Dissertation

Submitted to the Combined Faculties for
the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the Degree of Doctor of Natural Science

Presented by

Durba Sengupta
Born in: Delhi, India
Oral Examination: July, 2005

Insights into the Energetics of Membrane-Bound Peptides

Towards an Understanding of the Structural Organisation of Membrane Proteins

Referees: Prof. Dr. Jeremy C. Smith

Prof. Dr. G. Matthias Ullmann

Abstract

A large number of key cellular functions such as signalling and transport are performed and regulated by membrane proteins. The focus of this thesis is on a sub-class of membrane proteins, namely the monotopic and bitopic peptides. Analysis of several aspects of these peptides such as their orientation, electrostatics and association is reported here. The environment of these peptides is represented by a continuum model that distinguishes between water, membrane core and head-group region. The orientation of membrane peptides such as glycophorin A and melittin is calculated and reproduces the experimentally-calculated tilt angles. The length dependence of synthetic peptides such as WALPs is reproduced and found to depend on the cost of cavity formation in the aqueous layer. It is shown that the solvent reaction field plays a crucial role in determining the orientation of polypeptides. The reaction field is also shown to screen the helix dipole of membrane-bound helices depending on the proximity and geometry of the aqueous phase relative to the helix termini. As a result, the helix dipole of transmembrane helices is found to decrease with peptide length. The analysis is extended to helices in soluble proteins and rules of thumb are established to estimate the effective helix dipole from visual inspection of protein structures. The association of glycophorin A helices is modelled and the decomposition of free energy of dimerisation shows a favourable contribution from the residues experimentally implicated to contribute to the process. The association of erythropoietin receptor transmembrane dimers and that of its mutants is also modelled. The inefficient lipid raft localisation of the T242N mutant is proposed to be related to its helix packing. The thesis provides insight into the structural organisation and energetics of membrane proteins.

Zusammenfassung

Eine große Anzahl zellulärer Schlüsselfunktionen, wie Signaltransduktion oder Stofftransport, werden von Membranproteinen ausgeführt und reguliert. Das Augenmerk dieser Arbeit ist auf eine Unterklasse der Membranproteine, die monotopen und bitopen Peptide, gerichtet. Verschiedene Eigenschaften dieser Peptide, im Besonderen ihre Orientierung, Elektrostatik und Assoziation, werden analysiert. Die Membranumgebung wird durch ein Kontinuummodell beschrieben, welches zwischen Wasser, Kopfregion der Lipide und innerem Membranbereich unterscheidet. Die berechneten Neigungswinkel von Membranpeptiden, z.B. von Glycophorin A und Melittin, sind in guter Übereinstimmung mit den experimentellen Werten. Die Abhängigkeit des Neigungswinkels von der Peptidlänge wird am Beispiel von WALPs ermittelt und es wird gezeigt, dass dieser von den Kosten der Hohlraumformation in der Wasserschicht abhängt. Das Reaktionsfeld des Lösungsmittels spielt eine wichtige Rolle bei der Bestimmung der Orientierung von Polypeptiden. Dieses Reaktionsfeld schirmt die Helix-Dipole membrangebundener Helices ab, was von der Umgebung und Geometrie der wässrigen Phase relativ zu den Helixenden abhängt. Der Helix-Dipol von Transmembranproteinen nimmt daher mit der Peptidlänge ab. Diese Analyse wird erweitert auf Helices löslicher Proteine, und Faustregeln für die Abschätzung des effektiven Helix-Dipols mittels visueller Untersuchung der Proteinstruktur werden aufgestellt. Weiterhin wird die Assoziation der Glycophorin A-Helices modelliert. Eine Zerlegung der freien Energie der Dimerisierung in ihre Komponenten liefert einen günstigen Beitrag für genau die von Experimenten implizierten Residuen. Die Assoziation von Erythropoietinrezeptortransmembrandimeren und ihrer Mutanten wird ebenfalls modelliert. Vermutlich ist die dichte Packung der Helices ein Grund dafür, dass der T242N-Mutant nicht in Lipid Rafts vorkommt. Die vorliegende Arbeit liefert Einblicke in die strukturelle Organisation sowie die Energetik von Membranproteinen.

List of Publications

"To get the right word in the right place is a rare achievement ... Anybody can have ideas- the difficulty is to express them without squandering a quire of paper on an idea that ought to be reduced to one glittering paragraph.", Mark Twain.

[1] Durba Sengupta, Lars Meinhold, Dieter Langosch, G. Matthias Ullmann, and Jeremy C. Smith. Understanding the Energetics of Helical-Peptide Orientation in Membranes. *Proteins*, **58**:913-922, 2005.

[2] Durba Sengupta, Raghu Nath Behera, Jeremy C. Smith and G. Matthias Ullmann. The α Helix Dipole - Screened Out?. *Structure*, **13**:849-855, 2005.

[3] Verena Becker, Robin Ketteler, Durba Sengupta, Marcel Schilling, Ute Baumann, Achim C. Heinrich, G. Matthias Ullmann, Jeremy C. Smith and Ursula Klingmüller. Close Packing of the Erythropoietin Receptor Transmembrane Domain Regulates Lipid Raft Localization and Selective Signal Amplification, *submitted*.

Acknowledgments

"Thank You Jeeves", The Jeeves Omnibus, PG Wodehouse.

I would like to begin by thanking all past and current members of the Computational Molecular Biophysics group for giving it its unique character and providing a relaxed and friendly place to work. The main driving force of the group is Prof. Jeremy C. Smith and I thank him sincerely for the innumerous corrections of my article manuscripts and for the suggestions, discussions and ideas. I am grateful to Prof. G. Matthias Ullmann who helped give the project a direction at every cross road and provided motivation during several seemingly endless stretches. The original research project was proposed by Prof. Dieter Langosch and Jeremy and I thank them both for the guidelines as well as the Daimler-Benz Stiftung for making it possible for me to work on the project.

Lars Meinhold collaborated on the analytical solution of the PB equation and I thank him for patiently answering my countless questions. Dr. Raghu Nath Behera developed the charge-fitting method and his help in making it CHARMM compatible is gratefully acknowledged. The collaboration with Dr. Ursula Klingmüller and Verena Becker was extremely interesting and I appreciate their efforts to make the project fruitful. Many heartfelt acknowledgements are also due to Bogdan Costescu for being an extremely patient and helpful system administrator.

I thank Edda Kloppmann who has been a constant source of answers on biology, electrostatics, books and manuscripts. I also thank Erika Balog and Alexander Tournier who helped me immensely at the start of my doctoral studies. Acknowledgments are also due to Thomas Splette for help with the german translations and to Torsten Becker for patiently reading first versions of manuscripts. Many thanks are due to Ileana Borja for her help with non-academic issues. I also thank Chaitanya Athale for proof-reading this thesis when required and doing nothing when not required.

I am indebted to my parents, Dr. Santanu Sengupta and Dr. Krishna Sengupta for their patience and strength. I am also grateful to my sister, Dr. Kheya Sengupta who has always been a source of inspiration.

Contents

Chapter 1

Bio-membranes and Membrane Proteins

"Where shall I start... I will start with a fairly recent event, ***sub specie alternitatis****... and work both ways: backwards and forwards", Thomas Gray, The Philosopher Cat, Philip J. Davis.*

The cell is the basic unit of all living organisms. It is a complex but highly organised entity separated from the outside by a sophisticated transport and signalling membrane system. The genetic information is stored in nucleic acids which act as templates for proteins, which in turn carry out the basic functioning of the cell. Two general cell types can be distinguished: prokaryotic and eukaryotic. Both types are enclosed by bio-membranes, but eukaryotic cells contain additional membrane-bound organelles and structures that carry out specialised functions [1].

Bio-membranes are bilayer structures that separate the cell from its surrounding and the eukaryotic cell into compartments. The bio-membrane is heterogeneous and is composed of a variety of lipids, cholesterol and proteins [1]. The proteins associated with the membrane form a unique class of proteins called membrane proteins, since the membrane environment plays an important role in modulating their structural and energetic properties. The lipids comprising the bio-membrane make the environment of these membrane proteins less polar compared to their cytosolic soluble counterparts [2].

Membrane proteins play a crucial role in the cell and their importance is reflected in the fact that one third of the human genome is estimated to encode membrane proteins. These ubiquitous proteins play diverse roles in biological processes, such as transport of ions and polar compounds, pH regulation, signal transduction and vesicle fusion [2–4]. Due to their role in such key cellular functions, membrane bound proteins represent approximately 60% of drug targets.

The ubiquity and importance of membrane proteins is contrasted by a considerable lack of experimental and theoretical data regarding their structure and function, since methods to investigate these proteins in their physiological environment are in general demanding and cost intensive. In particular, the difficulties in crystallising membrane proteins has resulted in less than 30 well-resolved unique structures of membrane proteins being available so far [4]. New approaches, both experimental and theoretical are therefore needed to determine the structure and understand the energetics of membrane proteins in their natural environment taking into account their interaction with the membrane. In this chapter, general structural and energetic aspects of membrane proteins will be reviewed along with details of the membrane. Furthermore, various membrane models and current theoretical approaches to study membrane proteins will be discussed.

1.1 The Cell Membrane

Bio-membranes enclose and compartmentalise the cell and cellular organelles [1]. The average thickness of the bio-membrane is 4 nm, though slight variations of this value are common. In bio-membranes, the fraction of lipids and proteins vary depending on the function of the membrane and the compartments it separates. For instance, the mass percentages of proteins to lipids vary from 76% protein in the inner mitochondria membrane to 18% protein in the myelin membrane [5]. A cartoon representation of a membrane, depicting lipids (phospholipids and sphingolipids), cholesterol and proteins is shown in Fig. 1.1.

1.1.1 The Bilayer Structure

In the bio-membrane, the lipids *i.e.*, phospholipids and sphingolipids are arranged in an unique bilayer structure with each monolayer facing the other. The composition of the two monolayers is however different [1]. The bilayer arrangement of the lipids leads to a single hydrophobic core comprised of the lipid fatty-acid tails and two hydrophilic outer layers comprised of lipid head groups. Thus, the membrane has a spatially dependent polarity which changes from non-polar to polar as one traverses from the membrane centre towards the aqueous layer.

The major phospholipids comprising the bilayer are phosphatidylethanolamines, phosphatidylserines and phosphatidylcholines. In biological representations, the structure of these phospholipids is often simplified into a hydrophilic, zwitterion head-group region (represented by circles in Fig. 1.1) and a hydrophobic alipathic tail region (represented

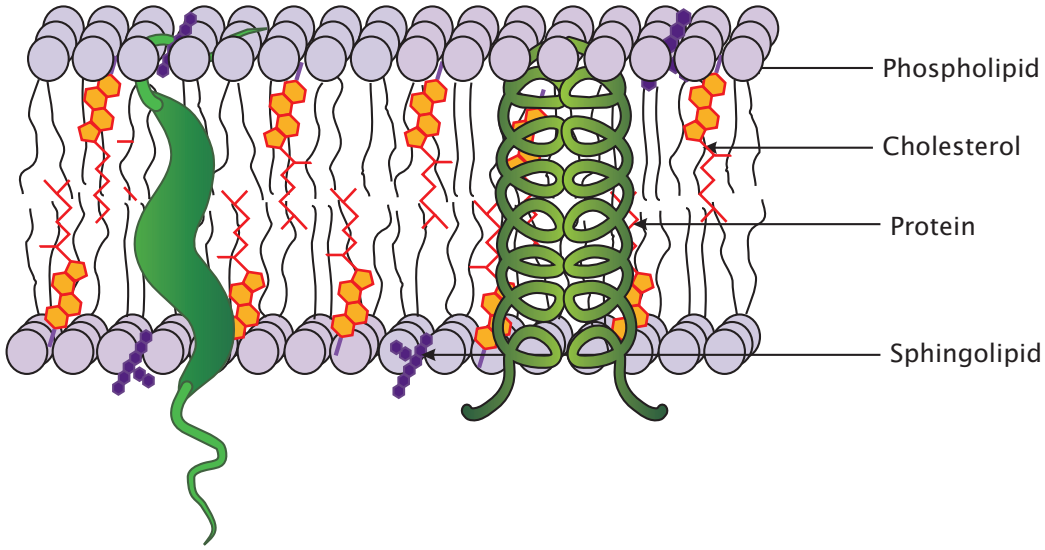

Figure 1.1: A cartoon representation of a bio-membrane. Membranes consist of phospholipids (depicted in lilac), cholesterol (yellow), sphingolipids (purple) and membrane proteins (green). In the figure, the head groups of phospholipids are represented by circles and fatty acid tails by lines. The aqueous layer surrounding the membrane is not represented for clarity.

by the two lines in Fig. 1.1). The phospholipid head groups interact with the surrounding aqueous environment and form the two outer membrane head-group regions. The fatty-acid tails are sandwiched in between and form the membrane core.

Sphingolipids are also made of hydrocarbon chains which are buried in the membrane core and zwitterion groups which are located in the membrane head-group region. However, the sugar moieties of the sphingolipids extend into the aqueous layer. Cholesterol is another major component of bio-membranes and is located between the lipids in the bilayer. The hydrophobic part of cholesterol (the ring system and the extending alkane chain) is present in the membrane core and the hydroxyl group extends to the head-group region.

The bilayer structure of the membrane confines lipids and membrane bound proteins within the plane of the bilayer. However, both lipids and proteins have considerable amount of in-plane lateral diffusion. The essential features of the bio-membrane are described by the *fluid mosaic model* as proposed by Singer and Nicolson in 1972 [6]. The model describes the membrane as a two-dimensional fluid, in which the hydrophobic integral mosaic components, such as lipids and membrane proteins are constrained within the plane of the membrane, but are free to diffuse laterally. In spite of its simplicity, the model has proved accurate in explaining many aspects of bio-membrane and membrane proteins.

1.1.2 Fluidity and Mobility in the Membrane

At room temperatures, the membrane core is fluid-like with disordered hydrocarbon tails [7]. The lipids comprising the membrane have high rotational and translational freedom within the plane of the membrane [8]. The lateral movement of the lipid molecules is essential for cellular function, such as facilitating membrane fusion events [9] and signal transduction [10]. In contrast to the high degree of lateral movement, the flipping of a lipid molecule from one layer to the other is uncommon and needs to be facilitated by enzymes known as flipases [7].

Fluidity of the membrane is dependent on the degree of saturation of the fatty acid tails of the lipid molecules. The saturated lipid molecules are closely packed and therefore allow little lateral movement [7]. To maintain fluidity at suboptimal temperatures, organisms use phospholipids with increasing degrees of unsaturation in their fatty acid tails [11]. A direct correlation between the growth temperature and the degree of lipid unsaturation has been observed — the lower the temperature, the lower the degree of lipid saturation [12]. The membrane lipid composition of thermophilic yeasts is characterised by a high percentage (30%-40%) of saturated fatty acids, as compared with the mesophilic (10%-30%) and psychrophilic (less than 10%) yeasts.

Membrane fluidity is also modulated by its cholesterol content. At high temperatures, cholesterol tends to reduce membrane fluidity by inserting and interposing between the larger phospholipid and glycolipid molecules and interacting with the fatty acid chains [7]. At low temperatures, however, the interposed cholesterol helps prevent membranes from freezing and thus tends to maintain membrane fluidity. Interestingly, the distribution of cholesterol within the bilayer is not uniform. More ordered, dynamic assemblies with high-cholesterol content, known as lipid rafts are present within the liquid-disordered bilayer [13].

1.1.3 Order in the Membrane - Lipid Raft assemblies

Within the fluid-like membrane, distinct ordered phases are observed. These dynamic ordered assemblies are known as *lipid rafts* and are characterised by high cholesterol and sphingolipid content [10, 14]. Since lipids rafts introduce variable degrees of order within the membrane, the simple fluid mosaic model [6] does not strictly apply.

Some membrane proteins are seen to aggregate in the lipid rafts. The majority of the proteins that localise in lipid rafts are involved in signalling [10]. Dynamic partitioning of proteins in lipid rafts has been implicated to promote specific receptor clustering and signal transduction [10]. Membrane proteins involved in endocytosis [15], exocytosis [10],

synaptic vesicle traffic [16] and secretion [17] are also found in lipid rafts.

The mechanisms by which proteins and peptides are targeted to lipid rafts is still not clearly understood. An 'address motif', such as a di-leucine sequence present in the cytosolic tails is proposed as a mechanism by which membrane proteins are targeted to lipid rafts [13]. However, the targeting mechanism need not be restricted to addresses in the cytosolic domain but may also involve specific interactions of the transmembrane segment with the surrounding lipids and cholesterol [13]. For instance, residues of the exoplasmic side of the transmembrane segment of hemagglutinin have been shown to influence lipid-raft association [18]. Close packing of the transmembrane segments as well as surrounding lipid molecules may also be important since an increase in acyl-chain order is observed around a synthetic model hemagglutinin peptide that has been implicated to be present in lipid rafts [19]. Another factor influencing the ability of peptides to sort into lipid rafts is the length of membrane spanning domain [20].

1.2 Membrane Proteins

A set of membrane proteins are associated with each bio-membrane and enable the membrane to carry out its distinctive functions and activities. The nature of proteins attached to a membrane vary depending on cell type and subcellular location [21]. Their functions include transport of substances across membranes, enzymatic activity, signal transduction, cell-cell recognition, energy transduction and attachment to the cytoskeleton and the extracellular matrix [7]. Despite their diversity in function, membrane proteins share a high degree of structural similarities, since the membrane dictates their energetics and modulates their structural properties. The first complete sequence of a protein spanning the entire bilayer was that of glycophorin A in 1978 [22]. The study of sequence, structure and energetics of membrane proteins has thus become possible only in the last two decades.

1.2.1 The Role of Proteins in Membranes

Proteins associated with the membrane carry out a variety of functions. Many of the embedded membrane proteins function as transporters, since the apolar membrane core is impermeable to ions and small hydrophilic molecules, such as glucose that need to be transported across the membrane [7]. The transporter proteins function with thermal fluctuation without input of energy (facilitated or mediated diffusion), or are driven by electrochemical potential gradients of H^+ and Na^+ (rarely of K^+) or by various exergonic chemical and photochemical reactions [11].

Ion channels are transport proteins that transport specific ions, such as Na^+, K^+, H^+ and Cl^- down their chemical or electrochemical potential gradient [7]. Most channels exist in two conformational states, open and closed. The opening of ion channels can be accomplished (a) by an electric potential (potential-gated channels), (b) by ligand binding (chemically gated channels), (c) by mechanical stress or strain (mechanically gated channels). When in the open conformation, the specific site in the channel can transiently bind solutes from both sides of the membrane.

Ion pumps, such as the Na^+/K^+-ATPase constitute another class of transport proteins and pump ions across the membrane against concentration gradient and often against the membrane potential [7]. The membrane potential is the electrical potential difference on either side of a membrane and usually measured between the cytosolic and extracellular side of the membrane. The membrane potential arises since the membrane acts as barrier to the diffusion of ions and is usually negative under resting conditions. For instance, the membrane potential in nerve cells is about -60 mV.

A key membrane protein is aquaporin, a water channel that transports water across the membrane [2]. Even though water diffuses across bio-membranes *via* osmosis, a water channel is required since the rate of diffusion is very low.

The protein domains that are found on the extracellular membrane surface are generally involved in cell-cell signalling, such as nerve transmission, hormone binding and growth stimulation [21]. Signalling pathways often function by inducing conformational changes in the receptor protein upon ligand binding. Furthermore, most proteins associated with energy transduction are membrane bound, such as the photosynthetic complex and the respiratory chain proteins. Membrane proteins may contain domains with enzymatic activity *e.g.* cytochrome P450.

1.2.2 Structure of Membrane Proteins

The structure and location of membrane proteins is correlated to their various functions. A schematic diagram of common types of membrane proteins is depicted in Fig. 1.2. Membrane proteins which penetrate into or span the bilayer are known as *integral membrane proteins*. Examples of such proteins are ion channels and gates, permeases and ATPases. The hydrophobic domain of such a protein resides in the membrane core, while hydrophilic domains protrude into the aqueous environment. *Peripheral membrane* proteins are associated with membranes, but do not penetrate the hydrophobic core of the membrane. They are often found in association with integral membrane proteins. The peripheral membrane proteins may also be anchored to the lipid bilayer *via* an 'anchor' moiety, such as a

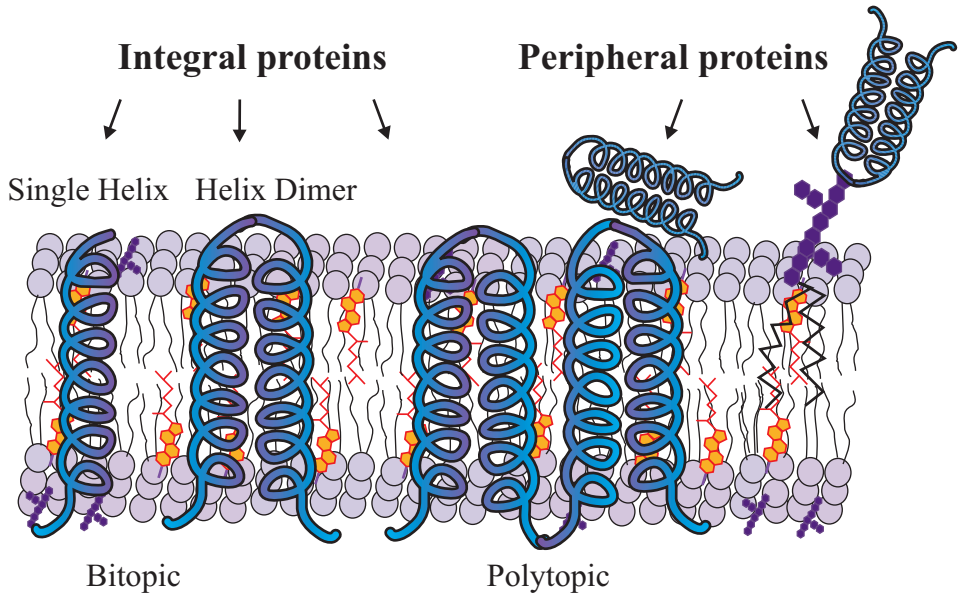

Figure 1.2: Different architectures of membrane proteins. Only α-helical proteins are depicted as typical examples, though β-barrel proteins are also present. Proteins containing membrane-bound peptides are known as integral proteins whereas proteins associated with the membrane but not membrane-bound are known as peripheral proteins. Bitopic proteins are integral membrane proteins with one membrane spanning (transmembrane) segment and polytopic proteins have more than one membrane spanning segment.

palmitate, a myristate fatty-acid or a GPI anchor. Examples of GPI anchored peripheral membrane proteins are uromodulin, dipeptidylpeptidase and alkaline phosphatase.

Integral membrane proteins are further classified into three categories, namely monotopic, bitopic and polytopic, according to their membrane-spanning topology (Fig 1.2). Monotopic proteins are those proteins which are surface aligned, *i.e.*, exposed to only one side of the membrane. Bitopic proteins have one membrane spanning (transmembrane) segment while polytopic proteins span the membrane more than once.

