) phase, usually associated with bilayers containing cholesterol. For a review on the properties of a fluid lipid bilayer see Bloom et al. [12], for lipid–cholesterol interactions see Ohvo-Rekilä et al. [69]. An extensive review of experimentally determined structural data of fully hydrated bilayers for many different lipids is given in a paper by Nagle and Tristram-Nagle [67]. Lipids can influence the structure and function of membrane proteins and in turn membrane proteins have an influence on the lipid dynamics [42, 49, 50, 60]. Therefore, it is important to be able to characterise the structure and dynamics of a lipid bilayer.

First of all, one should realise that there is no such thing as an atomic level structure of a bilayer in the liquid crystalline phase. The appropriate description of a bilayer is given by broad statistical distribution functions [79]. In addition to experiment, such distributions can be evaluated by molecular dynamics (MD) simulations. Simulations can also be used to provide additional information on an atomic level, in a resolution that cannot be obtained experimentally. However, simulations may suffer from uncertainties in the force field and relatively short simulation times. Since simulations result from a modelling of the actual interactions and forces in a bilayer, they should be validated by comparing to experimental data [67, 102]. Unfortunately, this is not as easy as it seems. Simulations and experiments have different timescales, and data obtained by these methods are fundamentally different [7]. Also, there is a significant amount of uncertainty in some experimental data [67].

The properties of a lipid bilayer are commonly described using the area per lipid, bilayer thickness and acyl chain order parameters. These properties are closely related, and a considerable amount of research has been dedicated to find the equations that relate them to each other [52, 67, 79, 80].

Lipid acyl chain order parameters are obtained easily from deuterium NMR quadrupole splittings. They can be directly measured by NMR [87], are model free [80], and reliable (reproducible). Deuterium NMR determination of acyl chain order parameter profiles has been reported for multi- and unilamellar vesicles [70, 81], bicelles [5] and lipid bilayers oriented at 90° with respect to the magnetic field [58], on either uniformly or specifically labelled lipids. Oriented bilayers provide the best resolution and direct access

to quadrupole splittings, while vesicles are easier to manipulate and can also provide quadrupole splittings at 90° with a good resolution, especially when a dePakeing algorithm [47, 62] is used.

This mini-review will focus on the comparison between experimental order parameters from deuterium NMR and those calculated from molecular dynamics simulations. It is not in the scope of this mini review to cover all the aspects of MD simulations of bilayers. Reviews on this subject can be found elsewhere [73, 84, 102]. However, some remarks about the choice of MD parameters will be made. We will also review the use of order parameters for calculation of the area per lipid, a property that is often used to check the equilibration of a simulation. In the final part of this mini-review, a case study of lipid-cholesterol interactions that was carried out in our lab will be presented.

3.2 Order parameters of the lipid chains

Lipids in a fluid bilayer are highly dynamic. Many movements on different timescales take place: rotation around chemical bonds and trans/-gauche isomerisations (picoseconds), rotation (axial diffusion) around the lipid axis (nanoseconds), wobbling (nanoseconds), lateral diffusion (microseconds), flip-flop across the bilayer (milliseconds–seconds) and undulations of the membrane (milliseconds–seconds). Most of these movements influence the order parameters of the acyl chains. It is of vital importance to compare the timescale of these movements to the timescale of the experimental method (about 10 μs for ^{2}H NMR). The characteristic timescale for ^{2}H NMR is generally considered to be long enough to sample the dynamic processes in the membrane that influence the order parameters.

Nowadays, most molecular dynamics simulations are done with small systems (several hundreds of lipids), for relatively short times (typically ranging from tens to hundreds of nanoseconds). Although these times allow for some lipid movement, the bilayer size of these systems is too small to adequately sample undulations and the time is too short for lipid lateral diffusion and slower processes to take place. Some longer simulations [107], and simulations of bilayer undulations in bigger systems [38, 54, 57] have been reported to investigate these undulations and/or obtain information on lipid lateral diffusion.

Lipid order parameters are a measure for the orientational mobility of the C–D bond and are defined as

$$S = \left\langle \frac{3\cos^2\Theta - 1}{2} \right\rangle \tag{3.1}$$

where Θ is the (time dependent) angle between the C–D bond vector and a reference axis. The angular brackets denote a time and ensemble average.

Order parameters do not only depend on the (dis)order of the system, but also on orientation. For example, an order parameter of zero can be either an unordered (isotropic) system or a perfectly ordered system oriented at the magic angle of 54.7° with respect to the magnetic field. A value of -0.5 indicates a perfectly ordered acyl chain in all-trans conformation, rapidly rotating around the bilayer normal. A complete description of lipid order parameters is given by Seelig [86]. In general, order decreases from the interface region to the bilayer centre.

3.2.1 Calculation of lipid order parameters from solid-state deuterium NMR spectra

It was first shown by Oldfield et al. [70] that deuteron resonance can be used to study the chain mobility in lipid systems. A detailed explanation of the physics involved is given in a review article by Seelig [86]. The order parameters of the C–D bond can be calculated directly from experimentally determined quadrupolar splittings $\Delta\nu_Q$ (equation 3.2a). A_Q is the quadrupolar coupling constant $A_Q = e^2qQ/h$, equal to 168 kHz for a C–D bond in lipid acyl chains [14]. Order parameters reflect the average orientation of internuclear C–D vectors with respect to the direction of the external magnetic field. The averaging occurs as a consequence of molecular rotations about different axes. In bilayers in a fluid phase, the lipids are rapidly rotating around the bilayer normal, which is thus taken as the reference axis. If the movements are independent, the general expression for order parameters can be factored into a product of individual contributions (figure 1 and equation 3.2b–3.2h). These individual contributions describe the orientation of the bilayer normal with respect to the magnetic field ($S_{bilayer}$), the fluctuations of the bilayer normal (S_{coll}), reorientation of the molecular director with respect to the bilayer normal (S_{mol}), and reorientation of the C–D vector with respect to the molecular director (S_{intra}). Note that a different usage of the symbol S_{mol} can be found in the literature, where it is used to describe the order parameter of the vector perpendicular to the CD_2 plane [87, 94].

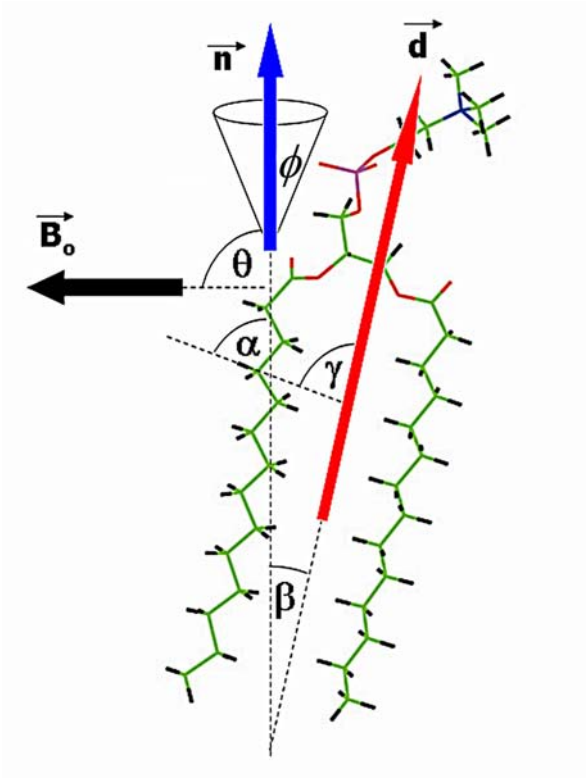

Figure 3.1: *Visual representation of different contributions to the observed order parameter S (Eq. 3.2a). Fast rotations of lipids about the vector normal to the bilayer permit to separate the contribution to the order parameter due to this movement from the contribution due to the overall positioning of the oriented bilayer with respect to the external magnetic field (Eq. 3.2b). The bilayer normal, symbolised by the vector $\vec{n}$, is oriented at an angle θ to the external magnetic field $\vec{B}_0$ (Eq. 3.2c). The observed CD vector is tilted at an angle α with respect to the bilayer normal (Eq. 3.2d). If the rotations of lipids about different axes are independent, the order parameter may be represented as the product of all individual contributions (Eq. 3.2e). The bilayer normal is subject to fluctuations (undulations), shown here as the angle ϕ (Eq. 3.2f). The molecular director $\vec{d}$ is at an angle β to the bilayer normal (Eq. 3.2g). The observed CD vector is at an angle γ with respect to the molecular director (Eq. 3.2h).*

$$\Delta\nu_Q = \frac{3}{2} A_Q \cdot S \tag{3.2a}$$

$$S = S_{\text{bilayer}} \cdot S_{\text{CD}} \tag{3.2b}$$

$$S_{\text{bilayer}} = \frac{3cos^2\theta - 1}{2} \tag{3.2c}$$

$$S_{\text{CD}} = \left\langle \frac{3cos^2\alpha - 1}{2} \right\rangle \tag{3.2d}$$

$$S_{\text{CD}} = S_{\text{coll}} \cdot S_{\text{mol}} \cdot S_{\text{intra}} \tag{3.2e}$$

$$S_{\text{coll}} = \left\langle \frac{3cos^2\phi - 1}{2} \right\rangle \tag{3.2f}$$

$$S_{\text{mol}} = \left\langle \frac{3cos^2\beta - 1}{2} \right\rangle \tag{3.2g}$$

$$S_{\text{intra}} = \left\langle \frac{3cos^2\gamma - 1}{2} \right\rangle \tag{3.2h}$$

3.2.2 Calculating order parameters from MD simulations

From simulations, when averaging by fast axial rotation is assumed, the order parameter S_{CD} is calculated according to equation 3.1. If a united atom force field is used (without hydrogen or deuterium atoms), the C–D bond vector needs to be reconstructed. To do so, the $C_{(i-1)} - C_{(i+1)}$ vector is usually taken to be the z-axis. The x- and y-axis are defined perpendicular to the z-axis and to each other, with the y-axis in the $C_{(i-1)} - C_{(i)} - C_{(i+1)}$ plane. Using this definition, $S = (2/3)S_{xx} + (1/3)S_{yy}$ and can be compared directly to ^{2}H-NMR data [26]. For lipids rotating rapidly around the z-axis, $S_{xx} = S_{yy}$ and $S_{xx} + S_{yy} + S_{zz} = 0$ [78], hence the equation becomes $S = -S_{zz}/2$. Details on this so called recursion relation can be found in Douliez et al. [24] and references therein. It is important to realise that this calculation only holds when the deuterons on a methylene are equivalent. This could be a problem in short (ps–ns) or small scale simulations where time and ensemble averaging are insufficient. One should also note that it is not valid around unsaturated bonds due to their different geometry.

It might be more useful to compare C_n–$C_{(n-1)}$ order parameters (S_{CC}), calculated from experimentally measured S_{CD} to simulated ones, because S_{CC} provides more information on the chain conformations [90].

3.3 Calculation of bilayer properties from order parameters

An extensive review of experimentally determined structural data of fully hydrated bilayers for many different lipids is given in a paper by Nagle and Tristram-Nagle [67].

3.3.1 Area per lipid, bilayer thickness, elasticity

Order parameters can be determined accurately and reproducibly by ^{2}H NMR. There is little disagreement between their values measured at different times, in different laboratories. However, there still is considerable uncertainty in the area per lipid calculated from experimentally determined order parameters [67]. Attempts have been made to improve this [80], but it is concluded that NMR order parameters can not (yet) quantitate the bilayer structure [79]. Nevertheless, the area per lipid in the tail region and the methylenic travel distance can be estimated with reasonable accuracy [28, 79]. It has been suggested that due to a difference in timescales, order parameters measured by NMR might not be the same as the ones calculated from MD simulations [79].

Based on ^{2}H NMR measurements of different lipids, at different temperatures, with and without cholesterol, Lafleur et al. [48] showed that localised changes at specific positions along the normal of the bilayer do not lead to localised changes in order parameters, but affect the entire curve. The length of the hydrocarbon chain and the hydrophobic thickness can be calculated from the average order parameter S_{CD} (Bloom et al. [12], Douliez et al. [23], Smondyrev and Berkowitz [90] and references therein). In a paper by Henriksen et al. [36] about the effect of sterols on POPC membranes, the first moment of the acyl chain order parameters is correlated with membrane rigidity and area expansion modulus. The elastic properties (compressibility) are also related to the order parameter, as was hypothesised by Bloom et al. [12] and confirmed by Henriksen et al. [36].

3.3.2 Differences between the sn-1 and sn-2 lipid chains

A correlation of the orientation of a lipid chain with that of its nearest neighbours has been demonstrated by Takaoka et al. [99], which is a good example for a property that is available from MD simulations, but has never been measured experimentally. Based on their simulations, de Vries et al. [22] suggest a lateral correlation length of 3–5 nm for hydrated liquid crystalline DPPC.

In a 2 nanosecond simulation by Shinoda et al. [88], it was shown that the sn-1 and sn-2 chain of DPPC are parallel to each other at 353 K. Given that the lipid headgroup is aligned (almost) parallel to the bilayer surface, the sn-2 chain has to be bent. It was indeed found that the gauche-ratio of the sn-2 C2–C3 bond was 30 %. However, this value is still too low to explain why the chains are parallel. Another effect contributing to the bending of the sn-2 chain is the fact that it was found to favour one specific stable conformation, while the conformations of the sn-1 chains were equally distributed over two stable states. In the same paper [88], the authors use NMR relaxation data to show that the disordered conformation of the tail carbons is due to movement, and not orientation. The calculated order parameter profile reproduces the shape of the profile from deuterium NMR, but the calculated profile is shifted to higher values. The observation of qualitative, but not quantitative agreement has been reported often: the simulated order parameters are usually lower than the experimental ones.

A study on the differences between the sn-1 and sn-2 chain by Smondyrev and Berkowitz [90] reveals that the sn-1 tail in DPPC is more ordered in the beginning of the chain, while the sn-2 chain is more ordered at its end. They also noted a stronger "odd–even" effect in the sn-2 chains, which is likely to be a result of the chain's tilt.

To explain the unusual quadrupolar splitting of C2 in the sn-2 chain as was reported by Seelig and Seelig [85], a histogram of instantaneous order parameters rather than an average over time is presented in a paper by Takaoka et al. [99]. In the histogram, a second peak was observed for β-C2, which possibly explains the experimentally observed splitting.

3.4 Parameters affecting molecular dynamics simulations

The importance of carefully choosing the parameters for an MD simulation has been stressed in many papers on MD methodology [1, 29, 63, 102]. A review on MD methodology by Norberg and Nilsson [68] describes the current state of MD simulations.

There has been a lively discussion on the ensemble that should be chosen and on the correct treatment of electrostatic interactions. Nowadays, most authors agree upon the use of particle mesh Ewald (PME) [27] to treat the electrostatics [76, 77] because the area per lipid and the other properties of bilayers are very sensitive to the electrostatic cut-off distance [1]. However, simulations by Zaraiskaya and Jeffrey [108] show no significant difference

between PME and a twin-range cut-off model. The authors suggest two possible reasons for this apparent contradiction: they simulate a larger system (1024 lipids), for a relatively short time (4 ns). It should be mentioned that other lattice-sum methods exist, but they seem to be used less frequently than PME. Therefore, and because using a cut-off distance for electrostatic interactions is generally considered "bad" this review will focus on PME simulations, even though it has been demonstrated that PME can introduce artificial periodicity (Tobias [104] and references therein). Note that most force-fields have been developed before it was common practice to employ PME. Next-generation force fields can be expected to address this issue.

As for the ensemble, a constant number of molecules, constant pressure and constant temperature (NPT) are most often used. In these simulations, the size of the simulation box is allowed to fluctuate in the xy-plane and z-direction separately (semi-isotropic), or all three dimensions are allowed to fluctuate independently (anisotropic). For simulations of systems that are large enough both choices are correct [1]. Ensembles with constant area per lipid (NPAT) or constant surface tension (NPγT) have been used as well. Although NPAT simulations tend to agree better with experimental data [15], they require as input an accurate value for the area per lipid from experiment, which is often a problem [67]. Note that some force-fields require NPAT simulations, as simulations carried out with these force-fields are known to yield unrealistic values for the area per lipid [42, 43]. In NPγT simulations, the area per lipid is sensitive to the surface tension γ [30, 57].

Constant volume simulations (NVT, NVE) often lead to artifacts [102]. However, it was shown by Suits et al. [95] that after an equilibration of the density by an NPAT simulation, constant volume simulations can be carried out, although the authors note that the confidence in their method rests on both the confidence in the experimental data used to calibrate the simulation and the strength of the relation between order and surface area. An investigation of the effect of system size under constant volume conditions was carried out by de Vries et al. [22], showing that 16 lipids per leaflet already reproduce many experimental properties after 3 nanoseconds, including the order parameters.

Several different force fields (and adaptations thereof) exist. Obviously, the force field affects the properties of the simulated system. An analysis of different force fields for different applications is given in a review by Norberg and Nilsson [68].

3.4.1 Setting up the system

In a paper by Suits et al. [95], the average NMR order parameter was used to set up the simulation parameters. The authors were able to reproduce the experimental order parameters for all atoms along the acyl chain. This seems to contradict the conclusion by Sternin et al. [93], who recommend using the shape of the order parameter profile.

Another approach was taken by Feller et al. [31]: four MD simulations (NPAT) with a different area per lipid were carried out, and the order parameters and electron densities were compared to experimental data. The simulation with the order parameters and electron density that matched the experiment were considered to have the best value for the area per lipid. This led to a value of 62.9 $Å^2$, close to the value of 64 $Å^2$ from Nagle and Tristram-Nagle [67]. Because Feller et al. [31] tested for values of 59.3, 62.9, 65.5 and 68.1 $Å^2$, this can be considered a good agreement between simulation and experiment.

A simulation by Takaoka et al. [99] was started from a crystal structure of DMPC. To pass the activation energy barrier of the phase transition to the L_α phase, the authors raised the temperature from 310 to 510 K for only 20 picoseconds, and then slowly lowered it to 310 K, leading to an average lipid chain tilt angle close to the experimental value of 0°. After additional equilibration at 360 K, the ^{2}H NMR order parameters from Seelig and Seelig [85] were successfully reproduced after 2.6 ns. Various other properties needed different times to equilibrate. The only way the authors could reproduce the experimental data while using PME was by choosing a constant surface-tension ensemble and an all-atom model.

In this work we have largely neglected the effect of undulations on calculated order parameters from the simulations, as in most simulations the membrane patches only allow minor undulations due to their limited size. Recent simulations [54, 57] have shown that significantly larger patch sizes are required to model undulations accurately, and that undulations take a long time to develop. It is unlikely, however, that the suppressed undulations due to this approximation explain the reported differences between experimental and simulated order parameters, as these differences mainly correspond to simulation order parameters being lower than the experimental ones.

3.5 Comparison of simulation and experiment

Table 3.1 on the next page presents a selection from the literature where NMR data and MD simulations are compared.

Table 3.1: *Selection of publications where order parameters from molecular dynamics simulations are compared to experimental order parameters from ^2H NMR. Note that the description (in the second column) does not necessarily correspond to the main subject of the article, but rather gives a description of the use of order parameters in the cited publication.*

lipid(s)	use of order parameter	reference
DMPC	comparison of DMPC (sim) to DPPC (exp), histogram of instantaneous order parameters	Takaoka et al. [99]
DMPC	decomposition of S_{CD} into S_{coll} Sintra and S_{mol}	Aussenac et al. [5]
DMPC	effect of cholesterol and ergosterol	Czub and Baginski [19]
DMPC	effect of SDS	Bandyopadhyay et al. [6]
DMPC	influence of hydration and temperature	Högberg and Lyubartsev [39]
DMPC	setting up the initial bilayer configuration	Tang et al. [100]
DPhPC	comparison of simulated and experimental average order parameter	Husslein et al. [40]
DPPC	Brownian dynamics, different field potentials	Pastor et al. [74]
DPPC	comparison of NMR relaxation data to simulated order parameters	Wohlert and Edholm [107]
DPPC	comparison of simulation and experiment	Egberts et al. [26]
DPPC	comparison of simulation and experiment	Tieleman et al. [102]
DPPC	differences between sn-1 and sn-2 chain	Smondyrev and Berkowitz [90]
DPPC	effect of cholesterol	Hofsäß et al. [38]
DPPC	effect of DMSO	Sum and de Pablo [96]
DPPC	effect of DMSO	Smondyrev and Berkowitz [89]
DPPC	effect of force-field parameters	Chandrasekhar et al. [16]
DPPC	effect of halothane	Koubi et al. [46]
DPPC	influence of different (constant) surface areas	Feller et al. [31]
DPPC	influence of system size and simulation time	de Vries et al. [22]
DPPC	temperature and disaccharide concentration dependence	Sum et al. [97]
DPPC	trans/gauche ratio and order parameters	Shinoda et al. [88]
DPPC	validation of force field	Berger et al. [10]
DPPC	validation of force field	Smondyrev and Berkowitz [91]
DPPC, DDPS	effect of asymmetry, order parameters for each bilayer leaflet separately	López Cascales et al. [56]
DPPC, DPPE	influence of different DPPC/DPPE ratios	Leekumjorn and Sum [52]
DPPC, POPC, PLPC, PAPC	effect of the number of double bonds	Hyvønen and Kovanen [41]
DPPG, GalCer	different GalCer/DPPG ratios	Zaraiskaya and Jeffrey [108]
POPC	analysis of lipid conformation and mobility	Heller et al. [34]
POPC	effect of NaCl	Böckmann et al. [13]

Continued on next page...

continued from previous page.

lipid(s)	use of order parameter	reference
SOPE	simulation set up using POPC experimental order parameters	Suits et al. [95]

3.5.1 Mixed lipids and asymmetric bilayers

Bilayers constituted of different lipids have different properties. In cells, membranes are composed of a complex mixture of lipids and proteins, unevenly distributed over the two monolayers.

In a simulation study on a mixed DPPC/DPPE bilayer [52], the DPPC order parameters agree with NMR data, but there is a poor agreement for DPPE, suggesting that a refinement of the DPPE force field is needed.

A simulation by López Cascales et al. [56] of an asymmetric DPPC/DPPS bilayer shows a good agreement with experimental data for both the area per lipid and order parameters. They find that the properties of both leaflets are different and that the DPPS that is present in only one of the leaflets does not influence the properties of the other (DPPC only) leaflet.

3.5.2 The effect of ions

In biological systems all sorts of different ions interact with the membrane. These interactions influence many processes, such as membrane fusion and the binding of anti-microbial peptides [44]. In MD simulations, ions are usually added to neutralise a system in order to prevent inaccurate calculation of the electrostatic interactions. But due to the relatively small amount of water molecules in these simulations, this quickly leads to high salt concentrations. For a recent review of simulation studies on the effect of ions and water on a lipid bilayer, see Berkowitz et al. [11].

An MD simulation of a POPC bilayer by Böckmann et al. [13], shows that Na^+ ions interact with the lipid carbonyl oxygen atoms, leading to a decrease in the area per lipid, a thickening of the bilayer and an increase of the order parameters. The addition of Na^+ and Cl^- ions decreased the self diffusion of the lipids. The self-diffusion agreed well with experimental data, suggesting that it takes place on a timescale below the simulation time (100 ns). Other simulations of POPC [44] and DPPC [72] confirm the increase in order parameters with increasing salt concentration. Kandasamy and Larson [44] find that the Na^+ ions bind to the glycerol ester group, deep in the interface region and they compare these findings to results from Sachs

et al. [83], who do not find this deep binding. This difference could be due to the shorter simulation time (50 ns and 5 ns respectively) or the different simulation conditions (force field, ensemble). Indeed, it has been shown that the binding of ions, especially divalent ions, is a slow process that takes on the order of a hundred nanoseconds to equilibrate [13].

On POPS lipids, Na^+ ions have a similar effect. Ions are found to bind to the carbonyl oxygen deep in the interface region, resulting in a small increase in the order parameters of the first atoms in the acyl chain [66]. Contrary to what one might expect, the area per lipid is unaffected. This is an important finding, because it shows that the area per lipid and the order parameter are not always straightforwardly related. It might —at least partly— explain the errors in the calculation of area per lipid from order parameters.

3.5.3 Hydration, the effect of water on bilayer simulations

Most MD simulations are carried out on fully hydrated bilayers. But since some experiments have been carried out at low hydration, simulations that investigate the effect of water on the properties of a bilayer are also interesting.

Molecular dynamics simulations of a DMPC bilayer show an increase in order parameters upon dehydration. This ordering effect of dehydration is confirmed by NMR data, although the increase is smaller in the simulations [39]. Two recent reviews explain in detail the effect of (de)hydration on model membranes [11, 64].

3.5.4 Small molecules

A comprehensive review article on the effect of small molecules has been published [51]. A few examples are given below.

Sucrose and trehalose decrease the mean square displacement of the lipid phosphorus atoms without affecting the DPPC area per headgroup and the order parameters [97].

In a paper by Sum and de Pablo [96] the effect of DMSO on a DPPC bilayer is studied by simulating a DPPC bilayer with DMSO from 0–100 mol%. DMSO appeared to accumulate below and above the headgroup, while it was excluded from the headgroup region itself. It was shown that low concentrations of DMSO have a large effect on the area per lipid, while the effect on the order parameter of the tails is very small, demonstrating (again) the complexity of the often assumed relation between order parameter and area per lipid.

A study of the effect of ethanol and methanol on lipid bilayers was carried out by Patra et al. [78]. They find a small increase in order upon adding 1 mol% of alcohol. The ordering increased close to the glycerol group, or below it, for methanol and ethanol respectively, in agreement with the mass density of these molecules. These results seemed to contradict those from Feller et al. [32], which was explained by the fact that alcohols have two modes of action: a local effect proportional to the alcohol concentration and a global effect as a consequence of the changes in surface tension. A more detailed analysis showed that the lipid order parameter decreased for the molecules bound to ethanol, but increased for the unbound lipids.

A mole fraction of 6.6% SDS does not significantly increase the ordering of DMPC lipids, while the order of the SDS tails is increased significantly, in agreement with experimental data [6].

3.5.5 Peptides and proteins

Among the many reviews on MD simulations of membrane proteins in lipid bilayers, two recent examples are those of Ash et al. [4] and Tieleman [101].

A simulation by Kandasamy and Larson [44] shows that the antimicrobial peptide magainin leads to disordering of POPC lipids in its vicinity. The binding of the peptide is stronger at lower concentrations of salt. The authors also note that their 50 ns simulation is not sufficient to equilibrate the position of the peptide in the bilayer.

In a simulation of a DPPC bilayer by Appelt et al. [3], ^{2}H NMR order parameters slightly increase in the presence of cyclo(RRWWRF) peptide. The authors compare this effect to the effect of Na^+ ions described by Böckmann et al. [13]. An NMR study by Vogel et al. [106] reveals that a peptide derived from N-ras protein drastically lowers DMPC order parameters. The DMPC packing however, is only slightly affected. A publication of a ^{2}H NMR study by Dave et al. [21] suggests no changes in order parameters when 6 mol% phospholamban binds to a POPC bilayer.

Peptides with a positive and negative hydrophobic mismatch influence the lipid order parameters differently, as demonstrated in a simulation of "KALP" model peptides [45].

These examples illustrate the complexity of peptide–lipid interactions. Different peptides and proteins may have different (or even opposite) effects on the order of lipid acyl chains.

3.5.6 Cholesterol and other sterols

Cholesterol increases the order in L_α phase bilayers. The amount of additional ordering induced by cholesterol differs, depending on the type of lipids. This effect has been investigated by deuterium NMR and MD simulations, where MD simulations are used to provide atomic detail on the cholesterol-lipid interactions. Reviews of interactions of cholesterol with phospholipids in membranes are given by Mouritsen and Zuckermann [65], Ohvo-Rekilä et al. [69].

A paper by Henriksen et al. [36] describes the correlation between the first moment of the ^{2}H-NMR spectra of a POPC bilayer and the membrane bending rigidity and area expansion modulus. They note that this correlation is the same for lanosterol, ergosterol and cholesterol, and independent of sterol concentration, even though these molecules have different effects on membrane thickness and sterol-lipid packing.

According to MD simulations, cholesterol increases the ordering in both the sn-1 and the sn-2 chain of DMPC, while in POPC it affects only the fragment of the sn-2 chain above the double bond [18, 82]. These simulations also show that the α and β face of cholesterol have a preferential interaction with saturated and unsaturated chains respectively.

It is beyond the scope of this publication to review all the work that has been carried out on lipid–cholesterol and protein–cholesterol interactions. Instead we present a study that was carried out in our lab, where experimental data and MD simulations were used to investigate the order parameters of cholesterol and lipid molecules in a DMPC bilayer.

3.6 Lipid-cholesterol interactions: a case study

Order parameter values have been obtained from measurements, performed on DMPC molecules with deuterated sn-2 chains. Also, for completeness, experimental results from other sources [5] have been included in the analysis. Cholesterol molecules have been selectively deuterated as shown in figure 2. High resolution ^{2}H spectra obtained with oriented bilayers and proton decoupling (figure 3) yielded low error bars ($< 2\%$) on the values of the experimental order parameters.

The simulation of molecular dynamics has been performed on a periodic 2:1 DMPC/-cholesterol system, composed of 256 DMPC molecules, 128 cholesterol molecules and 7310 H_2O molecules. The size of the system was 9.6 nm x 9.6 nm x 6.9 nm. The system was equilibrated until no changes in the energy were observed. For comparison, a simulation of pure DMPC was

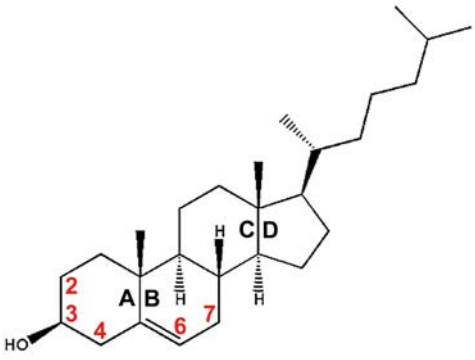

Figure 3.2: *Numbering of atoms labelled with deuterium (rings A and B)*

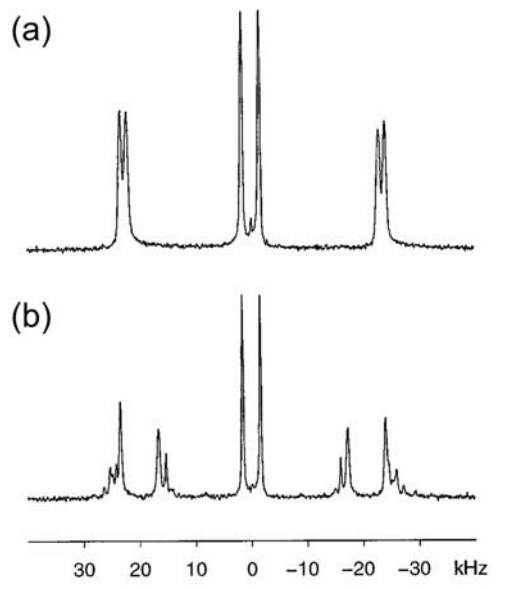

Figure 3.3: *^2H NMR spectra of cholesterol labelled with deuterium [58]. Label at atom: (a) 6, 7a, 7e, (b) 2a, 2e, 3, 4a, 4e, 6.*

carried out as well. This system contained 128 DMPC molecules and 1920 water molecules. For both systems, a 200 ns trajectory has been generated with the GROMACS software package [55, 105], using the Berger lipid force field [10]. The centre-of-mass motion was removed after each iteration. The Berendsen coupling [9] was used to keep the system at constant temperature (303 K) while the semi-isotropic Berendsen coupling kept the pressure at 1 bar [9]. The electrostatic interactions were calculated using the PME algorithm [20, 27]. The LINCS algorithm [37] was used to constrain bond lengths, allowing a time step of 2 fs. SPC was chosen as water model [8].

In previous simulations two types of DMPC molecules with respect to the dihedral angle θ_1 (based on atoms C1, C2, C3 and O in glycerol), referred to as type A and B, have been observed using InsightII and Discover 2000 software. The distribution of these angles, as reported in Aussenac et al. [5] shows two maxima, the major one at 60°(gauche−, type A) and one at 180°(trans, type B). In our recent simulations, using GROMACS, we saw one broad peak centred at ca. 20°, which corresponds to neither of the two aforementioned types. The major difference between the two approaches is the choice of the force field for the simulations.

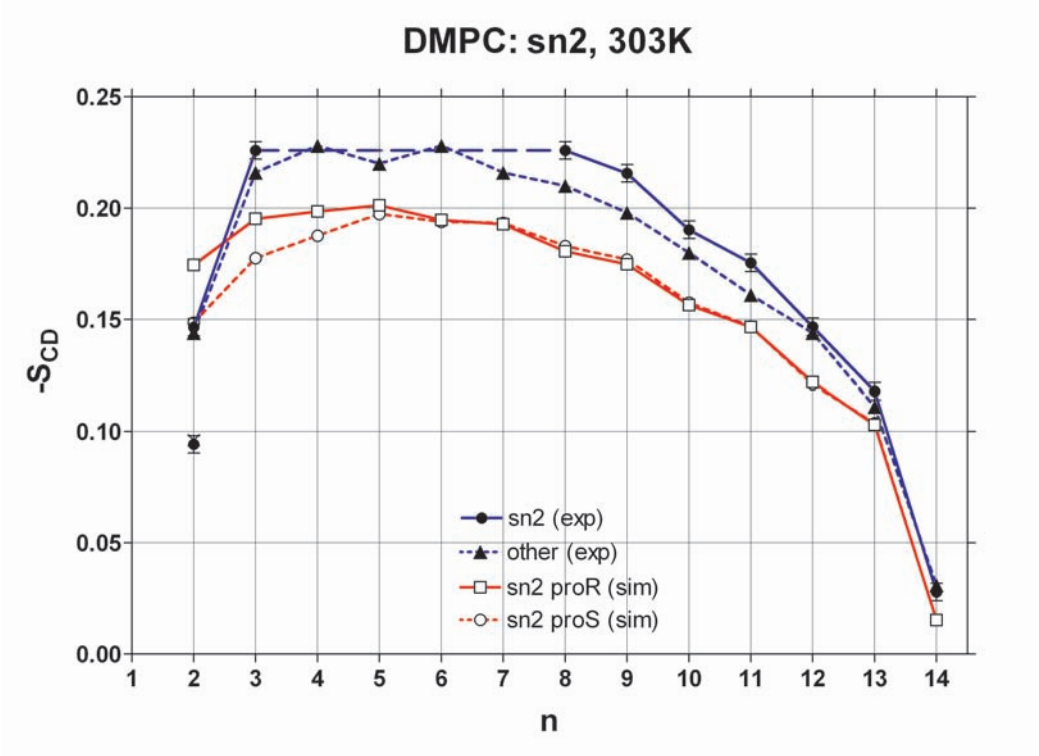

Figure 3.4: *Order parameter $-S_{CD}$ for the sn-2 chain of pure DMPC at 303 K as a function of the carbon atom index.*

The order parameters were calculated from the MD trajectories. Since the bilayers in the NMR experiment were oriented perpendicular to the magnetic field, $S_{bilayer}$ = -1/2. The fluctuations of the bilayer normal were found to be negligible in the simulations, since the value of S_{coll} is 0.9990 for DMPC, and increases to 0.9995 upon addition of 30% of cholesterol, which translate (using equation 3.2f) to fluctuations of 1.48° and 1.05° respectively.

S_{mol} (orientation of molecular director with respect to the bilayer normal) is 0.55 for pure DMPC and 0.90 for DMPC in the mixture, while the corresponding value for cholesterol molecules is 0.88. These values are in agreement with those found in the literature [58]. Figure 4 shows the order parameter $-S_{\mathrm{CD}}$, calculated for the sn-2 chain of pure DMPC as well as those measured experimentally. The agreement between experiment (filled symbols) and simulation (open symbols) is best at the end of the chain and acceptable for most carbon positions in the middle of it. However, for position 2 there is no agreement. The calculated values of the order parameter for deuterons proR and proS are similar, while the experimental values are

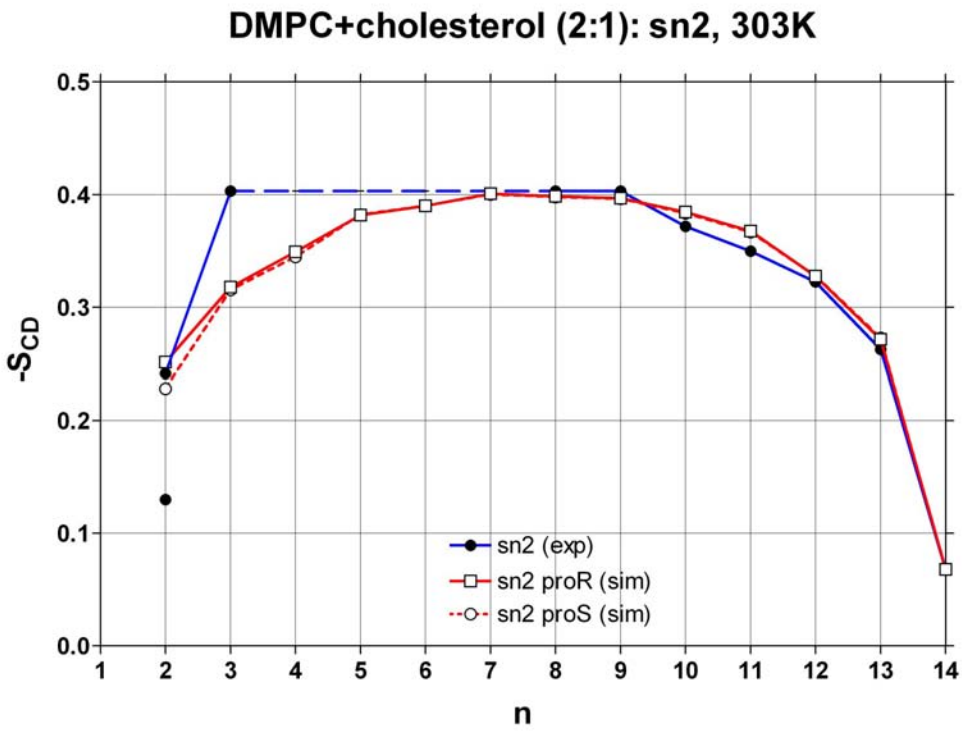

Figure 3.5: *Order parameter $-S_{CD}$ for DMPC in the DMPC-cholesterol 2:1 mixture at 303 K as a function of the carbon atom index.*

noticeably different. Similar conclusions can be drawn from figure 5, showing order parameters for the 2:1 DMPC/cholesterol mixture. The agreement for most of the carbon positions is excellent. However, the carbon position 2 causes similar problems, as the predicted proS and proR values are nearly identical, and the experimental ones are very different. The apparent plateau in the experimental data is due to lack of resolution and specific assignments.

The comparison of order parameters calculated from the trajectories with the available experimental data for the choline and glycerol shows differences up to one order of magnitude higher than the width of experimental error bars. However, a conversion of these differences into the average inclination of the C-^{2}H vector with respect to the bilayer normal yields tolerable differences, below 10°.

Figure 6 shows order parameters for cholesterol in the mixture. Again, the agreement is poor in the sense that the differences far exceed the limits of experimental uncertainties. However, the resulting differences in the inclination are small, as the maximal difference in orientation of individual

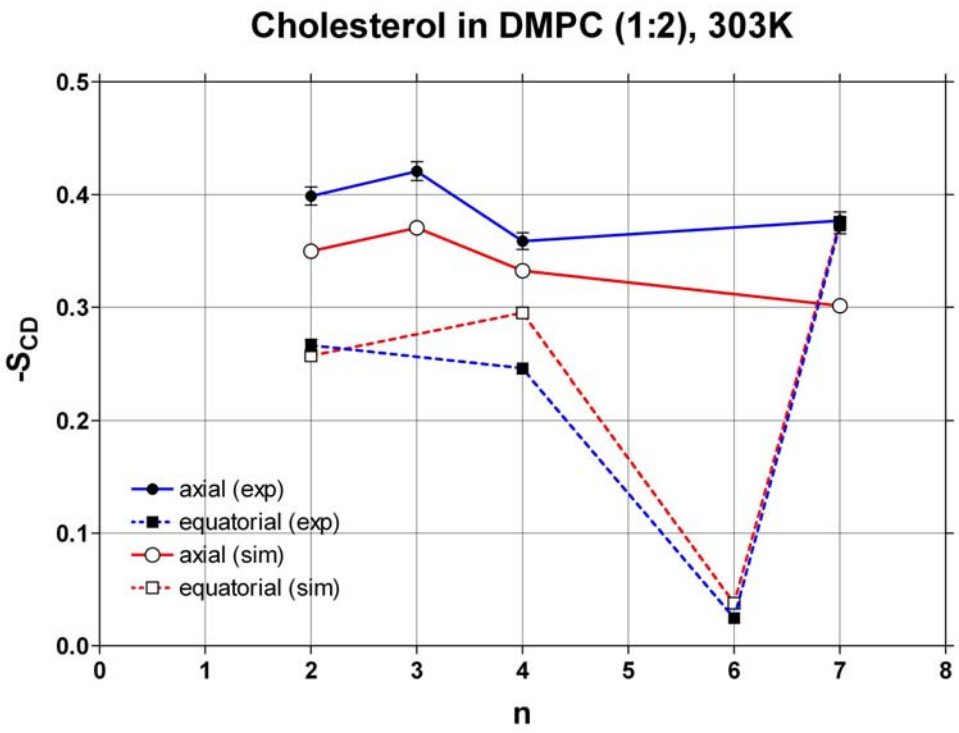

Figure 3.6: *Order parameters $-S_{CD}$ for cholesterol in the DMPC-cholesterol 2:1 mixture as a function of the carbon atom index.*

vectors amounts to 4°–5°.

A straightforward conversion of the observed differences between calculated and measured average vector orientations into quantitative corrections concerning the studied molecules is not feasible, since average molecular structures have no physical meaning. With this in mind, we may attempt to bring the calculated results in agreement with the experimental ones by calculating corrections to the orientations determined from NMR order parameters. This means (i) lifting the H2pro-S on sn-2 by $\sim 4°$, (ii) turning glycerol by $\sim 10°$ about the g2-g1 axis and (iii) lifting the entire choline unit by $\sim 6°$. For cholesterol in the 2:1 mixture, two corrections should be introduced: (i) rotation of the director of the molecule by $\sim 4°$ in the direction of C2 and (ii) tilting the molecule by $\sim 4°$ towards the protruding methyl groups.

Another interesting question is that of determining the average area per lipid (APL) [25, 33]. Molecular dynamics simulations have shown that the size of the simulation box remained practically constant after adding choles-

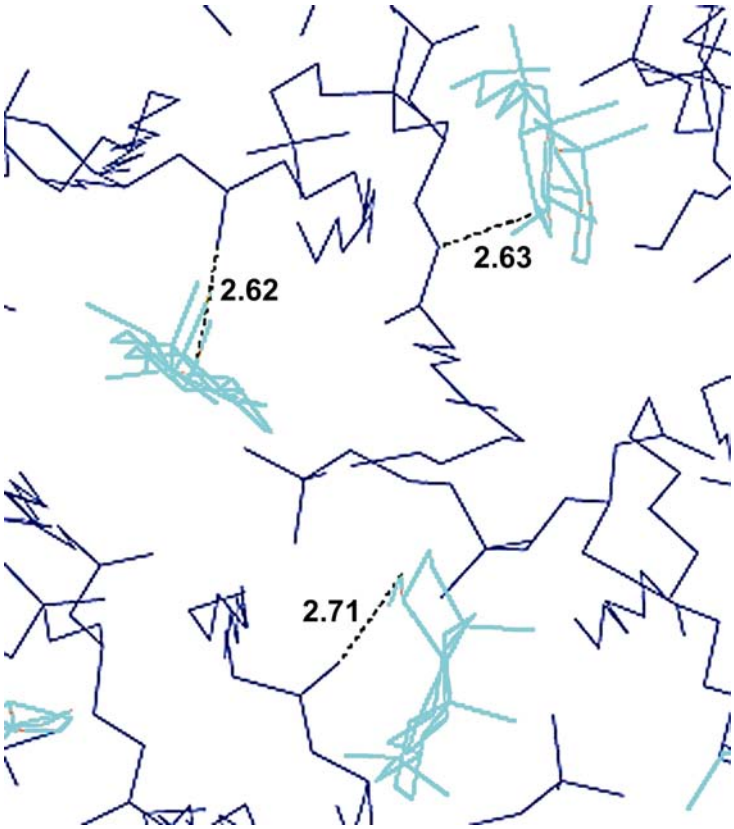

Figure 3.7: *Cholesterol molecules form hydrogen bonds with neighbouring lipids ($d_{O-O} = 2.6\,\mathring{A}$)*

terol molecules (APL = 63.13 Å^2 for DMPC and 63.02 Å^2 in the mixture). This finding is in agreement with experiments by Chapman et al. [17], where it was shown that cholesterol has little effect on condensed monolayers.

NMR order parameters have been used as basis for calculations of the APL [80], with limited success. Depending on the theoretical approach used, the results give widely differing numbers for the APL, often far from experimental values. Clearly, a different approach and a better theory are desirable.

An important conclusion, inferred from the MD simulations, is that no clustering of cholesterol molecules occurs within the time span of the trajectory. The analysis of intermolecular contacts leads to the conclusion that there is on average one transient hydrogen bond formed between the OH-group of cholesterol and the lipids, predominantly with the fatty-acyl chain esters, with the average distance between oxygen atoms (O3 of cholesterol and O2 of sn-1) of 2.6 Å (figure 7). In Soubias et al. [92] and Hénin and

Chipot [35] the hydrogen bonding patterns between cholesterol and its partners in lipid bilayers are discussed in detail, taking into account different rotameric states of cholesterol and bonding partners (acyl esters, phosphate groups, water).

Experimental results from neutron scattering data show that the hydrophobic thickness of the DMPC bilayer is 30.1 Å and that it increases to 34.6 Å upon addition of cholesterol [53]. Depending on the choice of reference atoms in defining the bilayer thickness, our calculations for DMPC give 27.3 Å for the distance between C2 (sn-1) atoms and 34.7 Å for the P–P distance. In case of the mixture we obtain the estimates of 34.1 Å and 42.2 Å.

In order to deduce the position of cholesterol in the bilayer, the neutron scattering studies used deuterated cholesterol and determined the distance between peaks in the corresponding density profiles. These peaks reflect the distance between the centres of gravity of the deuterated cholesterol fragments (rings A and B). The experiment yields 32.7 Å for the 2:1 mixture. Our calculations give the estimate of 29.3 Å between C4 atoms and 33.1 Å for O3 atoms. The latter value is in better agreement with the experiment, but does not correspond to the centre of gravity of the deuterated rings. The C4 atoms give a much better estimate of the centre of gravity.

3.6.1 Conclusions

The inner, hydrophobic region of the membrane is rather well represented in simulations. Upon approaching the interfacial region, more and more discrepancies with experiment can be observed. The differences may originate from the parametrisation of the unified-atom force field used in our studies. It will be interesting to see how the next-generation of all-atom lipid force fields performs in this respect. These force fields, that are currently being developed, use PME during calculations and are calibrated on and validated against state-of-the-art experimental data. In the region of the worst fit, high polarity is predominant and fluctuating charges would have to be taken into account to improve the overall agreement. It is therefore possible that perfect agreement can only be expected after including polarisability explicitly in the employed force fields [2, 71, 75, 98]

The differences between the calculated and experimental order parameters indicate that the orientations of molecules are described with an accuracy reaching $\pm$ 10° and often better. However, these differences are statistically significant because they far exceed the limits of the experimental uncertainties, including possible errors arising from the interpretation of the NMR data.

The question is whether these differences are important for biological

function. In many cases, the achieved level of accuracy is probably sufficient. However, there may be specific cases, such as interaction between a transmembrane protein and a ligand in a binding site located in the proximity of lipidic polar heads, where the reported differences may be critical. Hence, future development including continuous cross-validation between simulation and experiment will be required to address these issues.

Bibliography

[1] C. Anézo, A. H. de Vries, H.-D. Höltje, D. P. Tieleman, and S.-J. Marrink. Methodological issues in lipid bilayer simulations. *J Phys Chem B*, 107:9424–9433, June 2003.

[2] V. M. Anisimov, G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux, and J. Alexander D. MacKerell. Determination of electrostatic parameters for a polarizable force field based on the classical drude oscillator. *J Chem Theory Comput*, 1:153–168, 2005.

[3] C. Appelt, F. Eisenmenger, R. KÃijhne, P. Schmieder, and J. A. SÃűderhÃďll. Interaction of the antimicrobial peptide cyclo(RRWWRF) with membranes by molecular dynamics simulations. *Biophys J*, 89(4):2296–2306, Oct 2005.

[4] W. L. Ash, M. R. Zlomislic, E. O. Oloo, and D. P. Tieleman. Computer simulations of membrane proteins. *Biochim Biophys Acta*, 1666(1-2): 158–189, Nov 2004.

[5] F. Aussenac, M. Laguerre, J.-M. Schmitter, and E. J. Dufourc. Detailed structure and dynamics of bicelle phospholipids using selectively deuterated and perdeuterated labels. ^{2}H nmr and molecular mechanics study. *Langmuir*, 19:10468–10479, 2003.

[6] S. Bandyopadhyay, J. C. Shelley, and M. L. Klein. Molecular dynamics study of the effect of surfactant on a biomembrane. *J Phys Chem B*, 105:5979–5986, 2001.

[7] R. W. Benz, F. Castro-RomÃąn, D. J. Tobias, and S. H. White. Experimental validation of molecular dynamics simulations of lipid bilayers: a new approach. *Biophys J*, 88(2):805–817, Feb 2005.

[8] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. Interaction models for water in relation to protein hydration. In B. Pullman, editor, *Intermolecular Forces*, pages 331–342. D. Reidel Publishing Company, Dordrecht, 1981.

[9] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[10] O. Berger, O. Edholm, and F. JÃď hnig. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J*, 72(5): 2002–2013, May 1997.

[11] M. L. Berkowitz, D. L. Bostick, and S. Pandit. Aqueous solutions next to phospholipid membrane surfaces: insights from simulations. *Chem Rev*, 106(4):1527–1539, Apr 2006.

[12] M. Bloom, E. Evans, and O. G. Mouritsen. Physical properties of the fluid lipid-bilayer component of cell membranes: a perspective. *Q Rev Biophys*, 24(3):293–397, Aug 1991.

[13] R. A. Böckmann, A. Hac, T. Heimburg, and H. Grubmüller. Effect of sodium chloride on a lipid bilayer. *Biophys J*, 85(3):1647–1655, Sep 2003.

[14] L. J. Burnett and B. H. Muller. Deuteron quadrupole coupling constants in three solid deuterated paraffin hydrocarbons: C_2D_6, C_4D_{10}, C_6D_{14}. *The Journal of Chemical Physics*, 55(12):5829–5831, 1971.

[15] F. Castro-Román, R. W. Benz, S. H. White, and D. J. Tobias. Investigation of finite system-size effects in molecular dynamics simulations of lipid bilayers. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys*, 110(47):24157–24164, Nov 2006.

[16] I. Chandrasekhar, M. Kastenholz, R. D. Lins, C. Oostenbrink, L. D. Schuler, D. P. Tieleman, and W. F. van Gunsteren. A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field. *Eur Biophys J*, 32(1):67–77, Mar 2003.

[17] D. Chapman, N. F. Owens, M. C. Phillips, and D. A. Walker. Mixed monolayers of phospholipids and cholesterol. *Biochim Biophys Acta*, 183(3):458–465, 1969.

[18] S. Chiu, E. Jakobsson, and H. L. Scott. Combined monte carlo and molecular dynamics simulation of hydrated lipid–cholesterol lipid bilayers at low cholesterol concentration. *Biophys J*, 80:1104–1114, 2001.

[19] J. Czub and M. Baginski. Comparative molecular dynamics study of lipid membranes containing cholesterol and ergosterol. *Biophys J*, 90 (7):2368–2382, Apr 2006.

[20] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[21] P. C. Dave, E. K. Tiburu, K. Damodaran, and G. A. Lorigan. Investigating structural changes in the lipid bilayer upon insertion of the transmembrane domain of the membrane-bound protein phospholamban utilizing ^{31}P and ^{2}H solid-state NMR spectroscopy. *Biophys J*, 86 (3):1564–1573, Mar 2004.

[22] A. H. de Vries, I. Chandrasekhar, W. F. van Gunsteren, and P. H. Hünenberger. Molecular dynamics simulations of phospholipid bilayers: Influence of artificial periodicity, system size, and simulation time. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys*, 109(23): 11643–11652, Jun 2005.

[23] J. P. Douliez, A. LÃľonard, and E. J. Dufourc. Restatement of order parameters in biomembranes: calculation of C-C bond order parameters from C-D quadrupolar splittings. *Biophys J*, 68(5):1727–1739, May 1995.

[24] J. P. Douliez, A. Ferrarini, and E. J. Dufourc. On the relationship between C-C and C-D order parameters and its use for studying the conformation of lipid acyl chains in biomembranes. *J Chem Phys*, 109: 2513–2518, 1998.

[25] O. Edholm and J. F. Nagle. Areas of molecules in membranes consisting of mixtures. *Biophys J*, 89(3):1827–1832, Sep 2005.

[26] E. Egberts, S. J. Marrink, and H. J. Berendsen. Molecular dynamics simulation of a phospholipid membrane. *Eur Biophys J*, 22(6):423–436, 1994.

[27] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[28] E. Falck, M. Patra, M. Karttunen, M. T. Hyvönen, and I. Vattulainen. Lessons of slicing membranes: interplay of packing, free area, and lateral diffusion in phospholipid/cholesterol bilayers. *Biophys J*, 87(2): 1076–1091, Aug 2004.

[29] S. E. Feller. Molecular dynamics simulations of lipid bilayers. *Current Opinion in Colloid & Interface Science*, 5:217–223, 2000.

[30] S. E. Feller and R. W. Pastor. Constant surface tension simulations of lipid bilayers: the sensitivity of surface areas and compressibilities. *J Chem Phys*, 111:1281–1287, 1999.

[31] S. E. Feller, R. M. Venable, and R. W. Pastor. Computer simulation of a DPPC phospholipid bilayer: Structural changes as a function of molecular surface area. *Langmuir*, 13:6555–6561, Nov 1997.

[32] S. E. Feller, C. A. Brown, D. T. Nizza, and K. Gawrisch. Nuclear Overhauser enhancement spectroscopy cross-relaxation rates and ethanol distribution across membranes. *Biophys J*, 82(3):1396–1404, Mar 2002.

[33] A. I. Greenwood, S. Tristram-Nagle, and J. F. Nagle. Partial molecular volumes of lipids and cholesterol. *Chem Phys Lipids*, 143(1-2):1–10, Sep 2006.

[34] H. Heller, M. Schaefer, and K. Schulten. Molecular dynamics simulations of a bilayer of 200 lipids in the gel and in the liquid-crystal phases. *J. Phys. Chem.*, 97:8343–8360, 1993.

[35] J. Hénin and C. Chipot. Hydrogen-bonding patterns of cholesterol in lipid membranes. *Chem Phys Lett*, 425:329–335, 2006.

[36] J. Henriksen, A. C. Rowat, E. Brief, Y. W. Hsueh, J. L. Thewalt, M. J. Zuckermann, and J. H. Ipsen. Universal behavior of membranes with sterols. *Biophys J*, 90(5):1639–1649, Mar 2006.

[37] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.

[38] C. Hofsäß, E. Lindahl, and O. Edholm. Molecular dynamics simulations of phospholipid bilayers with cholesterol. *Biophys J*, 84(4):2192–2206, Apr 2003.

[39] C.-J. Högberg and A. P. Lyubartsev. A molecular dynamics investigation of the influence of hydration and temperature on structural and dynamical properties of a dimyristoylphosphatidylcholine bilayer. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys*, 110 (29):14326–14336, Jul 2006.

[40] T. Husslein, D. M. Newns, P. C. Pattnaik, Q. Zhong, P. B. Moore, and M. L. Klein. Constant pressure and temperature molecular-dynamics simulation of the hydrated diphtanolphosphatidylcholine lipid bilayer. *J Chem Phys*, 109(7):2826–2832, august 1998.

[41] M. T. Hyvønen and P. T. Kovanen. Molecular dynamics simulations of unsaturated lipid bilayers: effects of varying the numbers of double bonds. *Eur Biophys J*, 34(4):294–305, Jun 2005.

[42] M. Ø. Jensen and O. G. Mouritsen. Lipids do influence protein function — the hydrophobic matching hypothesis revisited. *Biochim Biophys Acta*, 1666(1-2):205–226, Nov 2004.

[43] M. Ø. Jensen, O. G. Mouritsen, and G. H. Peters. Simulations of a membrane-anchored peptide: structure, dynamics, and influence on bilayer properties. *Biophys J*, 86(6):3556–3575, Jun 2004.

[44] S. K. Kandasamy and R. G. Larson. Effect of salt on the interactions of antimicrobial peptides with zwitterionic lipid bilayers. *Biochim Biophys Acta*, 1758(9):1274–1284, Sep 2006.

[45] S. K. Kandasamy and R. G. Larson. Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: a systematic investigation of hydrophobic mismatch. *Biophys J*, 90(7):2326–2343, Apr 2006.

[46] L. Koubi, M. Tarek, M. L. Klein, and D. Scharf. Distribution of halothane in a dipalmitoylphosphatidylcholine bilayer from molecular dynamics calculations. *Biophys J*, 78(2):800–811, Feb 2000.

[47] M. Lafleur, B. Fine, E. Sternin, P. R. Cullis, and M. Bloom. Smoothed orientational order profile of lipid bilayers by ^{2}H-nuclear magnetic resonance. *Biophys J*, 56(5):1037–1041, Nov 1989.

[48] M. Lafleur, P. R. Cullis, and M. Bloom. Modulation of the orientational order profile of the lipid acyl chain in the L_α phase. *Eur Biophys J*, 19 (2):55–62, 1990.

[49] A. G. Lee. How lipids affect the activities of integral membrane proteins. *Biochim Biophys Acta*, 1666(1-2):62–87, Nov 2004.

[50] A. G. Lee. How lipids and proteins interact in a membrane: a molecular approach. *Mol BioSyst*, 1(3):203–212, Sep 2005.

[51] B. W. Lee, R. Faller, A. K. Sum, I. Vattulainen, M. Patra, and M. Karttunen. Sructural effects of small molecules on phospholipid bilayers investigated by molecular simulations. *Fluid Phase Equilib*, 228–229: 135–140, 2005.

[52] S. Leekumjorn and A. K. Sum. Molecular simulation study of structural and dynamic properties of mixed DPPC/DPPE bilayers. *Biophys J*, 90(11):3951–3965, Jun 2006.

[53] A. Léonard, C. Escrive, M. Laguerre, E. Pebay-Peyroula, W. Néri, T. Pott, J. Katsaras, and E. J. Dufourc. Location of cholesterol in DMPC membranes. a comparative study by neutron diffraction and molecular mechanics simulation. *Langmuir*, 17:2019–2030, 2001.

[54] E. Lindahl and O. Edholm. Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations. *Biophys J*, 79(1):426–433, Jul 2000.

[55] E. Lindahl, B. Hess, and D. Van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7: 306–317, 2001. Internet: `http://www.gromacs.org`.

[56] J. López Cascales, T. F. Otero, B. D. Smith, C. González, and M. MÃąrquez. Model of an asymmetric DPPC/DPPS membrane: effect of asymmetry on the lipid properties. a molecular dynamics simulation study. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys*, 110(5):2358–2363, Feb 2006.

[57] S. Marrink and A. Mark. Effect of undulations on surface tension in simulated bilayers. *J Phys Chem B*, 105:6122–6127, 2001.

[58] M. P. Marsan, I. Muller, C. Ramos, F. Rodriguez, E. J. Dufourc, J. Czaplicki, and A. Milon. Cholesterol orientation and dynamics in dimyristoylphosphatidylcholine bilayers: a solid state deuterium NMR analysis. *Biophys J*, 76(1 Pt 1):351–359, Jan 1999.

[59] D. Marsh. Lipid interactions with transmembrane proteins. *Cell Mol Life Sci*, 60:1575–1580, 2003.

[60] D. Marsh. Lipid-binding proteins: structure of the phospholipid ligands. *Protein Sci*, 12(9):2109–2117, Sep 2003.

[61] D. Marsh and T. Páli. The protein-lipid interface: perspectives from magnetic resonance and crystal structures. *Biochim Biophys Acta*, 1666 (1-2):118–141, Nov 2004.

[62] M. A. McCabe and S. R. Wassall. Rapid deconvolution of NMR powder spectra by weighted fast fourier transformation. *Solid State Nucl Magn Reson*, 10(1-2):53–61, Dec 1997.

[63] K. M. Merz. Molecular dynamics simulations of lipid bilayers. *Curr Opin Struct Biol*, 7(4):511–517, Aug 1997.

[64] J. Milhaud. New insights into water–phospholipid model membrane interactions. *Biochim Biophys Acta*, 1663:19–51, 2004.

[65] O. G. Mouritsen and M. J. Zuckermann. What's so special about cholesterol? *Lipids*, 39(11):1101–1113, Nov 2004.

[66] P. Mukhopadhyay, L. Monticelli, and D. P. Tieleman. Molecular dynamics simulation of a palmitoyl-oleoyl phosphatidylserine bilayer with Na^+ counterions and NaCl. *Biophys J*, 86(3):1601–1609, Mar 2004.

[67] J. F. Nagle and S. Tristram-Nagle. Structure of lipid bilayers. *Biochim Biophys Acta*, 1469(3):159–195, Nov 2000.

[68] J. Norberg and L. Nilsson. Advances in biomolecular simulations: methodology and recent applications. *Q Rev Biophys*, 36(3):257–306, Aug 2003.

[69] H. Ohvo-Rekilä, B. Ramstedt, P. Leppimäki, and J. P. Slotte. Cholesterol interactions with phospholipids in membranes. *Prog Lipid Res*, 41(1):66–97, Jan 2002.

[70] E. Oldfield, D. Chapman, and W. Derbyshire. Deuteron resonance: A novel approach to the study of hydrocarbon chain mobility in membrane systems. *FEBS Lett*, 16(2):102–104, Aug 1971.

[71] K. Palmo, B. Mannfors, N. G. Mirkin, and S. Krimm. Potential energy functions: from consistent force fields to spectroscopically determined polarizable force fields. *Biopolymers*, 68(3):383–394, Mar 2003.

[72] S. A. Pandit, D. Bostick, and M. L. Berkowitz. Molecular dynamics simulation of a dipalmitoylphosphatidylcholine bilayer with NaCl. *Biophys J*, 84(6):3743–3750, Jun 2003.

[73] M. Pasenkiewicz-Gierula, K. Murzyn, T. RÃşg, and C. Czaplewski. Molecular dynamics simulation studies of lipid bilayer systems. *Acta Biochim Pol*, 47(3):601–611, 2000.

[74] R. W. Pastor, R. M. Venable, and M. Karplus. Model for the structure of the lipid bilayer. *Proc Natl Acad Sci U S A*, 88(3):892–896, Feb 1991.

[75] S. Patel and C. L. Brooks. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J Comput Chem*, 25(1):1–15, Jan 2004.

[76] M. Patra, M. Karttunen, M. T. Hyvönen, E. Falck, P. Lindqvist, and I. Vattulainen. Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophys J*, 84(6): 3636–3645, Jun 2003.

[77] M. Patra, M. Karttunen, M. Hyvönen, E. Falck, and I. Vattulainen. Lipid bilayers drive to a wrong lane in molecular dynamics simulations by subtle changes in long-range interactions. *J Phys Chem B*, 108(14): 4485–4494, 2004.

[78] M. Patra, E. Salonen, E. Terama, I. Vattulainen, R. Faller, B. W. Lee, J. Holopainen, and M. Karttunen. Under the influence of alcohol: the effect of ethanol and methanol on lipid bilayers. *Biophys J*, 90(4): 1121–1135, Feb 2006.

[79] H. I. Petrache, K. Tu, and J. F. Nagle. Analysis of simulated NMR order parameters for lipid bilayer structure determination. *Biophys J*, 76(5):2479–2487, May 1999.

[80] H. I. Petrache, S. W. Dodd, and M. F. Brown. Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by ^{2}H NMR spectroscopy. *Biophys J*, 79(6):3172–3192, Dec 2000.

[81] M. Renault, V. Réat, M. Sugawara, P. Demange, É. Phez, J. Teissie, M. Piotto, and A. Milon. Giant vesicles as an efficient intermediate for ^{2}H NMR analyses of proteoliposomes in water suspension and in oriented bilayers. *C. R. Chimie*, 9:401–407, 2006.

[82] T. Róg and M. Pasenkiewicz-Gierula. Cholesterol effects on a mixed-chain phosphatidylcholine bilayer: a molecular dynamics simulation study. *Biochimie*, 88(5):449–460, May 2006.

[83] J. N. Sachs, H. Nanda, H. I. Petrache, and T. B. Woolf. Changes in phosphatidylcholine headgroup tilt and water order induced by monovalent salts: molecular dynamics simulations. *Biophys J*, 86(6):3772–3782, Jun 2004.

[84] H. L. Scott. Modeling the lipid component of membranes. *Curr Opin Struct Biol*, 12(4):495–502, Aug 2002.

[85] A. Seelig and J. Seelig. The dynamic structure of fatty acyl chains in a phospholipid bilayer measured by deuterium magnetic resonance. *Biochemistry*, 13(23):4839–4845, 1974.

[86] J. Seelig. Deuterium magnetic resonance: theory and application to lipid membranes. *Quarterly Reviews of Biophysics*, 10(3):353–418, 1977.

[87] J. Seelig and W. Niederberger. Deuterium-labeled lipids as structural probes in liquid crystalline bilayers. a deuterium magnetic resonance study. *J Am Chem Soc*, 96(7):2069–2072, April 1974.

[88] W. Shinoda, N. Namiki, and S. Okazaki. Molecular dynamics study of a lipid bilayer: convergence, strucutre, and long-time dynamics. *J. Chem. Phys*, 106(13):5731–5743, 1997.

[89] A. M. Smondyrev and M. L. Berkowitz. Molecular dynamics simulation of DPPC bilayer in DMSO. *Biophys J*, 76(5):2472–2478, May 1999.

[90] A. M. Smondyrev and M. L. Berkowitz. Molecular dynamics study of Sn-1 and Sn-2 chain conformations in dipalmitoylphosphatidylcholine membranes. *J Chem Phys*, 110(8):3981–3985, 1999.

[91] A. M. Smondyrev and M. L. Berkowitz. United atom force field for phospholipid membranes: Constant pressure molecular dynamics simulation of dipalmitoylphosphatidicholine/water system. *J Comput Chem*, 20(5):531–545, 1999.

[92] O. Soubias, F. Jolibois, V. Réat, and A. Milon. Understanding sterol-membrane interactions, part ii: complete ^{1}H and ^{1}3C assignments by solid-state NMR spectroscopy and determination of the hydrogen-bonding partners of cholesterol in a lipid bilayer. *Chemistry*, 10(23): 6005–6014, Nov 2004.

[93] E. Sternin, T. Zaraiskaya, R. Razavi, and R. M. Epand. Changes in molecular order across the lamellar-to-inverted hexagonal phase transition depend on the position of the double-bond in mono-unsaturated phospholipid dispersions. *Chem Phys Lipids*, 140(1-2):98–108, Apr 2006.

[94] G. W. Stockton, C. F. Polnaszek, A. P. Tulloch, F. Hasan, and I. C. Smith. Molecular motion and order in single-bilayer vesicles and multilamellar dispersions of egg lecithin and lecithin-cholesterol mixtures.

a deuterium nuclear magnetic resonance study of specifically labeled lipids. *Biochemistry*, 15(5):954–966, Mar 1976.

[95] F. Suits, M. C. Pitman, and S. E. Feller. Molecular dynamics investigation of the structural properties of phosphatidylethanolamine lipid bilayers. *J Chem Phys*, 122(24):244714, Jun 2005.

[96] A. K. Sum and J. J. de Pablo. Molecular simulation study on the influence of dimethylsulfoxide on the structure of phospholipid bilayers. *Biophys J*, 85(6):3636–3645, Dec 2003.

[97] A. K. Sum, R. Faller, and J. J. de Pablo. Molecular simulation study of phospholipid bilayers and insights of the interactions with disaccharides. *Biophys J*, 85(5):2830–2844, Nov 2003.

[98] M. Swart and P. van Duijnen. DRF90: a polarizable force field. *Mol Simul*, 32(6):471–484, 2006.

[99] Y. Takaoka, M. Pasenkiewicz-Gierula, H. Miyagawa, K. Kitamura, Y. Tamura, and A. Kusumi. Molecular dynamics generation of nonarbitrary membrane models reveals lipid orientational correlations. *Biophys J*, 79(6):3118–3138, Dec 2000.