1.2.3 Common Architecture of Integral Membrane Proteins

Integral membrane proteins exhibit either α-helical or β-sheet topologies, though the α-helical proteins are more common [2]. The structure and stability of integral membrane proteins is based on the principle that a large thermodynamic cost accompanies the transfer of charged or polar uncharged compounds into the membrane core. The first consequence of this thermodynamic cost is that most of the amino-acid side chains of transmembrane

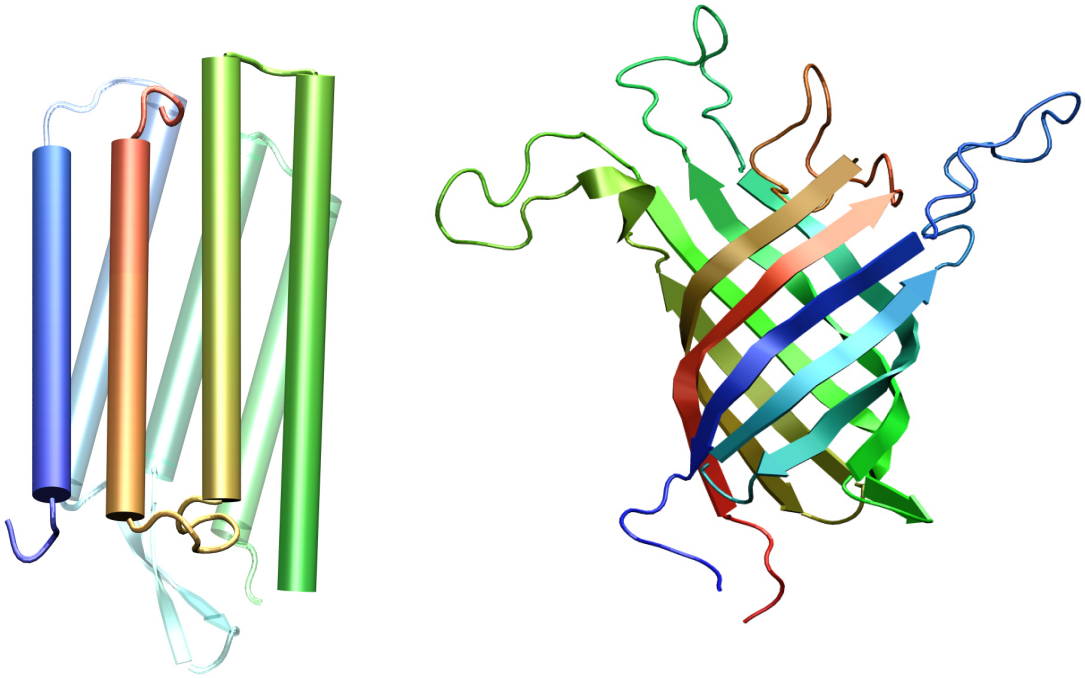

Figure 1.3: The α-helical bundle (left) and β-sheet (right) architectures of membrane proteins are seen in bacteriorhodopsin and OMPa respectively.

segments are non-polar (e.g. Ala, Val, Leu, Ile, Phe). The second effect is that the polar peptide bonds of the polypeptide backbone of transmembrane segments must participate in hydrogen bonds in order to lower the cost of transferring them into the hydrocarbon interior. This hydrogen bonding is most easily accomplished with α-helices for which all peptide bonds are H-bonded internally. It can also be accomplished with beta-sheets provided that the β-strands form closed structures, such as the β-barrel. Thus, α-helices and β-barrels are the two known structural motifs in integral membrane proteins. Fig. 1.3 shows the structure of two typical membrane proteins: bacteriorhodopsin (1C3W [23]), a seven helical bundle and the outer membrane protein (1G90 [24]), a β-barrel protein.

Single transmembrane α-helices represent the simplest category of membrane proteins and principles applying to them are general for all proteins. These helices are fundamental building blocks of membrane proteins and association of these helices in the membrane is a simple system for understanding membrane-protein architecture. The structural organisation and energetic properties of membrane helices and their self interaction inside the bio-membrane is the central focus of research in this thesis.

1.2.4 Folding and Assembly of Integral Membrane Proteins

The key finding of α-helical membrane protein folding and assembly is that helix folding usually precedes helix packing, *i.e.*, the association between helices of polytopic proteins or between dimers of single helices [8, 25, 26]. As a consequence, each transmembrane helix can be viewed as an independent folding domain. The net hydrophobicity of these segments precludes their dissociation from the membrane, while the cost of breaking hydrogen bonds within a low dielectric medium prevents the helices from unfolding [8].

The folding and association of membrane helices is currently described by the so called 'four-step association model' [8]. The four steps are partitioning, folding, insertion and association of the helices. The four steps are considered to be independent of each other and to proceed sequentially. The partitioning step is relevant for non-constitutive membrane peptides which first partition into the bilayer head-group region and then fold [8]. Constitutive membrane peptides are folded *via* the translocon complex, which provides a membrane like environment during folding. In the third step, the folded helix is inserted into the membrane. The driving force for insertion is derived mainly from the transfer of hydrophobic side chains from water to the apolar region of the bilayer [8, 25, 27]. The fourth step is the association of the inserted helices to form the native tertiary structure.

The free energies of the folding and association process have been estimated experimentally by various methods [8]. The experimental methods to measure the energetics of the processes often use synthetic model peptides with designed sequences ensuring a better handle on the variables and degrees of freedom involved in the processes. However, methods to measure the free energies of these processes association are not precise. Attempts have also been made to theoretically calculate the free energy for each of these steps and provide insight into these processes [8, 28–35].

1.2.5 Single Membrane Helices

The role of the single helix in the membrane is both structural and functional. The main structural role is to anchor the soluble domain of the protein to the membrane. Single membrane helices also play an active part in the functioning of membrane proteins, for example, in the transport of ions and polar moieties, in signal transduction and in vesicular fusion [4]. For many of these specialised roles, the transmembrane helices associate to bring soluble domains together and to aid in the function. Understanding the structure and energetics of the transmembrane domain monomer is thus important for studying its function and for analysing the driving force behind its association with other transmembrane segments.

When introduced into the bilayer, a single membrane peptide takes up the common α-helical fold and its characteristic orientation in the bilayer. The presence of the α-helical fold in single membrane peptides has been detected by Circular Dichroism [36], X-ray crystallography [8] and NMR [37]. The α-helical peptides may tilt and embed to varying extents in the membrane based on the peptide-lipid interactions [38–41]. Based on their orientation, the α-helical peptides are classified either as transmembrane or surface aligned. Transmembrane helices span the entire bilayer and interact with both monolayers of the membrane. A single transmembrane helical protein is thus termed as a bitopic integral membrane protein and a protein containing more than one transmembrane helix is a polytopic protein. A stretch of 20-30 hydrophobic residues is characteristic of transmembrane segments. Examples of transmembrane segments are glycophorin A [42], synaptobrevin [43] and the erythropoietin receptor [44]. In contrast, the surface-aligned peptides are present in the head-group region and interact with mainly one side of the bilayer. Surface-aligned peptides are also known as monotopic integral proteins and are usually amphipathic with the polar residues lining the helical face interacting with the aqueous layer and the non-polar residues interacting with the membrane core. Surface-aligned peptides have also been found to deviate from helical structures. Examples of surface-aligned peptides are melittin [45] and gramicidin A [46].

Certain positional preferences for amino-acid residues are seen in transmembrane segments. To characterise positional preferences of residues, the orientations of a number of model synthetic peptides, such as polyalanine [47], polyleucine [48], WALP [38–41] and KALP [38, 49] peptides, have been determined. WALP and KALP are peptides of alternating Ala and Leu residues with flanking Trp and Lys residues, respectively. Hydrophobic residues, such as Ala, Leu, and Ile are usually incorporated into the membrane core [47, 48, 50]. Aromatic residues, such as Trp and Tyr interact with the head-group region, often at the level of the lipid carbonyls [40, 51]. In contrast, positively-charged residues extend into the lipid phosphate region [38, 49]. From a functional point of view, it implies that Trp and Lys are often seen as flanking residues of transmembrane segments and influence the precise inter-facial positioning of membrane proteins. Negatively-charged residues, such as Asp and Glu are rarely found in transmembrane segments. However, these residues have been observed to participate in peptide association [44].

Spectroscopic studies on WALP and KALP peptides have shown the tilt angle *i.e.*, the angle between the helix axis and the bilayer normal to systematically increase with length [40, 48]. Peptides shorter than the bilayer thickness are oriented roughly parallel to the membrane normal and the longer peptides tilt appreciably. Bilayer thinning has been proposed to compensate the unfavourable hydrophobic mismatch between helices and lipid bilayers [41]

Single membrane α-helices have been the focus of experimental studies for investigating

interactions with the bilayer and the resulting intrinsic tilt of the peptide helix. Orientation of several synthetic model peptides in bilayers have been determined [38–41, 48, 49]. The single peptides represent a simple system in which peptide-lipid interactions are thought to dominate over peptide-peptide interactions in determining helix orientations [40]. However, the basic principles driving helix tilting and the energetics of the orientation of α-helices in membranes are still not well understood.

1.2.6 Helix Association

Helix association is a key event in the folding of polytopic membrane proteins [2]. Many important bio-energetic and signalling events involve the transient or permanent association of membrane proteins *via* their membrane-spanning domain [1]. Thus, helix-helix associations occur both as inter- and intramolecular processes. Intermolecular processes involve oligomerisation of single helices or association of polytopic membrane proteins to form oligomers. Intramolecular events are those in which the helices of a polytopic membrane protein associate during membrane protein folding.

Although the sequence of events during helix association is adequately described by the four-step model (see Section 1.2.4), the forces stabilising and driving specific helix-helix interactions is poorly understood [52]. However, the characterisation of the thermodynamic and kinetic properties of this assembly by both experimental [53] and theoretical [52] approaches is being attempted.

Several helix-helix interaction motifs have been identified, although, how they promote specific association is not well understood [27, 52, 54]. The specificity and stability of these helix pairs is possibly dictated by a precise fit between complementary surfaces and electrostatic interactions of backbone and side chains [8, 27]. Furthermore, certain organisational features, such as close packing of the residues at the interface and van der Waals interactions are considered to be important for helix association [27, 52, 54].

Non-polar residues, such as Gly, Ala and Leu, proposed to be involved in side-chain packing have been shown to mediate tight helix-helix contacts [55–57]. In fact, the GG4 motif (GxxxG) is known to mediate dimerisation in glycophorin A [55–59]. The GxxxG motif has also been observed in peptides, such as syndecan and the MI3 coat protein [59]. Interfaces of these pair of peptides adopt negative crossing angles and pack as a right-handed helix pair. A similar motif, SxxSSxxT, has also been shown to drive dimerisation in model peptides [55, 57]. A heptad repeat motif of Leu residues has been implicated to drive helix association in synaptobrevin and related SNARE proteins [43]. Furthermore, the leucine zipper like motif has been shown to drive self assembly of artificial transmembrane segments

in natural membranes and in detergent solution [57, 60]. In contrast to the GxxxG motif helices, these helices are proposed to adopt positive crossing angles and pack as left-handed helix pairs [43, 61].

Apart from the side-chain interactions discussed above, helix association is influenced by electrostatic interactions and hydrogen-bond formation [55]. Polar residues, such as Asn and Gln are not frequently found in the helix-helix interface, but have been shown to stabilise membrane association of helices [57] and even drive the association of model transmembrane peptides [44, 62–64]. However, another study indicated that polar residues mediate helix-helix association only when placed at appropriate positions [65].

An important question regarding the assembly of individual helices concerns the presence and importance of nonspecific driving forces for helix aggregation. The nature and origin of nonspecific forces are principally different from the specific driving forces discussed above. For example, it is speculated that removal of partially ordered lipid molecules from the helix surface gives rise to an unfavourable ordering of the lipid molecules, which in turn mediates helix-helix interaction. The effect is termed the *solvophobic effect* since it is similar to the well known *hydrophobic effect*, where the unfavourable ordering of water molecules drives protein folding and association to increase the entropy of the water molecules. However, the solvophobic effect has not been assessed experimentally and contradictory theoretical estimates of the associated energy are found in the literature [52, 66, 67].

The second nonspecific force that may drive helix-helix interaction is the electrostatic interaction between the dipoles of adjacent α-helices. The dipole moment of an α-helix in a lipid bilayer has not yet been determined experimentally. However, a theoretical study by Ben-tal and Honig suggests that the helix dipole is substantial only for helices shorter than the bilayer thickness, since the electrostatic interaction is screened by the solvent for longer helices [68]. A similar solvent screening has been reported for helix bundles in soluble proteins [69]. However, the factors modulating the helix dipole are not yet well established and are further studied in this thesis.

Glycophorin A is a homo-oligomeric transmembrane protein present in human red blood cells. It is used as a model system to study helix-helix interactions due to its symmetry and relative simplicity. Its dimer structure has been elucidated by x-ray crystallography [42] and by NMR spectroscopy [70]. The interface is defined by the GxxxG motif described earlier. In particular, the seven-residue motif, LIxxGVxxGVxxT, was implicated in dimerisation of glycophorin A [59, 71, 72]. Although the glycophorin A dimer serves as a model for analogous systems, several questions on the energetics of association processes of the helices remain unanswered. Membrane helix dimers of glycophorin A and the erythropoietin receptor transmembrane segment are modelled and studied in the present thesis.

1.3 Protein-Lipid Interactions: Theoretical Approaches

A stable, folded protein, regardless of the pathway by which it enters the membrane, resides in a free energy minimum [8, 25]. This free-energy minimum defines its structure and stability. Thus, membrane-protein structure is determined by the interactions of the peptide chains with each other and with the lipid bilayer and water [52]. Peptide-lipid interactions play a crucial role in membrane protein architecture but are not well characterised [8, 50]. The importance of the energetics of the lipid-peptide interactions is seen in the strong tendency of the bilayer interface to increase the propensity of helix formation [8, 73].

In theoretical studies, the complex membrane environment has to be simplified and thus, the accuracy of the lipid-peptide interactions depends on the approximations included in the membrane model [33, 52, 74]. In particular, the treatment of long-range electrostatics in such an anisotropic heterogeneous environment remains a challenge [33, 52, 74]. A wide variety of representations of the bilayer have been considered in the past. Examples of membrane representations used in theoretical calculations are the all-atom representations [75–79], the three-slab continuum model [28, 29, 31, 32, 34, 35] and the dipole lattice model [74]. The best representation of a biological membrane in computational studies strongly depends on the specific questions to be addressed and the limitations imposed by available computational resources.

1.3.1 All Atom Representations

Detailed theoretical work on membrane proteins has been performed treating the bilayer atoms and surrounding solvent explicitly. Calculations with explicit lipid bilayer and water atoms using molecular dynamics simulations have been performed on several membrane systems. Lipid-peptide interactions have been studied extensively for single peptides, such as gramicidin S [76, 79], melittin [29], alamethicin [77, 78], a bacterial KcsA K^+ channel [75], and the *Influenza* M2 channel [80, 81]. The structure and orientation of the helices have also been studied using computer simulation. In certain cases, such as in melittin and gramicidin S, the structure and orientation have been reported to vary marginally [29, 76]. In contrast, simulations of alamethicin have revealed hinge-bending motion that may correspond to conformational switches involved in signalling across a bilayer [77, 78]. Computer simulation of the KcsA K^+ channel have also shown a selectivity for K^+ ions [75].

Simulations on more complex systems, such as glycophorin [82], bacteriorhodopsin [83, 84], OmpF porin [85] and aquaporin [86] demonstrated the feasibility of computational studies on large and complex membrane proteins. It has been shown that in glycophorin A, the

length of the lipid hydrocarbon chains modulate its structural properties, such as helix-helix crossing angle and helix tilt angle [82]. Further, the selectivity of aquaporin and movement of water molecules through the channel has been demonstrated [86]. Simulations have thus helped to elucidate mechanisms by which the lipid bilayer influences the protein free energy landscape [80].

Although these calculations are detailed and have resulted in a rich variety of biophysical information, the use of an all-atom membrane environment is computationally expensive and limits the configurational space that can be explored. Moreover, only a few lipids (*e.g.* DMPC and DPPC) have been well parametrised for the commonly-used force fields till date. Thus, the simulations do not reflect the general characteristics of the bilayer. Simulations of mixed bilayers remain challenging because of the long timescale ($>> 1ns$) of mixing of lipid components, but attempts in this directions are being made [80, 87].

1.3.2 Continuum Methods

The lipid and solvent molecules can be also represented implicitly using continuum models. Commonly used methods for continuum representations of membranes are the Poisson-Boltzmann and the Generalised Born methods [88]. The implicit solvent models approximate the average influence of water, the membrane or both on a solute. In the implicit models applied so far the membrane has been represented as a slab of low dielectric embedded in a high dielectric aqueous phase (Fig. 1.4) [28, 29, 31, 32, 34, 35].

The continuum studies capture certain general features of the membrane, such as the hydrophobic nature of the membrane core and a hydrophobic-hydrophilic dielectric boundary. However, it is assumed that all locations are described adequately as either bulk hydrocarbon or bulk water. The studies have successfully accounted for various experimental data, such as tilt angles. However, comparing Fig. 1.4 with Fig. 1.1, it is obvious that the lipid head-group region is not represented by such models. Thus, a drawback of representing the membrane as a single slab of low dielectric is that the head-group energetics are not adequately modelled. One immediate consequence of this inadequacy is that the experimentally-observed preference of certain residues, such as Trp and Lys for the head-group region is absent [74].

Poisson-Boltzmann Representation of the Membrane

The Poisson-Boltzmann equation represents one of the common approaches in continuum electrostatics to compute solvation energies. It combines two fundamental equations of physics, namely the Poisson equation and the Boltzmann equation. The Poisson equation

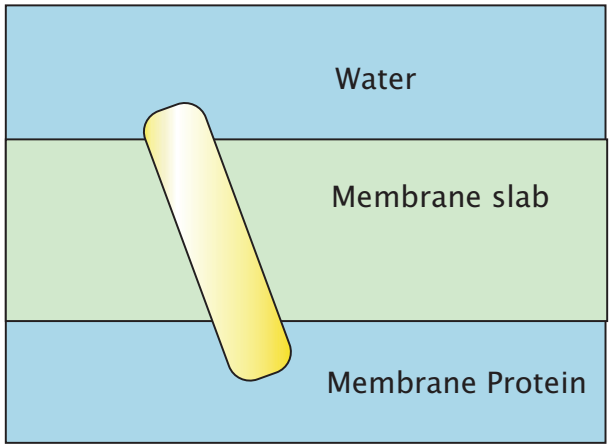

Figure 1.4: The membrane is often represented as a slab of low dielectric ($\epsilon = 2$) embedded in regions of high dielectric ($\epsilon = 80$). In such a model, the membrane headgroup properties are not modelled and the membrane-water interface is infinitely thin. The cylinder represents a membrane protein or α-helix within the bilayer.

describes the electrostatic potential that arises from a given fixed charge distribution and the Boltzmann relation describes the distribution of particles in response to any field. Together, the Poisson-Boltzmann equation relates the variation in the electrostatic potential, ϕ to the spatially dependent dielectric permittivity, $\epsilon(r)$ and the charge distribution, $\rho(r)$. Poisson-Boltzmann electrostatics is further discussed in Section 2.2.

The three-slab model in conjunction with the Poisson-Boltzmann equation to calculate the electrostatic energy has been applied to estimate the free energy of insertion of alamethicin [32] and oligoalanine [28]. The orientations of melittin [29], Src [34] and the FYVE membrane domains [30] have also been studied. The calculations have provided starting structures for detailed all-atom simulations [29, 80] and have provided insight into how the membrane modulates the structural and energetic properties of the membrane peptides. Electrostatic stabilisation of monovalent cations in ion channels has also been investigated [89]. Calculations based on the Poisson-Boltzmann equation are routinely used to determine the pKa of ionisable side chains in ion channels, such as bacteriorhodopsin [90] and the KcSA potassium channel [89]. The Poisson-Boltzmann equation has been extended to include membrane potential and to calculate the variations of the transmembrane voltage along the axis of KcsA [91]. Thus, the method has been useful for modelling how the lipid environment modulates the structural and energetic properties of membrane-bound peptides and proteins.

The main limitation of the Poisson-Boltzmann approach arises from the fact that the dielectric constant is a macroscopic concept describing intrinsically non-local electrostatic

interactions. For example, the electrostatic potential arising from the average solvent polarisation in the neighbourhood of a molecular solute may be considerably more complex than those expected from a continuum approximation [92]. Being aware of these limitations is necessary for the correct interpretation of these methods.

Generalised Born Methods

The Born equation relates the solvation energy, ΔG of a single spherical charge of magnitude q and radius a to the difference in energies for charging in two uniform media of dielectric constants ϵ_1 and ϵ_2, respectively [93]:

$$\Delta G = \frac{q^2}{8\pi a}\left(\frac{1}{\epsilon_1} - \frac{1}{\epsilon_2}\right) \tag{1.1}$$

The generalised Born method extends this theory by estimating a 'generalised Born radius', *i.e.*, the atomic radius necessary to produce the correct solvation energy (estimated experimentally or from Poisson-Boltzmann approaches) using the Born expression. These effective radii are then used to calculate the electrostatic solvation energy using the following semi-empirical expression [94]:

$$\Delta G_{solv} = -\frac{\epsilon_{water} - 1}{8\pi\epsilon_{water}} \sum_{i,j}^{N} \frac{q_i q_j}{\sqrt{r_{ij}^2 + a_i a_j \quad exp\left(\frac{-r_{ij}^2}{4a_i a_j}\right)}} \tag{1.2}$$

Adaptations of this method to represent the membrane by the three-slab model have been used to calculate the orientation of various single peptides, such as melittin [31], the influenza virus fusion peptide [35] and bacteriorhodopsin helical fragments [35] which were found to compare well with experimental data.

The main limitation of the Generalised Born method is the estimation of the generalised Born radius which is usually calculated from the Poisson equation [95]. Tilting of membrane helices will change the distance of an atom from the aqueous layer and may change the value of the Born radius of that atom. The definition of an unique born radius may also be difficult considering the anisotropic nature of the membrane.

1.3.3 Semi-empirical and Lattice Models

Solvent models that assume the solvation energy to be proportional to the exposed surface area have been applied to estimate solvation energies [33, 88, 96]. Membrane parameters have been calculated by modelling the membrane as a bulk hydrocarbon medium and has been used to study the membrane orientations of hemagglutinin [96] . In a similar study, the effective solvation free energy of glycophorin A and melittin was calculated in the membrane environment and correctly distinguished between the transmembrane and surface-aligned orientations of the two proteins, respectively [33]. However, there is no direct connection to the underlying electrostatics and the method usually introduces large errors in energy estimation compared to Poisson-Boltzmann or generalised Born calculations [88].

A semi-implicit model, in which the membrane-water system is modelled as a lattice of dipoles with varying dipole moments, has been able to reproduce the position of a 20-mer-WALP peptide and alamethicin in the membrane [74]. However, the model only qualitatively reproduces the lipid-protein interactions.

1.3.4 Choice of the Membrane Model

A variety of membrane models exist and the choice of an adequate membrane model is based on the requirement of the calculations and on the complexity of the system studied. Existing membrane models may also need to be modified for particular applications. For example, for a computationally-feasible modelling method, an all-atom representation is not tractable. On the other hand, care has to be taken not to oversimplify the energetics for the sake of computational efficiency. To understand the energetics of protein-protein and protein-lipid interactions within membranes, important structural features, such as the hydrophobic core as well as the polar head region have to be represented in the theoretical description. Polarity differences also need to be accounted for to accurately distinguish the energetics in bulk water and in the membrane. In this thesis, we present a five-slab membrane model which is an extension of the existing membrane models. The model accounts for the electrostatic properties of the polar head group region implicitly by introducing an additional dielectric layer.