[100] Y. Z. Tang, W. Z. Chen, C. X. Wang, and Y. Y. Shi. Constructing the suitable initial configuration of the membrane-protein system in molecular dynamics simulations. *Eur Biophys J*, 28(6):478–488, 1999.

[101] D. P. Tieleman. Computer simulations of transport through membranes: passive diffusion, pores, channels and transporters. *Clin Exp Pharmacol Physiol*, 33(10):893–903, Oct 2006.

[102] D. P. Tieleman, S. J. Marrink, and H. J. Berendsen. A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochim Biophys Acta*, 1331(3):235–270, Nov 1997.

[103] D. P. Tieleman, P. C. Biggin, G. R. Smith, and M. S. Sansom. Simulation approaches to ion channel structure-function relationships. *Q Rev Biophys*, 34(4):473–561, Nov 2001.

[104] D. J. Tobias. Electrostatics calculations: recent methodological advances and applications to membranes. *Curr Opin Struct Biol*, 11(2): 253–261, Apr 2001.

[105] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible and free. *J Comput Chem*, 26:701–1719, 2005.

[106] A. Vogel, C. P. Katzka, H. Waldmann, K. Arnold, M. F. Brown, and D. Huster. Lipid modifications of a Ras peptide exhibit altered packing and mobility versus host membrane as detected by ^{2}H solid-state NMR. *J Am Chem Soc*, 127(35):12263–12272, Sep 2005.

[107] J. Wohlert and O. Edholm. Dynamics in atomistic simulations of phospholipid membranes: nuclear magnetic resonance relaxation rates and lateral diffusion. *J Chem Phys*, 125(20):204703, Nov 2006.

[108] T. Zaraiskaya and K. R. Jeffrey. Molecular dynamics simulations and ^{2}H NMR study of the GalCer/DPPG lipid bilayer. *Biophys J*, 88(6): 4017–4031, Jun 2005.

4

Structural properties of a helix derived from H^+-V-ATPase subunit *a*

Contents

Summary

The 3D structure of a peptide derived from the putative transmembrane segment 7 from H^+-V-ATPase from *Saccharomyces cerevisiae* has been determined by solution state NMR in SDS. A stable helix is formed from L736 up to and including Q745, the lumenal half of the putative TM7. The helical region extends well beyond A738, as was previously suggested based on NMR studies of a similar peptide in DMSO. The pKa of both histidine residues that are important for proton transport was measured in water and in SDS. The differences that are found demonstrate that the histidine residues interact with the SDS polar heads. In detergent, circular dichroism data indicate that the secondary structure of the peptide depends on the pH and the type of detergent used. Using solid-state NMR, it is shown that the peptide is immobile in the presence of phospholipid bilayers, which means that it is probably not a single transmembrane helix in these samples. The environment is important for the structure of TM7, so in subunit *a* it is probably held in place by the other transmemebrane helices of this subunit.

This chapter has been accepted for publication as:
L.S. Vermeer, V. Réat, M.A. Hemminga and A. Milon, Structural properties of a peptide derived from H^+-V-ATPase subunit *a*. *Biochimica et Biophysica Acta — Biomembranes*, in press, February 2009.

4.1 Introduction

Vacuolar H^+-ATPase (V-ATPase) is an 830 kDa integral membrane protein responsible for the acidification of intracellular compartments. The pH of these cellular compartments plays an important role in many biological processes. The V-type ATPase protein has also been detected in the plasma membrane of specialised cells in various species, including humans. It is implicated in several diseases, such as osteoporosis and cancer. More detailed information can be found in reviews [2, 5, 11, 17, 19, 22, 31]. Much of the research has been carried out on V-ATPase from *Saccharomyces cerevisiae* (baker's yeast) and it has become a model for many eukaryotic V-ATPase studies.

V-ATPase consists of a cytoplasmic domain (V_1) and a transmembrane domain (V_0). Both domains contain several subunits. The V_0 transmembrane domain consists of subunits a, c, c′, c″ and d. The current consensus is that proton translocation takes place at the interface of subunit *a* and the rotating (c, c′, c″) subunits. Mutation of arginine-735 in putative transmembrane helix 7 (TM7) of subunit *a* completely inhibits proton transport [21]. The histidine residues at position 729 and 743 located in this helix have also been found to influence the proton pumping activity [22]. A single glutamic acid residue in each of the c subunits has been shown to be essential for proton translocation. Cysteine cross-linking studies suggest that TM7 from subunit *a* interacts with TM4 in subunit c and c′ [23] and with TM2 in subunit c″ [37]. By these effects, TM7 is likely to experience several important changes in its environment during proton translocation.

It has been shown that membrane spanning peptides can be used as a model for their structure and dynamics in the entire protein [7, 20, 28, 32, 34, 40, 41]. A similar approach has recently been used to collect structural data on the trans-membrane domain of subunit e from yeast F_1F_0-ATP synthase [40]. Several studies on peptides derived from V-ATPase have been published. Different peptides derived from TM7 were studied by NMR in DMSO [8, 9], by fluorescence in phospholipids [15], by ESR in SDS [36], by ESR and IR spectroscopy in phospholipid bilayers [26] and by CD and fluorescence in amphipols [10]. A high resolution structure of the entire subunit *a* has not yet been elucidated, but a structure of a peptide called sMTM7 (see table 4.1 on the facing page) derived from TM7 from subunit *a* was obtained by NMR studies in DMSO. A related peptide (called MTM7) was studied in SDS, but gave NOESY spectra with severely broadened cross peaks [9].

For the present work, we decided to use a different peptide derived from TM7, called KMTM7 (table 4.1). KMTM7 contains two non-native lysine residues at the N- and C-terminal end, to try and improve its reconstitution

Table 4.1: *Amino acid sequence of the putative TM7 from yeast V-ATPase subunit a, and some of the peptides derived from it that have been studied. Note that when most of the previous peptide studies were carried out, the putative TM7 was different [31] from the currently (revised) topology of TM7 [38]. The peptide used in this study (KMTM7) runs from S728 to S748 and thus contains the residues in TM7 that have been shown to be important for proton translocation. The current topological model places KMTM7 in the lumenal side of the membrane, where it may be lining the lumenal hemichannel. Two lysine residues were added at the N and C terminus to try and improve reconstitution in lipid bilayers. Throughout this chapter, the numbering is the same as in the native protein.*

	715 720 725 730 735 740 745 750	
putative TM7	IHQVIHTIEFCLNCVSHTASYLRLWALSLAHAQLSSVLWT	[31]
MTM7	IHTIEFCLNCVSHTASYLRLWALSLAHAQLSSVLWTM	[9, 26]
sMTM7	EFCLNCVSHTASYLRLWALSLAHAQ	[8, 36]
KMTM7 (this study)	KKSHTASYLRLWALSLAHAQLSSKK	[15]

in lipid bilayers. It also contains the essential arginine residue at position 735 as well as the two histidine residues at positions 729 and 743 from TM7 of V-ATPase subunit *a*. These residues are known to affect proton translocation. We used CD, liquid state NMR and solid state NMR spectroscopy of the peptide in different detergents and in the presence of phospholipid bilayers to assess its structural properties. A pH titration provided information about the effect of the protonation state of the histidine residues on the structure of the peptide.

4.2 Materials and methods

4.2.1 Peptide design and synthesis

Peptide KMTM7 (table 4.1) was designed based on its putative localisation in the lumenal hemichannel. Note that this is different from peptides that were designed earlier, because a recent chapter on the topology of subunit *a* proposed that it consists of 8 trans-membrane helices [38], in contrast to the now old model that assumed 9 trans-membrane segments [31]. Nearly the entire putative membrane spanning region has been studied in DMSO, but a structure calculation was unsuccessful in both DMSO and SDS [9]. Given that MTM7 is probably too long to span a membrane, KMTM7 was

designed to span the lumenal hemichannel only, and is expected to have a length similar to a POPC/POPG membrane.

To try and improve its reconstitution in lipid bilayers, two non-native lysine residues were added at the N- and C-terminus. Throughout this chapter the sequential numbering of amino acid residues of peptide KMTM7 is identical to the one used for subunit *a* in V-ATPase (table 4.1 on the preceding page). Peptide KMTM7 was produced on solid support using continuous flow chemistry by Pepceuticals Ltd., Leicester, UK. Its purity was tested by HPLC and mass-spectrometry and found to be greater than 90%. Peptide KMTM7 specifically labelled with ^{13}C on the carbonyl carbon of L734 was produced by Millegen biotechnology, Toulouse. Its purity was greater than 95%, as was demonstrated by HPLC and mass spectrometry. The carbonyl carbon of L734 was chosen as a labelling site because of its proximity to R735, the residue essential for proton translocation in V-ATPase.

4.2.2 Circular dichroism measurements

CD samples were prepared from a stock solution to make sure the concentration of peptide KMTM7 was equal in every sample (59 μM). For the CD studies in detergent, the SDS concentration was 82 mM and the octylglucoside concentration was 100 mM. The SDS concentration was chosen to be around 10 times the critical micelle concentration, 100 mM octylglucoside is about 3 times the CMC. Stock solutions of SDS and octylglucoside were prepared and adjusted with HCl or NaOH. After mixing the peptide with the stock solution, the pH was checked again. The final pH for each sample after having added the peptide is given in figure 4.1 on page 84.

CD spectra were recorded in the stepwise scanning mode from 190–350 nm on a Jasco J-815 CD spectrometer with a PTC-423S/15 temperature controller. The temperature was kept constant at 20°C. A cuvette with a path length of 1 mm was used. The sensitivity was set to 100 mdeg, pitch 0.5 nm, response 1 s, bandwidth 2 nm. Two spectra were accumulated and averaged.

Because of possible differences in the solubility of the peptide in the different detergents, the sample concentration was calculated by using the absorption at 280 nm and the theoretical extinction coefficient of the peptide [12], which was 6970 $cm^{-1}mol^{-1}$. CD spectra were corrected for the measured concentration and expressed in molar ellipticity (per residue) $[\theta]$ using $[\theta] = \theta/(10 \cdot C \cdot l \cdot n)$, where θ is the observed ellipticity in mdeg, C is the molar concentration, l the path length in cm and n the number of residues.

4.2.3 Solid state NMR spectroscopy

For solid-state NMR, peptide KMTM7 was specifically labelled with ^{13}C on the carbonyl carbon atom of L734. Peptide KMTM7 (2.3 mg) was mixed with 25.2 mg POPC and 6.3 mg POPG in 2 ml methanol, which led to a POPC/POPG ratio of 80/20 (mol/mol) and a peptide/lipid ratio of 1/50. The sample was shaken continuously overnight, under nitrogen atmosphere at room temperature. The methanol was evaporated off and the sample was placed under vacuum to remove any residual solvent. The dry sample was rehydrated with 500 μl, 10 mM sodium phosphate buffer at pH 7.1 and centrifuged at 12100 g for 10 min to precipitate the proteoliposomes. The pellet was put in a 3.2 mm magic angle spinning rotor and extra water was added to obtain full hydration (about 30 water molecules per lipid, as determined by weighing the sample). The sample was homogenised by centrifuging 8 times (3000 g, 1 min.), while turning the sample over after each centrifugation. ^{13}C solid-state NMR spectra were recorded the same day on a Bruker Avance 700 MHz spectrometer equipped with a 3.2 mm magic angle spinning triple resonance probe. A standard cross-polarisation pulse sequence was used, with a cross polarisation spin lock field of 40 kHz and a contact time of 1.4 ms. A TPPM proton decoupling pulse sequence was applied during the acquisition time, with a 58 kHz decoupling field. The relaxation delay was set to 2 s. ^{13}C MAS spectra of the labelled L734 residue (figure 4.2 on page 85) were recorded at a spinning frequency of 4 kHz. The temperature was kept constant at 302 K, which is well above the main gel to liquid crystalline phase transition temperature of the lipid mixture. For comparison, spectra of dry peptide powder were acquired as well. Before Fourier transformation, a Lorentzian line broadening of 50 Hz was applied. Spectra were referenced to DSS and the standard conventions were used in the analysis of the chemical shift anisotropy. The intensities of the observed spinning sidebands were subjected to a Herzfeld-Berger analysis [14] using the DMFIT software [30] to extract the anisotropy (δ_{CS}) and asymmetry (η) parameters of the chemical shift tensor.

4.2.4 Histidine titration by solution state NMR spectroscopy

Samples for the pH titration by liquid state NMR were prepared by adding 1 ml of either D_2O or SDS solution to 2 mg of peptide KMTM7, giving a peptide concentration of 718 μM. The SDS concentration was chosen to give a 1/250 peptide/detergent molar ratio in all the NMR experiments. In the CD experiments, this ratio had to be higher in order to be well above the

CMC. After vortexing, the samples were centrifuged for 10 min at 12100 g to remove any precipitated peptide. The supernatant was transferred to a different tube, and the pH was adjusted with DCl stock solution. DSS was added (30 μl, 5 mM) as standard to calibrate the spectra. Before recording an NMR spectrum, the pH was increased stepwise for each consecutive data point by adding the required amount of a 100 mM NaOD solution.

^{1}H NMR spectra were recorded on a Bruker Avance 600 MHz spectrometer with a 5 mm TCI cryoprobe. The sample temperature was kept constant at 298 K. The rf-field for the ^{1}H 90° pulse was 21 kHz. The relaxation delay was 1 s and the signal from residual water was suppressed by pre-saturating the water protons with a 31 Hz continuous wave pulse during this delay. For each data point, 512 scans were accumulated.

To calculate the pKa value from the experimental data, the GOSA computer program [1] was used to fit the data points to the following equation $CS_{\text{observed}} = (CS_1 + CS_2 \cdot 10^{(\text{pH-pKa})})/(1 + 10^{(\text{pH-pKa})})$. In this equation CS_1 and CS_2 are the chemical shift of the protonated and deprotonated state of the histidine residue. Fitting of three parameters (CS_1, CS_2 and pKa) resulted in a determination of the pKa values.

4.2.5 NMR structure determination

For the structure determination, ^{1}H NMR experiments were carried out. The peptide KMTM7 (1.8 mg) was dissolved in 1 ml of TFE. SDS-d25 was dissolved in 10 mM NaH_2PO_4 buffer (pH 3.0) to get a 160 μM solution, which was added to the peptide solution. After 5 min, 15 ml water was added and the sample was left under argon gas on a shaking table at room temperature overnight. The next day, the sample was frozen in liquid nitrogen and freeze-dried. The resulting white powder was dissolved in 636 μl buffer (10 mM NaH_2PO_4, pH 3.0, 10% D_2O), giving a final concentration of 1 mM peptide in 250 mM SDS. Because a white precipitate was visible, the sample was centrifuged at 21100 g for 2 min. The supernatant was transferred to an NMR tube and DSS was added (final concentration 0.3 mM) to calibrate the NMR spectra. The final pH of the sample was 5.0. The NMR tube was filled with argon gas and closed with a plastic cap and parafilm. NMR spectra indicated that the sample was stable for at least two months (no changes in the 1D and 2D ^{1}H spectra), when stored at room temperature.

NOESY and TOCSY spectra were recorded on a Bruker Avance 600 MHz spectrometer with a 5 mm TCI cryoprobe. The sample temperature was kept constant at 298 K or 313 K (two sets of experiments). The mixing time was 80 ms for the TOCSY spectra and 300 ms for the NOESY spectra. The spectral width was 12 ppm and the ^{1}H 90° pulse length was 12.5 μs. A WA-

TERGATE pulse sequence was used to suppress the water signal.

The data processing was done with Bruker Topspin software (version 1.3). The free induction decay was multiplied by a shifted-sine^2 window function. The line broadening was set to 10 Hz and a forward linear prediction was applied. After Fourier transformation, the spectra were phase corrected and a baseline correction was applied. The NMR spectra were calibrated using the DSS signal at 0 ppm.

Peak picking and integration were done using the software package Sparky 3.113 [13]. The calculation of the CSI and NOE contact tables were carried out using a Perl script that reads a Sparky peak list as input file. Four different reference sets for the H_α random coil chemical shifts were tested and all led to approximately the same result. Results presented in this chapter used the reference set from Arnold et al. [3].

The structure calculation was carried out using ARIA 2.2 software [33] with CNS 1.2 [6]. Only fully assigned NOEs were used in the structure calculation. No dihedral angle restraints were imposed. The default ARIA settings were used, but the number of trial structures was increased to 100 for each iteration step, of which the best 20 structures were used as input for the subsequent iteration. This led to a better convergence in the final 20 structures. No spin diffusion correction was applied during the calibration steps, because a structure calculation with corrections for spin diffusion resulted in structures that were similar to those without the correction. The manually assigned NOEs were defined as "trusted", meaning that ARIA was not allowed to change the assignments. After the last iteration, a final refinement step was carried out in water. The PROCHECK-NMR software [27] and a Perl script were used to check the quality of the structures.

4.3 Results

4.3.1 Peptide conformation depends on environment

The CD experiments (figure 4.1 on the next page) indicate that the α-helical content of peptide KMTM7 strongly depends on the pH and the detergent used. Note that "helical content" may have two different interpretations: either the structure of each molecule in the ensemble is less helical, or a part of the molecules are helical while another population contributes to the CD spectrum with a different conformation (such as for instance SDS-bound and unbound molecules).

In SDS, the line shape of the CD spectra indicated a helical content around 60% α-helix both at pH 3.3 and 9.0 (below and above the histidine

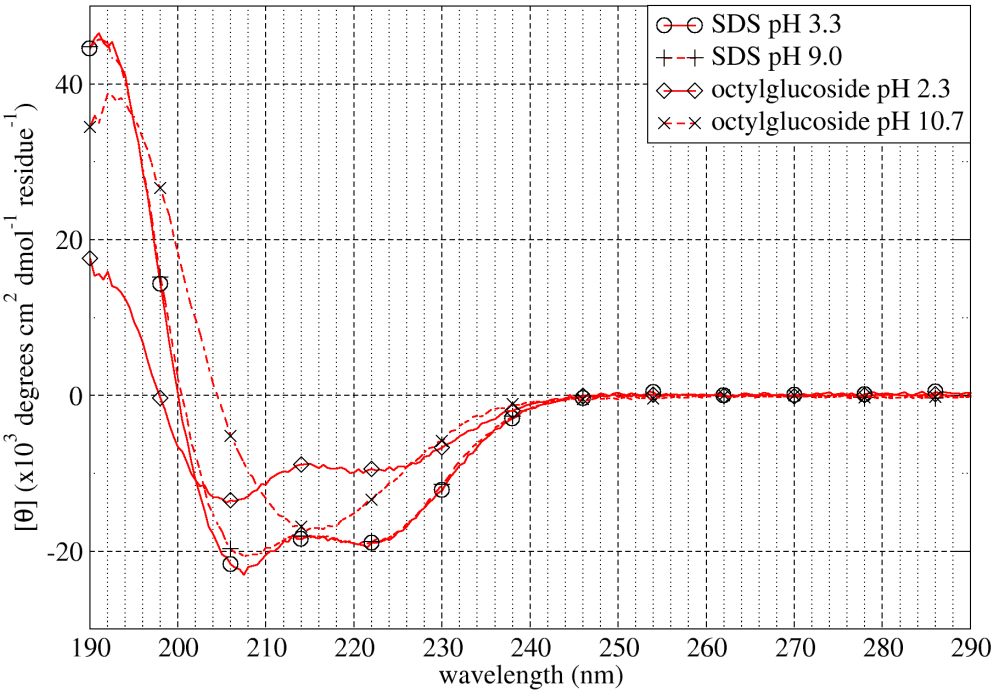

Figure 4.1: *Circular dichroism spectra of peptide KMTM7 in SDS and in octylglucoside at a pH above and below the histidine pKa*

pKa), in agreement with the NMR data. The intensity of both spectra was the same. Analysis with the CDPro software [35] reports 62% and 59% of helix, respectively, and the various deconvolution algorithms within CDPro (selcon3, cdsstr and continll) gave consistent results. In octylglucoside an α-helix is observed at pH 2.3, while at pH 10.7 a β-sheet is found. The low signal intensity in octylglucoside at pH 2.3 indicates that the α-helical content is lower than in SDS. The fitting software reports 34% helix at pH 2.3 and 43% helix at pH 10.7, a result which does not seem to agree with the lineshape of these spectra (see discussion of the results). An α-helical line shape with two minima at 207 and 222 nm was also observed in samples at pH 4, while at pH 5 the spectrum had only one minimum at 216 nm, typical for a β-sheet (data not shown).

The CPMAS NMR spectrum of the ^{13}C labelled carbonyl carbon of the L734 amino acid residue in peptide KMTM7 in the presence of hydrated POPC/POPG bilayers is shown in figure 4.2 on the facing page (top). We

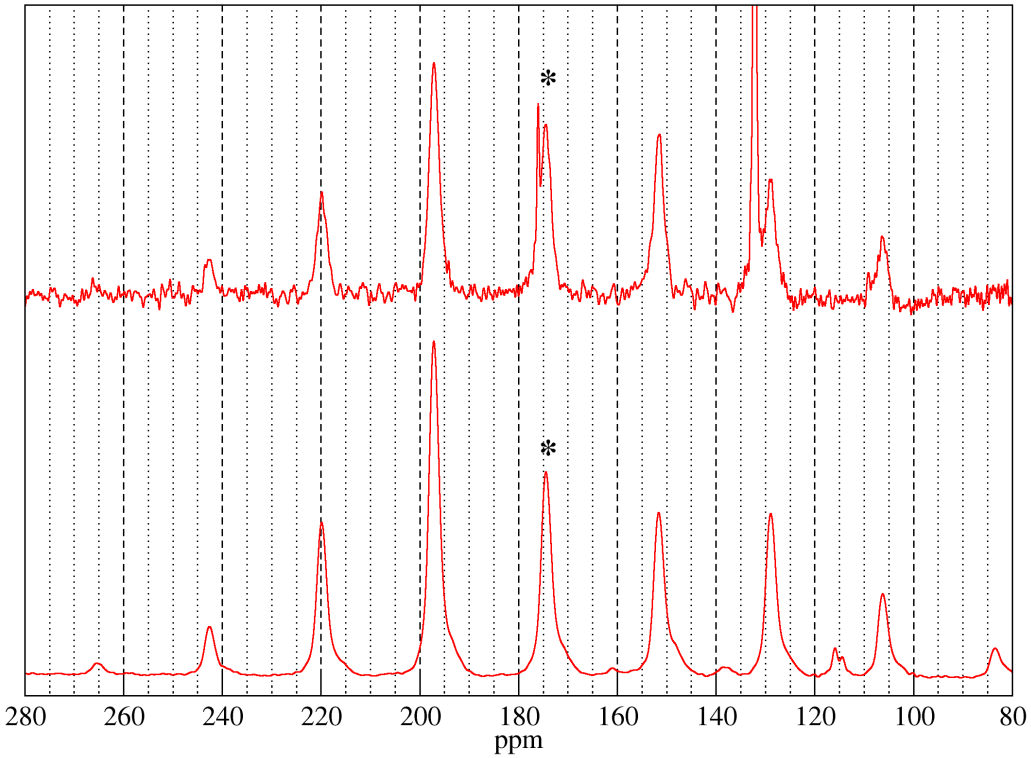

Figure 4.2: *top: 4 kHz CPMAS ^{13}C NMR spectrum of the ^{13}C labelled carbonyl carbon of the L734 residue in peptide KMTM7 in the presence of hydrated POPC/POPG bilayers. The isotropic peak at 174.5 ppm is marked with a star. The sharp peaks at 176.0 and 132.2 ppm are assigned to the ^{13}C natural abundance in the carbonyl carbon of the lipids and the double bond in the lipid tails, respectively. bottom: 4 kHz CPMAS ^{13}C NMR spectrum of dry peptide powder (without lipids), for comparison.*

do not believe that the peptide is reconstituted as a transmembrane α-helix, as will be discussed in section 4.4 on page 92. The sharp peaks at 176.0 and 132.2 ppm are assigned to the ^{13}C natural abundance in the carbonyl carbon of the lipids and the double bond in the lipid tails, respectively. Figure 4.2 (bottom) shows the CPMAS ^{13}C-NMR spectrum of dry peptide powder (without lipids). For the dry peptide powder, the anisotropy $\delta_{CS} = -84$ ppm and the asymmetry parameter $\eta = 0.60$. In the presence of hydrated lipid bilayers, $\delta_{CS} = -77$ ppm and $\eta = 0.75$.

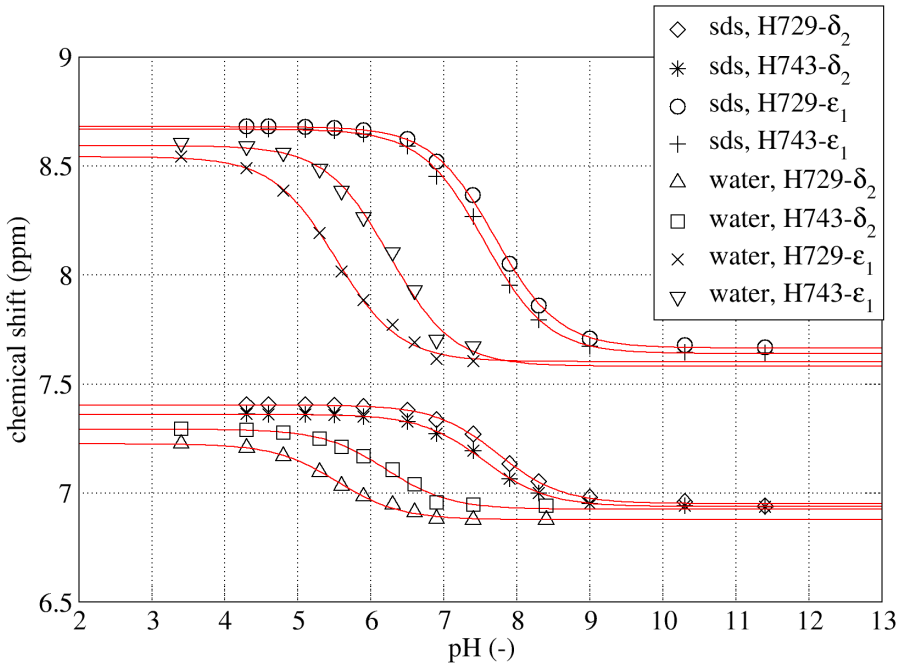

Figure 4.3: *Titration curves of H729 and H743 of peptide KMTM7 in 180 mM SDS and in water, at 298 K. The symbols are experimental data points, the curves are the best fit to the data, used to calculate the pKa values.*

4.3.2 The pKa of the Histidine residues

To determine the pKa of the histidine residues of peptide KMTM7, the chemical shift of its protons was followed during a pH titration [29]. It is not possible to measure the titrated δ_1 and ϵ_2 amine protons directly, because they exchange with the deuterons from the solvent. Therefore, a titration curve was measured by following the changes in ^{1}H chemical shift of the δ_2 and ϵ_1 protons. Their chemical shift changes because the shielding increases when the neighbouring nitrogen atom is deprotonated. The signal from the four protons of the two histidine residues was assigned based on the two characteristic singlets around the expected chemical shift values. This assignment was confirmed by monitoring the change of the peak positions when the pH was raised. In figure 4.3 the titration curves of H729 and H743 are shown. In SDS, the curve with the pKa that is slightly higher is assigned to H729 and the lower pKa curves to H743 (using the NOESY spectra). The pKa values

in SDS are 7.7 and 7.6. In water one histidine residue has pKa 6.3 and the other residue has a pKa of 5.5. It is likely that the residue with pKa 5.5 is H729, because the proximity of the two positively charged lysine residues probably destabilises the protonated state. The residue with pKa 6.3 would than be H743. A titration in octylglucoside was not successful, because the peptide precipitated above pH 5 leading to a loss of NMR signal. The CD results in octylglucoside agree with this observation.

4.3.3 Structure of the KMTM7 peptide in SDS

Analysis of the assignments

The H_α-H_N fingerprint region of the NOESY spectrum at 313 K is shown in figure 4.4 on the next page and the H_N-H_N region in figure 4.4 on the following page. Except for the first lysine residue, all of the residues could be (at least partly) assigned in the spectra at 313 K and 298 K (not shown).

In figure 4.5 on page 89 an overview of the some of the inter-residue contacts is shown. Sequential cross-peaks were found for nearly all residues. An almost continuous stretch of i–(i+3) contacts is observed between Y733 and Q745 at 313 K: a clear indication of a helical structure in that region. Although these contacts are not observed in all regions of the peptide, other typical contacts for a helix are visible. Only one long range contact (i–i+4) was observed. Most of the contacts that were present at 313 K could also be found in the spectrum at 298 K. Several contacts could be assigned in only one of the spectra. This is probably due to overlapping cross-peaks because many resonances had slightly shifted after the temperature change.

The chemical shift of the H_α protons strongly depends on the secondary structure of a protein or peptide. An up-field shift with respect to a random coil structure (a negative CSI value) indicates an α-helical region, while a down-field shift is indicative of β-sheets [39]. A difference that is more negative than -0.1 for 3–4 consecutive residues is usually considered an alpha helical region. Figure 4.6 on page 89 shows the H_α chemical shift difference from a reference set with random-coil chemical shifts.

Structure calculation

The structure calculation was carried out using a combination of both the spectra recorded at 298 K and 313 K. A superposition of the 20 best structures after minimisation in water is shown in figure 4.7 on page 90. The high convergence of these structures indicates an accurate and well defined structure. The average RMSD of the backbone atoms to the mean molecule

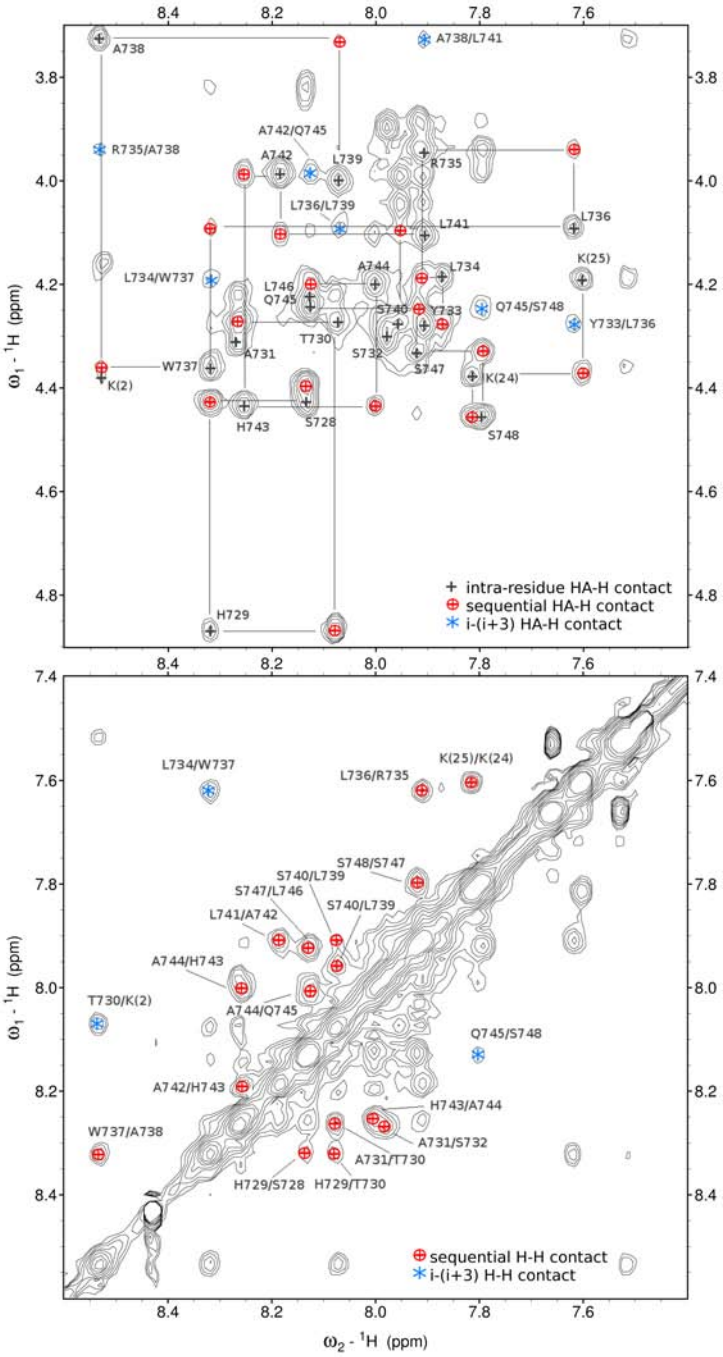

Figure 4.4: *top: The H_N-H_α fingerprint region of a NOESY spectrum of peptide KMTM7 in SDS at pH 5.0, 313 K, recorded at 600 MHz with a mixing time of 300 ms. Intra-residue contacts are labelled with black crosses, sequential HA–H contacts are shown as (red) crosses in circles and i–(i+3) contacts are labelled with (blue) stars. Other contacts are not labelled for presentation reasons, but all peaks that are visible in the figure were assigned. bottom: The H_N-H_N fingerprint region of the same spectrum. Sequential H–H contacts are labelled with (red) crosses in circles and i–(i+3) contacts are labelled with (blue) stars.*

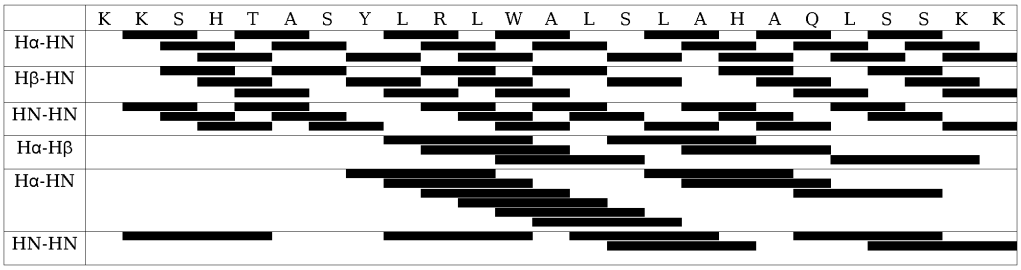

Figure 4.5: *Some of the inter-residue NOE connectivities. Only contacts that were kept (by ARIA) for the calculation of the final structure are shown. The first three rows show i–(i+1) contacts, the bottom three rows are i–(i+3) contacts. A list of the contacts that were used in the final structure calculation can be found in the BMRB database under number 11056.*

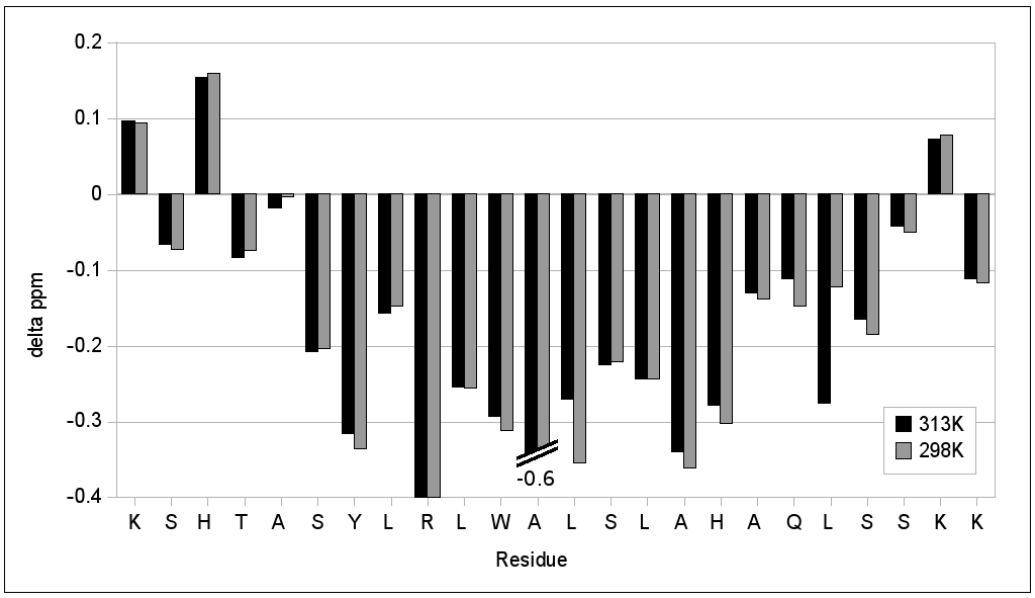

Figure 4.6: *Chemical shift differences between the H_α chemical shift in the peptide and reference values for a random coil [3].*

was 1.7 Å, calculated from residue 6–23. Other quality checks (table 4.2 on page 91) and the number of NOEs used (table 4.3 on page 91) confirm the quality of the result. The structures were deposed in the RCSB protein data bank under number 2RPW.

The side chain dihedral angles of H729 have largely different values in different structures, while R735 and H743 seem to be in a more defined con-

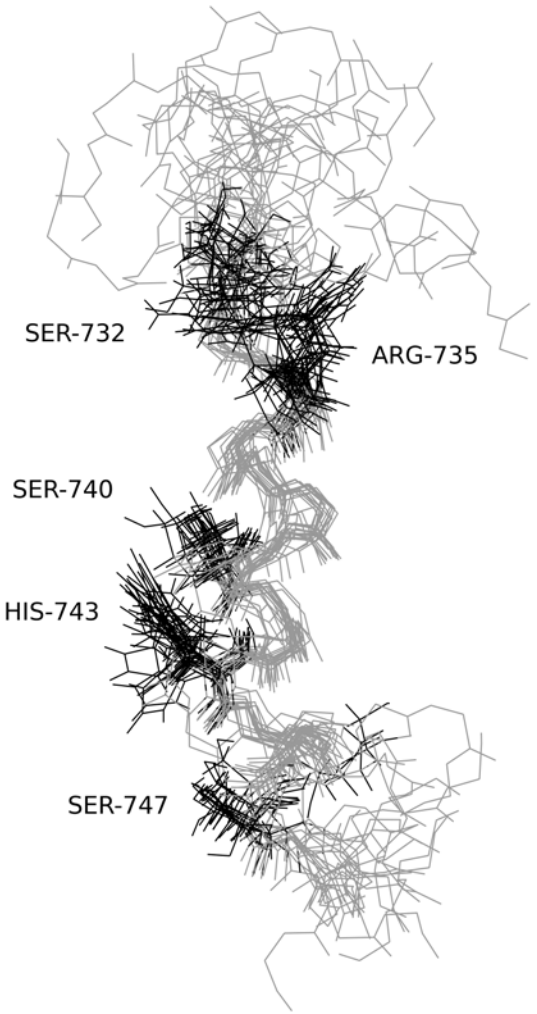

Figure 4.7: *Structure of peptide KMTM7 in SDS after refinement. A superposition of the 20 best structures is shown, The structures were aligned by pairwise fitting of the backbone atoms of residue 6–23 to the mean molecule, using the software package MOLMOL [25]. The RMSD of the fit was 1.7 Å (calculated on backbone atoms from residue 6–23). The trace of the backbone is shown as a thin grey line. The sidechains of the polar residues in the structured region of the peptide are shown in black. The structures were deposed in the RCSB protein data bank with reference number 2RPW.*

Table 4.2: *The quality of the 20 best structures before and after refinement. Percentages correspond to the number of dihedral angles that fall in the different regions of a Ramachandran plot.*

refinement:	no	water
favoured	52%	58%
allowed	35%	34%
generously al.	13%	7%
disallowed	1%	2%

Table 4.3: *Number of NOEs used for the structure calculation after the final iteration. The number of violations shown is the average number of violations > 0.5 Å on backbone atoms.*

total NOEs	270
intra-residue	141
inter residue:	129
sequential	83
i–(i+2)	10
i–(i+3)	35
long range	1
short (≤2.5 Å)	177
medium (≤3.5 Å)	82
long (>3.5 Å)	11
violations	5.7

formation. Multiple sidechain NOE contacts were assigned for these residues and very similar side chain dihedral angles were found before and after refinement, indicating that they are not an artifact of the refinement in water.

The N-terminus up to residue A731 shows a broad distribution of backbone dihedral angles. A similar effect is observed for residue S748 up to the C-terminus. This is not unexpected, because peptides usually show a certain amount of flexibility on the termini. From Y733–S747, most residues have small deviations from the average angle, and the average values are close to those of an ideal α-helix with $\phi = -57.8$ and $\psi = -47.0$. The dihedral angles are consistent with the CSI data (figure 4.6 on page 89).

Figure 4.8 on the next page shows an analysis of the secondary structure using the extended Kabsch & Sander method [18], as carried out by the procheck_nmr software [27]. An α-helix is assigned from L736 to L741 in most structures. The regions around the α-helical middle of the peptide tend to exhibit 3_{10} helical structures or hydrogen bonded turns. Although the N-terminus of the peptide seems rather flexible, it appears to be capable of forming helices, given the large number of structures that form hydrogen bonded turns in this region of the molecule.

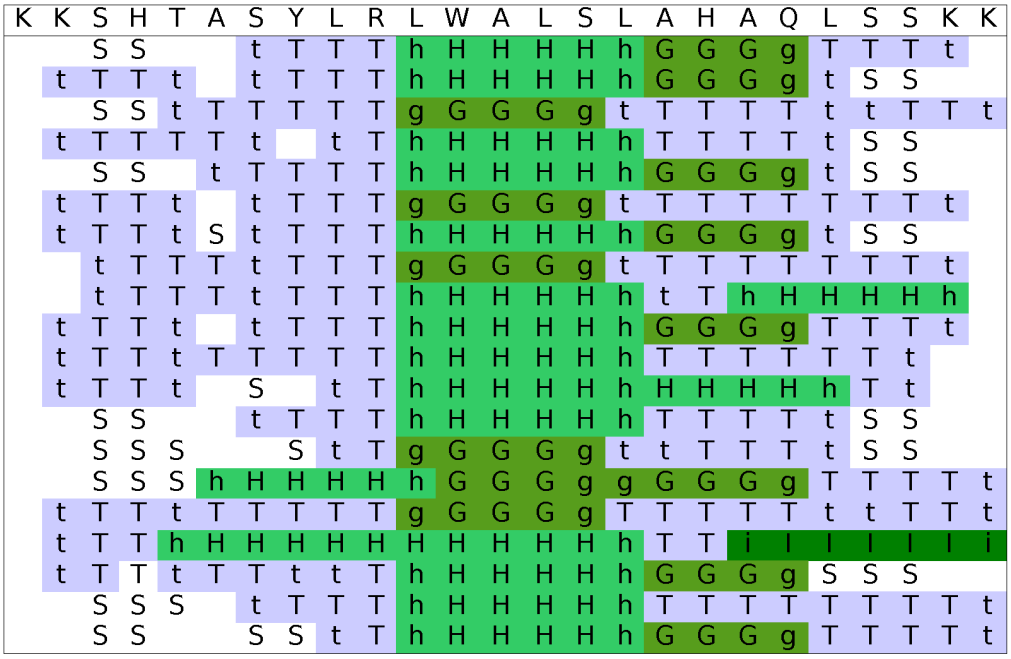

Figure 4.8: *DSSP code (Kabsch & Sander) secondary structure assignment of the twenty minimised structures shown in figure 4.7 on page 90. The structure assignment was carried out with procheck_nmr [27]. Capital letters are the original Kabsch and Sander definitions, lower case letters indicate the slightly extended assignment that is used by procheck. H/h = α-helix, G/g = 3_{10}-helix, I/i = π-helix, T/t = hydrogen bonded turn, S = bend.*

4.4 Discussion

The structural properties of a peptide derived from V-ATPase subunit *a* TM7 were studied. The peptide is different from the ones used in previous structural studies [8, 9], but it also contains the R735 residue that is essential for proton transport. In addition, two lysine residues were added to each of the termini. NMR experiments in a membrane mimicking environment gave better spectra than in previous studies, allowing a structure determination in SDS. Furthermore, the use of SDS instead of DMSO (that was used previously [8, 9]) allowed us to control the pH of the sample. NMR experiments enabled us to study the structure around residue R735 in detail, while CD studies in SDS and octylglucoside provided information on how the structure of the peptide is influenced by its environment. The role that the histidine residues play in these structural changes was investigated by determining

the protonation state (pKa) of the histidine residues in SDS and in water.

The CD experiments shown in figure 4.1 on page 84 indicate that the structure of the KMTM7 peptide depends on the pH and on its environment. An α-helical structure is observed in SDS, regardless of the pH. In octylglucoside, the peptide loses its α-helical structure at high pH in favour of a β-sheet conformation, indicating aggregation. The aggregation in octylglucoside is ultimately visible as a white precipitate in the sample. Possibly, the charged histidine residues prevent the peptides from aggregating at a low pH. When uncharged, the peptides interact and aggregate to β-sheets. In SDS, the negatively charged surface of the micelles may prevent the peptides from coming close enough to interact and aggregate, even at high pH. The difference in intensity that is observed between the CD spectra in octylglucoside and SDS may be due to a lower helical content, but uncertainties in the concentration prevent an accurate determination of the helical content. It is well known that the percentage of helicity calculated from CD spectra is often inaccurate, and very sensitive to the accuracy of the sample concentration. The percentages of helix in SDS agree with the NMR results, but based on the lineshape of the spectra the calculation of the percentage of helix in octylglucoside seems inaccurate, which is probably a result of inaccuracies in the sample concentration due to precipitation and the formation of micro-aggregates of peptides.

Solid-state NMR measurements of peptide KMTM7 in the presence of phospholipid vesicles indicated that it is more mobile in the presence of lipid membranes than as a dry powder ($|\delta_{\mathrm{CS,lipids}}| < |\delta_{\mathrm{CS,powder}}|$), and its asymmetry ($\eta = 0.75$) does not correspond to the expected symmetry of a single trans-membrane α-helix with a fast axial diffusion around the bilayer normal (which would have given $\eta = 0$). From this information we conclude that we did not succeed to reconstitute the peptide as a transmembrane helix in lipid bilayers. Several lipid mixtures and different reconstitution protocols have been tried, all with the same, negative, result. We conclude that this peptide does not behave like a typical transmembrane peptide, as is illustrated by the solid-state NMR spectrum in figure 4.2 on page 85. Furthermore, the ^{13}C isotropic chemical shift of L743 is not typical for an α-helical conformation. These solid-state NMR results in the presence of phospholipid bilayers do not fit with the previously shown α-helical trans-membrane conformation from fluorescence data [15]. This may be due to the high concentration and low hydration required for these MAS NMR experiments. Since in similar experimental conditions, typical α-helical behaviour has been observed for other peptides, we can conclude that KMTM7 has a strong tendency of structural polymorphism, as has already been observed by circular dichroism.

Depending on the environment, the pKa of the two histidine residues

changes markedly (figure 4.1 on page 84). Obviously, the pKa that was measured in our samples cannot be directly compared to the pKa in the native environment of the peptide. Especially the artificial lysine residues near H729 are expected to influence its pKa. However, the environmentally induced pKa differences in peptide KMTM7 help to explain the complex behaviour of the peptide in the different detergents. For the structure determination by 2D NMR experiments, it also allowed us to choose a pH where the protonation state of the histidine residues is known. The difference in pKa between the sample in water and in SDS is higher for H729 than for H743. The protonated state of H729 is destabilised in water, probably due to the presence of the non-native lysine residues. In SDS it has almost the same pKa as H743 because both are surrounded by negatively charged SDS headgroups. Therefore, these data clearly indicate that both histidine residues interact with the SDS headgroups. They are not located in a hydrophobic environment inside a micelle, which would have destabilised the charged state, and given a lower pKa in SDS. It has been previously demonstrated that charged histidine residues preferably interact with the negatively charged headgroup region, while uncharged histidine residues may take up a location in a hydrophobic environment [4].

Using peptide KMTM7 instead of MTM7 or sMTM7 greatly improved the quality of the NMR spectra in SDS (figure 4.4 on page 88), which allowed us to determine the structure in this detergent. The lysine, histidine and arginine residues were all charged under the experimental conditions used for the 2D NMR experiments. The N-terminus of the peptide shows more difference in the conformation of the 20 best structures than the C-terminus, which can be interpreted as more structural flexibility. In general, the C-terminus of peptides tends to be more flexible than the N-terminus [16], but this peptide shows more flexibility in its N-terminus, which we interpret as an effect of the primary structure.

The charged lysine residues probably interact with the negative charges that are present in the interface region, which gives us an idea of the location of the SDS headgroups. This is in agreement with the results of the pKa measurement in SDS. The R735 side chain points towards the N-terminus of the peptide, presumably also to the interface region. This behaviour is common for arginine residues, and has previously been suggested in the case of the KMTM7 peptide in phospholipid bilayers [15]. The end of the R735 side chain is nearly located on the same side of the helix as H743. Although a helical wheel plot (not shown) does not clearly indicate a hydropilic side of KTMT7 that could line the lumenal hemichannel of V-ATPase subunit *a*, the fact that the arginine side chain is long enough to move to the same side of the helix as H743 may indicate that it is accessible through a hemi-channel that

is lined by the residues on that side of the helix (R735–S740–H743–S747).

KMTM7 shows a helical structure in SDS from L736 up to and including Q745 (figure 4.8 on page 92). The region around the α-helical core also displays helical characteristics. In most of the structures, hydrogen bonded turns are observed from S728 up to the N terminus. The CSI values (figure 4.6 on page 89) and CD spectra (figure 4.1 on page 84) indicate a predominantly α-helical structure and are therefore consistent with the NMR structural data. Only one long range (i–(i+4)) NOE contact was observed (between the tyrosine HA and the HD1 proton from the tryptophan sidechain), indicating that the peptide does not aggregate or fold back on itself in a hairpin or coil structure.

In previous studies of MTM7 in DMSO, three helical regions were identified [9], spanning C723–A731, Y733–L739 and A742–V749, and it was suggested that TM7 is a 32 residue helical segment spanning from T719 to W751. Shortly afterwards, also in DMSO, two helical regions were identified more precisely in sMTM7: C726–H729 and S732–A738 [8]. The current results suggest that the helical region runs at least up to S748, and that TM7 can form a more stable and better defined helical structure than previous results in DMSO suggest. A similar observation has been made for a peptide derived from F_1F_0-ATPase [40]. A combination of the new and old experimental data indicates that TM7 can form helices consisting of mainly hydrophobic residues in a region that is much longer than is needed to span a biological membrane. This strongly suggests that TM7 interacts with other helices that are also longer than needed to span the membrane, which would shield the hydrophobic TM7 residues from the water phase. It may prove hard to accurately predict the transmembrane region of proteins with these properties using the common methods based on solvent accessibility and hydrophobicity scales. The data presented in this chapter show that the helix may even extend up to S748. We do not believe that the non-native lysine residues that were added to the termini have greatly influenced the structure of the peptide. Firstly, the tendency to form helices was also demonstrated in DMSO for a similar peptide without lysine residues, and secondly, we do not know of any examples from the literature where the addition of charged residues to the termini forces the entire peptide to adopt a certain structure. All available data indicate that peptide KMTM7 is capable of helix formation, but is not the archetype of a helical transmembrane peptide in lipids as compared to WALP [24]. Therefore we conclude that peptide KMTM7 probably requires its native environment, consisting of subunit *a* and *c* to be held in place in the native protein. The recent suggestion that transmembrane helix 7 consists of almost 40 amino acids [38] corresponds well with this picture.

4.5 Acknowledgements

The author wishes to acknowledge financial support from the European Marie Curie program (BIOMEM). The NMR spectra were recorded on spectrometers financed with the help of European Structural funds, Région Midi-Pyrenées and CNRS. I would like to thank Olivier Saurel and Pascal Ramos for their help with the NMR measurements and Afonso Duarte, Pascal Demange and Virginie Gervais for their comments and advice.

Bibliography

[1] GOSA: A curve fitting software for non linear regressions. `http://www.bio-log.biz`.

[2] Y. Arata, T. Nishi, S. Kawasaki-Nishi, E. Shao, S. Wilkens, and M. Forgac. Structure, subunit function and regulation of the coated vesicle and yeast vacuolar (H^+)-ATPases. *Biochim. Biophys. Acta*, 1555(1-3): 71–74, Sep 2002.

[3] M. R. Arnold, W. Kremer, H.-D. Lüdemann, and H. R. Kalbitzer. ^{1}H-NMR parameters of common amino acid residues measured in aqueous solutions of the linear tetrapeptides Gly-Gly-X-Ala at pressures between 0.1 and 200 MPa. *Biopysical Chemistry*, 96(2–3):129–140, 2002.

[4] B. Bechinger. Towards membrane protein design: pH-sensitive topology of histidine-containing polypeptides. *J Mol Biol*, 263(5):768–775, Nov 1996.

[5] K. W. Beyenbach and H. Wieczorek. The V-type H^+ ATPase: molecular structure and function, physiological roles and regulation. *J Exp Biol*, 209(Pt 4):577–589, Feb 2006.

[6] A. T. Brunger. Version 1.2 of the crystallography and NMR system. *Nature Protocols*, 2(11):2728–2733, 2007.

[7] F.-X. Ding, H. Xie, B. Arshava, J. M. Becker, and F. Naider. ATR-FTIR study of the structure and orientation of transmembrane domains of the *saccaromyces cerevisiae* α-mating factor receptor in phospholipids. *Biochemistry*, 40:8945–8954, 2001.

[8] A. M. Duarte, E. R. de Jong, R. Wechselberger, C. P. van Mierlo, and M. A. Hemminga. Segment TM7 from the cytoplasmic hemi-channel from V_0-H^+-V-ATPase includes a flexible region that has a potential role in proton translocation. *Biochimica et Biophysica Acta*, 1768:2263–2270, 2007.

[9] A. M. Duarte, C. J. Wolfs, N. A. van Nuland, M. A. Harrison, J. B. Findlay, C. P. van Mierlo, and M. A. Hemminga. Structure and localization of an essential transmembrane segment of the proton translocation channel of yeast H^+-V-ATPase. *Biochimica et Biophysica Acta*, 1768: 218–227, 2007.

[10] A. M. Duarte, C. J. Wolfs, R. B. Koehorst, J.-L. Popot, and M. A. Hemminga. Solubilization of V-ATPase transmembrane peptides by amphipol A8-35. *Journal of Peptide Science*, 14:389–393, 2008.

[11] M. E. Finbow and M. A. Harrison. The vacuolar H^+-ATPase: a universal proton pump of eukaryotes. *Biochem J*, 324 (Pt 3):697–712, Jun 1997.

[12] S. C. Gill and P. H. von Hippel. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, 182(2):319–326, Nov 1989.

[13] T. D. Goddard and D. G. Kneller. Sparky 3. URL: http://www.cgl.ucsf.edu/home/sparky/, 2007. University of California, San Francisco.

[14] J. Herzfeld and A. Berger. Sideband intensities in NMR spectra of samples spinning at the magic angle. *J. Chem. Phys.*, 73:6021–6030, 1980.

[15] R. W. Hesselink, R. B. Koehorst, P. V. Nazarov, and M. A. Hemminga. Membrane-bound peptides mimicking transmembrane Vph1p helix 7 of yeast V-ATPase: a spectroscopic and polarity mismatch study. *Biochimica et Biophysica Acta*, 1716(2):137–145, 2005.

[16] B. K. Ho, A. Thomas, and R. Brasseur. Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the α-helix. *Protein Science*, 12:2508–2522, 2003.

[17] K. C. Jefferies, D. J. Cipriano, and M. Forgac. Function, structure and regulation of the vacuolar (H^+)-ATPases. *Arch Biochem Biophys*, 476 (1):33–42, Aug 2008.

[18] W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, Dec 1983.

[19] P. M. Kane. The where, when, and how of organelle acidification by the yeast vacuolar H^+-ATPase. *Microbiol Mol Biol Rev*, 70(1):177–191, Mar 2006.

[20] M. Katragadda, J. L. Alderfer, and P. L. Yeagle. Assembly of a polytopic membrane protein structure from the solution structure of overlapping peptide fragments of bacteriorhodopsin. *Biophysical Journal*, 81:1029–1036, 2001.

[21] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Arg-735 of the 100-kDa subunit a of the yeast V-ATPase is essential for proton translocation. *Proc. Natl. Acad. Sci. U.S.A.*, 98(22):12397–12402, Oct 2001.

[22] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Proton translocation driven by ATP hydrolysis in V-ATPases. *FEBS Lett*, 545(1):76–85, Jun 2003.

[23] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Interacting helical surfaces of the transmembrane segments of subunits a and c′ of the yeast V-ATPase defined by disulfide-mediated cross-linking. *J Biol Chem*, 278 (43):41908–41913, Oct 2003.

[24] J. A. Killian and T. K. M. Nyholm. Peptides in lipid bilayers: the power of simple models. *Curr Opin Struct Biol*, 16(4):473–479, Aug 2006.

[25] R. Koradi, M. Billeter, and K. Wüthrich. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graphics*, 14: 51–55, 1996.

[26] Z. Kóta, T. Páli, N. Dixon, T. P. Kee, M. A. Harrison, J. B. Findlay, M. E. Finbow, and D. Marsh. Incorporation of transmembrane peptides from the vacuolar H^+-ATPase in phospholipid membranes: Spin-label electron paramagnetic resonance and polarized infrared spectroscopy. *Biochemistry*, 47:3937–3949, 2008.

[27] R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein, and J. M. Thornton. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*, 8(4): 477–486, Dec 1996.

[28] F. Li, H. Li, L. Hu, M. Kwan, G. Chen, Q.-Y. He, and H. Sun. Structure, assembly, and topology of the G185R mutant of the fourth transmembrane domain of divalent metal transporter. *Journal of the American Chemical Society*, 127:1414–1423, 2005.

[29] J. L. Markley. Observation of histidine residues in proteins by means of nuclear magnetic resonance spectroscopy. *Accounts of Chemical Research*, 8:70–80, 1975.

[30] D. Massiot, F. Fayon, M. Capron, I. King, S. Calvé, B. Alonso, J.-O. Durand, B. Bujoli, Z. Gan, and G. Hoatson. Modelling one and two-dimensional solid-state NMR spectra. *Magn Reson Chem*, 40:70–76, 2002.

[31] T. Nishi and M. Forgac. The vacuolar (H^+)-ATPases–nature's most versatile proton pumps. *Nat. Rev. Mol. Cell. Biol.*, 3(2):94–103, Feb 2002.

[32] V. Y. Orekhov, K. V. Pervushin, and A. S. Arseniev. Backbone dynamics of (1–71)bacterioopsin studied by two-dimensional ^{1}H-^{15}N NMR spectroscopy. *Eur. J. Biochem.*, 219:887–896, 1994.

[33] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, and M. Nilges. ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3):381–382, 2007.

[34] S. Soulié, J.-M. Neumann, C. Berthomieu, J. V. Møller, M. le Maire, and V. Forge. NMR conformational study of the sixth transmembrane segment of sarcoplasmic reticulum Ca^{2+}-ATPase. *Biochemistry*, 38:5813–5821, 1999.

[35] N. Sreerama and R. W. Woody. Computation and analysis of protein circular dichroism spectra. *Methods Enzymol*, 383:318–351, 2004.

[36] W. L. Vos, L. S. Vermeer, and M. A. Hemminga. Conformation of a peptide encompassing the proton translocation channel of vacuolar H^+-ATPase. *Biophysical Journal*, 92:138–146, 2007.

[37] Y. Wang, T. Inoue, and M. Forgac. TM2 but not TM4 of subunit c″ interacts with TM7 of subunit a of the yeast V-ATPase as defined by disulfide-mediated cross-linking. *J Biol Chem*, 279(43):44628–44638, Oct 2004.

[38] Y. Wang, M. Toei, and M. Forgac. Analysis of the membrane topology of transmembrane segments in the c-terminal hydrophobic domain of the yeast vacuolar ATPase subunit a (vph1p) by chemical modification. *J Biol Chem*, 283(30):20696–20702, Jul 2008.

[39] D. Wishart, B. Sykes, and F. Richards. The chemical shift index: A fast and simple method for the assingment of protein secondaray structure through NMR spectroscopy. *Biochemistry*, 31:1647–1651, 1991.

[40] H. Yao, R. A. Stuart, S. Cai, and D. S. Sem. Structural characterization of the transmembrane domain from subunit e of yeast F_1F_0 ATP synthase: A helical GXXXG motif located just under the micelle surface. *Biochemistry*, 47:1910–1917, 2008.

[41] P. L. Yeagle and A. D. Albert. Use of nuclear magnetic resonance to study the three-dimensional structure of rhodopsin. *Methods Enzymol.*, 343:223–231, 2002.

5

Solvation by SDS of a peptide derived from H^+-V-ATPase subunit a, studied by molecular dynamics simulations

Contents

Summary

In a previous study, the 3D structure of a peptide derived from the H^+ translocating V-ATPase, subunit *a*, TM7, was determined by NMR spectroscopy in SDS [28]. In that study, it was also shown that the secondary structure of this peptide was pH dependent in octylglucoside, but not in SDS.

In this chapter, the reasons behind those observations are studied in atomistic detail by molecular dynamics simulations. The effect of the pH is simulated by running two simulations: one with uncharged histidine residues and one with charged histidine residues. These simulations are carried out both in SDS and octylglucoside, for 20 ns each. One more simulation was run under the conditions of the previous NMR experiment, with the NOE restraints in place. This fifth simulation can be seen as a further refinement of the NMR structure, but it is also used as a reference to compare to the other, unrestrained, simulations.

In SDS, the peptide takes a position in the hydrophobic interior of the micelle if the histidine residues are uncharged, but it aligns itself to the SDS headgroups on the micelle surface at low pH. In octylglucoside, the peptide does not take a position inside a micelle, regardless of the histidine protonation state. From these data, it is concluded that at high pH, the peptide is protected from aggregation by the SDS micelle, but not in octylglucoside, in agreement with experimental results presented earlier [28]. Furthermore, it is observed that the structure of the peptide is different in the simulation with the NOE restraints which indicates a bias towards α-helices by the other simulations. The structural differences are small, but seem responsible for the increased amount of water contacts that are observed in the simulation with the NOEs.

Finally, the residues that are in contact with the water are identified, in order to locate the face of TM7 that lines the lumenal water channel in V-ATPase. Those residues are Y733, S740, W737 and H743. The R735 residue which is essential for proton transport is located on the other, hydrophobic side of the helix. The hydrophobicity of this face of TM7 fits well with the current opinion that it interacts with the hydrophobic subunit *c*.

A publication of this chapter is in preparation:
L.S. Vermeer, M.A. Hemminga, A. Milon and J. Czaplicki,
Solvation by SDS of a peptide derived from H^+-V-ATPase subunit *a*.
Biophysical Journal

5.1 Introduction

In a previous study, the structure of a peptide derived from V-ATPase subunit *a*, TM7 was studied by NMR and circular dichroism spectroscopy. It was found to be 60% α-helical in SDS, while in octylglucoside it tended to form a less helical structure, or aggregate as β-sheets at high pH [28]. Because the pH influences the charges on the two histidine residues in the peptide, it may also be solvated differently by the detergent at low and high pH. In the previous studies, it was concluded that the charged histidines interacted with the SDS headgroups. This conclusion was based on the fact that the protonated state of the histidine residues was stabilised in SDS, as was determined by measuring the histidine pKa with an NMR titration experiment. In this chapter, we study the solvation of the sidechains by MD simulations. We will compare the results from the simulations to the published experimental data [28]. The primary structure of the peptide is given in table 5.1, the detailed secondary structure may be found in the protein data bank with PDB ID "2RPW". For more details, please refer to our previous study [28] and chapter 4 on page 75.

Table 5.1: *Amino acid sequence of the putative TM7 from yeast V-ATPase subunit a. It has been proposed that this peptide lines the proton channel on the lumenal side of V-ATPase [32]. More details are given in section 4.2.1 on page 79. The residues are numbered as in V-ATPase from Saccheromyces cerevisiae, subunit a, but note that the two lysine residues on both ends are non-native. Residues that are (potentially) positively charged are shown in blue.*

	730	735	740	745		
peptide KMTM7	KKSHTASYLRLWALSLAHAQLSSKK	see also	[10, 28]			

To study the atomistic detail of the interactions between the peptide and its environment, molecular dynamics simulations were carried out. MD simulations have been successfully used in the past to study protein–lipid interactions [19]. Several examples exist in the literature where a detailed description of the interactions between a peptide and detergent was obtained. In a simulation of the antimicrobial peptide RP-1, it was shown that the peptide binds deeper inside an SDS micelle than in dodecylphosphocholine (DPC), and the SDS micelle was severely distorted by the peptide [4]. In another study, the partial unfolding of the Mistic protein has been observed in lauryldimethylamine-oxide (LDAO) micelles, and it was suggested that both

self-assembly and pre-formed micelles were suitable systems for the study [20]. The same study also shows that rather large conformational changes may be observed in the relatively short timescales of MD simulations. A particularly relevant example for this work is the publication with MD simulations of protein–detergent interactions by Bond et al. [3], where the self-assembly of detergent around different types of membrane proteins is studied. They observed that the number of interactomic contacts between detergent and protein did not change significantly after 15 ns. The stabilisation of the system took slightly longer and was estimated to take place between 10 and 45 ns, depending on the protein. Their results give us confidence that protein–detergent interactions can be studied, even on the relatively short timescales of MD simulations. A simulation of a different peptide derived from V-ATPase, called sMTM7, also exists [8]. In that study, the solvation of the peptide sidechains by DMSO was studied in detail. It was argued that DMSO is a suitable membrane-mimicking solvent for helices such as TM7 from V-ATPase that may be in contact with both a hydrophobic and a hydrophilic environment in the native protein.

For the current study, we chose to place the KMTM7 peptide in a box of randomly distributed detergent, so that after several tens of nanoseconds the solvent molecules self-assemble around it. The self-assembly method has been shown to correctly reproduce experimental results about the location of peptides in lipid bilayers [9] and predict the position of transmembrane proteins [26]. The same method is used here to determine how the peptide is solubilised by detergent micelles.

We investigate if the simulations agree with the experimentally determined peptide structure, study the solvation of the peptide by different detergents under different conditions, and determine if the histidine pKa values that were measured in SDS can be explained by the histidine–detergent interactions. Finally, the results are discussed in the light of the putative biological function of TM7 in V-ATPase.

5.2 Materials and methods

5.2.1 Setup of the system

A pool of 99 different detergent molecules was generated by carrying out 50 ps simulations of a detergent molecule in vacuum. From this pool, 128 detergent molecules were randomly chosen (so some conformations were necessarily chosen at least twice) and placed in a box of 6x6x6 nm around the lowest energy structure of the KMTM7 peptide that has been determined by NMR

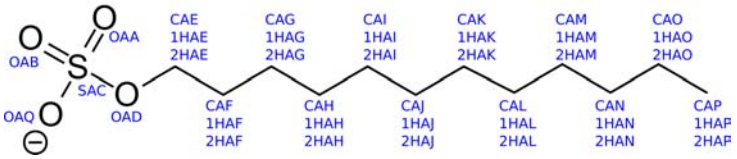

Figure 5.1: *SDS atom names as in table 5.2.*

Table 5.2: *Partial charges of the SDS molecule as calculated by AM1-BCC, compared to values from the literature. Note that the force fields that were used in the cited publications were different from the one used here. Because the charges slightly differ for all atoms in our topology file, only one representative value is given in this table. Further details may be found in appendix D on page 185, where the entire topology file for the detergent is shown.*

atom name	refs [5, 23, 25]	this study
OAA, OAB, OAQ	-0.654	-0.765
SAC	+1.284	+1.698
OAD	-0.495	-0.609
CAE	+0.137	+0.208
CAF–CAO	0.000	-0.079
CAP	0.000	-0.092
hydrogen atoms	0.000	0.040

spectroscopy in SDS (see [28] and chapter 4 on page 75).

5.2.2 Choice of force field and simulation parameters

Although the GROMACS utilities [27] were used for the initial preparation of the starting system as described above, all subsequent steps were carried out with the AMBER software and force fields. It was chosen to carry out the simulations itself using the AMBER 9 software package [6], because the combination of AMBER with topology files generated by antechamber is well established [1, 31].