1.4 Outline of the Thesis

This thesis aims at understanding the structural and electrostatic properties of α-helical membrane proteins. The following aspects are discussed:

- Design of a suitable membrane model to describe protein-lipid interactions
- Understanding the energetics of helical peptide orientation in membranes
- Calculating the macromolecular electrostatic properties of α-helices in membranes
- Modelling the association of transmembrane α-helices

In this thesis, the energy landscape of biomolecules is described by a Molecular Mechanics/ Poisson-Boltzmann approach as outlined in Chapter 2. The membrane environment is described here by a five-slab continuum dielectric model that distinguishes between the solvent and the membrane head group and core regions. In Chapter 3, the membrane model is introduced and the calculations performed to validate the model are reported.

In Chapter 4 , the five-slab membrane model is used to examine the factors determining the position and orientation of single membrane helices. The energetics of the position and orientation of a dipole in a membrane is calculated. Extending the analysis to N-methylacetamide and polypeptides shows that dipole energetics orient the axes of single helices perpendicular to the plane of the membrane. Furthermore, the length dependence of helix tilt angles and fluctuations in the tilt angles are examined. Cavity formation in the aqueous solvent is found to play an important role in determining tilt angles of helices longer than the membrane width. Experimentally-determined tilt angles of several membrane--spanning helices including glycophorin A, melittin and the WALP peptides are reproduced by the five-slab model.

The effect of the helix environment in screening and modulating the helix dipole of membrane α helices is reported in Chapter 5. The effective dipole moment is calculated from the electrostatic potential generated by the helix, which is represented at atomic detail, in the given environment which is represented implicitly by a continuum model. The dependence of effective helix dipole on the length of the α-helix is calculated in vacuum, aqueous solution and lipid bilayers. The analysis is extended to helices in protein interiors for membrane proteins as well as soluble proteins. The effective dipole moment is found to vary strongly with the orientation and position of the helix relative to the aqueous medium. For instance, the effective dipole moment of transmembrane helices decreases with peptide length. The results are at first glance surprising, but can be rationalised in terms of the

shielding of the helix termini. A set of simple rules is established for estimating helix dipole from experimental structures.

Association of helices in the membrane is modelled in Chapter 6. The structure of the glycophorin A dimer is modelled and compared with experimentally-determined structures to validate the methodology. The contribution of each residue in the dimer interface is calculated and reproduces the experimental results. In the second part of Chapter 6, the structures of the wild-type erythropoietin receptor transmembrane dimer and its mutants are modelled. The erythropoietin receptor has been experimentally found to locate in lipid rafts, whereas, its T242N mutant is impaired in lipid raft localisation. A substantial increase in inter-helical distance and volume is observed for the T242N mutant compared to the wild-type erythropoietin receptor dimer. The results suggest that absence of close packing may lead to exclusion of the T242N mutant from lipid rafts. However, the modelling does not rule out that additional motifs or a combination of motifs may be required for sorting into lipid rafts.

This thesis provides insights into the structure and energetics of α-helical membrane proteins. Structural aspects of single helices, such as orientation and their association to form larger complexes have been studied. Further, the relationship between structure and function has been briefly addressed for the erythropoietin receptor. This thesis is thus a step towards a better understanding of the energetics of the structural organisation of membrane proteins.

Chapter 2

The Energy Landscape

"They'd believe it; they'd believe it. Things have to make sense to be disbelieved.", Life of Galileo, Bertold Brecht.

The energy landscape of a protein is described by its potential energy in every degree of freedom. The structure of a protein is determined by this energy landscape such that the native folded protein is at a minimum. In this thesis, the potential energy is described by a Molecular Mechanics/ Poisson-Boltzmann approach. These two methods are discussed below.

2.1 Molecular Mechanics

The energy of large biomolecules is usually calculated by empirical potential energy functions *i.e.*, by molecular mechanics force fields [97]. Molecular mechanics is used as an alternative to quantum chemistry, which remains unfeasible for multi-atomic systems, such as proteins. The empirical force fields were developed to describe molecular structures and properties in an accurate yet computationally efficient manner. Certainly, the approximations introduced in this method lead to limitations compared to quantum chemical calculations. For instance, no bond making or breaking events can be modelled *via* molecular mechanics since electrons are not considered explicitly, but are assumed to find their equilibrium distribution once the positions of the nuclei are known. The above assumption is based on the Born-Oppenheimer approximation of the Schrödinger equation. The Born-Oppenheimer approximation states that nuclei timescales are much slower than electronic motions. Thus, nuclear motions, vibrations and rotations can be studied independently from electronic fluctuations since the electrons are assumed to adjust *instantly* to any

movement of the nuclei.

In molecular mechanics, the nuclei and electrons are treated together as spherical atom-like particles possessing a net point charge [97]. The radii and the net charges are measured experimentally or obtained from high-level quantum calculations. The interactions between atoms are based on harmonic approximations or classical potentials and determine the spatial distribution of atoms and their corresponding energies. The potential energy function used to calculate the energy and geometry of a molecule is called the force field. Further, atom types have to be defined to describe the atoms in a molecule. Parameters for bond lengths, bond angles, *etc.*, are also required.

Current generation force fields provide a reasonably good compromise between accuracy and computational efficiency. Their ability to reproduce experimentally-measured physical properties has been extensively tested. These properties include structural data obtained from x-ray crystallography and NMR, dynamic data obtained from spectroscopy and inelastic neutron scattering as well as thermodynamic data. Among the most commonly used potential energy functions are the AMBER [98], CHARMM [99] and GROMOS [100] force fields.

The energies of the biomolecules studied in this thesis were calculated using the CHARMM force field. In the CHARMM force field, the potential energy is calculated as a sum of bonded and nonbonded terms energy terms. The bonded terms, E_{bonded} describe the bonds, angles and bond rotations in a molecule as explained in Fig. 2.1. The non-bonded terms, $E_{nonbonded}$ account for interactions between nonbonded atoms *i.e.*, atoms separated by 3 or more covalent bonds. The total energy, E_{total} is given by:

$$\begin{aligned} E_{total} &= E_{bonded} + E_{nonbonded} \\ &= \underbrace{E_{bonds} + E_{angles} + E_{impr} + E_{UB} + E_{dihedrals}}_{E_{bonded}} + \underbrace{E_{vdW} + E_{elec}}_{E_{nonbonded}} \end{aligned} \tag{2.1}$$

where, E_{bonds} is the bond stretching energy term, E_{angles} is the angle bending energy term, $E_{dihedrals}$ accounts for rotation along a bond, E_{impr} is the distortion energy term, E_{UB} is the Urey-Bradley term, E_{vdW} the van der Waals energy and E_{elec} the electrostatic energy.

E_{bonds} is given by a harmonic potential representing the interaction between atom pairs separated by one covalent bond (*i.e.* 1,2-pairs). The term accounts for the energy of a bond of bond length, b as a function of displacement from the ideal bond length, b_0. The force constant, K_b, determines the strength of the bond. Both, the ideal bond length, b_0 and the force constant, K_b depend on the chemical type of the atom-constituents and are

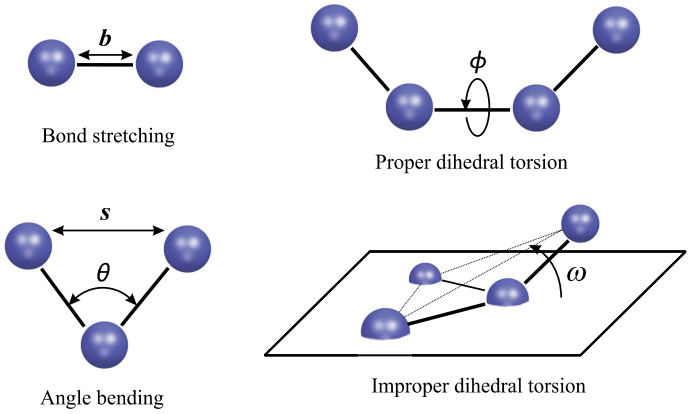

Figure 2.1: Schematic representation of the bonded interaction terms contributing to the force field: bond stretching, angle bending, proper and improper dihedrals.

specific for each pair of bound atoms.

$$E_{bonds} = \sum_{bonds} K_b \left(b - b_0\right)^2 \tag{2.2}$$

E_{angles} relates the alteration of bond angles, θ from ideal values, θ_0 and is also represented by a harmonic potential.

$$E_{angles} = \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 \tag{2.3}$$

The Urey-Bradley term, E_{UB} accounts for interactions between the atom pairs involved in 1,3-configurations *i.e.* atoms bound to a common atom

$$E_{UB} = \sum K_{UB}(s - s_{eq})^2 \tag{2.4}$$

where, K_{UB} is the Urey-Bradley force constant and s the distance between the two atoms in consideration.

$E_{dihedrals}$ represents the torsion angle potential function which models the presence of torsional barriers between atoms separated by 3 covalent bonds (*i.e.* 1,4-pairs). The term describes the rotation by a dihedral angle, ϕ around a given bond with force constant, K_ϕ,

phase, γ and multiplicity, n.

$$E_{dihedrals} = \sum_{1,4-pairs} K_\phi \left(1 + cos\left(n\phi - \gamma\right)\right) \tag{2.5}$$

K_b, K_{UB}, K_{UB}, K_θ and K_ϕ, are usually evaluated from studies of small model compounds and comparing the geometry and vibrational spectra in the gas phase (IR and Raman spectroscopy), supplemented, if necessary, with ab initio quantum calculations.

E_{impr} is an energy term used to maintain chirality and planarity around the torsional angle, ω.

$$E_{impr} = \sum K_\omega(\omega - \omega_{eq})^2 \tag{2.6}$$

$E_{nonbonded}$ has two components, the van der Waals interaction energy, E_{vdW} and the electrostatic interaction energy, E_{elec}. In the CHARMM force field, the hydrogen bond interactions are accounted for by the electrostatic and van der Waals interactions.

E_{vdW} is commonly described by a Lennard-Jones 12-6 potential. The Lennard-Jones potential arises from a balance of two terms, a short range repulsive and a slower decaying attractive force. The attractive $1/r^6$-term arises from spontaneous dipoles inducing opposing dipoles in nearby atoms. The repulsive force is predominant at short distances where the electron-electron repulsion is strong. The Lennard-Jones potential is determined by two parameters: the collision parameter, σ and the depth of the potential, A. The collision parameter is the distance between two atoms at which the Van-der-Waals energy is zero.

$$E_{vdW} = \sum_{i,k} A \left(\frac{\sigma_{ik}}{r_{ik}^{12}} - \frac{\sigma_{ik}}{r_{ik}^{6}}\right) \tag{2.7}$$

E_{elec}, the electrostatic interaction between a pair of atoms is represented by the Coulomb potential. The Coulomb potential is inversely proportional to ϵ, the effective dielectric function for the medium and r_{ik}, the distance between two atoms with charges q_i and q_k.

$$E_{elec} = \sum_{i,k} \frac{q_i q_k}{\epsilon r_{ik}} \tag{2.8}$$

The Coulomb potential is usually adequate to describe the interaction between two charged species in a homogeneous medium such as in vacuum. However, the solvent surrounding a biomolecule, usually water, has a fundamental influence on the energetics of the biomolecule. One of the most important effects of the solvent is the screening of electrostatic interactions. A simple approximation to include the screening effect into calculations is to assign a dielectric constant of 80 for the solvent in the electrostatic term of the force field. However, this is a crude approximation and does not compare well with experimental data since the polarisation in the protein environment is not represented. To obtain a better, physically-relevant description of a biomolecule in a solvent methods, such as the Poisson-Boltzmann model have been successfully applied in conjunction with the CHARMM forcefield.

2.2 Poisson-Boltzmann Electrostatics

Electrostatics plays a key role in biological processes. The methods that have been used to simulate electrostatics in biological systems may be broadly classified into those which explicitly simulate all molecules of the system, including solvent and dissolved salts, which are by far the more demanding, and those which simulate the solvent and salts by a continuum model. Among the latter, the Poisson-Boltzmann equation has been widely and successfully used [101, 102].

The Poisson-Boltzmann equation is derived from the Poisson equation which describes the electrostatic potential, $\phi(\mathbf{r})$ in vacuum at point, $\mathbf{r}$ from the charge density, $\rho(\mathbf{r})$

$$\Delta\phi(\mathbf{r}) = -4\pi\frac{\rho(\mathbf{r})}{\epsilon_0} \tag{2.9}$$

where, ϵ_0 is the dielectric constant in vacuum and equals 1. The Laplace operator, Δ is the sum of the second partial derivatives with respect to the spatial coordinates or equivalently the divergence of the gradient.

The electric field vector, $\mathbf{E}$ is given by the negative gradient of $\phi(r)$.

$$\mathbf{E} = \nabla\phi(\mathbf{r}) \tag{2.10}$$

The Nabla operator, ∇ is the gradient operator with respect to spatial coordinates and its square is the Laplace Operator, $\Delta = \nabla^2 = \nabla \cdot \nabla$.

The electrostatic interactions between charges in an uniform medium are usually weakened compared to those between the same charges in vacuum. An external electrostatic field will induce a dipole moment in atoms or molecules placed in it. If the molecules already have a

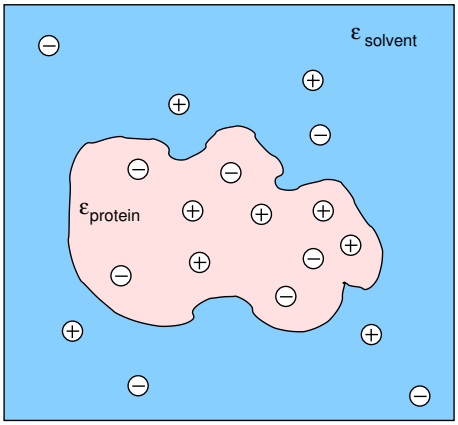

Figure 2.2: A protein of dielectric, $\epsilon_{protein}$ with fixed partial charges in a solvent of permittivity, $\epsilon_{solvent}$ with mobile charges (ions).

net dipole moment, these permanent dipoles will be aligned by the field. The electric field generated by the induced and the aligned dipoles is directed opposite to the inducing field. As a consequence, the overall field is weakened. The weakening of the external field can be described implicitly by reducing the electrostatic potentials everywhere by a constant factor known as the relative permittivity or dielectric constant, ϵ_r of the medium. On substituting the Nabla operator and the relative permittivity, the Poisson equation takes the following form:

$$\nabla \cdot \nabla \phi(\mathbf{r}) = -4\pi \frac{\rho(\mathbf{r})}{\epsilon_r} \tag{2.11}$$

Obviously, the potentials that result from this equation are reduced by $\frac{1}{\epsilon_r}$ relative to their values in vacuum for the same charge distribution.

A protein in aqueous solution is not well described by an uniform permittivity, because the dielectric properties of the protein differ significantly from those of water. By defining a relative permittivity, $\epsilon(r)$ which depends on r, the Poisson equation becomes:

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] = -4\pi \rho(\mathbf{r}) \tag{2.12}$$

In this approach, the macromolecule is treated as a low dielectric region with embedded atomic partial charges. The dielectric constant of the protein, $\epsilon_{protein}$ is typically set between 2 and 4 to take into account electronic polarisation and the limited flexibility of the macromolecule. The effects of the solvent molecules, whose motions are much faster than those of the solute molecule are taken into account on average through a continuum of high dielectric constant. The high dielectric constant of the solvent, $\epsilon_{solvent}$ usually set to 80

also accounts for the much weaker electrostatic interactions in water compared to a protein interior. Mobile ions in the solvent are also accounted for in this approach by a Boltzmann distribution. Fig. 2.2 illustrates such a system with a protein of dielectric constant, $\epsilon_{protein}$ and containing embedded charges in a solvent of permittivity, $\epsilon_{solvent}$.

The charge density, $\rho(r)$ of the above system can be thought to consist of two parts, the fixed charges of the protein, $\rho_{protein}(r)$ and the mobile charges of the solvent, $\rho_{ions}(r)$.

$$-\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi\,(\mathbf{r})] = 4\pi\left(\rho_{protein}(\mathbf{r}) + \rho_{ions}(\mathbf{r})\right) \tag{2.13}$$

The ion distribution in the solvent can be described by a Boltzmann distribution:

$$\rho_{ions}(\mathbf{r}) = \sum_{i=1}^{N} c_{(i,bulk)} q_i exp\left(-\frac{q_i\phi(\mathbf{r})}{RT}\right) \tag{2.14}$$

where, N is the number of different ion types, $c_{(i,bulk)}$ the concentration of ion type i in the bulk and q_i its charge. R is the gas constant and T the temperature.

The Poisson-Boltzmann equation is obtained from Eqs. 2.13 and 2.14:

$$-\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi\,(\mathbf{r})] = 4\pi\left(\rho_{prot}(\mathbf{r}) + \sum_{i=1}^{N} c_{(i,bulk)} q_i \ \exp\left(-\frac{q_i\phi(\mathbf{r})}{RT}\right)\right) \tag{2.15}$$

The Poisson-Boltzmann equation can be linearised if the electrostatic potential is small, *i.e.*, ($\frac{\phi(\mathbf{r})}{RT} << 1$). The exponential is expanded up to the linear term which when multiplied by the pre-factor gives the quadratic term:

$$-\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi\,(\mathbf{r})] = 4\pi\left(\rho_{prot}(\mathbf{r}) + \sum_{i=1}^{N}\left(c_{(i,bulk)} q_i - c_{(i,bulk)} q_i^2 \frac{\phi(\mathbf{r})}{RT}\right)\right) \tag{2.16}$$

The first term in the expansion of the Poisson-Boltzmann equation $\left(\sum_{i=1}^{N} c_{(i,bulk)} q_i\right)$ equals zero since the total charge of the mobile ions in solution is zero.

For small electrostatic potentials as they are generally found in proteins, the linear Poisson-Boltzmann equation is a good approximation.

The ionic strength, I of a medium is defined as:

$$I = \frac{1}{2}\sum_{i=1}^{N} c_{(i,bulk)} q_i^2 \tag{2.17}$$

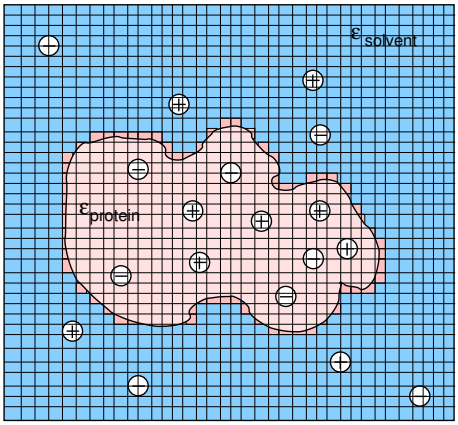

Figure 2.3: In the finite-difference method, a grid is superposed on the protein-solvent system to solve the Poisson-Boltzmann equation numerically. Electrostatic potentials and charges are defined at the grid points, the electrostatic potential at any off-grid location can then be obtained by interpolation from the nearest grid points.

The Poisson-Boltzmann equation can be extended to model a membrane environment by including additional dielectric slabs. Hence, instead of two dielectric regions, a low dielectric protein region and a high dielectric water region, additional slabs corresponding to the membrane are included in the calculations. In previous calculations, a single slab of low dielectric was used to model the membrane core. In this thesis, a new five-slab membrane model is applied and tested for its accuracy and validity, the details of which are given in Chapter 3.

Analytical solutions of the Poisson-Boltzmann equation exist only for simple, symmetric systems such as spherical bodies. Hence, for complex macromolecules with irregular shapes, Eq. 2.16 is usually solved numerically by finite difference methods in which the molecular charges and dielectric constants are discretised on a grid (Fig. 2.3). The discretisation procedure has some disadvantages. For instance, the free energy of the system is largely dependent on the relative position of charges on the grid and on the dimension of the mesh. The grid-dependent self energy of charges must be accounted for while computing free energies [103]. Further, the mesh must be fine enough not to merge opposite charges (dipoles) at the same node. fine grid is also required to ensure a proper representation of the macromolecule-solvent interface and thus, to prevent artefacts in the surface potential representation.

Charges are assigned to each grid point and the corresponding electrostatic potentials are then defined at those grid points. The electrostatic potential at any off-grid location can be obtained by interpolation from the eight nearest grid points. A charge q, a distance a, b and c from the nearest vertex points (shown in Fig. 2.4) is assigned to the grid with grid

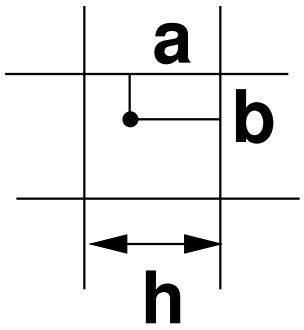

Figure 2.4: Assignation of charge, q that is a distance a and b from the nearest vertex points to the grid with grid spacing, h. The distance c is not shown but represents the distance in the z-direction.

spacing h in the following manner:

$$q_{\text{grid}} = q\left(1-\frac{a}{h}\right)\left(1-\frac{b}{h}\right)\left(1-\frac{c}{h}\right) \tag{2.18}$$

Of course, the split charges interact with each other and result in high interactions energies and are termed "grid artefacts". Taking an energy difference in two media of varying dielectric constants, such as in vacuum and water, cancels out the grid artefact.

Eq. 2.16 may now be solved by integrating over the grid.

$$\int \nabla[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})]dr + 4\pi\int \rho_{prot}(\mathbf{r})dr - \int\sum_{i=1}^{N} c_{(i,bulk)}q_i^2\frac{\phi(\mathbf{r})}{RT}dr = 0 \tag{2.19}$$

By integrating over the entire volume we obtain,

$$\int[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})]dA + 4\pi q_0 - h^3\sum_{i=1}^{N} c_{(i,bulk)}q_i^2\frac{\phi(\mathbf{r})}{RT} = 0 \tag{2.20}$$

where, ϕ_0 is the electrostatic potential at the central point **A**.

However, the surface integral can be defined as

$$\int[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})]dA = \sum_{i=1}^{6} h\epsilon(i)\left(\phi_i - \phi_0\right) \tag{2.21}$$

where, ϕ_i is the electrostatic potential at the nearest grid points i to the point **A**.

Thus, from Eq. 2.20 and 2.21 we obtain:

$$\sum_{i=1}^{6} h\epsilon(i)\left(\phi_i - \phi_0\right) + 4\pi q_0 - h^3 \sum_{i=1}^{N} c_{(i,bulk)} q_i^2 \frac{\phi(\mathbf{r})}{RT} = 0 \tag{2.22}$$

If we solve Eq. 2.22 for ϕ_0 we get:

$$\phi_0 = \frac{\left(\sum_{i=1}^{6} h\epsilon(i)\phi_i\right) + 4\pi q_0}{\left(\sum_{i=1}^{6} h\epsilon(i)\right) + h^3 \sum_{i=1}^{N} c_{(i,bulk)} q_i^2 \frac{1}{RT}} \tag{2.23}$$

ϕ_0 is solved in an iterative cycle until self consistent. The main problem problem of this method is that the charge cannot be extrapolated at the boundaries since there are less than the adequate number of neighbours. Techniques such as 'focussing' are used to overcome this problem. At first, a low-resolution large grid is used and the electrostatic potential at the boundary is calculated by the Born model or the Debye-Hückel theory. In the second step, the electrostatic potential from the larger grid is used to initialise the potential at the boundaries of a smaller high-resolution grid. Such focussing steps are repeated to obtain the potential for a very high-resolution grid.

In this thesis, the Poisson-Boltzmann routine [104] in CHARMM was used to calculate the electrostatic potentials. The parameters used in the calculations, such as the ionic strength and the temperature, are given in each chapter along with the results.

Chapter 3

The Membrane Model

"We've been looking for someone to test them on all summer...", Harry Potter and the Goblet of Fire, J K Rowling.

The structure and stability of membrane proteins are determined by interactions of the membrane-bound peptides with each other, the lipid bilayer and solvent molecules [8, 52]. Peptide-lipid interactions, though least characterised, play a dominant role in membrane protein architecture [8, 50, 52]. Thus, to study membrane proteins and their energetics, one needs to account for the membrane-protein interactions by a suitable membrane model. Firstly, the model should allow accurate calculation of interaction energies of the membrane and molecules embedded in it. Furthermore, a membrane model should capture essential features of membrane structure, such as heterogeneity and the changing polarity along the bilayer normal. Finally, the model must be computationally tractable, *i.e.*, the computational time required should be reasonable depending on the questions asked. For example, molecular modelling and exploring large conformational space demands a computationally inexpensive model, while detailed analysis of the dynamics and fluctuations in the membrane protein require an exact all-atom representation of the membrane. In this chapter, the membrane model used in this thesis is introduced and the basic tests performed are discussed. The model accounts for the membrane implicitly and distinguishes between the membrane head-group region, membrane core and aqueous regions. The orientation, position and energetics of ions and peptide side-chain analogues are chosen to access the range of validity of the proposed membrane representation.

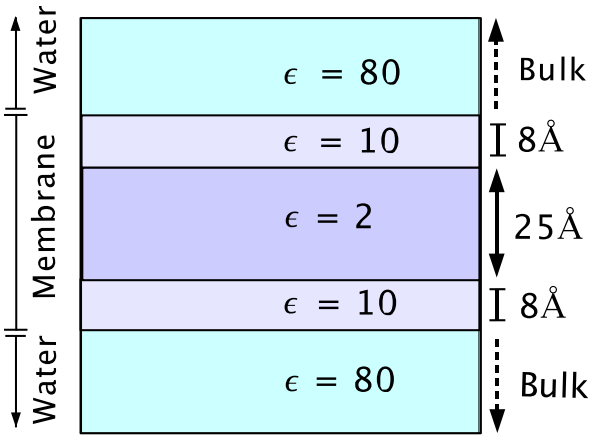

Figure 3.1: Five-slab continuum electrostatic model of a biological membrane environment. The membrane is represented as three slabs corresponding to the two head-group regions and the core region. The head-group region is modelled as a 8 Å slab on either side of the low dielectric core region. The two outer slabs correspond to bulk water. The dielectric constants assigned to the regions are shown in the figure.

3.1 The Five-Slab Membrane Model

In this thesis, the membrane is accounted for implicitly by a continuum dielectric approach. The membrane is represented by a five-slab model as shown in Fig. 3.1. The unique feature of the model is the distinction between the different dielectric regions: solvent, head-group and core regions. The outermost slabs correspond to the bulk aqueous phase, and the innermost slab corresponds to the membrane core. The two slabs in between correspond to the two lipid head-group regions of the membrane.

The outermost slabs of the model, *i.e.*, those representing the aqueous phase are assigned a dielectric constant of 80, accounting for the high polarity of bulk water.

The innermost slab represents the membrane core which is comprised of the non-polar long-chain fatty acids [105, 106]. Correspondingly, the relative dielectric constant of the innermost slab is set to 2, representing the non-polar nature of the fatty-acid tails [105, 107]. A typical width of the core region is 25 Å (*c.f.* Fig. 3.1). Variation of this model parameter, however, is straightforward and calculations resembling bilayer perturbations are presented in this thesis.

Inclusion of the head-group region of biological membranes in a continuum model is ambiguous since the head-group region is heterogeneous with varying fractions of phosphatidylethanolamines, phosphatidylserines, phosphatidylcholines, sphingomyelin, glycolipids and cholesterol. Estimates of the dielectric constant of the head-group region have yielded

varying results depending in part on the lipid moieties [105–110]. The polarity of the head-group region in phosphatidylcholine and phosphatidylserine bilayers has been found to be intermediary to that of bulk water and the bilayer core [108, 110]. In the present case, a dielectric constant of 10 was used to represent the permittivity of the head-group region. The slab width was set to 8 Å. A value of 10 for the dielectric constant for the head-group slab has been found to accurately reproduce experimental studies of adsorption of small molecules in the head-group region [109]. Variation of this parameter within reason did not affect the results presented.

3.2 Calculation of the Solvation Energy

The five-slab membrane model is used to investigate the orientation and energetics of various biologically relevant molecules such as membrane-embedded ions and polypeptides. The most-probable orientation of a membrane-bound molecule is determined by the minimum of the solvation free energy of the molecule in the membrane environment. The solvation free energy of the molecule, ΔG_{solv} in a membrane system is the free energy for transferring the molecule from vacuum into a given position and orientation in the membrane [97]. The solvation energy is calculated as described below, using an all-atom model for the molecule considered together with the five-slab continuum dielectric membrane model. In Chapter 4, the orientations of helical peptides in a membrane are determined by calculating the solvation free energies of the peptides in a similar fashion.

ΔG_{solv} was calculated according to the thermodynamic cycle shown in Fig. 3.2. ΔG_{solv} is considered to be the sum of two contributions:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{np} \tag{3.1}$$

where, ΔG_{elec} is the electrostatic component of the solvation free energy and ΔG_{np} the non-polar component of the solvation free energy.

ΔG_{elec} is the electrostatic energy required to transfer the peptide from vacuum to a given position in the membrane. It is given by the difference of the energy required to charge the peptide in the membrane, ΔG_{elec}^{mem} and in vacuum, ΔG_{elec}^{vac}:

$$\Delta G_{elec} = \Delta G_{elec}^{mem} - \Delta G_{elec}^{vac} \tag{3.2}$$

ΔG_{elec} was calculated using the linearised Poisson-Boltzmann equation (Eq. 2.16), relating the variation in the electrostatic potential, ϕ to the spatially dependent dielectric permit-

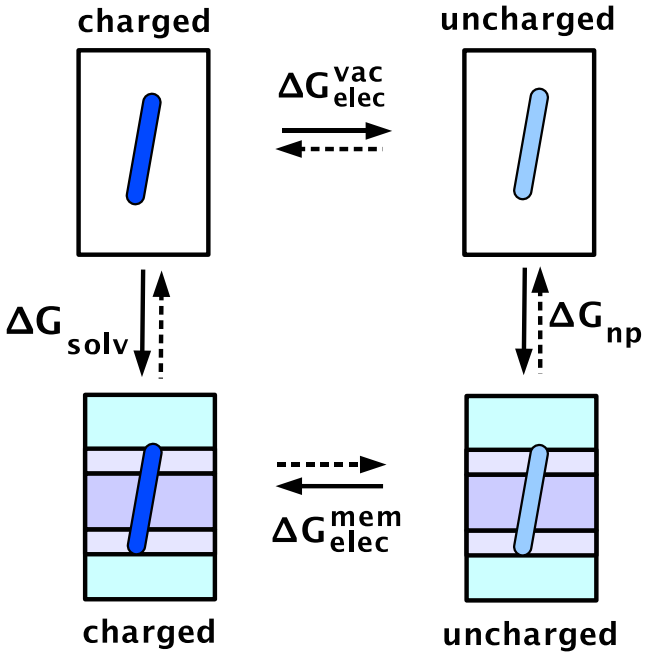

Figure 3.2: Four-step thermodynamic cycle to calculate ΔG_{solv} for a molecule (depicted as a cylinder) in a membrane. ΔG_{elec}^{mem} and ΔG_{elec}^{vac} correspond to the energy of charging the peptide in the membrane and in vacuum, respectively. The non-polar contribution, ΔG_{np} is the cost of cavity formation in the surrounding solvent.

tivity, $\epsilon(\mathbf{r})$ and the charge distribution, $\rho(\mathbf{r})$. For complex systems, the linearised Poisson-Boltzmann equation is solved numerically. The details of the calculations are given in the next section.

ΔG_{np} is the cost of cavity formation in the aqueous phase and is estimated to be linearly proportional to the water-accessible surface area of the molecules [111]. ΔG_{np} is estimated as [97]:

$$\Delta G_{np} = \gamma A + b \tag{3.3}$$

where, A is the solvent-accessible surface area of the peptide in the aqueous phase and γ and b are constants. The proportionality constant, γ is commonly referred to as 'surface tension coefficient' and represents the contribution to the solvation free energy per unit surface area [32, 112]. The constant, b is the free energy of hydration for a point solute ($A = 0$).

In this thesis, the values used are: γ = 0.0278 kcal/mol Å^2 and b = -1.71 kcal/mol. These values of γ and b were derived from the partitioning of alkanes between liquid alkane and

water [112]. Thus, the non-polar interactions between the lipid and the solute molecule are considered implicitly. The values of γ and b have been used earlier for calculations on membrane proteins and compare well with experimental results [28, 32, 34].

3.2.1 Methods

Model construction

The monovalent ions, Na^+ and Cl^- were modelled as spherical charges of unit charge and radii 1.02Å and 1.81Å, respectively. The ions were positioned at various depths of membrane insertion, v and ΔG_{solv} was calculated.

All-atom models of indole, butane and imidazolium molecules were constructed from the CHARMM parameter set [99]. The molecules were positioned along the membrane normal and ΔG_{solv} was calculated. The depth of membrane insertion, v was calculated as the distance of the centre of mass of the molecule from the centre of the membrane. Indole was oriented in the membrane with the ring perpendicular to the bilayer normal. Imidazolium was oriented with the ring along the bilayer normal and the positively-charged Nitrogen atom pointing outward towards the aqueous phase.

Calculation of ΔG_{solv}

ΔG_{solv} was calculated as a function of v for the above molecules according to the thermodynamic cycle shown in Fig. 3.2. The two components, ΔG_{elec} and ΔG_{np} were calculated as described in the following sections.

Calculation of ΔG_{elec}

ΔG_{elec} was obtained numerically by solving the linear Poisson-Boltzmann equation using a finite difference method implemented in the CHARMM software package (version 29a2). The ionic strength was set to 0.15 M. The electrostatic potentials were calculated using the focussing technique [104] on a grid with grid spacing 1.0 Å, 0.5 Å and 0.25 Å. In all calculations, the distance between the grid boundaries and the molecular surface was set to at least 15 Å. The atomic radii of the atom types were taken as the calculated Born radii [113]. A cardinal b-spline was used for distributing charges over the grid points. The temperature was set to 300 K and no membrane potential was applied. The electrostatic potentials in the membrane environment and in vacuum ($\epsilon = 1$) were calculated. The corresponding solvation free energies of the system in the membrane, ΔG_{elec}^{mem} and in vacuum,

ΔG_{elec}^{vac} were computed from the above potentials.

Calculation of ΔG_{np}

ΔG_{np} was calculated from Eq. 3.3. The solvent-accessible surface area, A was calculated using the Lee and Richards algorithm [114] implemented in the CHARMM software package. A probe sphere of radius 1.4 Å was used in all calculations.

3.3 Tests of the Membrane Model

Similar to every new model, simple tests were performed to validate the five-slab membrane model. The tests were performed by calculating the solvation energy of a few systems such as ions and side-chain analogues for which experimental data was available.

3.3.1 Ions

The simplest test of an electrostatic model is its treatment of a spherical charge. Thus, as a first test of the membrane model the solvation free energy of ions, such as Na^+ and Cl^- is calculated at various positions in the membrane. The ΔG_{elec} values are compared to the energies derived from the Born equation (Eq. 1.1) for homogeneous media. Fig. 3.3 shows ΔG_{elec} for spherical charges corresponding to Na^+ and Cl^- ions, at various depths of membrane insertion, v. The calculations were performed in the absence of mobile solvent ions to directly relate the energies to the Born energies.

Ion	Position	Born Energy[kcal/mol]	ΔG_{elec}[kcal/mol]
Na^+	Membrane Core	-81.5	-86.4
Na^+	Water	-161.0	-167.7
Cl^-	Membrane Core	-45.8	-51.0
Cl^-	Water	-90.5	-92.2

Table 3.1: The electrostatic solvation free energy for monovalent ions in the membrane model. The membrane core corresponds to a region of dielectric constant ϵ=2 and the water layer corresponds to ϵ=80. The corresponding Born energies in these homogeneous media, i.e., the transfer from vacuum to the given medium is also shown.

Both ions have a lower solvation energy in the aqueous medium compared to the membrane core. The head-group region also stabilises the ions compared to the membrane core. The

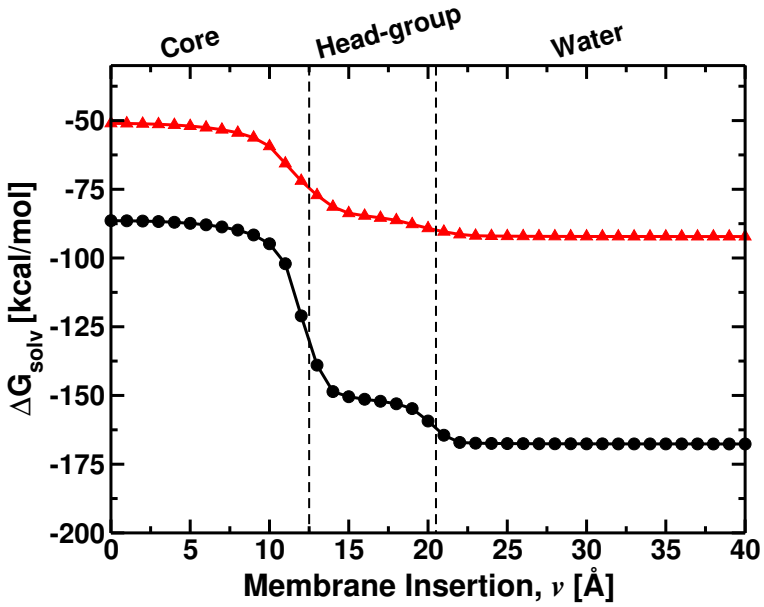

Figure 3.3: ΔG_{solv} *vs.* membrane insertion for sodium (•) and chloride (▲) ions in the membrane. Both ions have a lower solvation energy in the aqueous media compared to the membrane core. The head-group region also stabilises the ions compared to the membrane core. The core, head-group and aqueous regions of the membrane environment are marked.

effect of the dielectric boundary is seen even before the ion enters the lower dielectric region. Further, due to its larger radius, the chloride ion has a lower energy compared to the sodium ion, since its negative charge is delocalised over a larger volume.

ΔG_{elec} is reported for the two ions in the membrane core and in bulk water. The energies of the ions in the two media (*i.e.*, the membrane core and water) compare well to those calculated by the Born equation for a homogeneous media of $\epsilon = 2$ and $\epsilon = 80$, respectively (Table 3.1).

3.3.2 Side-chain Analogues

The next step in validating the model was to calculate solvation energy profiles for various side-chain analogues. Three molecules, butane, imidazolium and indole were chosen to investigate the relative importance of the free energy terms for different solutes. The three molecules are the side-chain analogues of Leu, His (at low pH) and Trp, respectively.

Fig. 3.4 shows the ΔG_{solv} *vs.* v for butane, imidazolium and indole. For butane, ΔG_{solv} is roughly constant inside the membrane and increases in the aqueous region. ΔG_{np} increases as the molecule enters the aqueous phase and the total increase in ΔG_{solv} in the aqueous

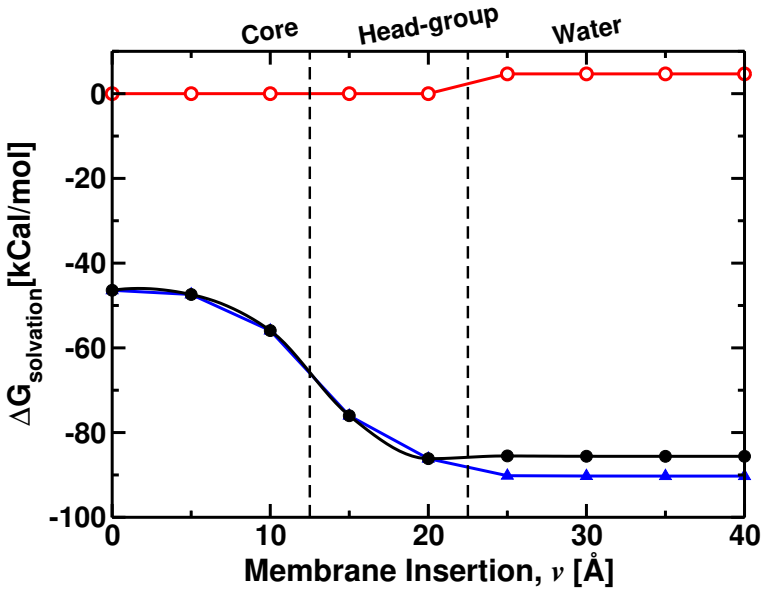

Figure 3.4: ΔG_{solv} *vs.* depth of membrane insertion, v for indole (•), imidazolium (▲) and butane (◦) ions in the membrane. The most probable position for butane is the membrane core and that for imidazolium is the aqueous phase. For indole, ΔG_{solv} is the minimum in the head-group region as a result of compensation between ΔG_{elec} and ΔG_{np}.

phase is attributed to this increase in ΔG_{np}. The partitioning of butane in the membrane core is as expected, since butane is non-polar and favours the non-polar membrane interior.

In contrast, for imidazolium, ΔG_{solv} is lowest in the aqueous region (Fig. 3.4, blue). Imidazolium, a charged ion, is known to prefer the aqueous medium. Similar to butane, ΔG_{np} of imidazolium increases as the molecule enters the aqueous phase. However, for a charged species such as imidazolium, the decrease in ΔG_{elec} in the aqueous phase is much larger than the increase in ΔG_{np}. Thus, increase in ΔG_{np} does not balance the larger decrease in ΔG_{elec} and consequently ΔG_{solv} decreases in the aqueous phase.

ΔG_{solv} *vs.* v for indole is shown in Fig. 3.4 (black circles). ΔG_{solv} obtains a minimum in the head-group region as a result of compensation between ΔG_{elec} and ΔG_{np} (see Fig. 3.5). The compensation between ΔG_{elec} and ΔG_{np} decides the final position of the molecule, since ΔG_{elec} favours polar environments while ΔG_{np} opposes entering the aqueous layer. The result agrees with recent NMR measurements showing that indole partitions into the head-group region of lipid bilayers [51].

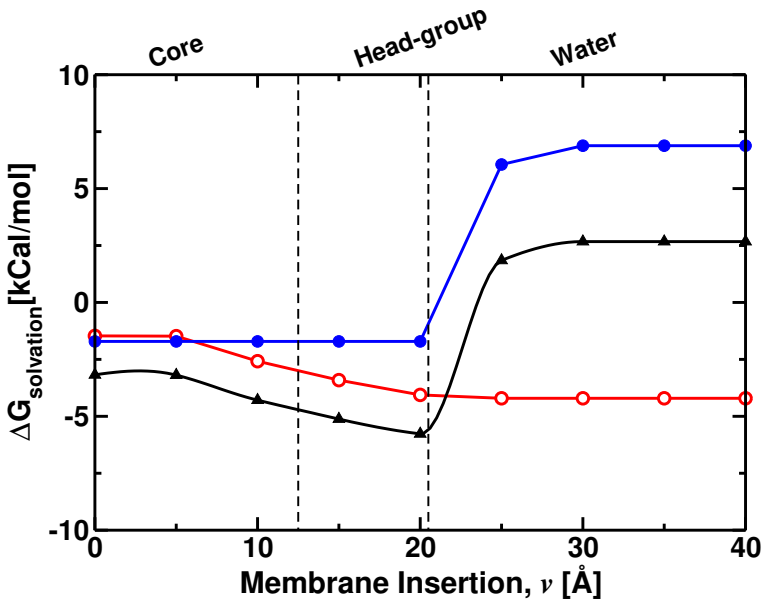

Figure 3.5: ΔG_{solv} (▲) as a function of v for indole in the membrane. The electrostatic component. ΔG_{np} (●) and the non-polar component, ΔG_{elec} (○) are also shown. ΔG_{solv} is minimum in the head-group region as a result of compensation between ΔG_{elec} and ΔG_{np}. The membrane core, membrane head-group and aqueous regions are marked.

3.3.3 Discussion

The five-slab continuum membrane model is introduced and shown to be validated by calculating the solvation energies of ions and side-chain analogues. The model is shown to account well for the heterogeneous environment of the membrane by implicitly incorporating a head-group region. The solvation energies for Na^+ and Cl^- ions located in the centre of the membrane core and the aqueous region of the model are calculated and compare well to the corresponding Born energies in homogeneous media. Thus, the membrane model accurately reproduces the electrostatic component of the solvation free energy. All subsequent calculations presented in this thesis represent the membrane environment using this model.

Furthermore, the five-slab membrane model accounted for the partitioning of various side-chain analogues at different positions along the membrane. Similar positional preferences for residues are also seen in transmembrane peptides. Butane, a non-polar molecule is found to partition in the membrane core. The favourable partitioning of butane in the membrane core corresponds to hydrophobic amino-acid residues such as Leu and Ile that are known to be incorporated in the membrane core [47, 48, 50]. In contrast, imidazolium, a charged species, is seen to clearly prefer the aqueous phase. The preference of imidazolium for the

aqueous phase is analogous to charged species such as His side chains which partition in the aqueous phase. Indole, a side-chain analogue of Trp, is seen to prefer the head-group region. Aromatic residues such as Trp have been seen to interact with the head-group region, often at the level of lipid carbonyls [40, 51]. Thus, the tests using side-chain analogues showed that the model is able to reproduce many known lipid-peptide interactions.

The model introduced here is found to accurately calculate solvation energies of biologically important ions. The membrane model also reproduces the fine balance between the various lipid-peptide and peptide-peptide interactions as seen in the calculations of the side-chain analogues. Further, the five-slab continuum model is computationally not very expensive and therefore allows the sampling of a large conformational space. Thus, the model is found suitable for studying the orientation of larger membrane peptides. In Chapter 4, the model is used to study the energetics of helical peptide orientations in the membrane. The electrostatics of membrane peptides and helices in membrane proteins is investigated with the presented model in Chapter 5. The five-slab membrane model is then used to model the structures of transmembrane dimers in Chapter 6.

Chapter 4

Energetics of Helical Peptide Orientation in Membranes

"Sway and sway in the summer air...", Her Voice, Oscar Wilde.

4.1 Orientation of Helical Peptides in Membranes

Helical peptides are the main building blocks of membrane proteins and determine membrane protein architecture by their association. The principles determining helical orientations therefore also decide the structure of the more complex membrane proteins. Though the orientation of various membrane peptides has been determined experimentally, [37, 45, 48, 115–117] the energetics of helical orientation in membranes are poorly understood. The main focus of research has been to determine helix propensity and the role of the various amino-acid residues. The orientation of melittin [29, 31, 33], the Src membrane domain [34], the FYVE membrane domain [30], the influenza virus fusion peptide [35], helical fragments of bacteriorhodopsin [35], a 20-mer WALP peptide [74], alamethicin [74] and glycophorin A [33] have been reproduced theoretically. However, none of these studies attempted to decompose the different contributions to the membrane-peptide interaction or to calculate the characteristic tilt angles of the helix axis.