Structure (pdb) files for SDS and octylglucoside were obtained from the prodrg2 server [24], and the AMBER topology files were created with antechamber [31], using the AM1-BCC [12] method for the calculation of the atomic partial charges. Although values for partial charges of both SDS [25] and octylglucoside [16] exist, there is no a-priori reason why these values are better or worse than our newly calculated ones. We decided to use the partial

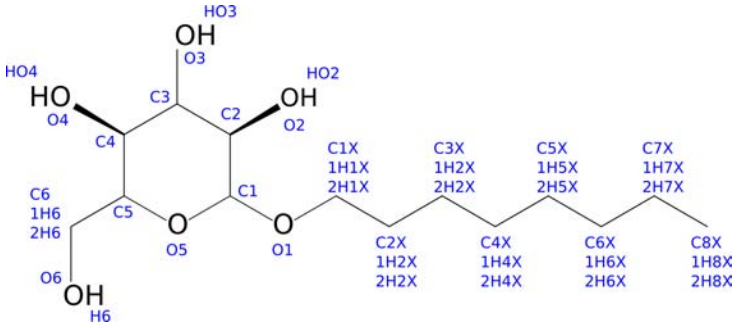

Figure 5.2: *octylglucoside atom names as in table 5.3.*

Table 5.3: *Partial charges of the octylglucoside molecule as calculated by AM1-BCC, compared to values from the literature. Note that the force fields that were used in the cited publications were different from the one used here. Because the charges slightly differ for all atoms, only one representative value is given in this table. All details may be found in appendix D on page 185.*

atom	refs [16, 17]	this study
O1	-0.40	-0.4244
C1	0.20	0.3059
O5	-0.40	-0.4329
C2,C3,C4	0.14	0.1029
C5	0.25	0.1065
O2,O3,O4,O6	-0.66	-0.5973
HO2,HO3,HO4,H6	0.43	0.4255
C6	0.05	0.1440
C1X	-0.01	0.1342
C2X-C7X	-0.18	-0.0756
C8X	-0.27	-0.0922
all other hydrogen atoms	0.09	0.0385

charges from antechamber because this program is intended to be used with AMBER [1], while values from the literature were calculated for different force fields. The ff03 AMBER force field [1, 7, 18] was used for the protein, in combination with the GAFF force field [30] for the detergent molecules. The partial charges of SDS and octylglucoside are given in table 5.2 on the facing page and table 5.3. Water molecules (TIP3P) were added to fill the box and some of them were replaced with Na^+ or Cl^- ions to neutralise the total system charge, leading to a system containing 1 peptide molecule, 128

Table 5.4: *List of the simulations presented in this chapter. Octylglucoside is abbreviated as OG in the table.*

simulation	detergent	NOEs	HIS charge	total peptide charge	number of ions	number of waters
HIS+SDS NOE	128 SDS	yes	2 x +1	+7	121 Na^+	4366
HIS+SDS	128 SDS	no	2 x +1	+7	121 Na^+	4366
HISnSDS	128 SDS	no	0	+5	123 Na^+	4365
HIS+OG	128 OG	no	2 x +1	+7	7 Cl^-	4867
HISnOG	128 OG	no	0	+5	7 Cl^-, 2 Na^+	4865

detergent molecules, and $\sim$ 5000 water molecules. After a minimisation, the system was equilibrated by first equilibrating the temperature using an NVT ensemble during 100 ps and subsequently equilibrating the pressure (box size) with a 1 ns NPT simulation. During these steps, the temperature was kept constant at 300 K and Cartesian position restraints were applied to the peptide. All simulations were carried out in a cubic box with isotropic pressure coupling, a suitable system for simulations of isotropic systems such as detergent micelles in water. Subsequently, a 20 ns production run was carried out using an NPT ensemble at 300 K without Cartesian position restraints. The different simulations are listed in table 5.4, and all details about the simulations are given in appendix D on page 185.

5.2.3 Analysis

The data from the MD simulations were analysed using the VMD software [11], the AMBER utility "ptraj" and several Perl scripts.

5.2.4 Other considerations

A production run with NOE restraints from the NMR structure determination was carried out in SDS. The run with the NOE restraints only makes sense under the same conditions as in the NMR experiment, so it was only carried out in SDS with charged histidine residues. This simulation can be seen as a refinement of the NMR structure, and the results will be compared to the unrestrained simulations and to the structure calculation [28] that was carried out with the ARIA software package [21].

The effect of the histidine charges on the peptide conformation and solvation was studied by four simulations: two in SDS and two in octylglucoside. In both detergents, one simulation was done with histidine residues that were

protonated on both the δ_1 and the ϵ_2 position (and thus positively charged), and one simulation with neutral histidine residues that were only protonated on the ϵ_2 position. The choice to protonate ϵ_2 and not δ_1 was arbitrary (since both forms consist in solution in tautomeric equilibrium), but we worked with the hypothesis that the exact site of protonation does not influence the simulations too much.

We investigated the influence of the histidine charges and the detergent on the peptide structure, its position with respect to the detergent micelle, the solvation of the sidechains and in particular of the histidine sidechains. These results were compared to the available experimental data, and discussed in the light of the position of this peptide within V-ATPase subunit *a*.

5.3 Results

5.3.1 Equilibration and detergent aggregation

To follow the aggregation of the detergent, the average diffusion coefficient of all the detergent molecules is shown in figure 5.3 on the following page. In the first nanosecond, a very steep curve indicates that the free detergent molecules quickly form larger aggregates. The diffusion rate continues to decrease up to a simulation time of around 10 ns, after which the curve starts to flatten out, although in some of the simulations it continues to decrease slightly up to around 15 ns. Note that the final diffusion coefficient is not the diffusion rate of the micelle, but the average diffusion rate over all the detergent molecules. A snapshot of the simulations after 20 ns is shown in figure 5.13 on page 127. The octylglucoside detergent molecules form non-spherical aggregates with the polar heads exposed to the water phase and the hydrophobic tails on the inside. According to experimental and simulated data, octylglucoside is expected to form large cylindrical shaped micelles [16, 33].

The SDS molecules aggregate in a micelle-like structure in the presence of the peptide with the uncharged histidines. The middle of the peptide is located in the hydrophobic interior. In the SDS simulation where the peptide is charged, a cylindrical, almost bilayer-like shape is formed. A distortion of micelles due to the interaction with a peptide has been observed before [4]. More details about the shape of the micelles can be found in the discussion.

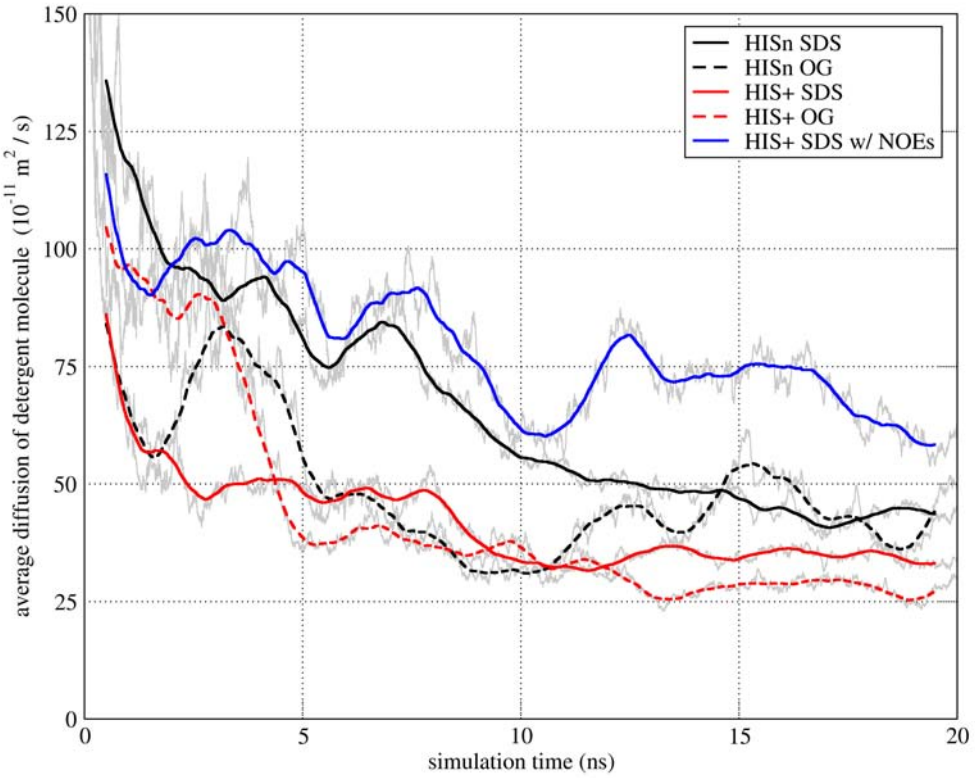

Figure 5.3: *Diffusion coefficient of the detergent, shown as an average over all detergent molecules in the simulation. A very steep decrease in the diffusion coefficient (from* $\sim 600 \cdot 10^{-11}\, m^2/s$*) during the first picoseconds is not visible due to the scale of the figure. The thick lines are running averages over 1 ns, the grey background shows the original data points.*

5.3.2 Peptide structure and dynamics

In figure 5.4 on the facing page, the RMSD of the protein from its starting structure is shown. The RMSD shown was calculated for the backbone residues in the structured region of the protein. In the simulation with the NOE restraints, the peptide structure remains somewhat closer to the initial structure than in the other simulations. This could be expected, because the initial structure was also calculated based on these restraints. The RMSD rises to an average value of around 1.1 Å in the first nanosecond and then remains stable. In the other simulations, an average RMSD between 1.4 and 1.9Å is observed. The large structural changes that are observed in the *HIS+SDS* simulation between 0 and 8 ns are considered part of the equili-

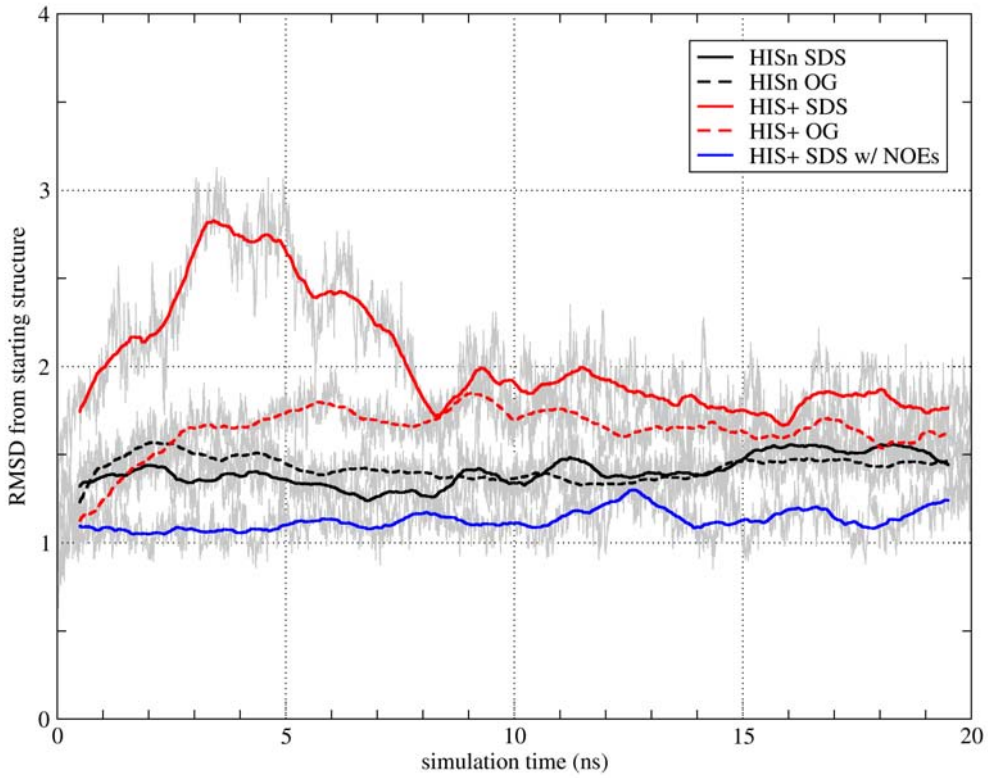

Figure 5.4: *RMSD from the starting structure during the 20 ns simulations. The RMSD was calculated on the backbone atoms of the structured region of the peptide (residue S732–S748). The grey background shows a trace of all the snapshots saved, the smooth lines in the foreground are the running averages over 1 ns of this data.*

bration of the system, and are as such not analysed, but it is interesting to note that such a large change can occur on the relatively short timescale of the simulations. After 8 ns, the RMSD has stabilised in all simulations. All analyses were carried out on the trajectory from 10–20 ns (unless stated otherwise) where the RMSD remains stable. Note that the similar RMSD values during all the simulations do not necessarily indicate that the structures are similar as well. The RMSD of the final structures (snapshots after 20 ns of simulation) is shown in table 5.5 on the next page, as compared to their average structure.

To compare the quality of the structure from the simulations to the initial structure (which was the lowest energy structure from the NMR data presented in chapter 4 on page 75), the procheck_nmr software was used. The

Table 5.5: *The number of NOE violations (> 0.5 Å) involving backbone atoms in the final structures after the different MD simulations. The RMSD (in Å) of residue S732–S748 from the average structure (average over the final snapshot taken from the five simulations) is shown to illustrate the differences between the structures. The last four columns show the percentage of residues that have dihedral angles in the various areas of the Ramachandran plot.*

simulation	NOE	RMSD	favour	allow	gen.al.	disal.
starting structure	6	1.1	52%	35%	13%	1%
HISnOG	15	1.1	96%	4%	0%	0%
HISnSDS	11	0.7	96%	4%	0%	0%
HIS+OG	16	1.1	87%	13%	0%	0%
HIS+SDS	15	1.2	87%	13%	0%	0%
HIS+SDS w/ NOEs	3	1.6	48%	39%	9%	4%

results from that analysis are shown in table 5.5. The slightly lower number of NOE restraint violations in the "new" simulation indicates that the structure is in better agreement with the NMR experimental data. The Ramachandran plot statistics are comparable between the starting structure and the simulation with the active NOE restraints. The non-restrained simulations show statistics that are different from the starting structure and restrained simulation, but close to each other. In the simulations without NOEs, no dihedral angles in the unfavourable regions are observed. These snapshot structures after 20 ns violate around 11–16 NOEs (backbone, > 0.5 Å), indicating that the simulations favour a different conformation than the one obtained with the experimental constraints in place.

When comparing the different simulations, the unordered region of the peptide shows different conformations, probably due to its mobility. This region is indeed the most mobile, as can be seen in figure 5.5 on the facing page where the B-factor (calculated over the last 10 ns of the trajectory) is shown. The structured region of the peptide is relatively rigid, as judged from the small B-factor values for residue Y733–L746 (a typical average B-factor for buried atoms in X-ray structures is $\sim 0.2 \cdot 10^3$ Å^2).

A more quantitative representation of the secondary structure is given by the Kabsch & Sander DSSP secondary structure analysis [13]. The percentage of time in which the peptide adopted a certain conformation is shown in figure figure 5.6 on page 116. Especially noteworthy is that the secondary structure of the ordered region forms α-helices in all simulations without NOE restraints, while mostly hydrogen bonded turns are observed in the NOE restrained simulation.

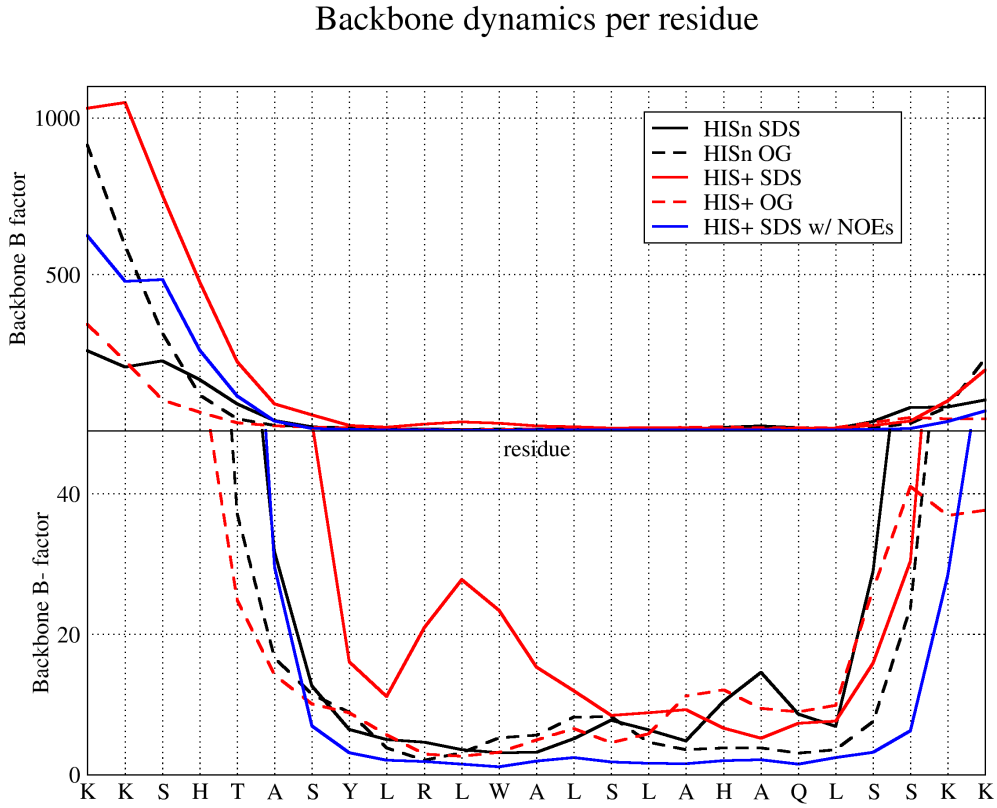

Figure 5.5: *The B-factor (in $\AA^2$ calculated over the last 10 ns of the production run. The B-factor is a measure for the dynamics of the backbone atoms (the squared fluctuations in the atom position multiplied by $(8/3) \cdot \pi^2$). The data shown in the top and bottom figure are the same, with a different scale on the y-axis.*

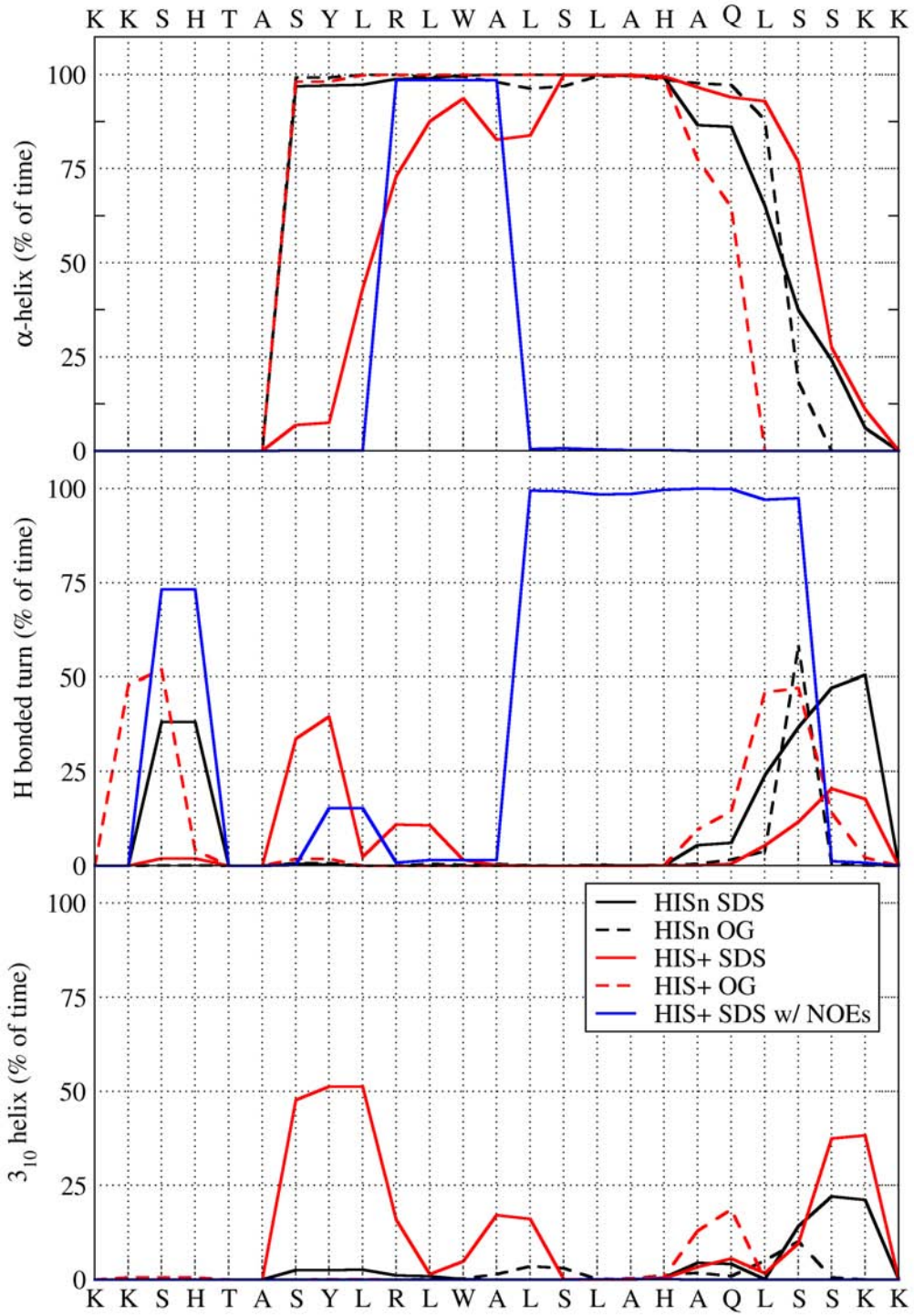

Figure 5.6: *The percentage of time that the peptide had a certain structure during the last 10 ns production run of the different simulations. The secondary structure was assigned using the DSSP Kabsch & Sander protocol [13].*

Effect of histidine charge

As can be seen in figure 5.6 on the facing page, in SDS, the presence of a charge on the histidine residues destabilises the α-helix in favour of a 3_{10} helix and hydrogen bonded turns in the region from residue A731–L739. In octylglucoside, the structure is relatively stable, regardless of the histidine charge. The structure of the C-terminus is slightly different in all simulations.

5.3.3 Solvation of the peptide

Solvation by SDS

Although the solvation of the peptide and its location in the bilayer may be derived from figure 5.13 on page 127, a more quantitative approach is desirable. In figure 5.7 on the following page the average number of water molecules located within 5 Å of each residue is shown. In all simulations, both termini of the peptide are water-exposed. In the simulation in SDS with the uncharged histidine residues, the peptide takes a position where it is shielded from the water, as evidenced by the small number of water molecules in contact with residues A731–A745. It is interesting to note that this region is the structured region of the peptide (figure 5.6 on the preceding page). In the other two SDS simulations, the charged H743 residue has a higher number of contacts with water molecules (figure 5.7 on the following page). Most other residues in the structured region of the peptide are more water-exposed than in the simulation with uncharged histidine residues. The charge of the histidine residues clearly influences the solvation of the peptide in SDS. Residue H729, however, is in contact with many water molecules regardless of its charge, probably because of its location close to the N-terminus of the peptide.

Some similarities between the *HIS+* simulations with and without the NOEs may be observed in the periodicity of the water exposed residues, indicating that the same residues are water-exposed, despite the structural differences. Aside from this similarity, the simulations with and without NOE restraints show some important differences. In the region from A738–Q743, more water molecules are in contact with the peptide in the simulation with the NOE restraints. Interestingly, this is the same region that has a different structure (figure 5.6 on the preceding page). The deviation from an α-helical structure of TM7 appears to create a larger hydrophilic surface.

A closer look at the water exposed sidechains reveals that the charged R735 residue is in contact with water molecules in every simulation, which means that it "snorkels" to the interface region in the *HISnSDS* simulation. Residues that show a significantly higher amount of water exposure than

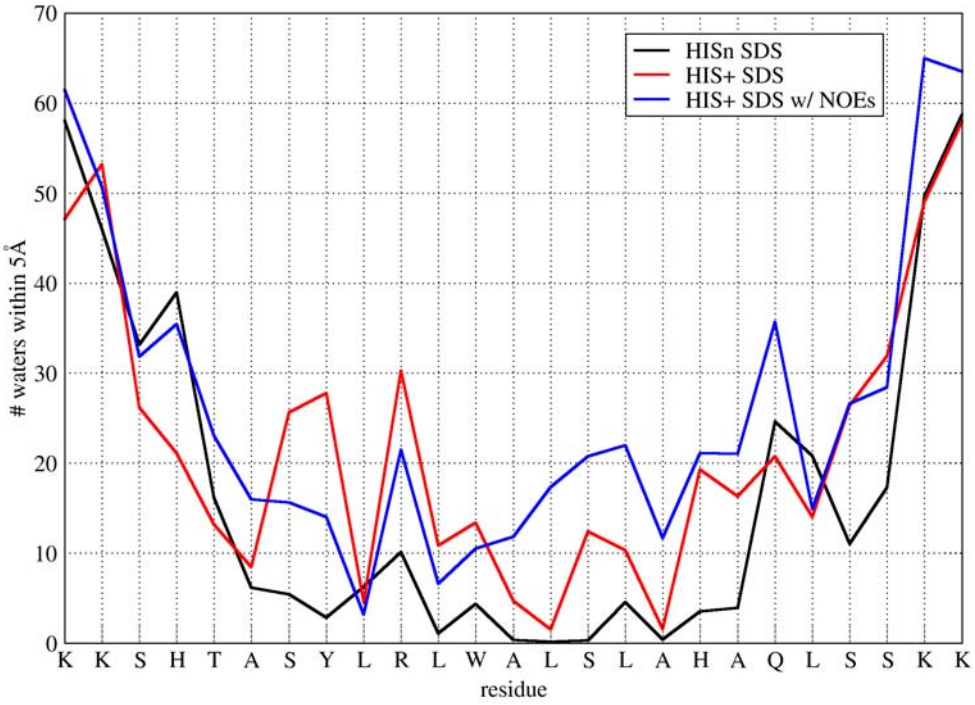

Figure 5.7: *Average number of water molecules within a radius of* 5 *Å from the sidechain of each residue. Averages are calculated over the last 10 ns of the simulations.*

their neighbours in the *HIS+* simulations, are R735, S740, L741, H743 and Q745. In figure 5.8 on the facing page, the solvation of the peptide by the SDS headgroup oxygen atoms is shown. The R735 sidechain is clearly in contact with SDS headgroups in every simulation. The contacts with SDS headgroups follow the same pattern as the water contacts, exposing residues S732, R735, S740 and H743 to more SDS headgroups than their neighbours. The charged histidine residues have more contacts with SDS headgroups that the uncharged ones. The solvation of the peptide with uncharged histidine residues by the SDS headgroups indicates that the structured region of this peptide is located in the hydrophobic micelle interior, with the arginine residue snorkeling to the SDS headgroups, and both termini in the interface region. The observation that the number of contacts with SDS headgroups increases towards the termini indicates that they are not sticking into the

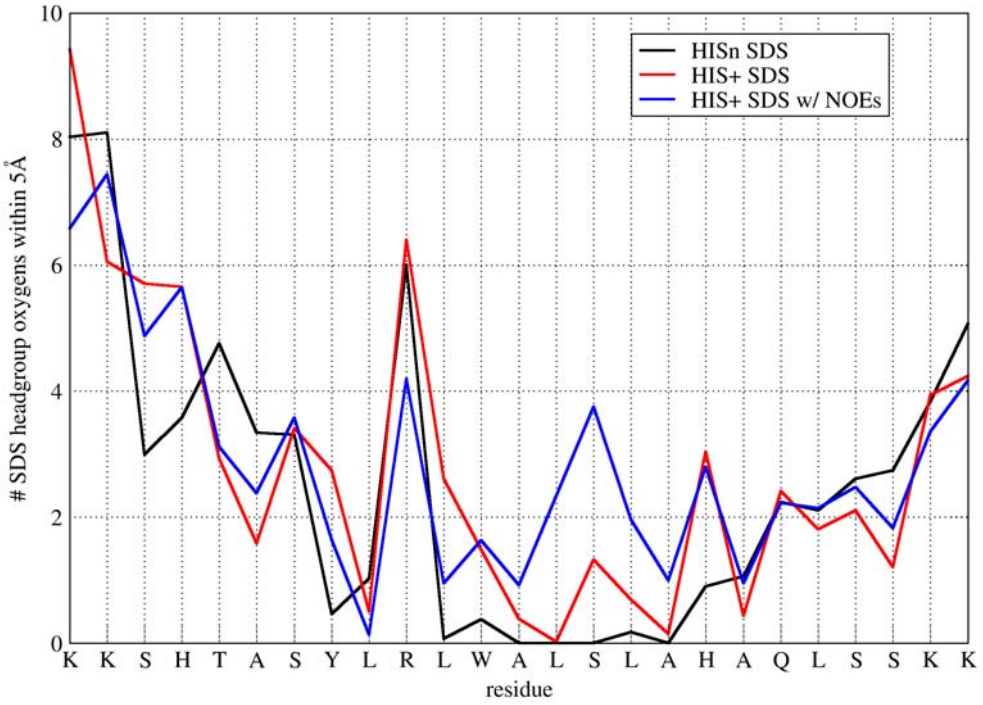

Figure 5.8: *Average number of SDS headgroup oxygen atoms within a radius of 5 Å from each residue. Averages are calculated over the last 10 ns of the simulations.*

water phase, but remain close to the micelle surface.

Solvation by octylglucoside

In octylglucoside, H743 is hydrated by more water molecules (figure 5.9 on the next page) and octylglucoside headgroups (figure 5.10 on page 121) when charged. Residue H729 shows more contacts with detergent headgroups when charged, but not with waters. The same phenomenon was observed in the simulations in SDS. The periodicity in the number of water–residue contacts is comparable to that in SDS.

An important difference with the simulations in SDS is that the uncharged peptide is not located in a hydrophobic micelle interior in octylglucoside. In the simulations where the histidine residues are charged, the number of contacts with water is similar in octylglucoside and SDS. Nearly all residues

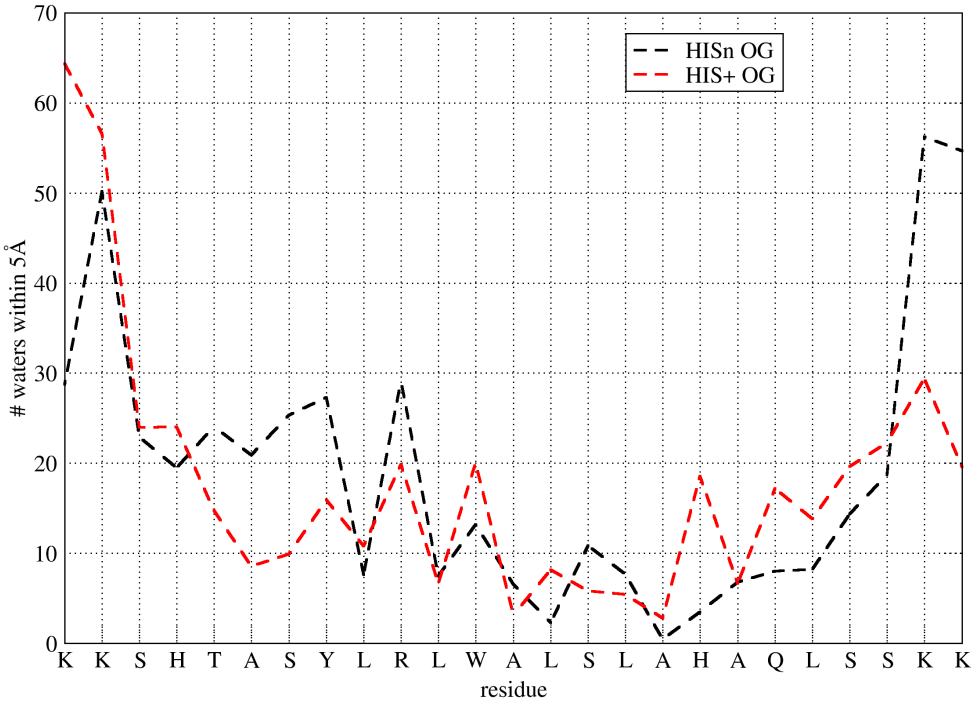

Figure 5.9: *Average number of water molecules within a radius of* 5 *Åof each residue. Averages are calculated over the last 10 ns of the simulations.*

are in contact with 5 or more water molecules, and only residue L739–A744 seem to be somewhat shielded from the water. All residues are within 5 Å of at least two hydroxyl-groups from the octylglucoside, except for residue A731, S732 and L734 in the *HISn* simulation, and residue L736 in the *HIS+* simulation. The increase in peptide–headgroup contacts towards the termini that was observed in SDS in every simulation is not visible in octylglucoside. Finally, it is noteworthy that the peaks (that indicate a higher number of interactions with octylglucoside headgroups than the neighbouring residues) occur at both hydrophobic and hydrophilic residues. This indicates that the octylglucoside headgroups can solvate both residue types.

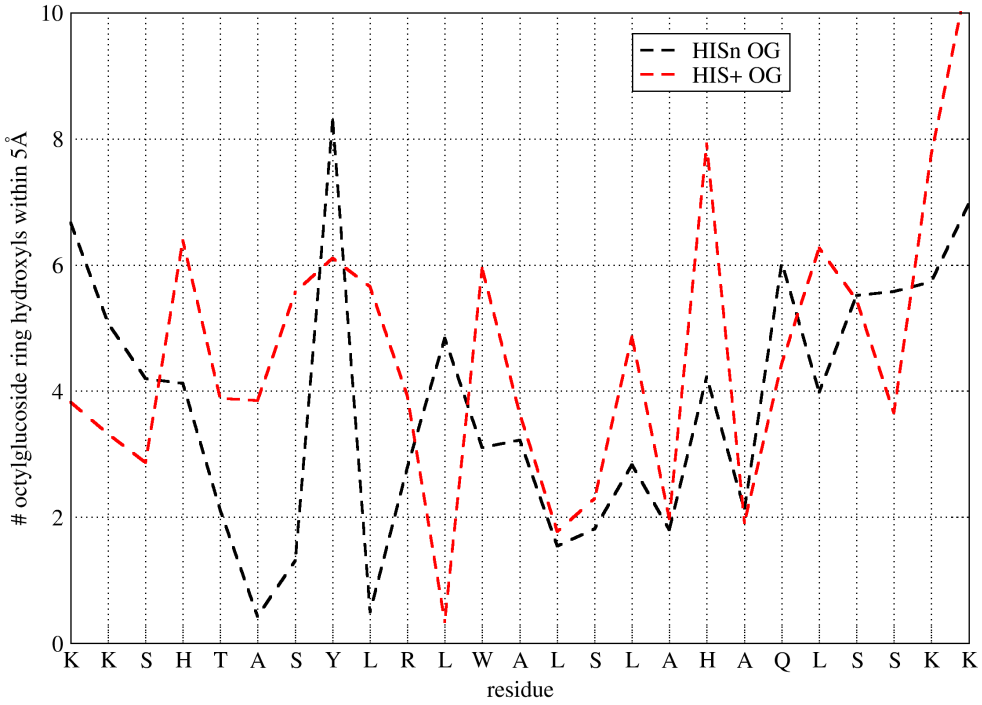

Figure 5.10: *Average number of octylglucoside ring hydroxyl atoms within a radius of* 5 *Å of each residue. Averages are calculated over the last* 10 *ns of the simulations.*

5.4 Discussion

5.4.1 Peptide structure

In our simulations, both the peptide structure and detergent diffusion rate stabilise after around 10 ns, and it has been shown before that 15 ns are enough to study protein–detergent interactions [3]. Although the simulations are too short to allow a complete unfolding or different refolding of the peptide, some structural differences were observed between the different simulations.

The simulation with the NOE restraints may be seen as a further refinement of the NMR structure. The structural differences between the simulation with AMBER and the previous simulation with ARIA [28] could be due to the change of force field and simulation software, but are most prob-

ably also a consequence of the fact that the new simulation is carried out in SDS instead of water, and thus better reproduces the real NMR sample. In any case, the analysis of the dihedral angles and NOE violations indicates that the "new" structure may be considered an improvement (table 5.5 on page 114) with respect to the old one. The refinement of an NMR structure with simulations that explicitly include the micellar environment during a NOE-restrained MD simulation has not often been published [22].

The two biggest differences observed in the peptide structure (figure 5.6 on page 116) between the different simulations, are (1) the conformational change from an α-helix to a 3_{10} helix and hydrogen bonded turns in the region from A731–L739, that occurs in the *HIS+SDS* simulation, and (2), the fact that the secondary structure of the peptide with NOE restraints shows hydrogen bonded turns, while all unrestrained simulations favour an α-helical structure. Later on in the discussion we will further analyse the structural differences between the simulation in SDS with charged and uncharged histidine residues. As for observation (2): apparently, the force field favours α-helical structures, while the NMR-determined constraints suggest a distorted helix (hydrogen-bonded turns). This raises the question if the force field used is biased towards helical peptides [2], even though the ff03 force field [7, 18] is known to have a decreased preference for helical configurations [1] compared to the older ff99 force field [29].

The mobility of the peptides in the simulations was represented by the B-factor calculated over the last 10 ns of the trajectory (figure 5.5 on page 115). The region from Y733–L746 remains stable in all simulations, but the mobility of the N-terminal region differs somewhat between the simulations. There is no clear correlation between this mobility and the detergent or the histidine charge.

5.4.2 Peptide solvation by detergent

In all of the simulations, both termini of the peptide were water exposed. This could be expected because of the presence of the two positively charged lysine residues. The structured region of the peptide takes a position inside the hydrophobic interior of the SDS micelle at high pH, while at low pH when the histidine residues are charged, the peptide positions itself in the interface between the SDS headgroups and water (see figure 5.13 on page 127, figure 5.7 on page 118 and figure 5.8 on page 119). In octylglucoside however, many residues remain water exposed in both simulations (figure 5.9 on page 120).

Circular dichroism measurements [28] have shown that the peptide was 60% helical in SDS, regardless of the pH. This is in very good agreement with the simulations. In octylglucoside, the percentage of helix was lower

than in SDS, and at high pH the peptides aggregated as β-sheets. That no β-sheets were observed in the simulations is not surprising, because β-sheets require peptide–peptide interactions and the system contains only one peptide molecule. An explanation for the differences observed by CD may be that the peptide with uncharged histidine residues (high pH) is located in the hydrophobic interior of the SDS micelle (figure 5.7 on page 118), while in octylglucoside it remains water exposed (figure 5.9 on page 120), and peptide–peptide interactions can still occur. The SDS micelle protects the peptides from aggregating as beta sheets.

5.4.3 Micelle formation

In figure 5.13 on page 127 it is visible that the detergent molecules aggregate, but they do not always form nice, round micelles. This has been observed before for both octylglucoside[16] and SDS micelles in the presence of peptides [4].

In the simulations with SDS, micelle-like structures are formed in the presence of the uncharged peptide, but in the simulation with the charged peptide the SDS aggregate is cylindrical. This cylindrical shape interacts with itself though the periodic boundary conditions, and possibly this is the only way for it to remain stable. Because of the interaction through the periodic boundary conditions, the structure in the *HIS+SDS* simulation must be considered an artifact, but it is not likely to influence the packing of the detergent around the peptide.

A titration of the histidine residues that was carried out in water and in SDS indicated that the pKa of H729 and H743 was nearly equal in SDS (7.6 and 7.7), while the pKa values in water were 5.5 and 6.3[28]. We concluded that in water the residue with pKa 5.5 is H729, because of its position close to the two positively charged lysine residues on the N-terminal end of the peptide. In the simulations, it was observed that both histidine residues have an increased number of contacts with SDS headgroups when charged, which explains their stabilised protonated state and thus their higher pKa in SDS. The average number of SDS oxygen atoms within 5 Å of the charged histidine residues is 5.7 for H729 and 3.0 for H743 (figure 5.8 on page 119). The negative SDS headgroup charges near H729 probably compensate for the proximity of the positively charged lysines, and the polarity of the environment of residue H729 and H743 becomes equal, leading to an identical pKa.

The data in figure 5.7 on page 118 indicate that the peptide is in contact with water when the histidine residues are charged, while with uncharged histidines the entire helical region from Y733–H743 is only in contact with

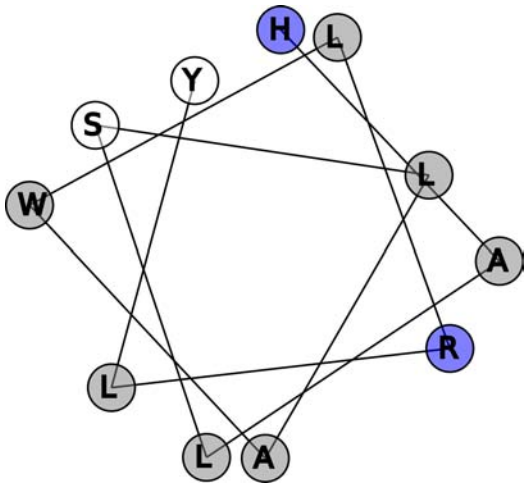

Figure 5.11: *Helical wheel representation of the ordered region of the structure-snapshot after 20 ns of simulation in SDS with charged histidine residues and NOE restraints. The peptide was manually aligned to the z-axis, and the C_α atoms in the xy-plane are drawn as circles. Positively charged residues are shown in blue, hydrophobic residues in grey.*

the hydrophobic tails of SDS. When the peptide structure is restrained by the NOEs however, it becomes even more water-exposed, possibly due to the large angle between the positive charges on R735 and H743 (where the angle is defined in the *xy*-plane, between the C_α atom of the two residues, when the principal axis of the helix is taken as the *z*-axis).

5.4.4 The lumenal proton channel of V-ATPase

With the exception of the HISnSDS simulation, a clear periodicity is observed in the peptide–water and peptide–detergent contacts. This seems to indicate a distinct preference of certain residues to be in contact with water. The fact that in most simulations the same residues are water exposed, could suggest that these residues are also water exposed in the native V-ATPase protein. It is likely that residues that are exposed to the hydrophobic detergent tails are in contact with other hydrophobic helices in the native protein, while residues that are water exposed in the simulations are probably in contact with other polar residues or water. Because TM7 from subunit *a* is believed to line the lumenal hemichannel that allows water to access the glutamic acid residue on subunit *c*, it would be interesting if we could identify the residues that form this channel as far as TM7 is concerned.

To study the amphipaticity of the KMTM7 peptide, the helical wheel rep-

resentation was drawn and compared to the information about the solvation. The helical wheel is shown in figure 5.11 on the preceding page. A Perl/PDL script was written to generate these images from pdb-files (see appendix A on page 151), so the helical wheel does not represent the idealised version of an α-helix, but the actual structure after the simulations. In figure 5.12 on the next page, the solvation of the peptide in the *HIS+SDS* simulations with NOEs is shown as a rendered image of the peptide and its environment within 5 Å. From that figure, and figure 5.11 on the facing page, it can be seen that the hydrophilic residues Y733, S740 and H743 are located on the same side of the helix. From the NMR structure, it is evident that residue W737 is also located on this side, and that its ring is bent away from the face of the helix that contains the polar residues, creating a hydrophilic face. The essential R735 residue [14] is located on the other side of the helix, surrounded by hydrophobic residues. This agrees with the observation that R735 interacts with the hydrophobic *c*-subunit [15].

This is the first time that the solvation of TM7 from V-ATPase subunit *a* is studied in SDS, a study which was based on the high resolution NMR structure of this peptide [28]. The finding that a hydrophilic face of the helix is located on the opposite side of the essential R735, which is surrounded by hydrophobic residues, is interesting. TM7 has been proposed to line the lumenal hemichannel, and the hydrophilic face of the helix we identified is a likely candidate for this job. It suggests that the access of protons to the glutamic acid residue on subunit *c* is not created by a direct proton channel along the side of TM7 that contains R735 and faces subunit *c*. The channel possibly involves the hydrophilic residues that are located on the same side of the helix and solvated by water: Y733–R735–W737–S740 and H743.

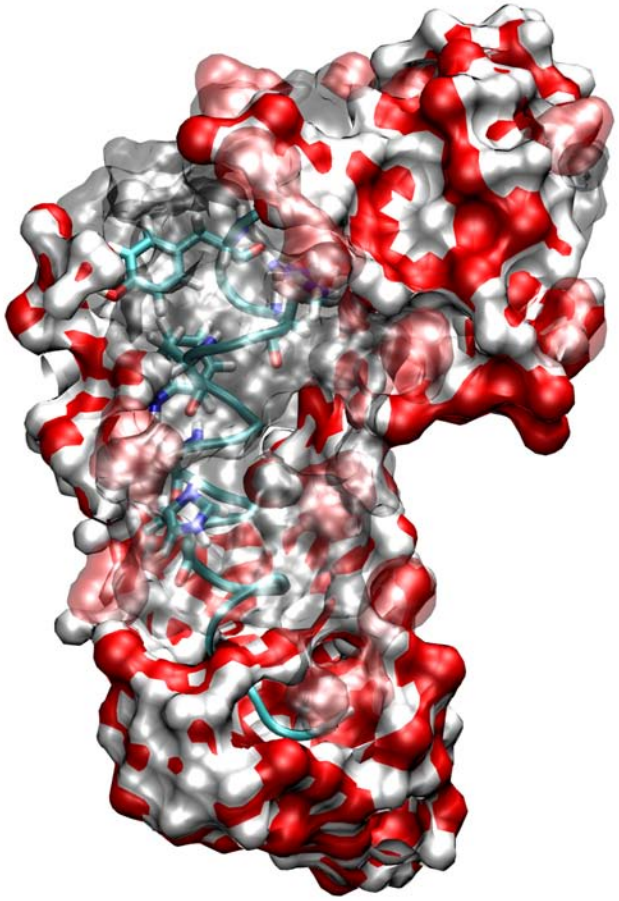

Figure 5.12: *Rendered image (prepared with VMD [11], rendered with POVray), of the solvation of the KMTM7 peptide. The structure shown is the final NMR structure after 20 ns of refinement with NOE restraints and charged histidine residues, in SDS. Residues shown are Y733, R735, W737, S740 and H743. Hydrophobic SDS tails are visible in translucent grey, SDS headgroups oxygen atoms are translucent red and water molecules are shown in a red/white surface representation. The N-terminus of the peptide is not visible because it is located in the water on top of the image. All SDS and water within 5 Å of the peptide is shown.*

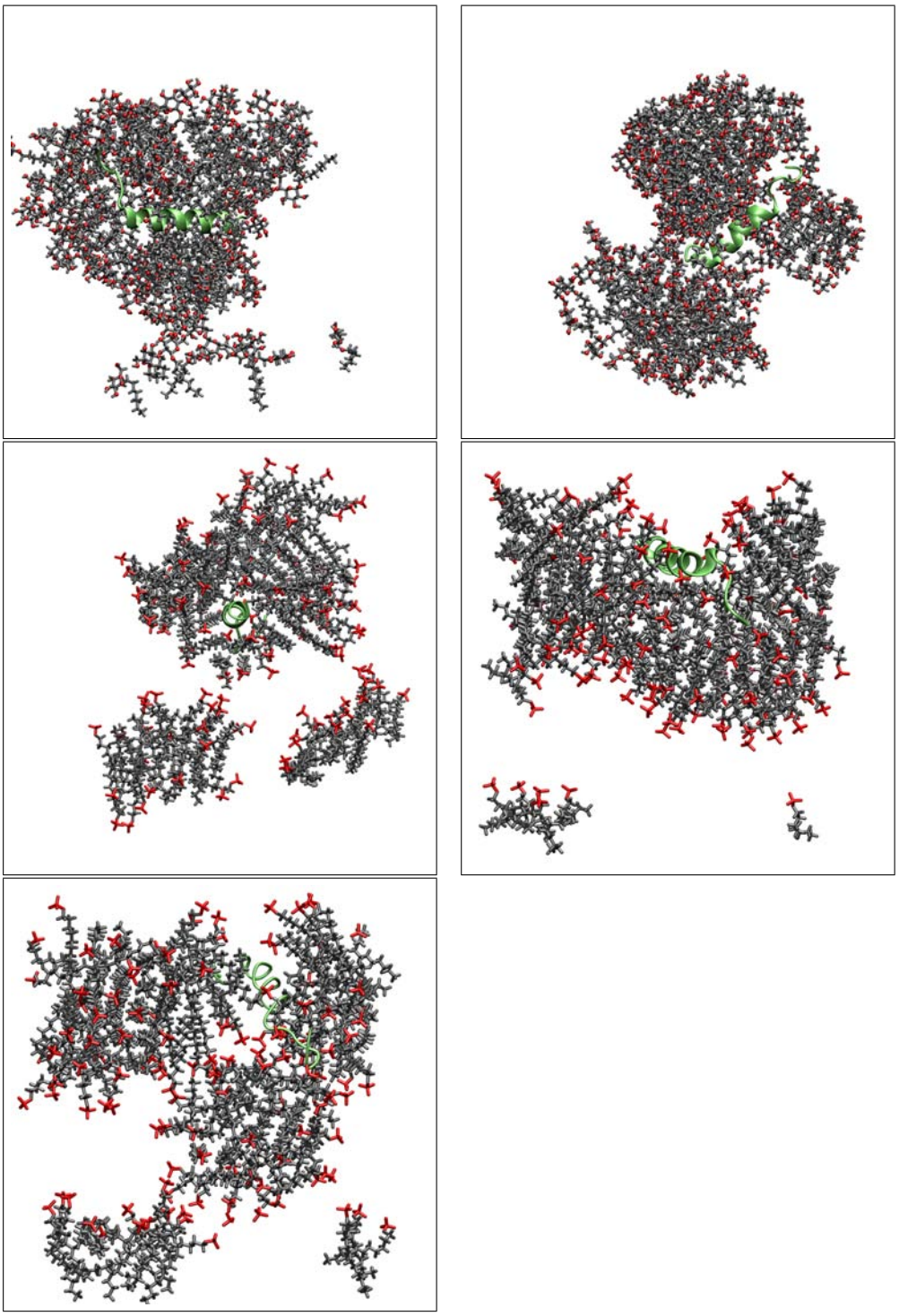

Figure 5.13: *Snapshots of the peptide and detergent molecules after 20 ns of simulation, shown to give an indication of micelle shape and size. Top left: HISnOG, top right: HIS+OG, Middle left: HISn SDS, middle right: HIS+SDS, Bottom: HIS+SDS with NOE restraints. Note that the simulations used periodic boundary conditions, which is why some detergent molecules appear to be "lonely" on the other side of the image, while in fact they are part of the detergent aggregate.*

Bibliography

[1] Amber 9 users' manual.

[2] R. B. Best, N.-V. Buchete, and G. Hummer. Are current molecular dynamics force fields too helical? *Biophys J*, 95(1):L07–L09, Jul 2008.

[3] P. J. Bond, J. M. Cuthbertson, S. S. Deol, and M. S. P. Sansom. Md simulations of spontaneous membrane protein/detergent micelle formation. *J Am Chem Soc*, 126(49):15948–15949, Dec 2004.

[4] S. Bourbigot, E. Dodd, C. Horwood, N. Cumby, L. Fardy, W. H. Welch, Z. Ramjan, S. Sharma, A. J. Waring, M. R. Yeaman, and V. Booth. Antimicrobial peptide RP-1 structure and interactions with anionic versus zwitterionic micelles. *Biopolymers*, 91(1):1–13, Jan 2009.

[5] C. D. Bruce, M. L. Berkowitz, L. Perera, and M. D. E. Forbes. Molecular dynamics simulation of sodium dodecyl sulfate micelle in water: Micellar structural characteristics and counterion distribution. *J. Phys. Chem. B*, 106(15):3788–3793, 2002.

[6] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *J Comput Chem*, 26(16):1668–1688, Dec 2005.

[7] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem*, 24(16):1999–2012, Dec 2003.

[8] A. M. S. Duarte, C. P. M. van Mierlo, and M. A. Hemminga. Molecular dynamics study of the solvation of an alpha-helical transmembrane peptide by DMSO. *J Phys Chem B*, 112(29):8664–8671, Jul 2008.

[9] S. Esteban-Martín and J. Salgado. Self-assembling of peptide/membrane complexes by atomistic molecular dynamics simulations. *Biophys J*, 92 (3):903–912, Feb 2007.

[10] R. W. Hesselink, R. B. Koehorst, P. V. Nazarov, and M. A. Hemminga. Membrane-bound peptides mimicking transmembrane Vph1p helix 7 of yeast V-ATPase: a spectroscopic and polarity mismatch study. *Biochimica et Biophysica Acta*, 1716(2):137–145, 2005.

[11] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[12] A. Jakalian, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J Comput Chem*, 23(16):1623–1641, Dec 2002.

[13] W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, Dec 1983.

[14] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Arg-735 of the 100-kDa subunit a of the yeast V-ATPase is essential for proton translocation. *Proc. Natl. Acad. Sci. U.S.A.*, 98(22):12397–12402, Oct 2001.

[15] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Interacting helical surfaces of the transmembrane segments of subunits a and c′ of the yeast V-ATPase defined by disulfide-mediated cross-linking. *J Biol Chem*, 278 (43):41908–41913, Oct 2003.

[16] P. Konidala, L. He, and B. Niemeyer. Molecular dynamics characterization of n-octyl-β-D-glucopyranoside micelle structure in aqueous solution. *J Mol Graph Model*, 25(1):77–86, Sep 2006.

[17] M. Kuttel, J. W. Brady, and K. J. Naidoo. Carbohydrate solution simulations: producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J Comput Chem*, 23(13): 1236–1243, Oct 2002.

[18] M. C. Lee and Y. Duan. Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins*, 55(3): 620–634, May 2004.

[19] E. Lindahl and M. S. P. Sansom. Membrane proteins: molecular dynamics simulations. *Curr Opin Struct Biol*, 18(4):425–431, Aug 2008.

[20] E. Psachoulia, P. J. Bond, and M. S. P. Sansom. MD simulations of Mistic: conformational stability in detergent micelles and water. *Biochemistry*, 45(30):9053–9058, Aug 2006.

[21] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, and M. Nilges. ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3):381–382, 2007.

[22] S. Rodziewicz-Motowidło, P. Czaplewska, E. Sikorska, M. Spodzieja, and A. S. Kołodziejczyk. The arctic mutation alters helix length and type in the 11-28 beta-amyloid peptide monomer-CD, NMR and MD studies in an SDS micelle. *J Struct Biol*, 164:199–209, Aug 2008.

[23] M. Sammalkorpi, M. Karttunen, and M. Haataja. Structural properties of ionic detergent aggregates: a large-scale molecular dynamics study of sodium dodecyl sulfate. *J Phys Chem B*, 111(40):11722–11733, Oct 2007.

[24] A. W. Schüttelkopf and D. M. F. van Aalten. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 8):1355–1363, Aug 2004.

[25] K. Schweighofer, U. Essmann, and M. Berkowitz. Simulation of sodium dodecyl sulfate at the water–vapor and water–carbon tetrachloride interfaces at low surface coverage. *Journal of Physical Chemistry B*, 101: 3793–3799, 1997.

[26] K. A. Scott, P. J. Bond, A. Ivetac, A. P. Chetwynd, S. Khalid, and M. S. P. Sansom. Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure*, 16(4):621–630, Apr 2008.

[27] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible and free. *J Comput Chem*, 26:701–1719, 2005.

[28] L. S. Vermeer, V. Réat, M. A. Hemminga, and A. Milon. Structural properties of a peptide derived from h^+-V-ATPase subunit *a*. *Biochimica et Biophysica Acta, Biomembranes*, *in press*, 2009.

[29] J. Wang, P. Cieplak, and P. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21: 1049–1074, 2000.

[30] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J Comput Chem*, 25(9):1157–1174, Jul 2004.

[31] J. Wang, W. Wang, P. A. Kollman, and D. A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model*, 25(2):247–260, Oct 2006.

[32] Y. Wang, M. Toei, and M. Forgac. Analysis of the membrane topology of transmembrane segments in the c-terminal hydrophobic domain of the yeast vacuolar ATPase subunit a (vph1p) by chemical modification. *J Biol Chem*, 283(30):20696–20702, Jul 2008.

[33] R. Zhang, P. Marone, P. Thiyagarajan, and D. Tiede. Structure and molecular fluctuations of n-alkyl-β-D-glucopyranoside micelles determined by x-ray and neutron scattering. *Langmuir*, 15(22):7510–7519, 1999.

6

Preparation of MscL in oriented bilayers for solid-state NMR experiments

Contents

Summary

Around 300 mg of uniformly ^{15}N labelled MscL has been produced. It has been successfully purified and reconstituted in lipid bilayers, in order to prepare oriented samples for solid-state NMR studies. So far, no oriented samples have been obtained, as is shown by ^{31}P solid-state NMR. Some suggestions are made for possible improvements in the preparation of oriented samples with MscL in lipid bilayers.

6.1 Introduction

Although MscL has been studied extensively [4] by patch-clamping techniques [9, 21], different fluorescence methods [15], X-ray crystallography [3] and electron spin resonance [12], at the time of writing no NMR studies of MscL exist.

An important difficulty in NMR studies of membrane proteins is the sample preparation, although much progress has been made recently [1]. Because these proteins are generally not water-soluble, they have to be studied in membrane-mimicking systems [13] such as detergent micelles [22], or bicelles [11, 18]. Another interesting possibility is the study of MscL in oriented lipid bilayers by solid-state NMR [2, 7]. Especially for a protein like MscL, that is known to gate by directly sensing protein–lipid interactions [5, 17], studies in lipid bilayers are important. Solid state NMR is the only method currently available that is capable of obtaining high resolution 3D structural information from membrane proteins in lipid bilayers [10].

In this chapter, the possibility for the preparation of oriented samples with MscL, reconstituted in lipid bilayers is explored. Unfortunately, the method described here resulted in non-oriented samples.

6.2 Materials and methods

6.2.1 Production

Tb-MscL was produced using *E. coli* bacteria that were generously donated by Dr. A.M. Powl and Prof. A.G. Lee. These bacteria were BL21(DE3)pLysS, containing a pET-19b plasmid (Novagen) with the MscL gene inserted between the NdeI and BamHI restriction sites. This construct places the protein under control of a T7 promoter and permits to induce protein expression with isopropyl β-D-1-thiogalactopyranoside (IPTG). The protein is expressed with a tag of 10 histidine residues at the N-terminus, which may be removed by using an Enterokinase cleavage site. In all experiments we used the F80W mutant of Tb-MscL, because this allowed us to easily check the reconstitution in lipid vesicles by fluorescence measurements. Further details can be found in the literature [14].

MscL without ^{15}N labelling was produced in rich medium for testing the purification protocol. For the production of unlabelled MscL, a preculture of 50 ml "Terrific Broth" medium (Euromedex), containing chloramphenicol (25 μg/ml) and ampicillin (100 μg/ml) was inoculated with 50 μl of the glycerol stock. The inoculated medium was incubated overnight at 37°C in a shaking incubator with the shaking speed set to 190 rpm. The following day, typically 5 two litre flasks with 800 ml TB medium each, were inoculated with the preculture so that the optical density at 600 nm was 0.08. The production cultures contained the antibiotic ampicillin. They were incubated at 37°C in a shaking incubator at 210 rpm. Protein production was induced when the optical density of the cultures reached a value of 1.0, by addition of IPTG to a final concentration of 1 mM. A typical growth curve of a production culture is shown in figure 6.1 on the next page.

6.2.2 Purification

All centrifugations mentioned were carried out at 4°C. After two hours of induced protein expression, the cells were harvested by centrifuging for 20 minutes at 5000 g, and the cell pellets were stored at -80°C overnight. The next day, the cell pellets were re-suspended in Buffer 1 (20 mM HEPES, 100 mM KCl, pH 7). We used 10 ml buffer per 10 g wet cell weight, with additionally added: 10 mM EDTA and 250 μM TCEP. The cell suspension was homogenised using an ultra thurax sample homogeniser and broken by two passes through a French press at a pressure of 25000 PSI. The broken cells were centrifuged at 100000 g for 30 minutes to precipitate the membrane fragments. The supernatant was discarded and the membrane fragments were

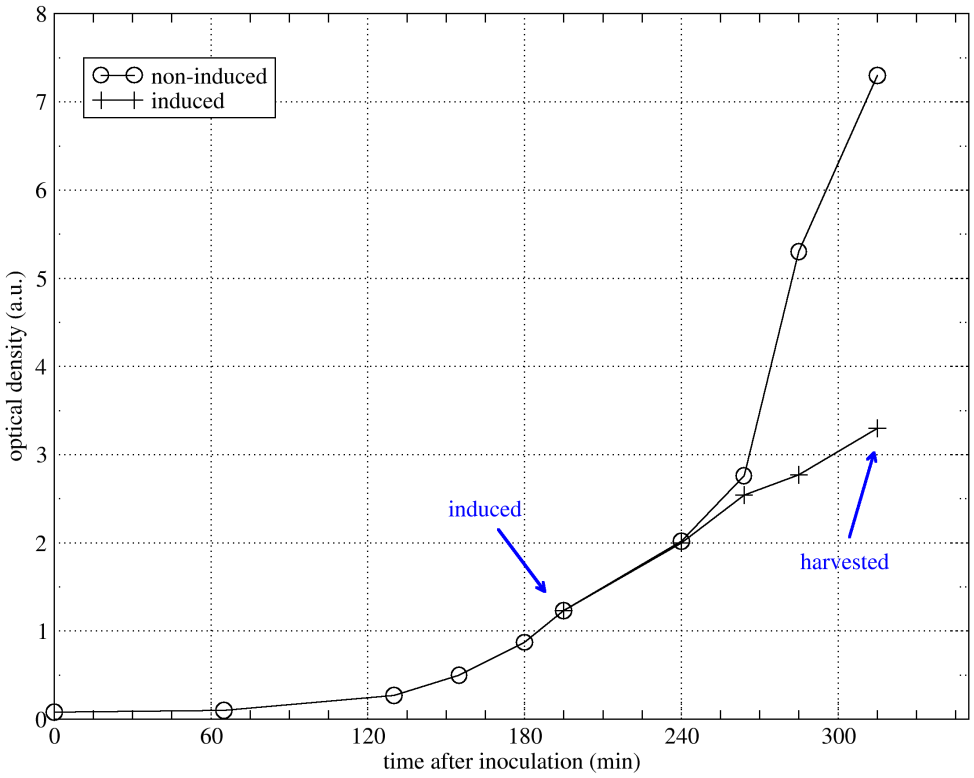

Figure 6.1: *Typical growth curves of a MscL production culture in rich medium, with and without induction of the protein expression.*

resuspended in 100 ml of Buffer 1 with 40 mM octylglucose. After homogenisation of this mixture, it was left overnight at 4°C to solubilise the membrane fragments. The next day, the sample was centrifuged at 28000 g for 30 minutes to precipitate the non-soluble fraction. The pellet of the non-soluble fraction was small compared to the pellet after breaking of the cells. It was discarded after analysis on gel.

Quiagen Ni-NTA agarose (5 ml) was washed with buffer and added to the supernatant, which was then left on a stirring plate overnight at 4°C. The next day, the Ni-NTA was filtered out and poured into a glass tube with a diameter of 1 cm. The Ni-NTA agarose column was washed with 20 ml Buffer 1, containing 40 mM octylglucoside and 35 mM imidazole, to remove any non-specifically bound protein. Finally, the MscL was eluted with 10 ml Buffer 1 containing 40 mM octylglucoside and 400 mM imidazole. Fractions of 1 ml were collected. To avoid aggregation, the fractions were collected in

2 ml tubes that already contained 1 ml of Buffer 1 with 40 mM octylglucoside, in order to dilute the protein immediately. The purity of the sample and possible loss of MscL during the purification process were verified by SDS-PAGE, as shown in figure 6.2 on the facing page and figure 6.3 on page 140. The fact that two-bands are visible around 20 kDa is characteristic for the channel, and has not been explained [3, 6, 20]. The protein concentration in the fractions was measured by recording the absorption of UV-light at 280 nm with a nanodrop spectrophotometer. Protein-containing fractions were mixed and frozen at -80°C until use, usually within the next week.

A similar protocol was used for the production of ^{15}N labelled MscL, with the following changes:

- M9 minimal medium [19] was used for the production culture
- The production culture was grown in a 50 l fermentation chamber, by GTP technologies, Toulouse
- The harvested cells were stored at -80°C instead of overnight, until used

6.2.3 Reconstitution in lipid vesicles

Successful reconstitution of MscL in closed lipid vesicles has been shown previously [16], and was adapted for the experiment described here. POPC-d31 and POPG were mixed in chloroform, in a 80/20 molar ratio. The lipids were dried down using a stream of Nitrogen gas, and possible traces of solvent were removed by leaving the samples at least 2 hours under a vacuum. One millilitre of 20 mM HEPES buffer with 40 mM octylglucoside, at pH 7.0, was added to the lipids, and the lipids were solubilised by sonication in a bath sonicator. An amount of the MscL fractions resulting from the purification was added to the lipids to get a protein/lipid molar ratio of 1/50. This has been shown to be the maximum ratio for successful reconstitution [14]. Typically, the addition of about 5 ml of the fractions was needed. The resulting sample was left to equilibrate for one hour at 4°C, which is above the gel to liquid-crystalline phase transition temperature for POPC/POPG mixtures. After that, 0.7 mg of bio-beads (Bio-rad SM2) was added to remove the detergent, and after one hour the bio-beads were removed by carefully pipetting off the solution. Another 0.7 mg bio-beads was added, removed again after one more hour of interaction, and centrifuged at $165 \cdot 10^3$ g for 30 minutes, to precipitate the proteoliposomes in order to concentrate the sample.

The pellet was resuspended in 500 μl pure (milliQ) water and dialysed overnight against 4 litre milliQ water at 4°C to remove possible traces of

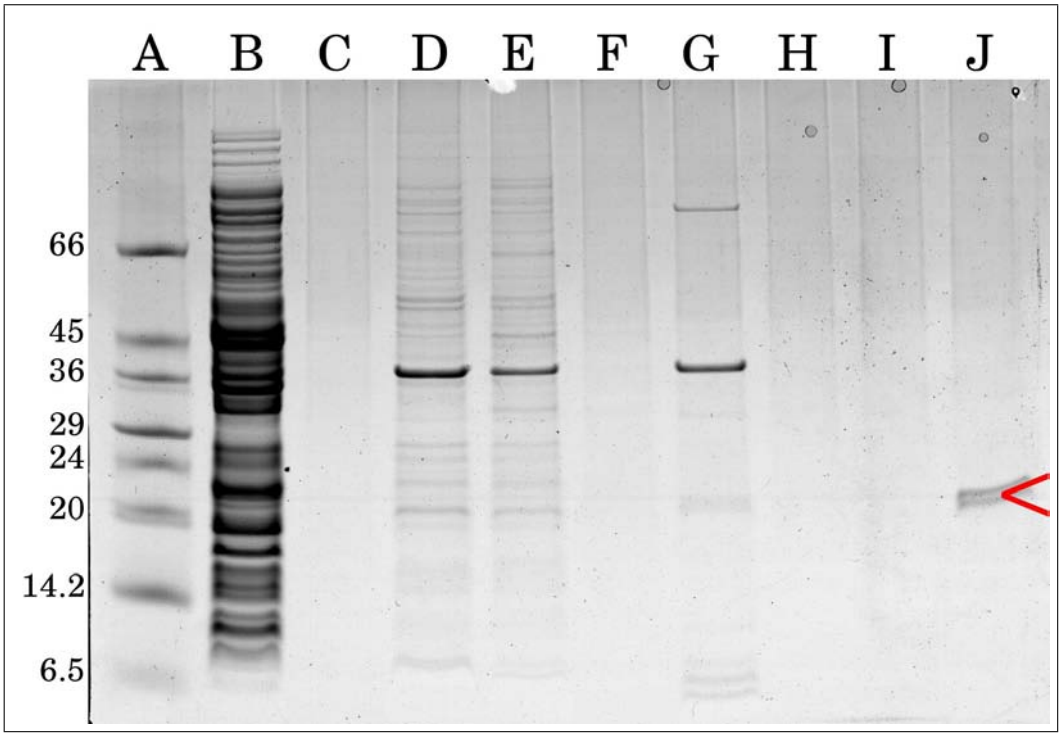

Figure 6.2: *A coomassie blue stained SDS-PAGE TRIS/Tricine gel showing the protein content in some of the the different steps during the MscL purification as described in the text. The molecular weight marker sizes are shown in kDa, on the left. The band that was assigned to MscL is marked with a red arrow. (lane A): Molecular weight marker (lane B): Sample taken after cells were broken with French press (lane C): Smaller sample of broken cells (proteins not visible) (lane D): Flow-through after interaction with Ni-NTA column (lane E): Washing of column with 40 mM imidazole, after first 20 ml (lane F): Washing of column with 40 mM imidazole, after 32 ml (lane G): Non-soluble fraction after interaction with octylglucoside overnight (lane H): Fraction 3 (after elution with 400 mM imidazole) (lane I): Fraction 4 (lane J): Fraction 5*

detergent. Successful incorporation of the MscL protein in the liposomes was checked by the blue shift observed in fluorescence measurements, which showed a tryptophan residue in a hydrophobic environment ($\lambda_{max} = 315\,nm$, superimposed with a second component with a maximum at $\lambda_{max} = 331\,nm$).