In this study, the energetics of the orientation of membrane-bound α-helices is investigated by calculating the solvation energy of various peptides in a membrane system, *i.e.*, the free energy for transferring the helix from vacuum into a given position and orientation in the membrane. The characteristic orientation of a number of transmembrane and surface aligned peptides is calculated and the factors determining their position and orientation are examined. The different contributions to the membrane-peptide interactions in membrane-bound helices are calculated. Simple models, such as dipole and peptide analogues are used

to describe the energetics of the membrane helices, thereby, decomposing the complex energetics of membrane peptides into simpler terms. The membrane environment is represented by the five-slab model described in Chapter 3 (Fig. 3.1). The solvation energy is calculated using an all-atom model for the peptide together with the Poisson-Boltzmann equation for the electrostatic solvation energy and a term representing the cost of cavity formation in the aqueous solvent. The most probable orientation is determined by the minimum of the solvation free energy of the peptide in the membrane environment.

At first, the orientation of a dipole in a membrane is examined analytically and shown to be a good approximation to the energetics of N-methylacetamide, which contains a single peptide group. The peptide dipole energetics are found to tend to orient the axes of single helices perpendicular to the membrane plane and the calculations indicate that this effect dominates the positioning of polyalanine helices. The length dependence of helix tilt angles and fluctuations in the tilt angles are examined. Further, cavity formation in the aqueous solvent is found to play a crucial role in determining tilt angles of helices longer than the membrane width.

The model reproduces the experimentally-determined tilt angles of several membrane--spanning helices including glycophorin A, melittin and the WALP peptides. Further, the length-dependent fluctuation in the tilt angles of the WALP peptides is predicted and may rationalise the non-zero tilt of the shorter peptides. The effect of bilayer thinning on the orientation of the peptides has also been studied and shown not to significantly affect the results. Potential applications of the model in the analysis of peptide-membrane interactions and membrane-protein structure prediction are discussed.

4.2 Methodology

The most probable orientation of a membrane helix is that with the lowest solvation free energy, ΔG_{solv}, *i.e.*, the free energy for transferring the helix from vacuum into a given position and orientation in the membrane. Here, the position and orientation of a membrane peptide are described by two parameters: the tilt angle, θ, and the depth of membrane insertion, v. θ and v are depicted in Fig. 4.1. θ is the angle between the helix axis and the membrane normal and describes the orientation of the helix in the membrane. For instance, $\theta \sim 0^\circ$ for a transmembrane helix and $\theta \sim 90^\circ$ for surface aligned peptides. The position of the peptide is described by v, the distance of the centre of mass of the helix from the centre of the membrane. The value of v depends on the hydrophobicity of the peptides and for symmetric transmembrane peptides, v is expected to equal zero.

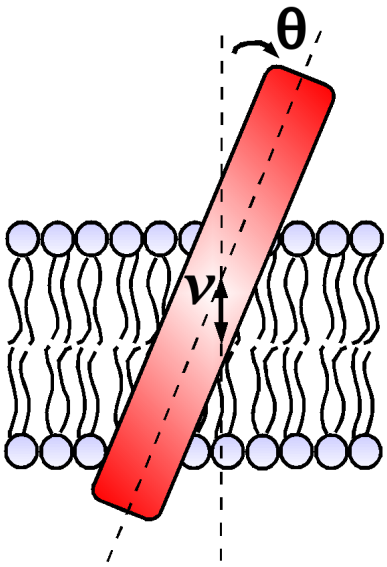

Figure 4.1: Parameters describing the orientation of a single helix in the membrane: tilt angle, θ and depth of membrane insertion, v. θ is the angle between the membrane normal and the helix axis and v is the distance between the centre of the membrane and the centre of mass of the helix.

4.2.1 Model Construction

N-methylacetamide

N-methylacetamide is a commonly used peptide-bond analogue. The structure of N-methylacetamide is depicted in Fig. 4.2. Atomic coordinates for the molecule were constructed using CHARMM [99] molecular parameters for peptide models.

Polyalanine Peptides

Polyalanine peptides with 20, 25, 28, 30, 35 and 40 Ala residues, were modelled as ideal α-helices with ψ=-47° and ϕ=-57° [2] using CHARMM to assign the molecular parameters. The N- and C-termini were blocked with acetyl and N-methyl acetyl groups, respectively. In addition, helices with 20 and 30 residues were modelled with charged ends.

WALP Peptides

WALP peptides are helices of alternating Ala and Leu residues with flanking Trp residues. The peptides were modelled as ideal α-helices with ψ=-47° and ϕ=-57° using CHARMM

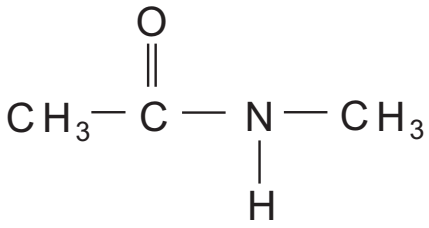

Figure 4.2: The structure of N-methylacetamide, a commonly used peptide-bond analogue. The molecule was used in the thesis to investigate the energetics of the peptide bond in a membrane.

to assign the molecular parameters. Peptides with 19, 20, 25, 30, 35 and 40 residues were constructed with blocked ends.

Melittin

The melittin peptide backbone structure, 2MLT [118], was taken from the Protein Data Bank [119] (PDB) and the N- and C- termini were charged so as to represent the naturally-occurring peptide. The side chains were constructed using the software SCWRL [120] (version 2.0), which determines the most probable rotameric conformations from a backbone-dependent rotamer library [121]. Steric clashes are relieved systematically by a combinatorial search of rotamers in an order defined by the rotamer library and calculated interaction energies.

Glycophorin A

The backbone structure of the transmembrane segment of the glycophorin A monomer, code 1AFO [42] residues 73 - 100, was taken from the PDB. The N-terminus was blocked with an acetyl group and the C-terminus with an N-methyl group. The side chains were again constructed using the software SCWRL.

All the above peptide models were energy minimised in a dielectric medium of ϵ=2 using the CHARMM force field [99] (version 29a2) with 1000 steps of Steepest Descent followed by 1000 steps of Newton-Raphson minimisation with 1kcal/mol harmonic constraints on the backbone atoms.

4.2.2 Calculation of Solvation Energy

ΔG_{solv} was calculated as a function of v and θ for the above-modelled peptides according to the thermodynamic cycle shown in Fig. 3.2. The two components ΔG_{elec} (electrostatic) and ΔG_{np} (non-polar) were calculated as described below.

Calculation of ΔG_{elec}

ΔG_{elec} was evaluated numerically by solving the linear Poisson-Boltzmann equation to calculate the electrostatic potentials in the membrane environment and in vacuum. The corresponding electrostatic solvation free energies of the system in the membrane, ΔG_{elec}^{mem} and in vacuum, ΔG_{elec}^{vac} were computed from the calculated electrostatic potentials. The parameters used were identical to those described in Section 3.2.1.

Calculation of ΔG_{np}

ΔG_{np} was calculated using Eq. 3.3. The solvent-accessible surface area, A was calculated using the CHARMM software package. As in Section 3.2.1, a probe sphere radius of 1.4 Å was used to define the surface.

Analytical Solution of the Poisson Equation for a Dipole in a Membrane

To calculate ΔG_{elec} for a dipole in a membrane the Poisson equation was solved analytically for a dipole in the five-slab continuum model. The analytical solution was derived in collaboration with Lars Meinhold.

The Poisson equation is:

$$\nabla \cdot [\epsilon(r) \nabla \phi(r)] = -4\pi\rho(r) \tag{4.1}$$

where, ϕ is the electrostatic potential within a medium of uniform dielectric constant ϵ and ρ is the charge distribution. The distance between the two dielectric boundaries was considered to be large enough for the dipole charges to experience only one dielectric boundary at a time. This approximation was validated by a numerical solution in which a dipole of the same dipole strength traversed the five-slab membrane model. The numerical solution (*c.f.* Fig. 4.5) showed that towards the centre of the head-group region ΔG_{elec} was constant indicating a homogeneous field and that at either dielectric boundary the change of ΔG_{elec} was independent of the other dielectric boundary.

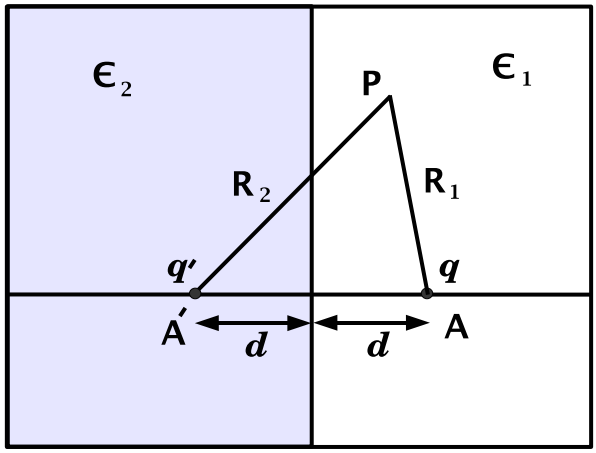

Figure 4.3: A single charge, q, and its mirror charge, q' located at points **A** and **A'**, respectively. Both charges are placed a distance, d away from the boundary between media with dielectric constants, ϵ_1 and ϵ_2, respectively. The electrostatic potential, ϕ is calculated at point **P** which is at distance, R_1 from q and distance R_2 from q'. If the point **P** lies in the dielectric region with dielectric constant, ϵ_2 then only a single image charge q'' placed at the same position as q is experienced.

The total electrostatic potential Φ can be written as:

$$\Phi = \phi_{q_a} + \phi_{q_b} \tag{4.2}$$

where ϕ_{q_a} and ϕ_{q_b} are the potentials at point $\mathbf{r}$ due to charges q_a and q_b respectively. For a dipole $q_a = -q_b$. The electrostatic potential, ϕ_q at point, P in cylindrical coordinates (ρ, ϕ, z), due to a single charge, q placed a distance, d away from a dielectric boundary between media with dielectric constants, ϵ_1 and ϵ_2 is [93] (see Fig. 4.3):

$$\begin{aligned} \phi_q &= \tfrac{1}{4\pi\epsilon_1}\left(\tfrac{q}{R_1} + \tfrac{q'}{R_2}\right) \quad z > 0 \\ \phi_q &= \tfrac{1}{4\pi\epsilon_2}\tfrac{q''}{R_1} \qquad\qquad\quad z < 0 \end{aligned} \tag{4.3}$$

where, $R_1 = \sqrt{\rho^2 + (d-z)^2}$ and $R_2 = \sqrt{\rho^2 + (d+z)^2}$. The image charges q' and q'' are chosen to satisfy the boundary conditions for the electric field $\mathbf{E}$ *i.e.*,

$$\lim_{z\to 0^+} \begin{Bmatrix} \epsilon_1 E_z \\ E_\rho \end{Bmatrix} = \lim_{z\to 0^-} \begin{Bmatrix} \epsilon_2 E_z \\ E_\rho \end{Bmatrix} \tag{4.4}$$

Using $\mathbf{E} = -\nabla\phi_q$ and Eq. 4.3 yields

$$\begin{aligned} q' &= \tfrac{\epsilon_1-\epsilon_2}{\epsilon_2+\epsilon_1} q \\ q'' &= \tfrac{2\epsilon_2}{\epsilon_2+\epsilon_1} q \end{aligned} \tag{4.5}$$

The electrostatic field energy of the dipole is then given by

$$W = \int_V \epsilon(\mathbf{r})\mathbf{E}^2 \, dV \tag{4.6}$$

$$= \int_V \epsilon(\mathbf{r})[\underbrace{(\nabla\phi_a)^2 + (\nabla\phi_b)^2}_{\text{self energy}} + \underbrace{\nabla\phi_a \cdot \nabla\phi_b}_{\text{interaction energy}} \;] \, dV \tag{4.7}$$

where, ϕ_a and ϕ_b are obtained from Eq. 4.3. Problems arise from the divergent self energy terms when integrating Eq. 4.7. These difficulties were avoided by integrating on a plane parallel to the dipole axis, thus excluding the singularities. This procedure corresponds to a remapping of the energy density of the electric field from three to two dimensions.

4.2.3 Computational Requirements

Each system setup took about one hour on four 800MHz processors in a Linux cluster. Each ΔG_{solv} calculation took about one hour on a single processor (2400 MHz, 2048 MB RAM). The total CPU time required for the calculations was ~3000 h.

4.3 Peptide Dipoles

We begin our analysis by calculating the electrostatic energy of peptide dipoles in membranes. Consider first a simple dipole, consisting of charges $q = \pm 1.0\ e$ placed 0.5 Å apart, crossing the five-slab membrane depicted in Fig. 3.1. The Poisson equation is solved analytically for this system. We compare two orientations of the dipole: θ=0° and θ=90°, where θ is the angle between the dipole axis and the membrane normal (see Fig. 4.1). The difference in ΔG_{elec} between the 0° and 90° orientations is the barrier to flipping, E_{flip} and gives the preference of one orientation over the other.

Plots of ΔG_{elec} *vs.* the depth of membrane insertion, v for the two orientations of the simple dipole are shown in Fig. 4.4. ΔG_{elec} decreases when traversing from the membrane core to the head-group region to the aqueous medium since ΔG_{elec} is inversely proportional to ϵ, which increases in this direction. Further, the difference in ΔG_{elec} between the membrane

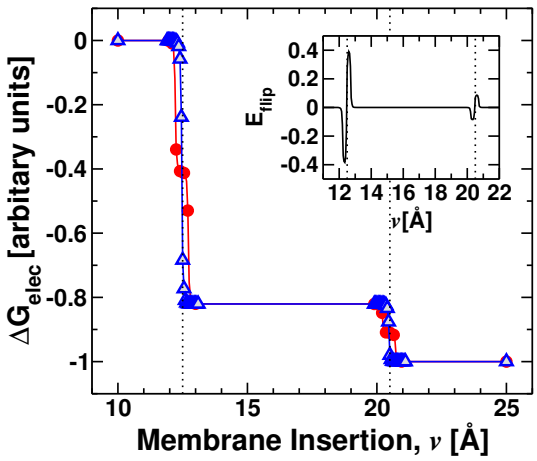

Figure 4.4: ΔG_{elec} *vs.* membrane insertion, v for a dipole (dipole moment = 2.4 D) calculated by the analytical solution of the Poisson equation. Two orientations, θ=0^o (•) and θ=90^o (△), are considered. The energy of flipping in the membrane, E_{flip} is given by the difference of ΔG_{elec} of the two orientations at a given value of v. The two dotted lines depict the boundaries of the head-group region. E_{flip} is non-zero only near the boundary between two media giving rise to a preference of one conformation over the other.

core and head-group region is approximately four times the difference in ΔG_{elec} between head-group region and water. This difference in ΔG_{elec} arises due to the variation of ϵ (and hence the potential) in the three media and equals $(\frac{1}{\epsilon_1} - \frac{1}{\epsilon_2})(\frac{1}{\epsilon_2} - \frac{1}{\epsilon_3})^{-1}$ where, ϵ_1, ϵ_2 and ϵ_3 are the dielectric constants of the membrane core, head-group region and water, respectively.

In a homogeneous medium, such as the membrane interior, the centre of the head-group region or the aqueous region, there is no orientational preference of ΔG_{elec} and hence no tendency of the dipole to orient itself in a particular direction. In contrast, a difference in ΔG_{elec} of the two orientations is seen near the dielectric boundaries, *i.e.*, when the dipole traverses from the membrane core to the head-group region or from the head-group region to the aqueous layer. Near a dielectric boundary, if the dipole axis is parallel to the membrane normal then a reaction field countering the charge closest to the boundary builds in the medium with higher polarity, lowering ΔG_{elec}. In contrast, if the dipole axis is perpendicular to the membrane normal, the reaction fields due to the two charges cancel each other and there is no net lowering of ΔG_{elec}. Thus, near a boundary between the dipole-containing medium and a higher-polarity medium, the orientation with the dipole axis parallel to the membrane normal is favoured over the perpendicular orientation. However, this picture reverses once the mid-point of the boundary is reached. On crossing this point, the dipole with θ=90° is completely immersed in the more-polar medium and is hence more favourable.

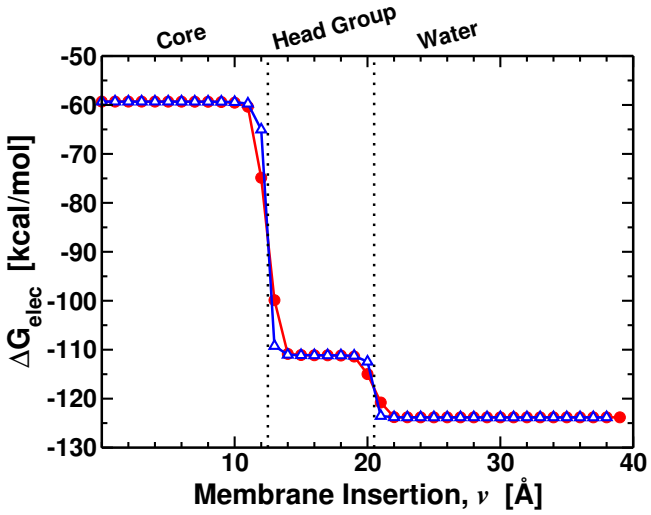

Figure 4.5: ΔG_{elec} *vs.* membrane insertion, v for a Dipole (dipole moment = 2.4D) calculated numerically as described in Section 4.2. Two orientations θ=0^o (•) and θ=90^o (△) are considered. The energy of flipping in the membrane, E_{flip} is given by the difference of ΔG_{elec} of the two orientations at a given value of v. The two dotted lines depict the boundaries of the head-group region. E_{flip} is non-zero only near the boundary between two media giving rise to a preference of one conformation over the other.

As a control calculation, ΔG_{elec} is also calculated numerically as described in Section 4.2. A non-zero value of radii was assigned to the dummy atoms to avoid infinite self energies. Fig. 4.5 shows the plot of ΔG_{elec} *vs.* v at θ=0° and θ=90°. The plot is similar to Fig 4.4 and shows that ΔG_{elec} is constant in the interior of the head-group region.

4.3.1 Orientational Preference of N-methylacetamide

The analysis is now extended to a molecule containing a single peptide bond: N-methylacetamide. For this molecule, ΔG_{elec} was calculated numerically as described in Section 4.2. Fig. 4.6 shows the plot of ΔG_{elec} *vs.* v at θ=0° and θ=90°. The results are qualitatively similar to the simple dipole model of Fig. 4.4. E_{flip} is seen to depend on the v as in the case of a simple dipole. Thus, the energetics of the orientational dependence of N-methylacetamide is well described by that of a simple dipole.

4.4 Polyalanine Peptides

In the next step, polyalanine peptides of varying lengths were investigated. Polyalanine peptides have been shown to adopt predominantly helical conformations in lipid bilayers [47].

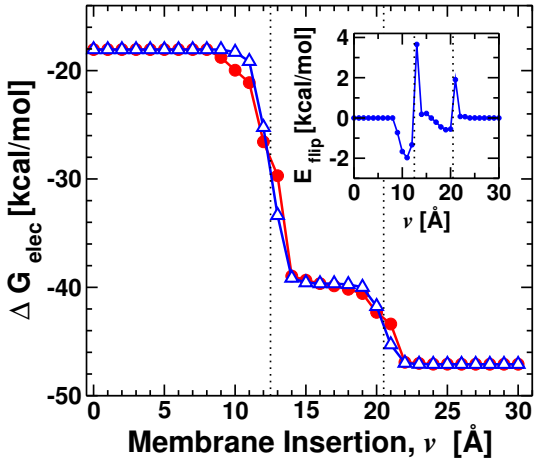

Figure 4.6: ΔG_{elec} *vs.* membrane insertion, v for N-methylacetamide calculated numerically as described in Section 4.2. Two orientations θ=0^o (•) and θ=90^o (△) are considered. The energy of flipping in the membrane, E_{flip} is given by the difference of ΔG_{elec} of the two orientations at a given value of v. The two dotted lines depict the boundaries of the head-group region. E_{flip} is non-zero only near the boundary between two media giving rise to a preference of one conformation over the other.

However, due to experimental limitations such as peptide aggregation, polyalanine peptide orientation in membranes has not been studied so far. Here, the results of a decomposition of the various factors governing the orientation these simple helices in a membrane are presented.

4.4.1 Effect of Different Classes of Charged Groups on ΔG_{elec}

The chemical groups in a helix can be divided into the peptide backbone, the side chains and the N- and C-terminal groups. The contributions of each of these three classes to ΔG_{elec} in a 30-mer polyalanine peptide are shown in Figs. 4.7 A and 4.7 B. Two types of end groups: charged (NH_3^+ and CO_2^-) and blocked (acylated N-terminus and N-Methylated C-terminus), are also examined.

To calculate the contribution of the above classes to the total ΔG_{elec} for the blocked end-group 30-mer peptide, the partial charges of all other atoms except those constituting the selected groups were set to zero (Fig. 4.7A). The figure shows that the side chains contribute very little to the total ΔG_{elec} barrier. The blocked peptide end groups make the second lowest contribution. The peptide backbone contributes the most, about three times more than the end groups. Thus, it is indeed the peptide group dipoles that play the major role in orienting the polyalanine helix axis perpendicular to the membrane surface.

The presence of charged termini increases the energy difference between the orientation with

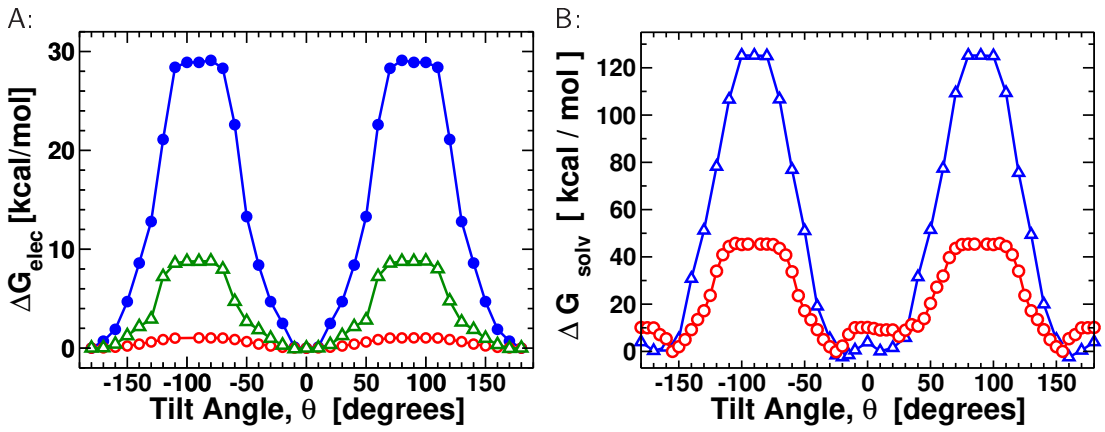

Figure 4.7: A: ΔG_{elec} *vs.* tilt angle, θ for a 30-mer polyalanine peptide with blocked ends with partial charges only on the side-chain atoms (○), N- and C- termini (△) or backbone (●). The contribution of the backbone atoms to the energy of flipping in a membrane, E_{flip} is the largest. The total energy cannot be expressed as a sum of the different contributions due to the interaction terms. B: ΔG_{solv} *vs.* θ at $v = 0$ for a 30-mer polyalanine peptide with charged ends (△) and blocked ends (○). The energies in both A and B are relative to the lowest energy conformer.

$\theta = 0°$ and $90°$. This increase arises from the fact that it is energetically less favourable to bury the charged end groups than the blocked ones. In contrast, there is little difference in the minimum-energy tilt angle between the blocked and charged peptides (Fig. 4.7B).

4.4.2 Effect of Peptide Length

We now further develop the analysis by calculating the solvation energy, ΔG_{solv}, *i.e.*, by adding in the non-polar term, ΔG_{np}, and by examining the dependence of the minimum-energy tilt angle on peptide length. The minimum-energy tilt angles for polyalanine peptides with 20, 25, 28, 30, 35 and 40 residues are plotted in Fig 4.8. The membrane thickness is 41 Å, corresponding approximately to a helix of 27 residues. The length of the 20-mer polyalanine peptide is thus shorter than the bilayer thickness. Consequently, the lowest-energy orientation is parallel to the membrane normal *i.e.*, $\theta = 0^o$, so as to optimise the above-characterised electrostatic energy of the peptide dipoles.

For peptides longer than the bilayer thickness, the cost of cavity formation in the aqueous layer (ΔG_{np}) is large enough to cause tilting of the helices so as to decrease the surface area exposed to the aqueous medium. Thus, for the 28-mer polyalanine peptide, which is longer than the membrane thickness, a non-zero tilt is observed. A further increase in tilt angle is observed with increasing number of residues. The minimum-energy tilt angle increases roughly linearly with the peptide length, as is evident in Fig. 4.8. Approximating the α-helix as a cylinder of length, l and radius r, the exposed surface area at θ=0° is given

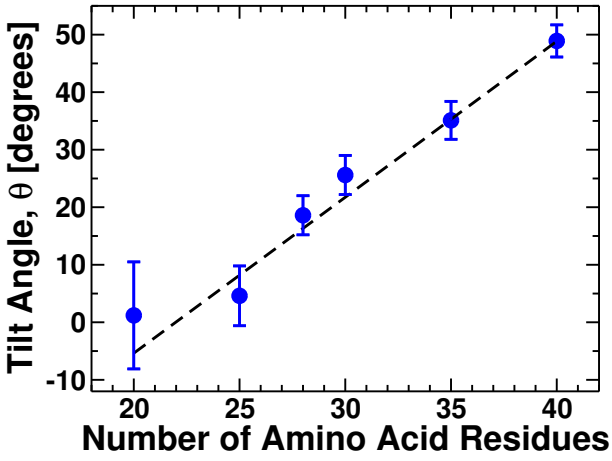

Figure 4.8: Calculated orientations of polyalanine peptides in a membrane environment Minimum-energy tilt angles *vs.* number of residues for polyalanine peptides at v=0Å. The thermal fluctuations given by the bars were calculated by fitting a harmonic potential to the energy profile of each peptide.

by $A = 2\pi r(l - d)$ where d is the width of the bilayer. Since ΔG_{np} is proportional to the exposed surface area, the contribution of ΔG_{np} to the total ΔG_{solv} increases approximately linearly.

To estimate the dynamic fluctuation of the tilt angle, a harmonic function was fitted to the solvation free energy profiles in the low-energy region. The resulting thermal fluctuations of the tilt angles, calculated from the fitted harmonic potentials, are also shown in Fig. 4.8, as bars. The fluctuations of the peptides decrease with increasing helix length. Large fluctuations are possible in shorter peptides such as the 20-mer since for $v = 0$, ΔG_{elec} is relatively broad and the peptides are never exposed to the aqueous layer for any value of θ. In contrast, fluctuation of the helix axis for the longer peptides leads to stronger variations in ΔG_{np} and ΔG_{elec}. Thus, the longer peptides are locked in steeper potentials. The above aspects of the free energy profiles are evident in Fig. 4.9 in which ΔG_{solv}, ΔG_{elec} and ΔG_{np} are plotted *vs.* θ for the 40-mer polyalanine peptide. This decomposition also illustrates how the tilting of the longer polyalanine peptides is due mainly to the contribution of ΔG_{np}.

The results in Fig. 4.8 suggest a simple geometric model in which the tilt angle adopted by a helix is the smallest possible while ensuring that the length of the helix can be fully incorporated within the bilayer. According to this model, $d = l\cos\theta$, where d is the width of the bilayer and l the length of the helix. However, as shown in Fig. 4.10 this model overestimates the tilt angles. Thus, the balance between ΔG_{elec} and ΔG_{np} is not well represented by the simple geometric model.

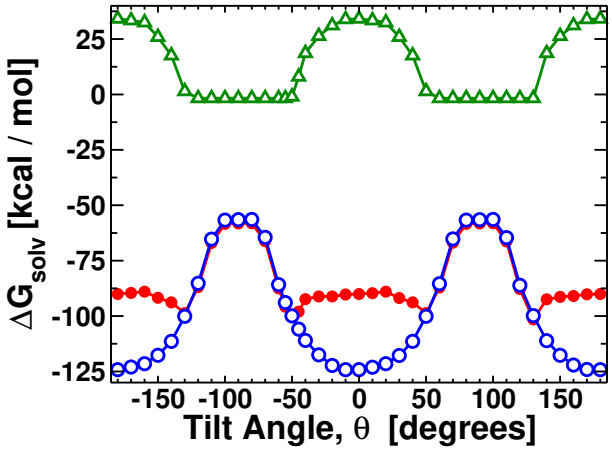

Figure 4.9: ΔG_{solv} *vs.* θ for a 40-mer with blocked ends at v=0Å. The electrostatic, ΔG_{elec} (∘) and non-polar, ΔG_{np} (△) components of the total ΔG_{solv} (•) are also plotted. The most probable orientation of the peptide is $\theta = 45°$.

4.4.3 Bilayer Thinning and Sensitivity of Model Parameters

It has been proposed that, for short peptides, bilayer thinning occurs to compensate for the hydrophobic mismatch between the membrane thickness and peptide length [39]. Reducing the membrane core thickness from 25 Å to 20 Å did not lead to a change in the tilt angle for the 20-mer polyalanine peptide. Further, to test whether the dielectric constant of the non-polar membrane core affects the results, a control calculation, increasing the dielectric constant of the core region to ϵ=4, was carried out for the 40-mer. However, the tilt angle did not change, remaining at 50°. Thus, we see that in the present case, neither bilayer thinning nor permittivity changes greatly influence the peptide orientations.

No theoretical or experimental estimates of ΔG_{np} for the membrane head-group region exist. To test the cost of cavity formation in the head-group region in present membrane model, control calculations were performed on polyalanine peptides. A term similar to ΔG_{np} for the aqueous layer but lower in magnitude was incorporated in the calculations of the tilt angles. However, the location of the free-energy minimum did not change in these calculations.

4.5 WALP Peptides

The WALP peptides are of particular interest as experimental NMR [40] and spectroscopic [38] data exist on their tilt angles. Therefore, calculations similar to those discussed above for the polyalanines were performed for WALP peptides with varying numbers of

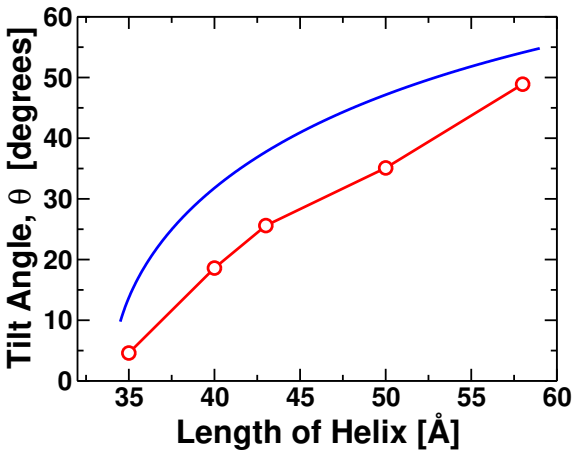

Figure 4.10: The minimum-energy θ (○) and that predicted by a geometric model (—) is plotted against peptide length, l. The geometric model assumes that the value of θ adopted by a peptide is the maximum such that the peptide just spans the bilayer (see text).

amino-acid residues. The resulting tilt angles and corresponding fluctuations are plotted in Fig. 4.11.

As for the polyalanine helices, the tilt angle for the 19-mer and 20-mer WALP peptide is 0°, the peptides orienting perpendicular to the membrane plane. The tilt angle for the longer peptides again deviates from zero, increasing with the number of residues. The tilt angles are found to be in good agreement with experiment.

A further point of interest is that in the experimental determination of the tilt angle of the 19-mer WALP peptide it was suggested that this peptide has a non-zero tilt angle but is nevertheless oriented roughly perpendicularly to the membrane plane [40]. The present analysis suggests that the relatively large fluctuations of the 19-mer WALP peptide tilt angle may have led to this observation.

To test whether the bilayer head-group region contributes significantly to the final tilt angle and corresponding fluctuations of a peptide, a control calculation of the orientation of a 20-mer WALP peptide was carried out with a three-slab membrane model as used in previous studies [32]. Due to the absence of a head-group region in the bilayer, the tryptophan residues are unable to anchor the peptide to the head-group region and the peptide was seen to tilt appreciably: the tilt angle calculated was 21° with fluctuation of ∼4°. These results are in contrast to the calculations with the five-slab model and also with the corresponding experimental results. Hence, including head-group energetics is necessary for calculating membrane-peptide orientations.

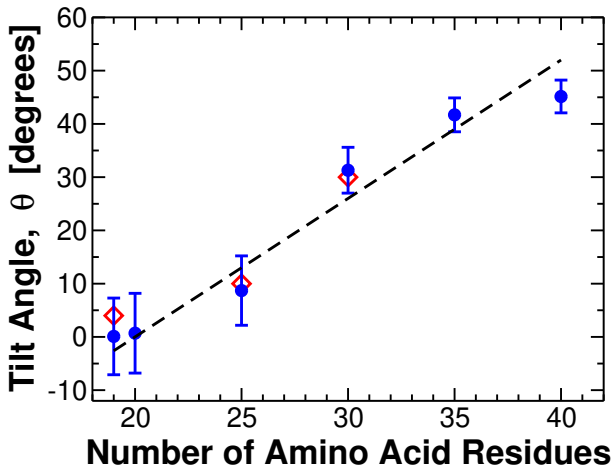

Figure 4.11: Minimum-energy tilt angle as a function of residue length for WALP peptides at v=0 Å. The theoretically-determined tilt angles are represented by (•) and the bars correspond to the thermal fluctuations, calculated by fitting a harmonic potential to the energy profile. The experimentally-determined tilt angles [38, 40] are represented by (⋄). The main contribution to the non-zero tilt angles arises from ΔG_{np}. Since ΔG_{np} increases linearly with the exposed surface area the graph is approximately linear.

4.6 Melittin

The model is further applied to two extensively-studied heteropolymeric membrane helices: melittin and glycophorin A. The membrane orientation of melittin has been a topic of much debate, especially as the various mechanisms proposed for melittin-mediated cell lysis depend on its membrane orientation [29, 45]. Previous studies have indicated both transmembrane and surface-aligned orientations depending on the experimental conditions [45, 115, 116]. At low concentrations, in phosphatidylcholine bilayers the predominant orientation of monomeric melittin determined by x-ray scattering was found to be surface aligned with the helical axis at the depth of the glycerol groups. The amphipathic nature of melittin makes it favourable for it to interact with both the dielectric boundary of the polar aqueous layer and the non-polar membrane core.

ΔG_{solv} was calculated for melittin orientations corresponding to $-180° \leq \theta \leq 180°$ and $0 \leq v \leq 25$Å. In a contour plot of ΔG_{solv} *vs.* θ and v (Fig. 4.12) the lowest energy orientation is labelled as Region I (lying between $14\text{Å} \leq v \leq 19\text{Å}$ and $70° \leq \theta \leq 90°$). This region corresponds to a surface-aligned orientation with the helix axis perpendicular to the membrane normal, consistent with previous theoretical [29] and experimental [45] studies.

In the crystal structure of melittin, a kink is formed in the helix backbone due to a Pro residue, giving rise to a short C-terminal hydrophobic tail [118]. Due to this kink

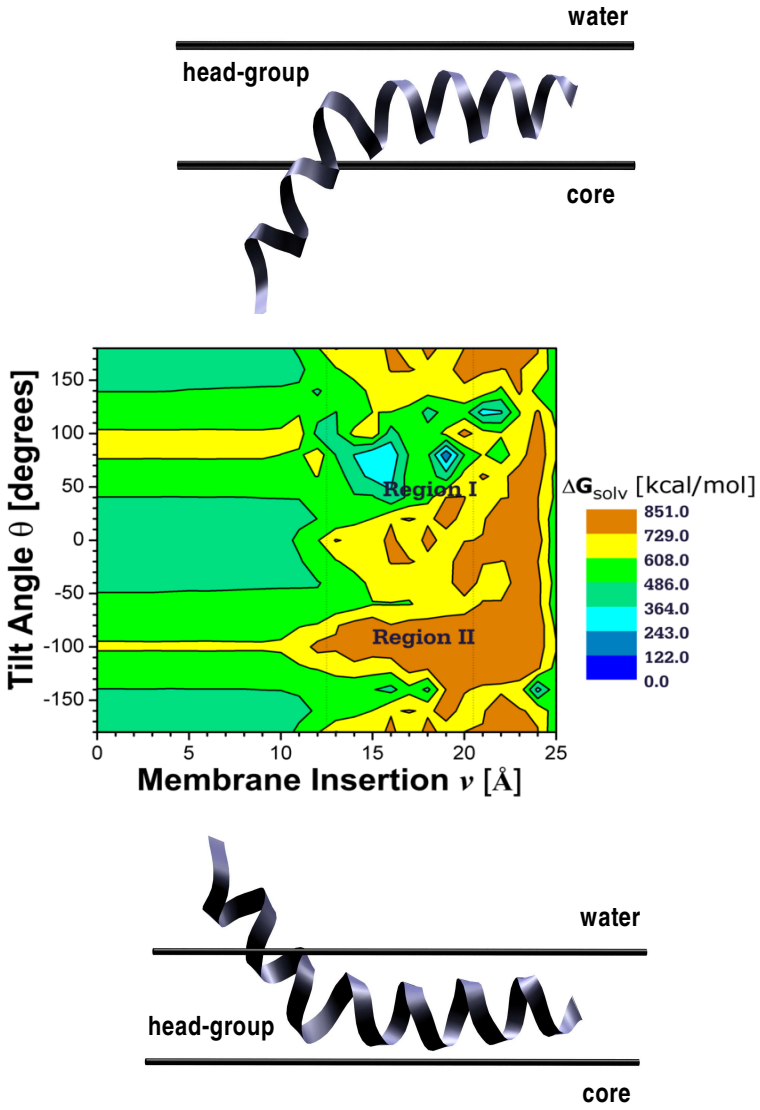

Figure 4.12: Solvation energy, ΔG_{solv} as a function of tilt angle, θ and membrane insertion, v for melittin monomers in a membrane environment. The most favourable orientation of melittin is region I, corresponding to a surface-aligned orientation with the C-terminal tail interacting with the membrane core (depicted on top). Region II also corresponds to a surface-aligned orientation but with C-terminal tail interacting with the water (depicted at the bottom) and is energetically highly unfavourable. The lowest energy conformers cluster around $80° \leq \theta \leq 90°$ and 18 Å $\leq v \leq$ 19 Å and all energies are relative to the lowest energy conformer. The head-group region extends between 12.5 Å and 20.5 Å.

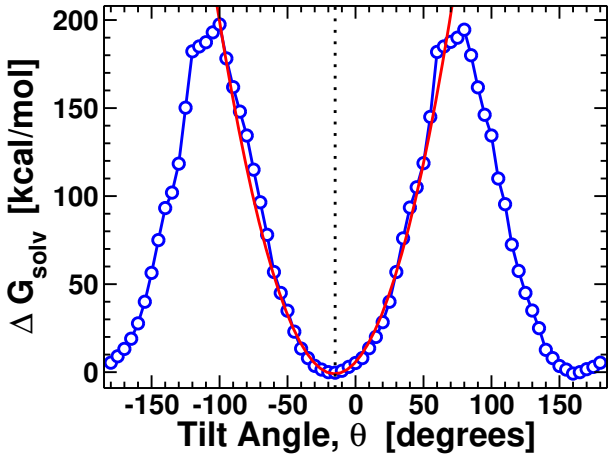

Figure 4.13: Solvation energy, ΔG_{solv} (○) *vs.* tilt angle, θ for the glycophorin A monomer in a membrane. ΔG_{solv} was calculated by the thermodynamic cycle shown in Fig 3.2. The lowest energy orientation is at θ=-15° and v=0Å. The energies are relative to the lowest energy conformer. A harmonic potential (—) is fitted on the energy profile between θ=-25° and θ=5° The resulting tilt angle is 14° with thermal fluctuations of ∼5^o. The experimentally-determined tilt angle is 20° [81].

the orientation corresponding to Region I allows the C-terminal tail to interact with the core region of the membrane and is energetically favourable. In contrast, in Region II, *i.e.*,$-90° \leq \theta \leq -70°$, the non-polar tail interacts with the aqueous layer increasing ΔG_{solv} substantially and making it highly unfavourable.

4.7 The Glycophorin A Monomer

In Fig. 4.13, the solvation energy, ΔG_{solv} of glycophorin A is plotted as a function of θ at $v = 0$. The orientation with the lowest energy is at $\theta = 15^o$ with thermal fluctuations of ∼5°. The value is consistent with the experimental tilt angle of ∼20° determined by solution NMR [42] and FTIR spectroscopy [122] of glycophorin A monomers. The non-zero tilt angle stems in this case mainly from the ΔG_{elec} contribution. The transmembrane segment of glycophorin A consists of several long chain and aromatic side chains and is thus more non-polar than peptides containing many Ala residues. Tilting of the helix incorporates more of these side chains into the non-polar membrane core region while allowing the flanking polar residues to interact with the head-group region. This gives rise to a non-zero tilt angle in glycophorin A in contrast to the zero tilt angle in WALP peptides of similar length.

4.8 Peptide Orientations: Overview and Applications

In the work presented in this chapter, the membrane orientations of single α-helices have been quantitatively examined using all-atom models of the helices embedded in a five-slab continuum dielectric environment representing the membrane and associated aqueous phase. The model has allowed the energetic factors determining the helix orientations to be quantitatively assessed. The approach taken was to start from considerations of a simple dipole in a membrane and then to build in additional effects as modelled by an all-atom peptide force field and burial of accessible surface area. The peptide dipoles are shown to favour the orientation perpendicular to the membrane plane *via* a solvent reaction field effect that is significant only for groups that approach the dielectric boundaries of the system. This reaction field is shown to be responsible for determining the membrane orientation of blocked polyalanine helices. Another important effect is solvent cavity formation that tends to minimise the surface area accessible to water. According to the model the cavity term plays an important role in influencing the steepness of the tilt-angle potential such that the longer peptides undergo smaller tilt-angle fluctuations than the shorter ones. The combination of solvation terms also allows the rationalisation of the length dependence of the peptide tilt angle on length. Furthermore, the presence of a separate dielectric zone for the head-group region is found to be important in determining the helix orientational properties.

Importantly, the model is found to correctly reproduce the helix tilt angles of peptides for which the tilt angles have been determined experimentally. These comprise the length-dependent helix tilt angles of the WALP peptides together with the membrane orientations of melittin and glycophorin A. The comparison with experiment provides evidence that the model quantitatively captures the main elements of the orientation-dependent membrane-helix solvation energy.

The potential usefulness of the present model is in three domains. Firstly, the electrostatic properties of peptides in membranes may be further studied using the model. Macromolecular electrostatics has been postulated to play a dominant role in the membrane environment [68]. In Chapter 5, the model has been used to calculate the macromolecular dipoles of transmembrane helices. Further, screening of the peptide atomic charges by the aqueous layer has also been calculated in membrane proteins such as aquaporin.

Secondly, the model can provide starting geometries for detailed molecular dynamics simulations of helices in membranes, in which all helix, membrane and water atoms are explicitly represented [29, 75–81]. All-atom simulations, although time-consuming, can provide a detailed picture of the helix/membrane system and include effects such as internal peptide

flexibility and explicit peptide-lipid interactions.

Thirdly, the model may provide a stepping stone towards the obtention of a theoretical framework for determining the structures of the many important proteins that contain helical transmembrane domains and that are difficult to crystallise. To do this it will be necessary to search relative helix translational and rotational angles in helix oligomers and to include backbone and side-chain flexibility. As a first step in this direction, the membrane association of helices into dimers have been studied in Chapter 6.

Chapter 5

The α-Helix Dipole: Screened Out?

"Curiouser and curiouser!", Alice in Wonderland, Lewis Carroll.

5.1 The α-Helix Dipole

In an α-helix, the alignment of the dipoles of peptide bonds leads to a "macroscopic" dipole parallel to the helix axis. The macroscopic helix dipole is discussed to play a crucial role in membrane peptides by determining the orientation of single helices or membrane protein architecture through association of the helices [68]. The helix dipole has been especially implicated in non-specific forces which contribute to helix-helix interactions in the membrane [52]. The helix dipole has also been shown to be of functional importance in membrane proteins such as aquaporin [123], the KcsA K^+ channel [89, 124] and the EcClC Cl^- channel [125].

In soluble proteins as well, the helix macrodipole has been implicated in protein function [126] and in stabilising structural motifs containing helix pairs [126–129]. Furthermore, the helix dipole moment has been suggested to influence pK_a values [130], absorption spectra [131] and electron transfer [132] and to stabilise the presence of charged residues at helix termini [133].

The strength of the helix dipole is given by the sum of the microscopic dipole moments, $\vec{\mu}_i$ [134, 135], arising from the individual peptide bonds, $i = 1, ...N$. The magnitude of the vacuum helix dipole is therefore proportional to the number of peptide bonds in the helix [135, 136] and is approximately equivalent to placing charges of $1/2e$ at the N- and C-termini [136–138].

For a peptide embedded in a membrane or in aqueous solution, solvent screening of the peptide-group charges is expected to lower the effective dipole of a helix. The screening

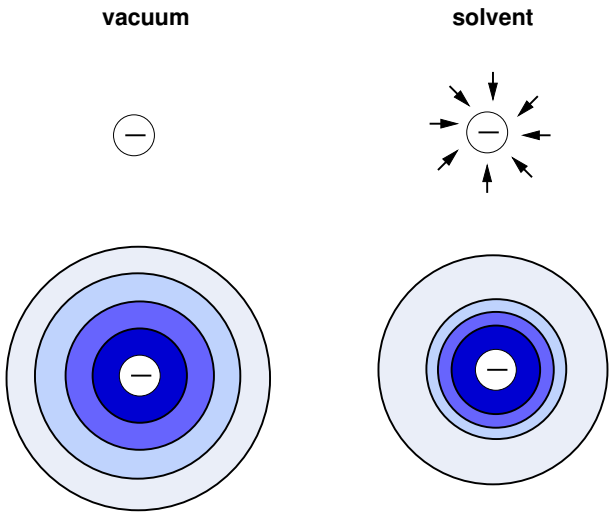

Figure 5.1: The equipotential lines of a single charge in vacuum and in aqueous solvent are depicted. The top left panel shows a single charge in vacuum and the bottom left panel the equipotential lines of that charge in vacuum. The top right panel shows the same charge in aqueous solvent. The arrows indicate the reaction field built up in the solvent. In the bottom right panel, the equipotential lines are depicted. Due the reaction field the effective charge is lowered and the electrostatic potential experienced same distance away from the charge is lowered. The equipotential spheres are represented with a colour scale map with higher potential corresponding to darker colour.

occurs due to the reaction field of the solvent which acts against the field generated by the vacuum dipole, leading to an effectively lower dipole moment [139]. The effect of the reaction field is depicted pictorially in Fig. 5.1 However, the magnitude of this screening, and consequently the effective dipole moment is unknown.