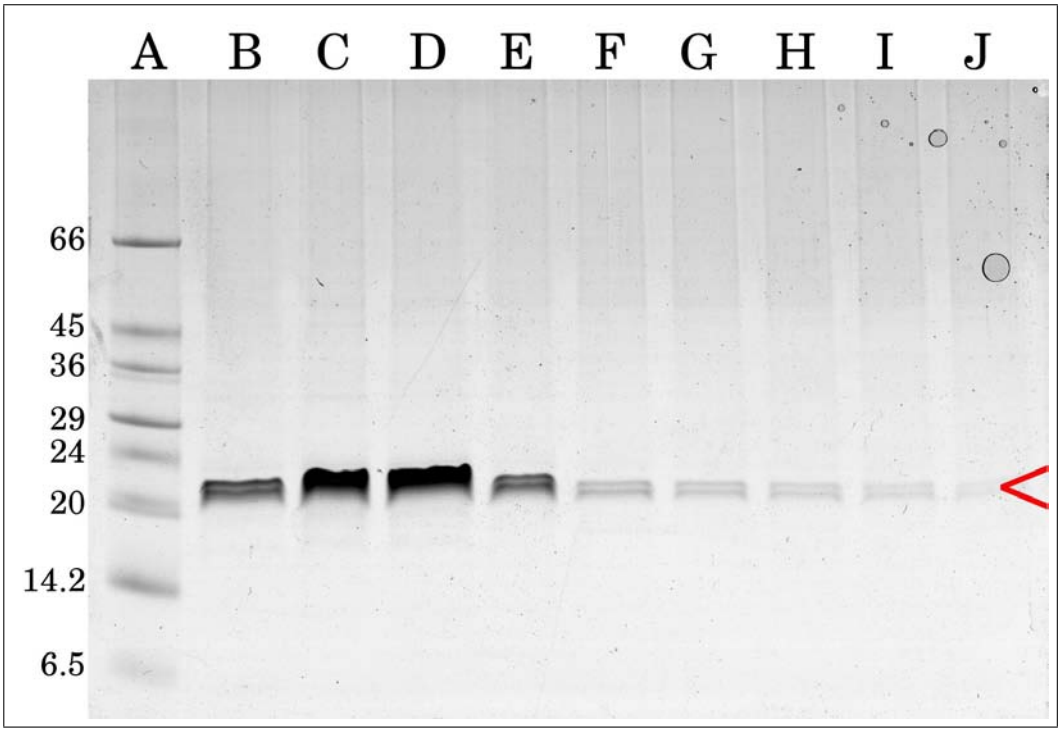

Figure 6.3: *A coomassie blue stained SDS-PAGE TRIS/Tricine gel showing the protein content in the different fractions after MscL purification on an Ni-NTA column. The molecular weight marker sizes are shown on the left, in kDa. The position of the bands that were assigned to MscL is marked with a red arrow. (lane A): Molecular weight marker (lane B–J): Fraction 6–14 (elution)*

6.2.4 Preparation of oriented samples

Glass plates were cut to size from cover glasses that are usually used in light microscopy. These glass plates were thoroughly cleaned by leaving them overnight in pure HNO_3, after which they were washed extensively with pure (milliQ) water. After the plates had dried on paper towels, 40 μl of the proteoliposome suspension was deposed on 12 plates, to obtain a surface density of slightly less than 1 mg lipid per cm^2. The glass plates were dried in a closed chamber that contained silicagel. After they appeared dry, the plates were rehydrated by replacing the silicagel by a saturated KCl solution at 40°C, (relative humidity of 85%) and leaving them to equilibrate overnight. Three such hydration–dehydration cycles were carried out, but for the dehydration steps after the first one, a vacuum was used instead of the silicagel. The last step was obviously a rehydration, after which the plates were stacked and

sealed inside a 7 mm NMR tube, and ^{31}P spectra were recorded to study the orientation of the lipids.

6.2.5 Solid state NMR

The ^{31}P spectra were recorded at 303,K on a Bruker 500 MHz spectrometer, equipped with a 10 mm single resonance probe with a solenoid coil which was oriented at 90° with respect to the static magnetic field. The 7 mm NMR tube was fixed in the middle of a 10 mm coil. The reason for this was that we have to carry out the ^{15}N NMR measurements using a 7 mm probe, and we chose to prepare the samples directly in a tube of this size. Spectra were acquired with a Hahn echo pulse sequence with a 90° pulse at 37 kHz, an echo delay of 30 μs, an acquisition time of 4 ms and a relaxation delay of 1 second. 25k scans were accumulated. Spectra were referenced to the isotropic peak of lyso-PC at 0 ppm. Spectra of lyso-PC were recorded under the same conditions as the other measurements.

6.3 Results & Discussion

In figure 6.4 on the following page the ^{31}P solid-state NMR spectra of reconstituted MscL in POPC/POPG lipid bilayers on glass plates are shown. Because the isotropic chemical shift for POPC and POPG is nearly identical, and the CSA as well [8], the spectra may be analysed as for samples containing a single lipid species. In an oriented sample we therefore expect to see a single spectral component that corresponds to the resonance frequency of a particular orientation ($\sim$ 28 ppm for samples oriented with the bilayer normal at an angle of 0° with respect to the magnetic field, and ~ -12 ppm for samples oriented at 90°[8]).

When the spectra are compared to the simulated spectra of oriented samples given in section 2.4 on page 30, it is clear that most of the the lipids are not oriented. The broad peak that is visible around 30 ppm in the spectrum oriented at 0° probably corresponds to a (very) small fraction of oriented lipids. The peaks at -16 and -20 ppm, are lipids oriented at 90° with respect to the magnetic field, and are probably part of a powder spectrum. The peaks at -2 and -6 ppm, are most likely caused by small lipid or proteolipid particles in the sample [8]. For our experiments, the most important conclusion is that the lipid bilayers are not oriented, which prevents the study of the structural properties of MscL such as the tilt angle of the trans-membrane helices.

Unfortunately, the first experiments to reconstitute MscL in oriented lipid

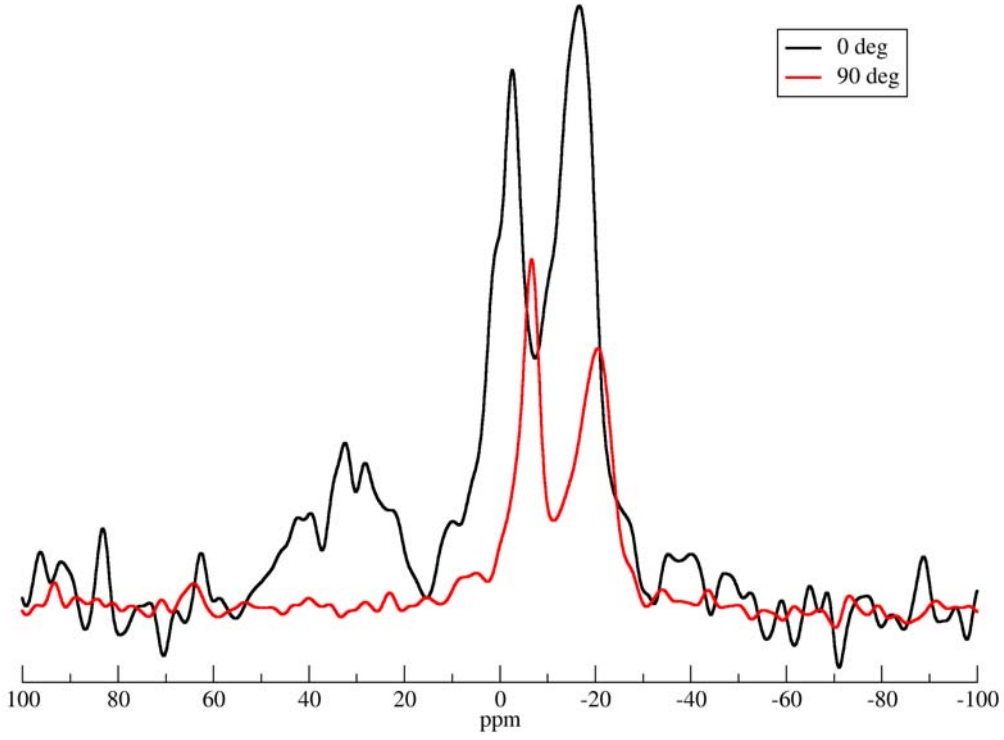

Figure 6.4: *^{31}P solid-state NMR spectra of MscL reconstituted in POPC/POPG lipids on glass plates, prepared as described in the text. Spectra with the normal to the glass plates oriented in the direction of the magnetic field (0 deg) and with the normal to the glass plates perpendicular to the magnetic field (90 deg) are shown.*

bilayers on glass plates were unsuccessful so far. There are many different protocols to try, and multiple parameters that may be important to correctly orient the sample. The first thing to try would probably be to decrease the protein/lipid ratio. Although this would mean less ^{15}N NMR signal once the protein will be studied, the first goal is to create samples in which the lipids are oriented. Other ideas for possible improvements are: decreasing the amount of lipids per glass plate, using different lipids or a different PC/PG ratio, using a different number of rehydration cycles, changing the temperature during the rehydration, cutting the HIS-tag off MscL or using MscL mutants with a smaller cytoplasmic part. Another, very promising option for performing solid state NMR experiments on MscL is the use of bicelles [7], particularly bicelles that orient perpendicular to the magnetic field [11].

Bibliography

[1] B. Bechinger. A dynamic view of peptides and proteins in membranes. *Cell Mol Life Sci*, 65(19):3028–3039, Oct 2008.

[2] M. F. Brown, M. P. Heyn, C. Job, S. Kim, S. Moltke, K. Nakanishi, A. A. Nevzorov, A. V. Struts, G. F. J. Salgado, and I. Wallat. Solid-state 2H NMR spectroscopy of retinal proteins in aligned membranes. *Biochim Biophys Acta*, 1768(12):2979–3000, Dec 2007.

[3] G. Chang, R. H. Spencer, A. T. Lee, M. T. Barclay, and D. C. Rees. Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. *Science*, 282(5397):2220–2226, Dec 1998.

[4] B. Corry and B. Martinac. Bacterial mechanosensitive channels: experiment and theory. *Biochim Biophys Acta*, 1778(9):1859–1870, Sep 2008.

[5] C. C. Häse, A. C. L. Dain, and B. Martinac. Purification and functional reconstitution of the recombinant large mechanosensitive ion channel (MscL) of *escherichia coli*. *J Biol Chem*, 270(31):18329–18334, Aug 1995.

[6] C. C. Häse, A. C. L. Dain, and B. Martinac. Molecular dissection of the large mechanosensitive ion channel (MscL) of E. coli: mutants with altered channel gating and pressure sensitivity. *J Membr Biol*, 157(1): 17–25, May 1997.

[7] R. Mahalakshmi, C. M. Franzin, J. Choi, and F. M. Marassi. NMR structural studies of the bacterial outer membrane protein OmpX in oriented lipid bilayer membranes. *Biochim Biophys Acta*, 1768(12):3216–3224, Dec 2007.

[8] R. Mani, J. J. Buffy, A. J. Waring, R. I. Lehrer, and M. Hong. Solid-state NMR investigation of the selective disruption of lipid membranes by protegrin-1. *Biochemistry*, 43(43):13839–13848, Nov 2004.

[9] B. Martinac, M. Buechner, A. H. Delcour, J. Adler, and C. Kung. Pressure-sensitive ion channel in Escherichia coli. *Proc Natl Acad Sci U S A*, 84(8):2297–2301, Apr 1987.

[10] R. C. Page, C. Li, J. Hu, F. P. Gao, and T. A. Cross. Lipid bilayers: an essential environment for the understanding of membrane proteins. *Magn Reson Chem*, 45(S1):S2–S11, Dec 2007.

[11] S. H. Park, C. Loudet, F. M. Marassi, E. J. Dufourc, and S. J. Opella. Solid-state NMR spectroscopy of a membrane protein in biphenyl phospholipid bicelles with the bilayer normal parallel to the magnetic field. *J Magn Reson*, 193(1):133–138, Jul 2008.

[12] E. Perozo, D. M. Cortes, P. Sompornpisut, A. Kloda, and B. Martinac. Open channel structure of MscL and the gating mechanism of mechanosensitive channels. *Nature*, 418(6901):942–948, Aug 2002.

[13] S. F. Poget and M. E. Girvin. Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better. *Biochim Biophys Acta*, 1768(12):3098–3106, Dec 2007.

[14] A. M. Powl. *Interaction of a Membrane Protein with its Surrounding Lipid Bilayer: Studies with the Mechanosensitive Channel MscL*. PhD thesis, School of Biological Sciences, University of Southampton, 2004.

[15] A. M. Powl and A. G. Lee. Lipid effects on mechanosensitive channels. *Current Topics in Membranes*, 58:151–178, 2007.

[16] A. M. Powl, J. M. East, and A. G. Lee. Anionic phospholipids affect the rate and extent of flux through the mechanosensitive channel of large conductance MscL. *Biochemistry*, 47(14):4317–4328, Apr 2008.

[17] A. M. Powl, J. M. East, and A. G. Lee. Importance of direct interactions with lipids for the function of the mechanosensitive channel MscL. *Biochemistry*, 47(46):12175–12184, Nov 2008.

[18] R. S. Prosser, F. Evanics, J. L. Kitevski, and M. S. Al-Abdul-Wahid. Current applications of bicelles in NMR studies of membrane-associated amphiphiles and proteins. *Biochemistry*, 45(28):8453–8465, Jul 2006.

[19] J. Sambrook, E. Fritsch, and T. Maniatis. *Molecular cloning — a laboratory manual*, volume 3. Cold Spring Harbor Laboratory Press, second edition edition, 1989.

[20] S. I. Sukharev, P. Blount, B. Martinac, F. R. Blattner, and C. Kung. A large-conductance mechanosensitive channel in E. coli encoded by mscL alone. *Nature*, 368(6468):265–268, Mar 1994.

[21] S. I. Sukharev, P. Blount, B. Martinac, and C. Kung. Mechanosensitive channels of Escherichia coli: the MscL gene, protein, and activities. *Annu Rev Physiol*, 59:633–657, 1997.

[22] G. Wang. NMR of membrane-associated peptides and proteins. *Curr Protein Pept Sci*, 9(1):50–69, Feb 2008.

7

Conclusions & Perspectives

Contents

7.1 A peptide derived from V-ATPase subunit *a*

The 3D structure of a peptide derived from the putative transmembrane segment 7 from H^+-V-ATPase from *Saccharomyces cerevisiae* has been determined by solution state NMR in SDS[6]. The peptide appeared to form a helical structure from L736–Q745, indicating that a longer helix can be formed than the previous results in DMSO suggested.

With the structure of the peptide now known, a good starting point for MD simulations was available. A refinement of the NMR structure by molecular dynamics simulations in SDS led to an improved structural model with respect to the structure calculation carried out by the ARIA software (as judged by procheck_nmr software).

The results from other, unrestrained, simulations indicated that one helical face of the peptide was solvated by the water, while a large portion of the peptide interacted with the SDS apolar tails. The MD simulations were able to help to explain some experimental observations. The fact that the peptide did not aggregate in SDS at high pH is probably due to the micelle protecting it. This protection was not observed in octylglucoside, and we concluded that this could be the cause for the aggregation.

The pKa of both histidine residues present in the peptide has been measured experimentally in SDS. Although we did not attempt to calculate an exact pH from the simulations, the fact that H729 is solvated by more negatively charged SDS headgroups than H743 is consistent with the larger difference in pKa between samples in water and SDS. Solid-state NMR data indicated that TM7 alone did not take a transmembrane position. We cannot exclude that this is a matter of sample preparation and/or lipid composition, but clearly it is not easy (we spent a long time trying!) to reconstitute this peptide as a transmembrane helix in a lipid bilayer.

We concluded that the native environment is probably important to hold TM7 in place. Current theories on the function of TM7 in the entire protein suggest that it is located in the interface between the hydrophobic subunit *a* and *c*, and that it lines the channel through which protons have access to the glutamic acid residues on the *c* subunit assembly. The organisation of water and SDS around the peptide shows that it indeed has a hydrophilic side, which we propose is part of the proton channel, and a hydrophobic side, which probably interacts with other transmembrane helices from subunit *a*. The essential R735 residue was found to be a single charged residue in the middle of a hydrophobic helical face, consistent with it being in contact with subunit *c* [1], which is also hydrophobic.

Studies similar to this one could be carried out on the other transmembrane helices from subunit *a*, and the structures could be docked together using computer simulations. This would allow the building of a model for the structure of subunit *a*, which could lead to more insight into the proton translocation mechanism of V-ATPase, or even to new specific inhibitors, a step in the right direction towards a medicine for osteoporosis. Because attempts to crystallise subunit *a* appear to have been unsuccessful, the "peptide-dissection" approach remains an interesting option. The combination between NMR in detergent and MD simulations is very well suited for this type of work. Another option would be to try and perform solid-state NMR on the entire V_0 section of V-ATPase. Although this is still a major challenge for a system as large as this one, this approach is becoming increasingly realistic[2, 5]. In this thesis, we focused our attention first on a more accessible (entire) membrane protein, MscL, by establishing a collaboration with the group of prof. A.G. Lee, who has been studying MscL extensively by fluorescence measurements[4].

7.2 Oriented samples of MscL

In order to prepare samples to study MscL by solid state NMR, we first established the protocols to express, purify and reconstitute MscL in the lab in Toulouse, using the protocols developed by Dr. A.M. Powl, in the research group of prof. A.G. Lee. A fermentation in a 50 litre fermentor was carried out with a ^{15}N labelled medium, which provided us with around 300 mg of fully ^{15}N labelled protein. Although the first tests led to non-oriented samples, possibilities for improvement have been presented. Particularly the use of biphenyl phospholipid bicelles [3] is promising for structural studies on membrane proteins such as MscL.

Bibliography

[1] S. Kawasaki-Nishi, T. Nishi, and M. Forgac. Interacting helical surfaces of the transmembrane segments of subunits a and c′ of the yeast V-ATPase defined by disulfide-mediated cross-linking. *J Biol Chem*, 278(43):41908–41913, Oct 2003.

[2] J. Korukottu, R. Schneider, V. Vijayan, A. Lange, O. Pongs, S. Becker, M. Baldus, and M. Zweckstetter. High-resolution 3D structure determination of kaliotoxin by solid-state NMR spectroscopy. *PLoS ONE*, 3(6): e2359, 2008.

[3] S. H. Park, C. Loudet, F. M. Marassi, E. J. Dufourc, and S. J. Opella. Solid-state NMR spectroscopy of a membrane protein in biphenyl phospholipid bicelles with the bilayer normal parallel to the magnetic field. *J Magn Reson*, 193(1):133–138, Jul 2008.

[4] A. M. Powl and A. G. Lee. Lipid effects on mechanosensitive channels. *Current Topics in Membranes*, 58:151–178, 2007.

[5] R. Schneider, C. Ader, A. Lange, K. Giller, S. Hornig, O. Pongs, S. Becker, and M. Baldus. Solid-state NMR spectroscopy applied to a chimeric potassium channel in lipid bilayers. *J Am Chem Soc*, 130(23):7427–7435, Jun 2008.

[6] L. S. Vermeer, V. Réat, M. A. Hemminga, and A. Milon. Structural properties of a peptide derived from h^+-V-ATPase subunit *a*. *Biochimica et Biophysica Acta, Biomembranes*, *in press*, 2009.

A

A program to draw a helical wheel plot from a pdb file

Contents

A.1 Calculation of a helical wheel from a pdb file

Many programs are available that draw helical wheels from a given peptide primary structure (sequence), assuming a perfect alpha helix. Sometimes it may be interesting to see what the helical-wheel plot of a non-ideal helix looks like.

Because I could not find such a program by an internet search, I created one myself. It is a Perl script that uses the PDL (Perl data language) module for the calculations and the SVG module to generate the image. You should have Perl and these two modules installed to use it. The script was only tested on Linux, but should work under other operating systems as well.

I have made the script available under the creative commons by-nc-sa licence, which means you may use and distribute it as long as you say you got it from me.

The program goes through the following steps to draw the helical wheel:

- Read the coordinates of the CA atoms from the pdb file
- Find the principal axis of those atoms (the longest axis)
- Move the centre of mass of the CA atoms to the origin
- Calculate the mass distribution tensor of the CA atoms
- Calculate the eigenvectors and eigenvalues of this matrix (the matrix of eigenvectors is the rotation matrix)
- Rotate the principal axis to coincide with the z-axis
- Draw the xy-plane, which is the helical wheel plot

Because the program creates an SVG image, it may be edited and resized without losing quality. It should be noted that it currently does not support pdb files that contain a chain identifier field. This is however easy to work around by removing the chain identifier from the pdb file or making a small change to the source code.

Listing A.1: *wheel.pl source code*

```
1 #!/usr/bin/perl
2 use warnings;
3 use strict;
4 use PDL;
```

```
5  use SVG;
6  use Getopt::Long;
7  use Pod::Usage;
8
9  ########## some definitions (you may want to change/add something) ##
10
11 # the letters that will be used for each residue:
12 my %short_residue_name = (
13     'ALA' => 'A', 'CYS' => 'C', 'ASP' => 'D', 'GLU' => 'E',
14     'PHE' => 'F', 'GLY' => 'G', 'HIS' => 'H', 'ILE' => 'I',
15     'LYS' => 'K', 'LEU' => 'L', 'MET' => 'M', 'ASN' => 'N',
16     'PRO' => 'P', 'GLN' => 'Q', 'ARG' => 'R', 'SER' => 'S',
17     'THR' => 'T', 'SEC' => 'U', 'VAL' => 'V', 'TRP' => 'W',
18     'TYR' => 'Y',
19     'HIP' => 'H', 'HIS+' => 'H','LYS+' => 'K', 'ARG+' => 'R',
20 );
21
22 # the definition of the background color for
23 # apolar, polar, positively and negatively charged residues
24 sub rescol {
25     my $checkres = shift;
26     my @apo = ('A','C','F','I','L','M','P','U','V','W','Y');
27     my @pos = ('H','K','R');
28     my @neg = ('D','E');
29     foreach (@apo) { if ($_ eq $checkres) { return 'grey'; } }
30     foreach (@pos) { if ($_ eq $checkres) { return 'blue'; } }
31     foreach (@neg) { if ($_ eq $checkres) { return 'red';  } }
32     return 'white';
33 }
34 my $picsize  = 600; # 600 works best, it is best to scale the image →
   →afterwards
35
36 ########## handle the command line options ##########################
37
38 # default values if no command line option is given:
39 my $filename = '';
40 my $outfile  = 'wheel-out.svg';
41 my $logfile  = 'wheel-log.txt';
42 my $startres =   0;
43 my $endres   =   0;
44 my $man      =   0;
```

```
45 my $help     =   0;
46 my $rotate   =   1;
47 my $number   =   0;
48 my $options  = GetOptions(
49     "filename=s" => \$filename,
50     "outfile=s"  => \$outfile,
51     "logfile:s"  => \$logfile,
52     "rotate!"    => \$rotate,
53     "number!"    => \$number,
54     "startres:i" => \$startres,
55     "endres:i"   => \$endres,
56     "help|?"     => \$help,
57     man          => \$man,
58 ) or pod2usage(-verbose => 0);
59 pod2usage(-verbose => 0)  if $help;
60 pod2usage(2) if $man;
61
62 if ( $filename eq '' ) {
63     die "\nNo input file specified. See \"wheel.pl -h\" for help.\n\n";
64 }
65
66 ########## read the input data ######################################
67
68 open LOGFO, ">$logfile" or die "Error opening logfile $logfile. Quit.\n\n→
   →";
69
70 # read CA atom coordinates into piddles, and the residue name into a perl→
   → array
71 my ( $x, $y, $z, $resnames_ref )
72     = rcols( $filename, 5, 6, 7, { INCLUDE=>'/^\s*ATOM.*CA/', PERLCOLS→
     →=>[3] } );
73 my $coord_pdl = pdl($x,$y,$z);
74 if ($endres==0) {
75     $endres = nelem($x)-1;
76 }
77 print LOGFO "Read data: $coord_pdl\n";
78
79 # keep only the helical part (as defined by the user)
80 $coord_pdl     = slice($coord_pdl,"$startres:$endres,:");
81 $x             = slice($coord_pdl,":,0");
82 $y             = slice($coord_pdl,":,1");
```

```
83  $z             = slice($coord_pdl,":,2");
84  @$resnames_ref = @$resnames_ref[$startres..$endres];
85  print LOGFO "Helical part (user defined): $coord_pdl\n";
86
87  ########## calculate the xy projection of the helix #################
88
89  # move mass midpoint to origin
90  $x -= average($x);
91  $y -= average($y);
92  $z -= average($z);
93  print LOGFO "Mass midpoint moved to origin: $coord_pdl\n";
94
95  if ($rotate) {
96
97      # calculate the mass distribution tensor ([Ixx,Ixy,Ixz],[Iyx,Iyy,Iyz→
        →],[Izx,Izy,Izz])
98      my $massdist = $coord_pdl x transpose($coord_pdl);
99      print LOGFO "Mass distribution tensor: $massdist\n";
100
101     # calculate eigenvectors
102     my ($eigenvectors,$eigenvalues) = eigens($massdist);
103     print LOGFO "Eigenvectors:  $eigenvectors\n";
104     print LOGFO "Eigenvalues:\n $eigenvalues \n\n";
105
106     # rotate molecule to principal-axis
107     $coord_pdl .= transpose($eigenvectors) x $coord_pdl;
108     # in the previous statement .= is used so that the slices $x,$y and →
        →$z change as well
109     print LOGFO "Molecule rotated to principal-axis: $coord_pdl\n";
110
111     # define the z-axis as the principal axis (the longest one)
112     if ( max($x)-min($x) > max($y)-min($y) ) {
113         my $tmp = $x;
114         $x = $y;
115         $y = $tmp;
116     }
117     if ( max($y)-min($y) > max($z)-min($z) ) {
118         my $tmp = $y;
119         $y = $z;
120         $z = $tmp;
121     }
```

```
122
123 }
124
125 ########## draw the image #################################################
126
127 # prepare an svg image
128 my $svg = SVG->new( width => $picsize, height => $picsize);
129
130 # caluclate the size of the drawing area dependent on the canvas size, →
    →leaving 20% margins
131 my $drawsize = ( $picsize - (0.2 * $picsize ) ) / 2;
132
133 # normalize x and y coordinates (for drawing)
134 my $max_x_y;
135 if ( $x->abs->max > $y->abs->max ) {
136     $max_x_y = $x->abs->max;
137 }
138 else {
139     $max_x_y = $y->abs->max;
140 }
141 $x .= $x / $max_x_y;
142 $y .= $y / $max_x_y;
143 print LOGFO "Normalized x and y coordinates for drawing: $coord_pdl\n";
144
145 # multiply normalized coordinates by $drawsize and dump in perl array
146 my @xc = list $x * $drawsize;
147 my @yc = list $y * $drawsize;
148 my $pic_mid = $picsize / 2;
149
150 # move the drawing origin to the middle and set some drawing properties
151 print LOGFO "Creating image:\n";
152 my $wheel_drawing = $svg->group(
153     id              => 'wheel',
154     transform       => "translate($pic_mid, $pic_mid)",
155     style           => {
156         stroke=>'black',
157         'stroke-width' =>2
158     }
159 );
160
161 # draw the residues with lines between them (last residue is drawn first)
```

```
162 for (
163     my $current_array_element = (nelem($x)-1);
164     $current_array_element > 0;
165     $current_array_element--
166     )
167 {
168 
169     my $current_residue_name
170         = $short_residue_name{ "$$resnames_ref[$current_array_element]" →
        →};
171 
172     # determine the color for this residue
173     my $c = &rescol($current_residue_name);
174 
175     print LOGFO
176         "Drawing line: $$resnames_ref[$current_array_element] ("
177         .$current_array_element
178         .") - $$resnames_ref[$current_array_element-1] ("
179         .($current_array_element-1)
180         .") Drawing residue: CA of $$resnames_ref[$current_array_element]→
        → ("
181         .$current_array_element
182         .")\n"
183     ;
184     $wheel_drawing->line(
185         'x1' => $xc[ $current_array_element     ],
186         'y1' => $yc[ $current_array_element     ],
187         'x2' => $xc[ $current_array_element - 1 ],
188         'y2' => $yc[ $current_array_element - 1 ],
189     );
190     $wheel_drawing->circle(
191         'cx'           => $xc[$current_array_element],
192         'cy'           => $yc[$current_array_element],
193         'r'            => $drawsize / 10,
194         'fill-opacity' => 0.5,
195         'fill'         => $c,
196     );
197     if ($number) {
198         $wheel_drawing->text(
199             'x'         => $xc[$current_array_element] - $drawsize / 12,
200             'y'         => $yc[$current_array_element] + $drawsize / 40,
```

```
201             'font-size' => $picsize / 30,
202             'color'     => 'white',
203             '-cdata'    => $current_residue_name.($current_array_element
204                            + $startres),
205         );
206     }
207     else {
208         $wheel_drawing->text(
209             'x'         => $xc[$current_array_element] - $drawsize / 20,
210             'y'         => $yc[$current_array_element] + $drawsize / 25,
211             'font-size' => $picsize / 20,
212             'color'     => 'white',
213             '-cdata'    => $current_residue_name,
214         );
215     }
216 }
217 # finally, draw the circle for the first residue on top (but no line)
218 my $current_residue_name = $short_residue_name{"$$resnames_ref[0]"};
219 my $c = &rescol($resnames_ref);
220 $wheel_drawing->circle(
221     'id'           => "r0",
222     'cx'           => $xc[0],
223     'cy'           => $yc[0],
224     'r'            => $drawsize / 10,
225     'fill-opacity' => 0.5,
226     'fill'         => $c,
227 );
228 if ($number) {
229     $wheel_drawing->text(
230        'x'         => $xc[0] - $drawsize / 12,
231        'y'         => $yc[0] + $drawsize / 40,
232        'font-size' =>  $picsize / 30,
233        'color'     => 'white',
234        '-cdata'    => $current_residue_name . $startres,
235     );
236 }
237 else {
238     $wheel_drawing->text(
239         'x'         => $xc[0] - $drawsize / 20,
240         'y'         => $yc[0] + $drawsize / 25,
241         'font-size' => $picsize / 20,
```

```
242         'color'     => 'white',
243         '-cdata'    => $current_residue_name,
244     );
245 }
246 print LOGFO "Drawing CA for residue $$resnames_ref[0] (0)\n";
247
248 # render the svg image and save to output file
249 open OUTFO, ">$outfile"
250     or die "Error opening output file $outfile. Quit.\n\n";
251 print OUTFO $svg->xmlify;
252 close OUTFO;
```

A.1.1 Manual for wheel.pl

```
SYNOPSIS

wheel.pl -f infile.pdb -o outfile.svg [OPTIONS]

  Options:
   -filename  -f    The name of the input file (pdb).
   -outfile   -o    The name of the output file (svg).
   -logfile   -l    The name of the log file (txt).
   -(no)rotate      Align the molecule along its principal axis or not.
   -(no)number      Show residue numbers or not.
   -startres  -s    The residue number to start from. The first residue
                    in the pdb file is always residue 0, the second 1, →
                    →etc.
                    regardless of the numbering in the pdb file.
   -endres    -e    The last residue number to use in the helical wheel
   -help      -h    Print a (hopefully) helpful message and exit.
   -man       -m    Print a more verbose description of this program.

OPTIONS

<-infile> [filename]

The name of the input pdb file. Use "-f -" to read the pdb file from
STDIN.

<-outfile> [filename]
```

```
The name of the output file (an svg image). If this option is omitted,
the default filename wheel-out.svg will be used. A possibly existing
file with this name will be overwritten without warning.  If you use
"-o -" the image contents will be printed to STDOUT.

<-logfile> [filename]

The filename of the log-file. Some intermediate results are saved in
this file.  It could be used to debug the program or to verify that the
resulting image is really what you think it is. If this option is omitted→
→,
the file will be called wheel-log.txt. Existing files will be overwritten
without warning.  You can use "-l -" to dump the log to STDOUT instead
of a file.

<-(no)rotate>

It is necessary to have the molecule aligned to the z-axis to draw a
useful helical wheel, but if the molecule is already aligned to the z-→
→axis
(using an external program for example) you may want to draw it directly,
without any further manipulations.

<-(no)number>

Use -number if you want to see the residue numbers alongside the residue →
→names.
The default is not to show the numbers.

<-startres> [number]

Residues smaller than the number given after -startres are ignored. Note
that the fist residue in the pdb file is always addressed as residue 0,
the second residue as 1, etc. If this option is not given, the →
→calculation
starts with residue 0. Usually, you'll want to use this option to select
a small region of a larger protein or to prevent a non-helical part of
the protein from being used in the calculation. It is not a good idea
to try and include non-helical regions in the plot, because the result
will likely be useless.
```

```
<-endres> [number]

Like -startres, but defines the residue where to stop. If the option is
omitted, it defaults to the last residue in the pdb file.

DESCRIPTION

wheel.pl reads the coordinates of the CA atoms from a pdb file to produce
a so-called helical-wheel plot. The centre of mass of the CA atoms is
first moved to the origin, after which the molecule is rotated so that
its principal axis points in the z-direction. After this, a projection
of the xy-plane is drawn, with the CA atoms represented as circles with
a letter inside, corresponding to the residue name. The residues are
coloured according to their properties: hydrophobic=grey, positively
charged: blue, negatively charged: red, other residues are drawn in white→
→.

AUTHOR

Louic S. Vermeer (online@louic.nl)
```

B

A program to convert 1D Bruker Topspin processed data to the ASCII text file format

Contents

B.1 Reading Bruker Topspin 1-dimensional data files

Although the Bruker Topspin software has many useful data-processing capabilities, its inability to easily create figures in a quality suitable for publication can be a problem.

The Bruker Topspin software saves its processed data in a binary format. Although software to read or convert this file-format to standard ASCII exists [1, 3, 4], this software uses a graphical user interface and cannot easily be used in scripts. Furthermore, the gsim and editnmr/dmfit software are not platform-independent, and there are some known issues with dmfit on Linux machines (using the Wine software). Although NMRpipe[2] contains several data-format converters, it uses unprocessed data as input files, and the data is output in NMRpipes own file format.

Of course, the software packages mentioned here do much more than just convert file-formats. But when only a simple conversion is required, an easy to use command-line utility may be preferable. Therefore a script was written to convert the Bruker Topspin data, once processed, to a simple ASCII-text file. The output file contains two tab-separated columns, and can be imported by almost any other program. At the moment, only one-dimensional data files are supported, and the middle of the spectrum is always set to 0 ppm, so you would need to shift the spectrum manually to the right frequency. This will be corrected in the next version of the script.