For helix pairs in membranes it has been shown that the electrostatic interaction between helices can be considerable if embedded in the low dielectric region of the membrane, but that the interaction is reduced significantly if the helix termini protrude out into the aqueous phase, even, by only a few Ångströms [68]. Similar calculations performed on solvent-exposed helical bundles, with the aqueous medium represented implicitly, have suggested that the helix dipole plays little or no role in stabilising the observed bundle geometry [69]. Thus, solvent screening is an important factor determining the strength of the interaction between α-helices themselves and between α-helices and other structural elements in various media.

In the present work, we calculate the effect of the helix environment in screening and modulating the helix dipole. The effective dipole moment, μ_{eff} of α-helices of varying lengths are calculated in vacuum, aqueous solution and lipid bilayers. The analysis is

extended for helices in soluble proteins and membrane protein interiors. μ_{eff} is calculated from the electrostatic potential generated by the helix which is represented at atomic detail, in the given environment which is represented implicitly by a continuum model. The helix dipole is relatively strong in vacuum. However, in aqueous solution the helix dipole may be drastically reduced due to an electrostatic reaction field generated by the solvent. Furthermore, whereas in vacuum μ_{eff} increases with helix length, the opposite is found to be the case for transmembrane helices. In soluble proteins, μ_{eff} is found to vary strongly with the orientation and position of the helix relative to the aqueous medium. μ_{eff} of helices in soluble proteins and membrane proteins has been calculated and common principles have been elucidated. The results are at first glance surprising, but can be rationalised in terms of the shielding of the helix termini. A set of simple rules is given for estimating μ_{eff} from experimental structures.

5.2 Calculating the Helix Dipole

Standard polyalanine α-helices [134] ($\phi = -57°, \psi = -47°$) of 10—34 residues length were modelled at atomic detail using CHARMM [99]. The N- and C-termini were blocked with acetyl and N-methyl groups, respectively. All the peptides were energy minimised in a dielectric medium of ϵ=2 using 1000 steps of Steepest Descent followed by 1000 steps of Newton-Raphson minimisation applying harmonic constraints on the backbone atoms (force constant 1 kcal/mol/Å^2). The environments of the all-atom helices were represented using continuum electrostatics. Four environments were examined: vacuum, water, protein and a lipid bilayer. $\vec{\mu}_{eff}$ was calculated in a three-step procedure:

1. Calculation of the Electrostatic Potential generated by the Helix
2. Fitting Atomic Point Charges to the Electrostatic Potential
3. Dipole Moment Calculation

5.2.1 Calculation of the Electrostatic Potential Generated by the Helix

The electrostatic potential around the α-helical peptide was calculated by solving the linearised Poisson-Boltzmann equation numerically using the PBEQ routine [104] in CHARMM. The vacuum and water dielectric constants were set to 1 and 80, respectively. The calculations were performed at 300 K and 0.15 M ionic strength. The electrostatic potentials were calculated on a grid with cell sizes 1.0 Å, 0.5 Å and 0.3 Å. The potentials of the coarser

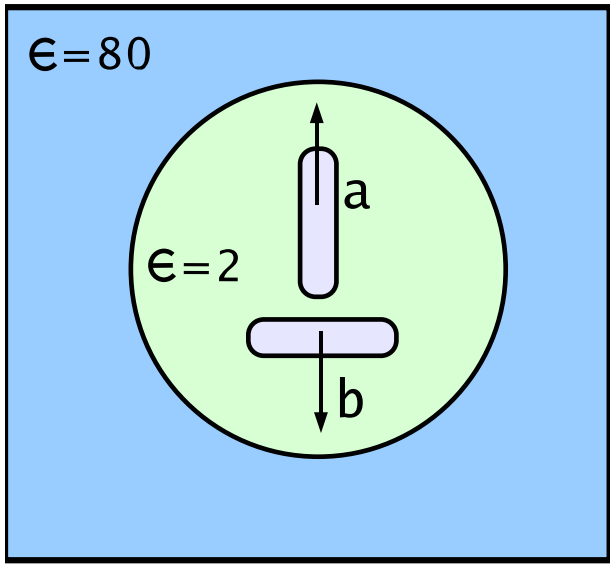

Figure 5.2: Protein models used in the calculations An α-helix in a globular protein depicted as a sphere of low polarity, ϵ=2. The two orientations in which the helix is placed in the protein correspond to the helix axis a) along and b) perpendicular to the radius vector.

grid were used in subsequent focussing onto the finer grids. The distance between the grid boundaries and the peptide surface was at least 12 Å. The charges were taken from the CHARMM force field and the atomic radii were taken as the Born radii [113].

A five-slab membrane model (Fig. 3.1), as introduced in Chapter 3, was used to calculate the electrostatic potential around helices in a lipid bilayer. In the five-slab model, the dielectric constants for the bulk water, head-group region and membrane core are set to 80, 10 and 2 respectively. The membrane core is 25 Å wide and both head-group regions are 8 Å in width.

Two models were used to calculate the helix dipole strengths in proteins. In the first, simplified model, a globular protein was represented as a sphere of low polarity with dielectric constant ϵ=2 (Fig. 5.2). Two sphere radii were examined: 25 Å corresponding to a protein of approximately 50 kDa, and 15 Å corresponding to a protein of ~10 kDa. Outside the spherical region the dielectric constant was set to 80, corresponding to bulk water.

In the second set of protein calculations the whole protein was represented at atomic detail and embedded in an ϵ=80 continuum phase. The partial charges for only the backbone atoms of the α-helix of interest were considered whereas the charges on the rest of the protein were set to zero. The dielectric constant of the protein interior was set to 2 and the

protein surface was defined as the water-accessible surface. The potential arising from the helix backbone atoms was calculated on grid points up to a distance of 4.5 Å from the helix and outside the van der Waals radii of the helix atoms. The proteins and their corresponding helices used for these all-atom calculations were as follows: **a:** Flap endonuclease (PDB [119] code: 1A76; residues 58-70) **b:** retinoblastomer protein (PDB code: 1GUX; residues 569-577) **c:** myoglobin (PDB code: 1A6N; residues 124-149) **d:** MyoD DNA binding domain (PDB code: 1MDY; residues 194-220) **e:** calmodulin (PDB code: 3CLN; residues 69-96). **f:** aquaporin (PDB code: 1FQY).

5.2.2 Fitting Atomic Point Charges to the Electrostatic Potential

Effective atomic point charges, q_i^{eff} of the helix atoms were obtained by adjusting the charges to reproduce the electrostatic potential generated in Step 1. The fitting was performed using a least-squares procedure similar to CHELPG [140]. The method was developed by Raghu Nath Behera and is described in detail in Section 5.2.5. The electrostatic potentials were mapped to a grid with 0.3 Å grid spacing. All points up to a distance of 4.5 Å from the molecule but outside the atomic van der Waals radii were included. The total charge of the helix was constrained to be zero.

5.2.3 Dipole Moment Calculation

In the third and final step, the magnitude and the direction of the effective dipole was calculated as $\vec{\mu}_{eff} = \sum q_i^{eff} \vec{r_i}$ where, $\vec{r_i}$ is the position of atom i of the helix. The centre of mass of the helix was used as the reference point for the calculation.

For systems with a net charge, the value of $\vec{\mu}_{eff}$ is dependent on the choice of the origin. However, the net charge of the α-helices considered was zero and therefore, $\vec{\mu}_{eff}$ does not depend on the choice of origin.

5.2.4 Reaction Field

Furthermore, the dipole vector originating from the reaction field was computed using the above discussed strategy. The reaction field potential was calculated as the difference between the electrostatic potentials around the helix in water and in vacuum. The helix atom point charges were then fitted to the solvent reaction field potential so as to obtain the effective charges best representing the reaction field.

5.2.5 The Charge Fitting Method

The effective atomic partial charges are derived from a least-squares fit to the molecular electrostatic potential. However, the derived charges have been found to be sensitive to small changes in molecular structure or in the sampling of the potentials, if proper fitting procedures are not used. The molecular potential of nearby points are highly correlated such that there is not enough independent data to assign charges to all the atoms of the molecule of interest. Thus, the system is *under determined* or becomes rank deficient. In such a situation, the method of singular value decomposition may be used to solve the problem.

A linear least-squares technique is used to find the partial atomic charges that best represent the molecular electrostatic potential. The fitted charges are obtained by minimising the function:

$$f\left(\mathbf{q}\right) \equiv f\left(q_1, q_2, ..q_{N_p}\right) = \sum_{i=1}^{N_p} \left(\phi^{PB}\left(\mathbf{r}_i\right) - \phi^{c}\left(\mathbf{r}_i\right)\right)^2 \tag{5.1}$$

where, $\phi^{PB}\left(\mathbf{r}_i\right)$ is the calculated Poisson-Boltzmann potential on N_p number of points. $\phi^{c}\left(\mathbf{r}_i\right)$ represents the electrostatic potential generated by the fitted charges located at N_{atom} number of atomic coordinates. The grid points selected to calculate the potentials should be of reasonable density (0.2 Å — 0.3 Å grid spacing) and not within atomic radii since it introduces unrealistic charges.

The electrostatic potential generated by the fitted charges is given by the classical Coulomb potential and hence:

$$f\left(\mathbf{q}\right) = \sum_{i=1}^{N_p} \left(\phi^{PB}\left(\mathbf{r}_i\right) - \sum_{j=1}^{N_a} \frac{q_j}{|\mathbf{r_i} - \mathbf{R_j}|} \right)^2 \tag{5.2}$$

Partial derivation with respect to q_k yields:

$$\frac{\partial f\left(\mathbf{q}\right)}{\partial q_k} = -2 \sum_{i=1}^{N_p} \Big(\phi^{PB}(\mathbf{r}_i) - \phi^{c}(\mathbf{r}_i) \Big) \Big(\frac{1}{|\mathbf{r_i} - \mathbf{R_j}|} \Big) = 0 \tag{5.3}$$

The system of equations for all q_k has the form $Aq - a = 0$. The set of these linear equations is solved using the singular value decomposition method. Constraints are introduced by the Lagrange Multipliers method. The total charge on the molecule was constrained to be zero.

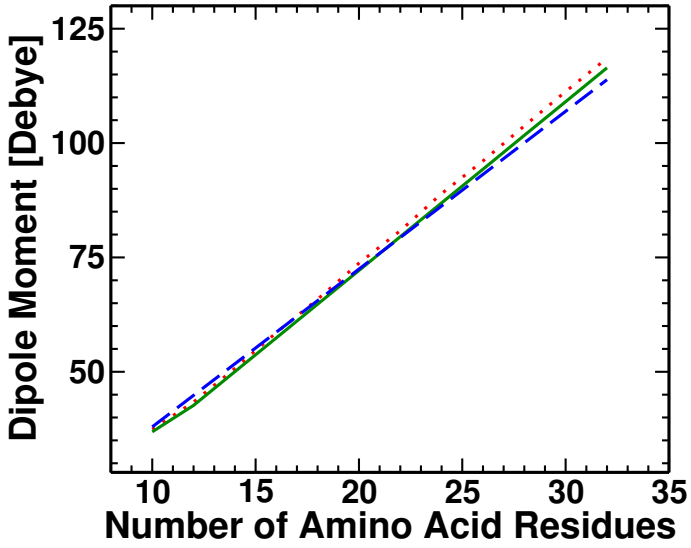

Figure 5.3: Effective dipole moment of polyalanine helices in vacuum as a function of peptide length. Dipole moments calculated using the atomic partial charges (- - -), calculated by multiplying the 3.45 D peptide bond dipole by the number of peptide bonds (—), and obtained as the dipole moment of the effective charges obtained from fitting the Poisson potential in vacuum to point charges (· · ·). Practically identical results were obtained using effective charges calculated in vacuum, thus demonstrating that the fitting procedure does not introduce unphysical artifacts.

5.3 Effective Helix Dipole in Homogeneous Media

5.3.1 Helix Dipole in Vacuum: Validity of the Fitting Method

μ_{eff} of polyalanine helices of various lengths is calculated in vacuum, water, a lipid bilayer and in proteins. In vacuum, the dipole moment of an α-helix increases linearly with peptide length. μ_{eff} (· · ·) is plotted against peptide length in Fig. 5.3. The ionic strength is of course zero, implying that the effective charges were obtained from fitting the Poisson potential in vacuum to point charges. μ_{eff} calculated from the fitted charges, and also that from the atomic partial charges (- - -), is similar to the product of the dipole moment of a peptide group (3.45 D) and the length of the helix (—), thus demonstrating that the fitting procedure does not introduce unphysical artefacts.

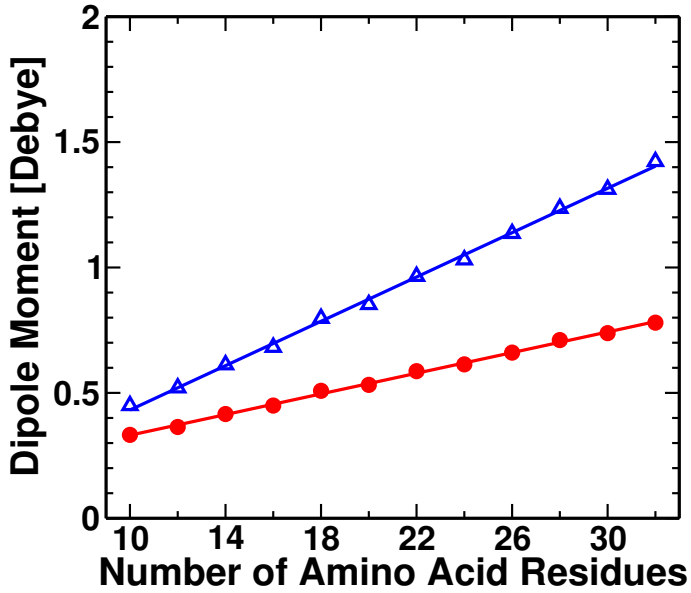

Figure 5.4: Dipole moments of α-helices in water derived from the effective charges calculated by fitting to the Poisson-Boltzmann potential in water at zero ionic strength (△) and at 0.15 M ionic strength (•). The solid and dashed lines are linear regression fits to the calculated points.

5.3.2 Effective Helix Dipole Moment in Water: Role of the Solvent Reaction Field

In aqueous solution, the solvent reaction field considerably lowers the effective electrostatic potential (Fig. 5.4). μ_{eff} again increases linearly with peptide length but with a significantly smaller slope than in vacuum. As expected, the screening is even stronger at non-zero ionic strength.

To estimate the magnitude of the reaction field dipole moment arising from the environment, the reaction field potential for helices in water was calculated as the difference of the electrostatic potentials around the helix in water and in vacuum. The helix atom point charges were then fitted to the solvent reaction field potential so as to obtain the effective charges best representing the reaction field. The dipole vectors of the helix in vacuum and the reaction field dipole vectors in aqueous solution are listed in Table 5.1. The magnitude of the reaction field dipole vector is similar to that of the dipole vector of the helix in vacuum, but points in the opposite direction. Therefore, the effective dipole vector is the sum of the two nearly-equal contributions and is thus very small. Additional calculations (Fig. 5.5) demonstrated that μ_{eff} scales with $1/\epsilon$. Thus, in a homogeneous medium all the charges are uniformly screened and the extent of the screening depends only on the polarity of the solvent.

No. of Residues	Dipole Moment [D]		Dipole Moment Vector [D]					
	Water	Vacuum	Vacuum			Reaction field		
10	0.45	37.4	0.0	-8.0	36.5	-0.0	8.0	-36.0
12	0.5	42.7	4.3	-3.9	42.3	-4.3	3.9	-42.2
14	0.6	50.7	-1.0	2.0	50.7	1.0	-2.0	-50.0
16	0.7	58.5	2.9	-3.9	58.3	-2.9	3.9	-57.6
18	0.8	65.8	2.0	1.0	65.8	-2.0	-1.0	-64.8
20	0.85	73.7	-0.5	-3.9	73.6	0.5	3.9	-74.4
22	0.95	80.8	2.4	-0.5	80.7	-2.4	0.5	-79.6
24	1.0	88.9	-1.4	-2.9	88.8	1.4	2.9	-87.4
26	1.15	96.0	2.9	-2.0	95.9	-2.9	2.0	-95.0
28	1.25	103.7	-1.4	-1.4	103.7	1.4	1.4	-102.2
30	1.3	111.2	2.4	-2.4	111.2	-2.4	2.4	-109.9
32	1.4	118.5	-1.4	-1.0	118.5	1.4	1.0	-116.6

Table 5.1: The dipole moments in vacuum and water of polyalanine helices with 10-32 residues are listed. The reaction field vector in water is also given.

5.4 Charge Screening in Membranes

In a heterogeneous environment, such as in a lipid membrane, the situation becomes more complex. μ_{eff} of the helices, now embedded in a membrane with the helix axes perpendicular to the membrane plane, are shown in Fig. 5.6. μ_{eff} falls off approximately linearly with helix length until the helix spans the membrane (28 residues), at which point the helix dipole moment is reduced to its aqueous solution value and does not change further with further increase in length. This interesting behaviour is the opposite of that in uniform media (Figs. 5.3 and 5.4), in which lengthening the helix increases the dipole moment.

The physical origin of the inverse length dependence of helices spanning the membrane is revealed by calculating μ_{eff} of a two-charge dipole. For this calculation, charges of $\pm 0.52\ e$, generating exactly the helix dipole of the 20-mer in vacuum, were placed at a distance corresponding to the length of the polyalanine helix modelled. A cylinder of diameter 5 Å and dielectric constant ϵ=2 was placed around the dipole to mimic the associated dielectric

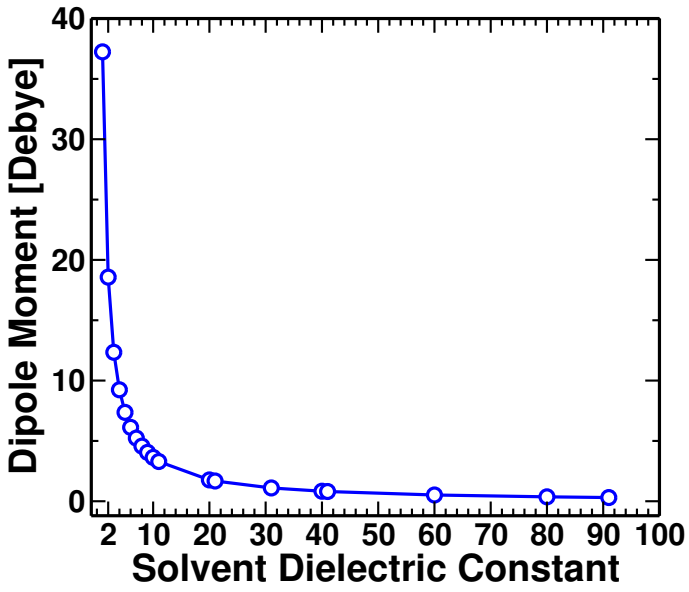

Figure 5.5: Effective dipole moment of a decaalanine peptide as a function of the polarity of the medium. The μ_{eff} decreases as $1/\epsilon$, where ϵ is the dielectric constant of the medium. Increasing the solvent polarity decreases the effective dipole moment because the reaction field increases and the charges are screened more effectively.

boundary. The charged cylinder was placed in the membrane model environment and the potentials calculated as above.

As shown in Fig. 5.6, the length dependence of μ_{eff} for the atomic-detail peptide is reproduced by the two-charge dipole. This observation indicates that the reversal of the slope of μ_{eff} *vs.* peptide length in a membrane is due to the inhomogeneity of the environment. As the peptide length increases, the terminal charges approach the high-dielectric solvent which then screens them. Thus, the reaction field in the surrounding water increases with increasing peptide length, decreasing μ_{eff}.

5.4.1 Effect of Helix Tilting in Membranes

Most transmembrane helices tilt to varying angles in the membrane. The tilt angle, θ is defined as the angle between the helix axis and the normal to the membrane surface. μ_{eff} *vs.* θ for the 10-mer, 20-mer and 30-mer polyalanine helices is shown in Fig. 5.7. $\theta = 0°$ is the orientation in Fig. 5.6, with the 10-mer and 20-mer buried in the membrane and the 30-mer extending into the aqueous layer. When the helix axis is placed parallel to the membrane plane, *i.e.* at $\theta = 90°$, the helices are completely embedded in the membrane core. Consequently, μ_{eff} is much higher than at θ=0°, and is approximately equal to μ_{eff} for the corresponding helix in a homogeneous medium of ϵ=2. As θ increases so does the

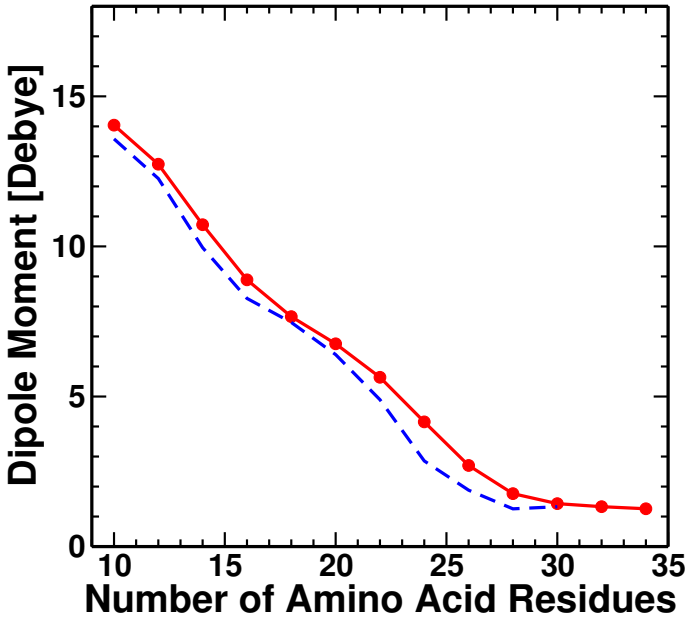

Figure 5.6: Effective dipole moment of polyalanine helices (—) centred in a five-slab membrane with helix axis parallel to the membrane normal as a function of the peptide length. The dotted line (···) corresponds to the effective dipole moment of a two-charge dipole.

shielding of the charges by the reaction field and hence μ_{eff} decreases. Interestingly, the 30-mer has the largest μ_{eff} when placed parallel to the membrane plane but the smallest μ_{eff} when placed along the membrane normal. Most transmembrane helices have relatively small tilt angles (∼0-40°) and therefore, their corresponding μ_{eff} values will be rather low.