Listing B.1: *bruk2tab.pl source code*

```
#!/usr/bin/perl
#===============================================================
#
#         FILE:  bruk2tab.pl
#
#        USAGE:  see ./bruk2tab.pl -h
#
#  DESCRIPTION:  Convert one-dimensional bruker files (1r) to normal,
#                readable tab separated text files. It no output file
#                is given, output goes to STDOUT.
#
#       AUTHOR:  Louic Vermeer, <online@louic.nl>
#      VERSION:  1.0
#      LICENCE:  GPL
#      CREATED:  02/04/2009 01:05:33 PM CET
```

```
16 #=====================================================================
17
18 use warnings;
19 use strict;
20 use Getopt::Long;
21 use Pod::Usage;
22
23 # get command line parameters
24 my $acqus_file    = '';
25 my $spectrum_file = '';
26 my $outfile       = '-';
27 my $quiet         = 0;
28 my $help          = 0;
29 my $man           = 0;
30 my $options = GetOptions(
31     "acqus=s"    => \$acqus_file,
32     "spectrum=s" => \$spectrum_file,
33     "outfile=s"  => \$outfile,
34     "quiet!"     => \$quiet,
35     "help|?"     => \$help,
36     "man"        => \$man,
37 ) or pod2usage (-verbose => 0);
38 pod2usage(1) if $help;
39 pod2usage(-verbose => 2) if $man;
40
41 # check command line parameters
42 if ( $acqus_file eq '' ) {
43     die "No acqus input file specified. See bruk2tab.pl -h for help\n";
44 }
45 if ( $spectrum_file eq '' ) {
46     die "No spectrum input file specified. See bruk2tab.pl -h for help\n"→
    →;
47 }
48
49 # open files
50 open (ACQUS, $acqus_file)    or die "*** Error opening input file: →
   →$acqus_file\n";
51 open (SPECF, $spectrum_file) or die "*** Error opening input file: →
   →$spectrum_file\n";
52 open (OUTFO, ">$outfile")    or die "*** Error opening output file: →
   →$outfile\n";
```

```
53
54 # read sweap witdh ($sw) from acqus file
55 my $sw;
56 while (<ACQUS>) {
57     if (/^##\$SW=/) {
58         (my $name, $sw) = split(' ');
59     }
60 }
61 # check if sweep width was read
62 if (!$sw) {
63     die "*** Error: could not find sweepwidth in file: $acqus_file\n";
64 }
65 close ACQUS;
66
67 my $npoints = 0; # number of data points that was read successfully
68 my @ydata;       # array that will hold (y-values of) data points
69
70 # read the data-file (only tested on bruker topspin 1r files)
71 until ( eof(SPECF) ) {
72     my $data;
73     # read data with block size of 8 bytes
74     my $nbytes = read(SPECF,$data,8);
75     # check if an entire 8-byte block was read
76     if ($nbytes == 8) {
77         # treat block as a signed long integer
78         my $yvalue = unpack "l", $data;
79         push @ydata, $yvalue;
80         $npoints++;
81     } else {
82         warn "Warning: found block with size different from 8 bytes,\n";
83         warn "         which is weird and should not have happened.\n";
84         warn"          Maybe there is something wrong with the input file→
        →?\n";
85         warn"      !!! You should check if the data in the output file is→
        → correct.\n";
86     }
87 }
88 close SPECF;
89
90 # calculate dwell time and x-axis data points
91 # ... and output the data points to the $outfile
```

```
92  my $dw = $sw/($npoints-1);
93  foreach my $n (0..$npoints-1) {
94      my $xvalue = ( $sw/2 - $dw*$n );
95      print OUTFO "$xvalue\t$ydata[$n]\n";
96  }
97  close OUTFO;
98
99  if (!$quiet) {
100     print "converted $npoints data points, output file is $outfile\n";
101 }
```

B.1.1 Manual for bruk2tab.pl

```
SYNOPSIS

bruk2tab.pl -a acqus -s 1r [OPTIONS]

  Options:
   -spectrum   -s   required    The spectrum input file, usually called →
   →"1r".
   -acqus      -a   required    The acqus input file, usually called →
   →acqus.
   -outfile    -o   optional    The name of the output file. If omitted,
                                output goes to the screen (STDOUT, →
                                →actually)
   -quiet      -q   optional    Prevents messages from being displayed.
   -help       -h   optional    Print a short helpful message.
   -man        -m   optional    Print a longer, more helpful message.

OPTIONS

-spectrum

This file contains the actual NMR data that will end up on the y-axis,
but no information about the x-axis, which is why the acqus-file is also
required as input for this script.

-acqus

This file contains some information about the spectral parameters. The
```

```
sweep-width is read from this file to calculate the values that are be
displayed on the x-axis.

<-outfile>

The name of the output file. Output is tab-separated, with the x-axis
data in the first column and the y-axis data on the second column. If
no output file is given, the data is sent to STDOUT (which is usually
the screen).

<-quiet>

If you do not wish to see the message about the number of converted data
points and the name of the output file, use this option.

AUTHOR

Louic S. Vermeer (online@louic.nl)
```

Bibliography

[1] GSim - free software tool for visualisation and processing of experimental and simulated nuclear magnetic resonance (NMR) spectra. `http://sourceforge.net/projects/gsim`.

[2] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, 6(3):277–293, Nov 1995.

[3] R. H. Fogh, W. F. Vranken, W. Boucher, T. J. Stevens, and E. D. Laue. A nomenclature and data model to describe NMR experiments. *J Biomol NMR*, 36(3):147–155, Nov 2006.

[4] D. Massiot, F. Fayon, M. Capron, I. King, S. Calvé, B. Alonso, J.-O. Durand, B. Bujoli, Z. Gan, and G. Hoatson. Modelling one and two-dimensional solid-state NMR spectra. *Magn Reson Chem*, 40:70–76, 2002.

C

NMR structure calculation: from Sparky to ARIA and procheck_nmr

Contents

Table C.1: *Settings that seem to work well for integrating peaks with Sparky*

integration method:	Gaussian fit
peak motion:	on
adjust linewidths:	on
fit baseline:	off
substract off fit peaks:	off
use data above lowest contour:	on
use data within rectangle:	on
group peaks in contour boundary:	on
group close peaks	on
max motion (ppm):	0.005 (for both dimensions)
linewidth min–max (Hz):	2.0–80.0 (for both dimensions)
grouping range (Hz):	30.0
maximum minimisation steps:	100000
minimisation tolerance (%):	0.1

C.1 Resonance assignment with Sparky 3.110

First of all, the reader may want to know that the Sparky software package is no longer being updated. It may therefore be interesting to consider using other software for the assignment of the resonances, such as CARA [3, 4]. The CCPNMR software [1, 2, 6] is also definitely worth a look. If you decide to use Sparky however, the short tutorial below may be useful.

- Make sure you use the IUPAC naming conventions [5] while assigning the resonances in Sparky, otherwise you will need to convert them later anyway.
- Integrate the peaks with Sparky. The settings given in table C.1 worked well for me.

C.2 Structure calculation with ARIA

- Make backup copies of all files before you do something!
- Use the sparky command “dr” to delete all ununsed resonances.
- Export a dyana/xeasy chemical shifts file (protein.prot) from Sparky, using the “xe”-command and the “write shifts” button. Do not check

any of the boxes (You want to export volumes, not heights. Do not export unintegrated or unassigned peaks. Do not export assignments without residue number, and do not export any notes) As an example of the file format, several lines of a .prot file are give in listing C.1.

- Export a dyana/xeasy peak list (protein.peaks), also usung the "xe" dialog box, see listing C.2 for an example of the file.

Listing C.1: *example of a .prot file as exported by Sparky*

```
1    9    8.137  0.001      H 3
2   10    4.428  0.001     HA 3
3   11    3.821  0.016    HB# 3
4   12    8.136  0.000      H 3
5   13    4.429  0.000     HA 3
6   14    8.318  0.002      H 4
```

Listing C.2: *example of a .peaks file as exported by Sparky*

```
1 # Number of dimensions 2
2    1  4.363  8.318  1 U  2.24e+06        0 e 0   71   69
3    2  4.429  8.136  1 U  1.87e+06        0 e 0   10    9
4    3  3.809  8.136  1 U  5.02e+05        0 e 0   11    9
5    4  3.832  8.137  1 U  1.01e+06        0 e 0   11    9
```

If you have two spectra at different temperatures, the .prot files are identical, but the .peaks files differ for the two spectra. They will list assignments with slightly differing chemical shift, but ARIA seems to handle that without problems.

- Make a sequence file (protein.seq), with a three-letter amino-acid code on each line, as shown in listing C.3.

Listing C.3: *example of a .seq file*

```
LYS
LYS
SER
HIS
THR
ALA
```

C.2.1 Compiling Aria, Cns, Procheck and Aqua

For these examples, cns_solve_1.21_all-mp, aria 2.2, procheck 3.5.4 and aqua 3.2 were used. Everything was compiled from source on a Gentoo Linux installation (without using the portage package manager). All the programs keep their files nicely in their own direcory, and no root access is required.

- Download and unzip Aria, Cns, Procheck and Aqua
- Copy the aria2.2/cns/src/* files to the cns_solve_1.21/source/ directory
- Compile Cns according to the instructions in the Readme file
- Install Aria according to the instructions in the README file. The CCPN data model was not used in this case.
- Do not forget to set the environment variables and aliases as stated in the README file

You should now have a working Aria setup. Now we will install procheck and aqua.

- Edit the procheck Makefile: FF77 = gfortran -w
- Compile, and make a symlink to the executable file:

```
# make
# ln -s procheck.scr procheck
```

- Set the environment variable "prodir" to the correct value for the procheck directory (e.g. in .bashrc)

Procheck should now be functional, and we will proceed to install Aqua:

- Go to the "src" directory
- Edit the makefile: fill in the value for MYROOT, set CC = /usr/bin/g-fortran, and replace "F77" by "$(CC)" (2 occurances).
- compile

```
# make
```

- Set the "aquaroot" environment variable
- Edit the location of the perl binary in scripts/aquanal.pl
- Go back to the "src" directory if you have left it

To use these programs, you need to be in a c-shell, for example tcsh. The scripts were not made for bash, so

```
# tcsh
```

Now check if all environment variables mentioned above have been set correctly, and prepare to run aqua by starting its setup script:

```
# $aquaroot/aqsetup
```

- To run some tests, execute the command:

  ```
  # make test
  ```

- ... and finally, to clean things up a bit:

  ```
  # make strip; make clean
  ```

You are now ready to run Aria. It will automatically use procheck to analyse its output, but you can additionally run aqua and procheck_nmr. This will be shown below.

C.2.2 using Aria 2.2

The ARIA software will automaticly calibrate the NOE restraints based on the volumes of the integrated peaks, so there is no need to calibrate the NOEs manually. The following files are needed: .seq, .prot and .peaks, but some require editing, as detailed below.

Listing C.4: *The conversion.xml file used in this study*

```
1 <!DOCTYPE conversion SYSTEM "conversion1.0.dtd">
2
3 <conversion>
4
5   <project name="p1-structure">
```

```
6       <output
7         filename="run1.xml"/>
8     </project>
9
10    <molecule
11      molecule_type="PROTEIN"
12      molecule_name="kmtm7"
13      molecule_segid="A"
14      first_residue_number="1">
15      <input
16        filename="peptide.seq"
17        format="seq"
18        naming_convention=""/>
19      <output
20        filename="peptide.xml"/>
21    </molecule>
22
23    <spectrum
24      spectrum_name="NOESY298K"
25      segids="A">
26      <chemical_shifts>
27        <input
28          filename="298.prot"
29          format="xeasy"/>
30        <output
31          filename="298-prot.xml"/>
32      </chemical_shifts>
33      <cross_peaks>
34        <input
35          filename="298__.peaks"
36          format="xeasy"
37          proton1="1"
38          hetero1=""
39          proton2="2"
40          hetero2=""/>
41        <output
42          filename="298-peaks.xml"/>
43      </cross_peaks>
44    </spectrum>
45
46    <spectrum
```

```
47      spectrum_name="NOESY313K"
48      segids="A">
49      <chemical_shifts>
50        <input
51          filename="313.prot"
52          format="xeasy"/>
53        <output
54          filename="313-prot.xml"/>
55      </chemical_shifts>
56      <cross_peaks>
57        <input
58          filename="313__.peaks"
59          format="xeasy"
60          proton1="1"
61          hetero1=""
62          proton2="2"
63          hetero2=""/>
64        <output
65          filename="313-peaks.xml"/>
66      </cross_peaks>
67    </spectrum>
68
69  </conversion>
```

- Create an empty conversion.xml file:

```
# aria2 --convert -t conversion.xml
```

- Fill in the values in conversion.xml (see the example file in listing C.4 on page 175)
- Edit all the .peaks files that were exported from Sparky as follows (example given for 313.peaks only):

```
# sed s/0 e 0/0e0 m 0   /g < 313.peaks > 313_.peaks
# sed s/$/ 0/g < 313_.peaks > 313__.peaks
```

Make sure that there are no empty lines at the end of the .peaks file You should end up with something like this:

Listing C.5: *The 313_ _.peaks file after the necessary edits*

```
# Number of dimensions 2 0
   1  4.363  8.318  1 U  2.15e+06        0e0 m 0      63   62  0
   2  4.429  8.136  1 U  1.49e+06        0e0 m 0      18   17  0
   3  3.809  8.136  1 U  5.17e+05        0e0 m 0      19   17  0
   4  3.832  8.137  1 U  1.02e+06        0e0 m 0      19   17  0
   5  4.199  8.002  1 U  3.45e+06        0e0 m 0      95   94  0
```

- Remove any "strange" peaks from the .prot files (for example if you made assignments with question marks, comments or other things that are not normal atom names)
- Check that the filenames in conversion.xml corrspond to the new, edited input files (.peaks, .prot, .seq)
- Convert the data to xml:

```
# aria2 --convert conversion.xml > conversion-log.txt
```

- Check if there were any warnings or errors by looking at the log file. correct the assignments as needed. It is easiest to make the corrections in Sparky and restart the procedure outlined above.

If all goes well, you should now have a couple of .xml files in your working directory. You are now ready to set up the ARIA structure calculation project.

- Use the Aria2 GUI to set up the structure calculation options:

```
# aria2 -g run1.xml
```

- Fill in the settings for the structure calculation protocol. You may want to increase the number of trial structures and the number of accepted ones. Fill in the following options as well
- PROJECT: fill in the working directory and temporary directory to be used for the ARIA output
- CNS: set the location of the CNS binary

- DATA/SPECTRA: set trust assigned peaks to "yes"
- ANALYSIS: set the location of the procheck binary, disable whatif and prosa checks
- STRUCTURE GENERATION/JOB MANAGEMENT: to use both cores on a core2duo machine, edit the first line: yes — tcsh — 2 — (the cns directory) — yes
- Have Aria2 prepare its directories:

```
# aria2 -s run1.xml
```

- Start the structure calculation

```
# aria2 run1.xml
```

- Have a coffee

C.3 Analysis of the calculated structures with Procheck_nmr and Aqua

Analysis with Procheck_nmr and aqua after a Cns or Aria structure calculation

- First, you need to set up the environment for aqua and procheck_nmr (if you have not already done so). Note that we use the tcsh shell again.

```
# tcsh
# setenv prodir /some/directory/procheck
# setenv aquaroot /some/other/directory/aqua3.2
# $aquaroot/aqsetup
```

- The analysis tools expect all the pdb files you want to analyse to be merged into a single file. A script to do so is available with the aqua software package. First, go to the directory called "refine", that contains the final structures after refinement. You may also want to run the analysis on the last iteration before refinement.

```
# $aquaroot/extras/joinpdb -o allpdb.pdb *.pdb
```

- To convert the (xplor) restraints to Aqua format with qconvert, they first need to be edited slightly for qconvert to understand them. You will have to remove the "segid" section:

```
# sed 's/segid \"   A\" and //g < unambig.tbl > unambig_.tbl
```

- Also, the lines starting with "or" should be deleted because aqua does not understand them. But doing so would remove parts of the ambiguous assignments (and only keep for example one of the three methyl protons, which leads to problems). To correct this, the atoms can be renamed to their "ambiguous" names. I did this by hand, but if there are a lot of them you may want to make a script. The changes to be made should become clear from the example files below, before and after the changes (listing C.6 and listing C.7 on the facing page)

Listing C.6: *unambig.tbl as output by aria*

```
assign (segid "   A" and resid   10 and name   HA) (segid "   A" →
→and resid   13 and name   HN) 3.288 1.352 1.352 weight 1.000 ! →
→spec=NOESY298K, no=202, id=199, vol=1.120000e+06
assign (segid "   A" and resid   19 and name  HB2) (segid "   A" →
→and resid   18 and name  HD2) 4.044 2.044 2.044 weight 1.000 ! →
→spec=NOESY298K, no=203, id=200, vol=3.240000e+05
    or (segid "   A" and resid   19 and name  HB1) (segid "   A" →
    →and resid   18 and name  HD2)
    or (segid "   A" and resid   19 and name  HB3) (segid "   A" →
    →and resid   18 and name  HD2)
assign (segid "   A" and resid   11 and name  HB2) (segid "   A" →
→and resid   12 and name   HN) 2.593 0.840 0.840 weight 1.000 ! →
→spec=NOESY298K, no=204, id=201, vol=4.660000e+06
assign (segid "   A" and resid   14 and name  HB2) (segid "   A" →
→and resid   14 and name   HA) 2.265 0.641 0.641 weight 1.000 ! →
→spec=NOESY298K, no=224, id=221, vol=1.050000e+07
assign (segid "   A" and resid   14 and name   HG) (segid "   A" →
→and resid   14 and name   HA) 2.265 0.641 0.641 weight 1.000 ! →
→spec=NOESY298K, no=226, id=223, vol=1.050000e+07
assign (segid "   A" and resid   14 and name HD12) (segid "   A" →
→and resid   14 and name   HA) 2.203 0.606 0.606 weight 1.000 ! →
→spec=NOESY298K, no=228, id=225, vol=1.240000e+07
```

```
    or (segid "   A" and resid   14 and name HD11) (segid "   A" →
    →and resid   14 and name   HA)
    or (segid "   A" and resid   14 and name HD13) (segid "   A" →
    →and resid   14 and name   HA)
    or (segid "   A" and resid   14 and name HD22) (segid "   A" →
    →and resid   14 and name   HA)
    or (segid "   A" and resid   14 and name HD21) (segid "   A" →
    →and resid   14 and name   HA)
    or (segid "   A" and resid   14 and name HD23) (segid "   A" →
    →and resid   14 and name   HA)
assign (segid "   A" and resid   11 and name   HG) (segid "   A" →
→and resid   11 and name   HA) 2.452 0.751 0.751 weight 1.000 ! →
→spec=NOESY298K, no=229, id=226, vol=6.520000e+06
```

Listing C.7: *unambig.tbl after the necessary changes*

```
assign (resid   10 and name   HA) (resid   13 and name   HN) 3.288 →
→1.352 1.352 weight 1.000 ! spec=NOESY298K, no=202, id=199, vol→
→=1.120000e+06
assign (resid   19 and name  HB#) (resid   18 and name  HD2) 4.044 →
→2.044 2.044 weight 1.000 ! spec=NOESY298K, no=203, id=200, vol→
→=3.240000e+05
assign (resid   11 and name  HB2) (resid   12 and name   HN) 2.593 →
→0.840 0.840 weight 1.000 ! spec=NOESY298K, no=204, id=201, vol→
→=4.660000e+06
assign (resid   14 and name  HB2) (resid   14 and name   HA) 2.265 →
→0.641 0.641 weight 1.000 ! spec=NOESY298K, no=224, id=221, vol→
→=1.050000e+07
assign (resid   14 and name   HG) (resid   14 and name   HA) 2.265 →
→0.641 0.641 weight 1.000 ! spec=NOESY298K, no=226, id=223, vol→
→=1.050000e+07
assign (resid   14 and name  HD#) (resid   14 and name   HA) 2.203 →
→0.606 0.606 weight 1.000 ! spec=NOESY298K, no=228, id=225, vol→
→=1.240000e+07
assign (resid   11 and name   HG) (resid   11 and name   HA) 2.452 →
→0.751 0.751 weight 1.000 ! spec=NOESY298K, no=229, id=226, vol→
→=6.520000e+06
```

- Convert the restraints

```
# $aquaroot/scripts/qconvert -r aquaout unambig_.tbl
```

- If your local settings use comma "," as a decimal separator, replace it by a dot "." in the output file. Note that the output file will not contain the residue names because they are unknown at the moment. That is not a problem, however.

```
# $aquaroot/scripts/aqpc -r aquaout allpdb.pdb
```

- now run procheck_nmr:

```
# $prodir/procheck_nmr.scr allpdb.pdb
```

- To analyse the results, you may also want to run:
  - Look at all the .ps files
  - Display a summary:

    ```
    # cat allpdb.nsm
    ```

  - Visualize the noe restraints (this creates .pdb files that contain the restraints as "bonds")

    ```
    # $prodir/viol2pdb.scr allpdb.pdb ML
    ```

  - Show only the noes bigger than a certain distance (and more...). Note that this script only runs from its own directory, hence the "cd" command.

    ```
    # cd \$aquaroot/scripts/aquanal.pl
    # ./aquanal.pl /path/to/file.nsv
    ```

Bibliography

[1] R. Fogh, J. Ionides, E. Ulrich, W. Boucher, W. Vranken, J. P. Linge, M. Habeck, W. Rieping, T. N. Bhat, J. Westbrook, K. Henrick, G. Gilliland, H. Berman, J. Thornton, M. Nilges, J. Markley, and E. Laue. The ccpn project: an interim report on a data model for the nmr community. *Nat Struct Biol*, 9(6):416–418, Jun 2002.

[2] R. H. Fogh, W. F. Vranken, W. Boucher, T. J. Stevens, and E. D. Laue. A nomenclature and data model to describe NMR experiments. *J Biomol NMR*, 36(3):147–155, Nov 2006.

[3] R. Keller. *The Computer Aided Resonance Assignment Tutorial (ISBN 3-85600-112-3)*. published on-line, 2004.

[4] R. Keller. *Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer Aided Resonance Assignment*. PhD thesis, Swiss Federal Institute of Technology Zurich (ETH Zï£¡rich), 2004/2005.

[5] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, and K. WÃijthrich. Recommendations for the presentation of nmr structures of proteins and nucleic acids–iupac-iubmb-iupab inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by nmr spectroscopy. *Eur J Biochem*, 256(1):1–15, Aug 1998.

[6] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. The ccpn data model for nmr spectroscopy: development of a software pipeline. *Proteins*, 59(4):687–696, Jun 2005.

D

Files used in the MD simulations

Contents

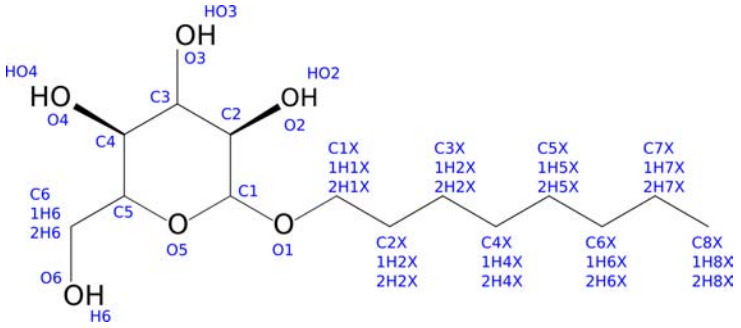

Figure D.1: *Octylglucoside atom names as used in this thesis.*

D.1 Octylglucoside topology

In table D.1, the partial charges, atom types of the GAFF forcefield and atom names are given for the entire molecule. The atom names can be identified on the structure using figure D.1.

Table D.1: *List of atom types and partial charges of octylglucoside.*

pdb-file atom name	prep-file atom type	GAFF atom type	partial charge
C8X	C8X	c3	-0.0922
1H8X	H8X1	hc	0.0310
2H8X	H8X2	hc	0.0318
3H8X	H8X3	hc	0.0328
C7X	C7X	c3	-0.0799
1H7X	H7X1	hc	0.0391
2H7X	H7X2	hc	0.0359
C6X	C6X	c3	-0.0794
1H6X	H6X1	hc	0.0385
2H6X	H6X2	hc	0.0432
C5X	C5X	c3	-0.0756
1H5X	H5X1	hc	0.0392
2H5X	H5X2	hc	0.0369
C4X	C4X	c3	-0.0803
1H4X	H4X1	hc	0.0327
2H4X	H4X2	hc	0.0478
C3X	C3X	c3	-0.0776

continued on next page...

. . . continued from previous page.

pdb-file atom name	prep-file atom type	GAFF atom type	partial charge
1H3X	H3X1	hc	0.0593
2H3X	H3X2	hc	0.0379
C2X	C2X	c3	-0.0810
1H2X	H2X1	hc	0.0506
2H2X	H2X2	hc	0.0462
C1X	C1X	c3	0.1342
1H1X	H1X1	h1	0.0289
2H1X	H1X2	h1	0.0553
O1	O1	os	-0.4244
C1	C1	c3	0.3059
H1	H1	h2	0.0534
O5	O5	os	-0.4329
C5	C5	c3	0.1065
H5	H5	h1	0.0669
C6	C6	c3	0.1440
1H6	H61	h1	0.0808
2H6	H62	h1	0.0359
O6	O6	oh	-0.5973
H6	H6	ho	0.4200
C4	C4	c3	0.1151
H4	H4	h1	0.0658
O4	O4	oh	-0.6089
HO4	HO4	ho	0.4311
C3	C3	c3	0.1092
H3	H3	h1	0.0647
O3	O3	oh	-0.6068
HO3	HO3	ho	0.4282
C2	C2	c3	0.0886
H2	H2	h1	0.0779
O2	O2	oh	-0.6043
HO2	HO2	ho	0.4255

D.2 SDS topology

In table D.2 on the next page, the partial charges, atom types of the GAFF forcefield and atom names are given for the entire molecule. The atom names can be identified on the structure using figure D.2.

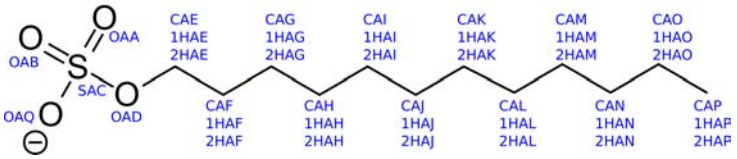

Figure D.2: *SDS atom names as used in this thesis.*

Table D.2: *List of atom types and partial charges of SDS.*

pdb-file atom name	prep-file atom type	GAFF atom type	partial charge
CAP	CAP	c3	-0.0916
1HAP	HAP1	hc	0.0262
2HAP	HAP2	hc	0.0350
3HAP	HAP3	hc	0.0309
CAO	CAO	c3	-0.0786
1HAO	HAO1	hc	0.0324
2HAO	HAO2	hc	0.0383
CAN	CAN	c3	-0.0787
1HAN	HAN1	hc	0.0393
2HAN	HAN2	hc	0.0387
CAM	CAM	c3	-0.0787
1HAM	HAM1	hc	0.0379
2HAM	HAM2	hc	0.0366
CAL	CAL	c3	-0.0800
1HAL	HAL1	hc	0.0379
2HAL	HAL2	hc	0.0438
CAK	CAK	c3	-0.0765
1HAK	HAK1	hc	0.0384
2HAK	HAK2	hc	0.0307
CAJ	CAJ	c3	-0.0793
1HAJ	HAJ1	hc	0.0434
2HAJ	HAJ2	hc	0.0383
CAI	CAI	c3	-0.0781
1HAI	HAI1	hc	0.0298
2HAI	HAI2	hc	0.0363
CAH	CAH	c3	-0.0822

continued on next page...

. . . continued from previous page.

pdb-file atom name	prep-file atom type	GAFF atom type	partial charge
1HAH	HAH1	hc	0.0556
2HAH	HAH2	hc	0.0363
CAG	CAG	c3	-0.0730
1HAG	HAG1	hc	0.0168
2HAG	HAG2	hc	0.0356
CAF	CAF	c3	-0.0902
1HAF	HAF1	hc	0.0417
2HAF	HAF2	hc	0.0417
CAE	CAE	c3	0.2078
1HAE	HAE1	h1	0.0078
2HAE	HAE2	h1	0.0104
OAD	OAD	os	-0.6098
SAC	SAC	s6	1.6985
OAA	OAA	o	-0.7657
OAQ	OAQ	o	-0.7647
OAB	OAB	o	-0.7387

D.3 Details of the MD simulations

The minimisation was carried out with Cartesian position restraints (ntr=1) of 1.0 kcal/mol (ntr=1) on the peptide (restraintmask=:1-25), with constant volume periodic boundary conditions (ntb=1), a nonbonded cutoff distance of 10 Å (cut=10),

Listing D.1: *sander (AMBER 9) input file for minimisation*

```
Minimization with Cartesian restraints on the 25AA protein
 &cntrl
  imin=1,
  ncyc=1000
  maxcyc=2500,
  ntb=1,
  cut=10,
  ntpr=5,
  ntr=1,
  restraint_wt=1.0,
  restraintmask=':1-25',
 /
```

After the minimisation, first the temperature was calibrated during 100 ps (nstlim=50000) with a 2 fs timestep (dt=0.002). The collision frequency was set to 1.0 ps^{-1} and Langevin dynamics were used with a Leapfrog integrator (gamma_ln=1.0, ntt=3). The initial temperature was set to 0 K (tempi=0.0), and it was raised to a final value of 300 K during the simulation (temp0=300). Pressure scaling was not active (ntp=0). Bonds involving hydrogen atoms were constrained with the SHAKE algorithm (ntc=2) and forces were only evaluated for the bonds not involving hydrogen atoms (ntf=2).

Listing D.2: *sander (AMBER 9) input file for temperature equilibration*

```
MD run: temperature equilibration: T coupling but no P coupling
 &cntrl
  imin=0,
  irest=0,
  ntx=1,
  cut=10,

  nstlim=50000,
  dt=0.002,

  ioutfm=0,
  ntwe=500,
  ntwx=500,
  ntpr=500,
  ntwr=500,

  iwrap=1,

  ntr=1
  restraint_wt=1.0,
  restraintmask=':1-25',

  tempi=0.0,
  temp0=300.0,
  ntt=3,
  gamma_ln=1.0,

  ntb=1,
  ntp=0,
  pres0=1.0,
```

```
  taup=1.0,

  ntc=2,
  ntf=2,

/
```

After the temperature calibration, the volume was equilibrated during 1 ns, still using a timestep of 2 fs (nstlim, dt). The equilibration of the volume was run with active pressure coupling (ntb=2, ntp=1) at 1 bar (pres0=1.0). The pressure relaxation time was set to 1.0 ps (taup=1.0), which is the default value.

Listing D.3: *sander (AMBER 9) input file for volume equilibration*

```
MD run: pressure (volume) equilibration: T coupling and P coupling
 &cntrl
  imin=0,
  irest=1,
  ntx=5,
  cut=10,

  nstlim=500000,
  dt=0.002,

  ioutfm=0,
  ntwe=5000,
  ntwx=5000,
  ntpr=5000,
  ntwr=5000,

  iwrap=1,

  ntr=1
  restraint_wt=1.0,
  restraintmask=':1-25',

  ntt=3,
  gamma_ln=1.0,

  ntb=2,
  ntp=1,
```

```
  pres0=1.0,
  taup=1.0,

  ntc=2,
  ntf=2,

/
```

For the productuon run, the same settings were used as for the pressure equilibration, except that the Cartesion restraints on the protein were removed (ntr=0). While for the equilibration runs the "sander" executable was used, the production run was carried out with the "pmemd" command because of its increaed performance.

Listing D.4: *pmemd (AMBER 9) input file for the production run*

```
Production run
 &cntrl
  imin=0,
  irest=1,
  ntx=5,
  cut=10,

  nstlim=10000000,
  dt=0.002,

  ioutfm=0,
  ntwe=5000,
  ntwx=5000,
  ntpr=5000,
  ntwr=5000,

  iwrap=1,

  ntr=0

  ntt=3,
  gamma_ln=1.0,

  ntb=2,
  ntp=1,
  pres0=1.0,
```

```
27   taup=1.0,
28
29   ntc=2,
30   ntf=2,
31
32 /
```

For the run with the active NOE restraints, all parameters were kept constant, but the restraints were added. The part of the input file that is equal to the file for the run without restraints is presented here by (...).

Listing D.5: *part of the sander (AMBER 9) input file for a production run with NOE restraints*

```
1   ...
2
3   nmropt=1
4 /
5  &wt type='REST', istep1=0, istep2=3000, value1=0.1, value2=1.0 /
6  &wt type='REST', istep1=3001, istep2=10000000, value1=1.0, value2=1.0 /
7  &wt type='END'  /
8 DISANG=./noe-restraints-sander.txt
9 LISTOUT=POUT
```

List of Figures

List of Tables

Listings

Author
Louic S. Vermeer

Title
NMR structure determination and MD simulations of membrane peptides and proteins: a peptide derived from H+-V-ATPase subunit a, and MscL

Title (French)
Détermination de structure par RMN et simulations dynamique moleculaire de peptides et protéines membranaires: un peptide derivé de la sous-unité a d'une H+-V-ATPase, et MscL

Summary
The structural properties of a peptide derived from the proton translocating H+-V-ATPase subunit a were studied. An NMR structure was obtained in SDS micelles. Circular dichroism measurements indicated that the peptide formed a beta-sheet at high pH in octylglucoside, while it was 60% alpha-helical in SDS. These findings were explained using molecular dynamics simulations, which indicated that at high pH the peptide took a transmembrane position in SDS but was located in the interface region in octylglucoside. The combination of the NMR structure and the MD simulations allowed us to identify the residues that line the lumenal proton channel.

As a second part of this thesis, the preparation of samples in oriented lipid bilayers of the mechanosensitive channel of large conductance (MscL) was attempted, in order to carry out solid-state NMR measurements.

Résumé
H+-V-ATPase est une protéine responsable de la translocation des protons. Les propriétés structurales d'un peptide dérivé de la sous-unité a de H+-V-ATPase, ont été étudiées. La structure dans des micelles de SDS a été déterminée par RMN. Il a été démontré par dichroisme circulaire que le peptide est structuré en feuillet beta dans l'octylglucoside à haut pH, tandis qu'il présente 60\% d'hélice alpha dans le SDS. Ces observations ont été expliquées par des simulations de dynamique moléculaire, qui ont démontré que le peptide est inseré dans la micelle de SDS à haut pH, alors qu'en présence d'octylglucoside il est en interaction avec la surface de la micelle. La structure RMN en combinaison avec les simulations nous a permis d'identifier les residues participant au canal périplasmique.

Dans la deuxième partie de la thèse, des essais de préparation d'échantillons du canal mécanosensible de haute conductance (MscL) dans des bicouches orientées de lipides ont été réalisés afin d'étudier cette protéine par RMN du solide.

Key words
NMR, solid-state NMR, membrane protein, membrane peptide, SDS, octylglucoside, molecular dynamics simulation, MD, cholesterol, DMPC, order parameters, MscL, mechanosensitive channel, H+-V-ATPase subunit a, proton translocation