5.5 Model Proteins: Principles of Helix Dipole Screening

Extending the analysis to calculate helix dipole strengths in proteins, we first model a globular protein as an uniform sphere of low dielectric constant (ϵ=2). Two such cases are examined: in one the helix is much shorter than the protein diameter and in the other the helix spans roughly the entire protein. For both cases, the helix is buried at varying depths in the protein in the two orientations shown in Fig. 5.2, *i.e.*, with the helix axis parallel or perpendicular to the sphere radius.

We first consider a decaalanine peptide (∼15 Å length) positioned in a protein of radius 25 Å, corresponding roughly to a molecular weight of 50 kDa. Fig. 5.8 shows μ_{eff} *vs.* the distance, d from the centre of the sphere. In both orientations considered, the reaction

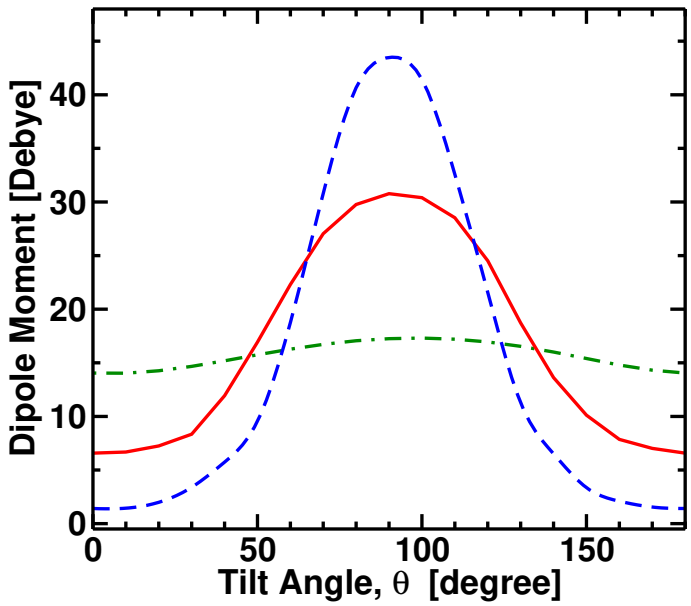

Figure 5.7: Effective dipole moment of polyalanine peptides centred at the core of a five-slab membrane as a function of the tilt angle, θ. μ_{eff} for polyalanines with 10 residues (-·-), 20 residues (—) and 30 residues (- - -) is shown. θ is defined as the angle between the helix axis and the normal to the surface of the membrane.

field significantly screens the charges of the helix even before it is exposed to the solvent. The screening starts when the helix termini are within ∼5 Å from the aqueous phase. In the interval 0 $\mathring{A}$ $< d <$ 17 Å the helix oriented along the radius vector is closer to the solvent than that oriented perpendicularly and hence the charges are shielded more by the reaction field (Case A, Fig. 5.8). However, on increasing the distance from the centre, one of the helix termini touches the surface of the sphere and the helix is subjected to a complex dielectric medium with the low dielectric protein at one end and high dielectric water at the other end. The asymmetric reaction field is not able to efficiently stabilise the buried dipole and the screening is lower, giving rise to a less steep slope between 17 Å and 25 Å (Case C, Fig. 5.8). In the other orientation, with the helix axis perpendicular to the radius vector, the termini of the helix perpendicular to the radius are equidistant from the aqueous phase at all values of d. The reaction field set up at each helix terminus is equal, and μ_{eff} decreases smoothly as the distance from the protein centre increases (Case B, Fig. 5.8) until it reaches its value in the aqueous phase (Case D, Fig. 5.8). Fig. 5.8 quantifies how the relative geometries of the helix and the aqueous phase are important in determining the reaction field set up and the extent of the screening of a protein helix.

In a further set of calculations, a 20-residue helix (∼ 30 Å length) was positioned in a sphere of radius 15 Å, corresponding roughly to a protein of 10 kDa. μ_{eff} *vs.* the distance from the

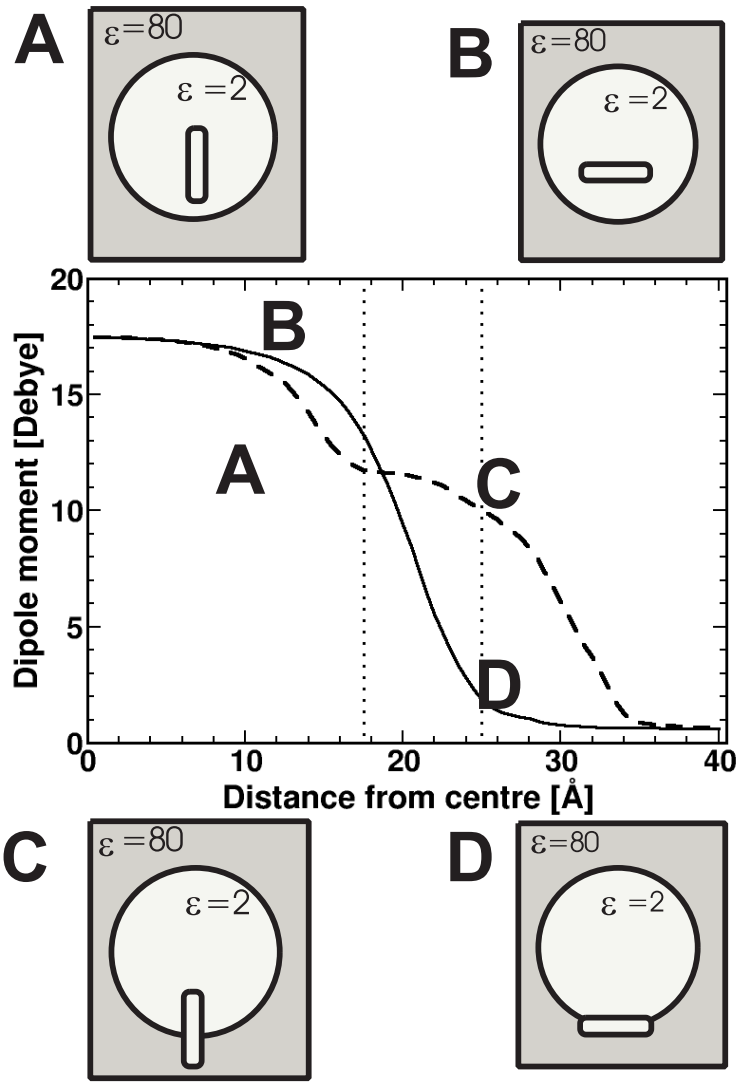

Figure 5.8: Effective dipole moment of a decaalanine α-helix as a function of the distance from the centre of a low-dielectric sphere of radius 25 Å. Two orientations are plotted: helix axis along (- - -) and perpendicular to (—) the radius. The dotted lines indicate the radius of the sphere (25 Å) and the distance at which the peptide that is along the radius touches the aqueous medium (17.5 Å). The radius also corresponds to the distance at which the centre of the helix placed normal to the radius touches the surface of the sphere.

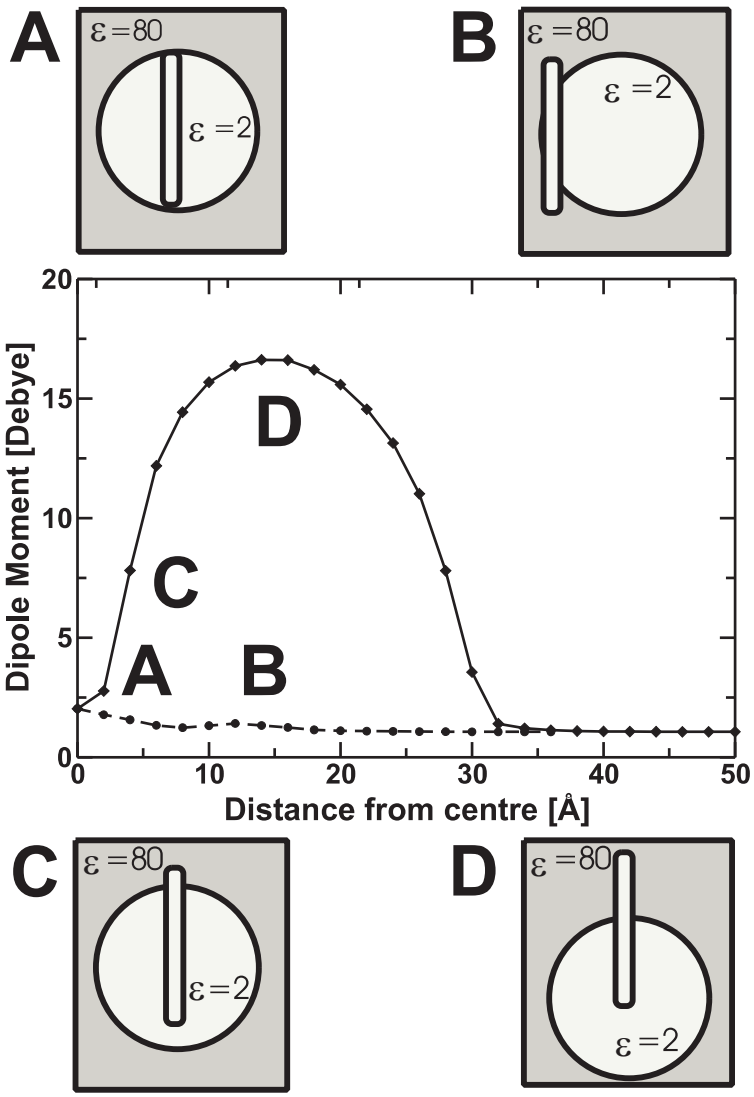

Figure 5.9: Effective dipole moment of a 20-mer polyalanine α-helix as a function of the distance from the centre of a low-dielectric sphere of radius 15 Å. Two orientations are plotted: helix axis parallel (—) and perpendicular (- - -) to the radius vector.

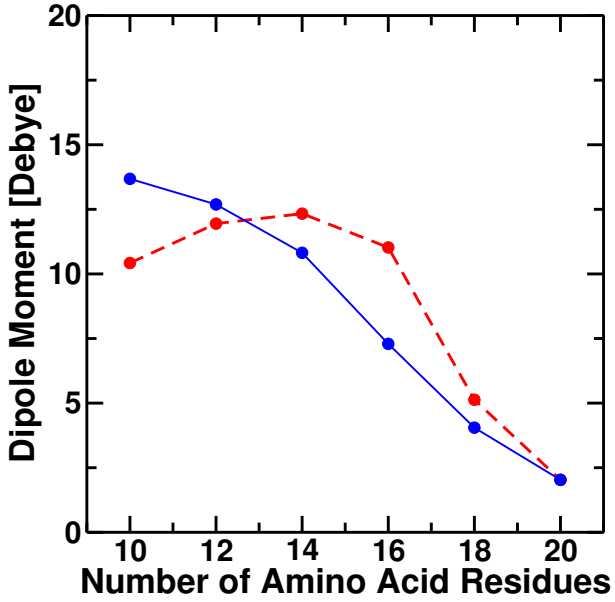

Figure 5.10: Effective dipole moment of polyalanine helices placed along the radius of a low-dielectric sphere of radius 15 Å. The helices are positioned in two ways. In the first the helix centre of mass coincides with the protein centre of mass (—). In the second the N-terminal of each helix is placed at the surface of the protein and the growing helix extends into the the protein interior (- - -).

centre of the sphere is plotted in Fig. 5.9 for the same two orientations of the peptide as in Fig. 5.8. The same principle holds as for Fig. 5.8, *i.e.*, that the geometry of the aqueous phase around the helix determines the extent to which the μ_{eff} is screened. However, the shape of the plot is different from that of Fig. 5.8. As in Fig. 5.8, the helix perpendicular to the radius vector is screened at both termini uniformly. However, μ_{eff} is roughly constant and significantly smaller since the termini are always very close to the aqueous layer (Cases A and B, Fig. 5.9). In contrast, the helix oriented along the radius vector exhibits complex screening behaviour. When the geometric centre of the helix coincides with the centre of the sphere, both the helix termini are close to the aqueous layer and the helix dipole is screened significantly (Case C, Fig. 5.9). However, when the geometric centre of helix is displaced from the centre of the sphere, one of the termini is embedded in the low dielectric sphere, and hence is poorly shielded. Thus, μ_{eff} increases until the helix terminus coincides with the sphere centre (Case D, Fig. 5.9) and then gradually decreases until it is fully exposed to water, when μ_{eff} of the helix is equal to its value in aqueous solution.

Polyalanine helices of varying lengths are now considered in the above globular protein model (radius = 15 Å). The effective dipole moment of the helices is plotted *vs.* peptide length in Fig. 5.10. Two positions are considered. In the first position, the helix centre

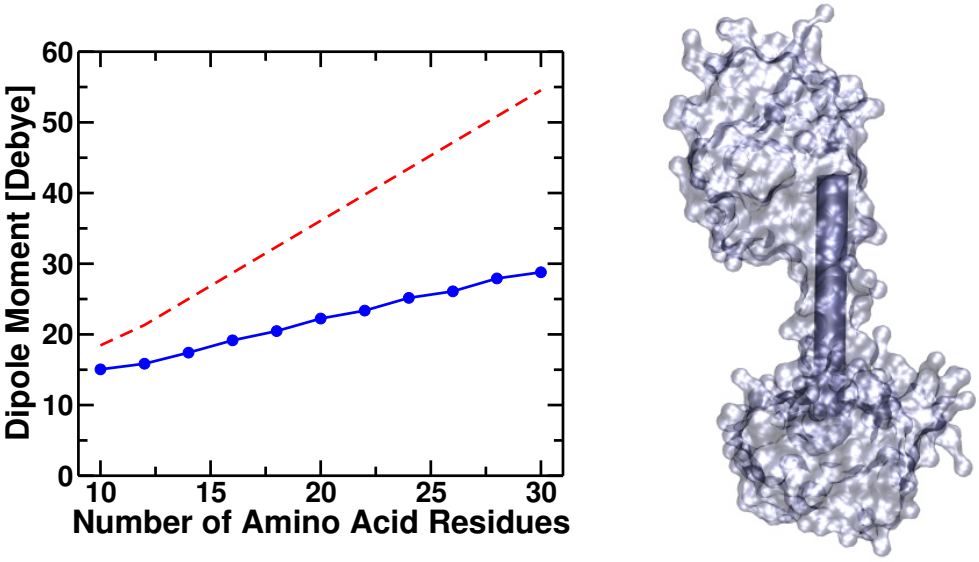

Figure 5.11: Effective dipole moment of polyalanine helical linkers of varying lengths(—). The corresponding dipole moments in a homogeneous medium of ϵ=2 are also shown (- - -). The cartoon representation of the α-helical linker in calmodulin (3CLN). In linkers, the terminal residues are buried whereas all the non-terminal residues are solvent exposed.

of mass is at the protein centre of mass. For this position, μ_{eff} decreases as a function of peptide length as the ends of the helices approach the aqueous medium. This behaviour is similar to that of helices centred in a membrane (Fig. 5.6). In the second position, the N-terminus of the helix is placed at the surface of the protein and the 'growing' helix extended into the protein interior along the radius. In this orientation, the decamer reaches the centre of the protein and the 20-mer spans the entire sphere. In this case, two opposite effects are in play: the increase of dipole moment with increasing number of peptide bonds and the increasing proximity to the aqueous medium of the C-terminus of the helix. For the 12-mer and 14-mer, the C-terminus is sufficiently far from the aqueous layer that the dipole moment actually increases as a function of peptide length. However, for the 16-mer and the 18-mer the ends are close enough to the aqueous layer to increase the strength of the reaction field and hence μ_{eff} decreases. For the 20-mer both the orientations in Fig. 5.10 coincide and μ_{eff} is the lowest of the cases studied above.

5.5.1 Linkers

A further interesting case is that of α-helical linkers, linking two domains or two subunits. α-helical linkers are important in the calcium binding proteins such as calmodulin. To examine these cases, two globular protein models each of radius 25Å were placed at either

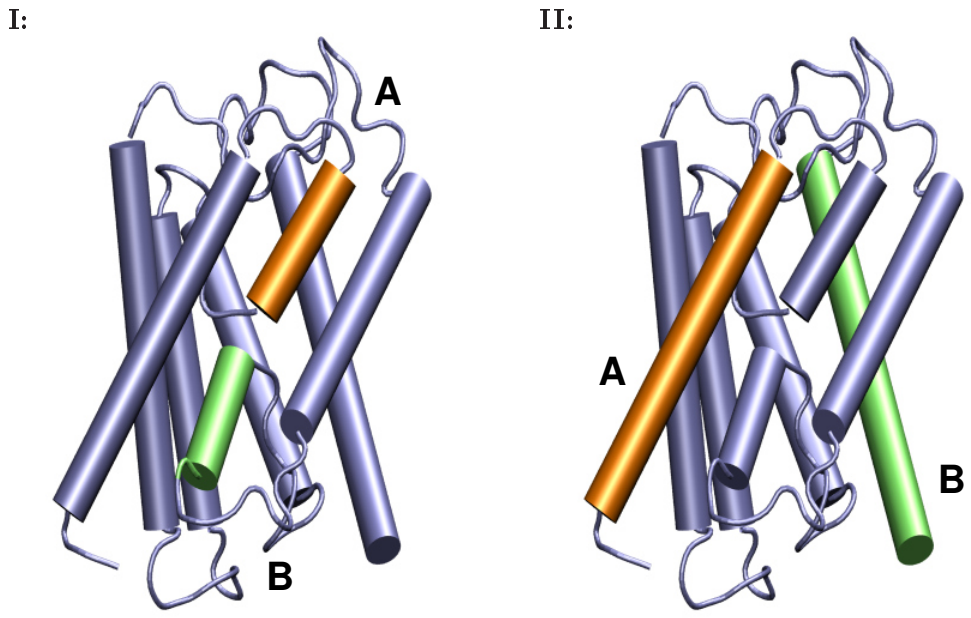

Figure 5.12: **I:** Membrane-embedded helices A and B in aquaporin. μ_{eff} for these helices is large. **II:** Transmembrane helices A and B in aquaporin. μ_{eff} for these helices is small.

end of polyalanine helices of increasing length. μ_{eff} *vs.* peptide length for helices comprising of 10-30 residues is shown in Fig. 5.11. μ_{eff} is seen to increase with peptide length, with magnitudes comparable to that in a low-polarity medium, albeit with reduced gradient relative to that of a homogeneous medium of ϵ=2. The situation of a linker is reverse of that of the earlier cases, since there is no preferential shielding of the termini, along the length of the helix. Thus, on adding more peptide groups, the terminal charges still experience a medium of low polarity and μ_{eff} increases.

5.6 Effective Helix Dipole Moments in Membrane Proteins

We examine atomic-detail model of aquaporin embedded in the five-slab membrane. Four helices of varying length were considered. First, μ_{eff} for the two short helices A (orange) and B (green) depicted in Fig. 5.12 I were considered. The two helices extend from the membrane core outwards and have been implicated in orientating the water molecules and in constraining proton transport. Helix A is 10 residue long, and is calculated to have a

high dipole moment of 12 D. Helix B is 9 residue long, and the calculated μ_{eff} is 10 D. The two helices are placed with opposite dipole directions such that there is a high positive potential inside the channel, thereby preventing H^+ to pass through it.

Next, μ_{eff} for the transmembrane segment A (orange) in Fig. 5.12 II was calculated. The helix is 22 residue long, and is shorter than the thickness of the model membrane. μ_{eff} is calculated to be 8 D. The second transmembrane helix B (green) in Fig. 5.12 II is 28-residue helix. The two termini of the helix are solvent-exposed and μ_{eff} is calculated to be 5 D.

5.7 Effective Helix Dipole Moments in Soluble Proteins

Finally, we examine atomic-detail models of selected proteins embedded in an ϵ=80 dielectric continuum. The proteins and the helices examined are illustrated in Fig. 5.13. The 12-residue α-helix in Flap endonuclease corresponds to Case A in Fig. 5.8. The value of μ_{eff} obtained is 17 D, which is about half the *in vacuo* value and 40 times the value in aqueous solution. This helix dipole is not affected by the aqueous phase and its electrostatic properties are similar to that in a homogeneous ϵ=2 continuum. Case C of Fig. 5.8 corresponds to the 9-residue helix of the retinoblastomer protein. Although this helix is strongly solvent exposed μ_{eff} = 10 D, a value thirty times higher than a fully solvent exposed helix.

Case A of Fig. 5.9 corresponds to proteins such as myoglobin in which the helix spans the entire protein. In this case, solvent screening of the termini considerably lowers the helix dipole of the 26-residue helix to μ_{eff} = 2 D which is less than twice the value in water. In contrast, when only one terminus is solvent-exposed, corresponding to Case D of Fig. 5.9, the screening is much lower. An example of this geometry is the 27-residue helix in the MyoD DNA binding domain for which μ_{eff} = 23 D. Extending the analysis to α-helical linkers, the μ_{eff} of the 28 residue linker of calmodulin (see Fig. 5.11) equals 17 D. In this case, only the terminal residues are buried whereas all the non-terminal residues are solvent exposed.

5.8 Rules of Thumb to Determine the Effective Helix Dipole

The dipole moment of a helix is a macroscopic property that can be used to determine electrostatic interactions at distances that are large compared to the dipole length. Near-field interactions, as would be important in estimates of association free energies are not accurately represented by the dipole. Thus, the helix dipole is likely to be inadequate

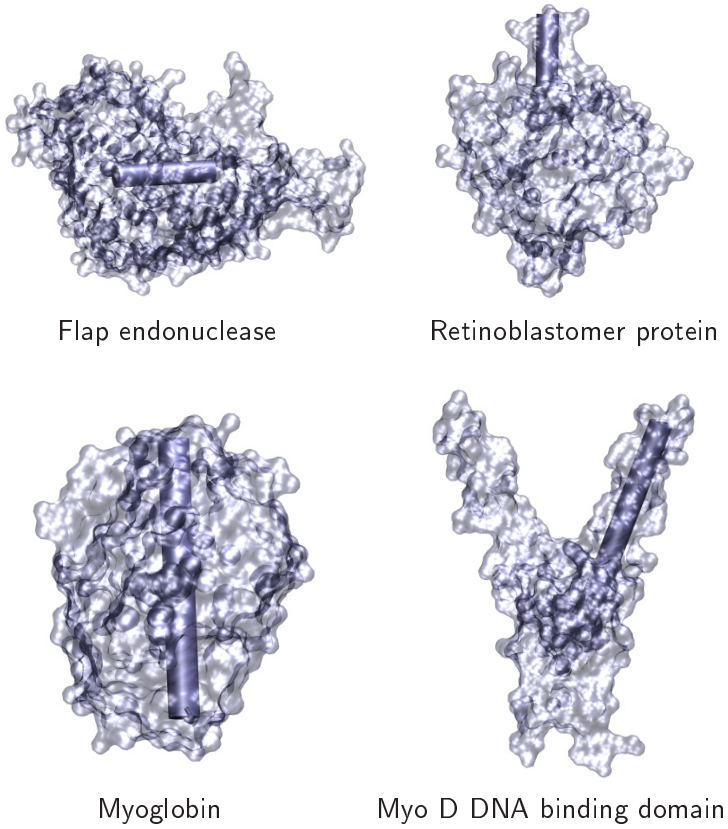

Figure 5.13: The cartoon representation of helices considered for calculations in the proteins 1.Flap endonuclease (1A76) 2.retinoblastomer protein (1GUX, chain A) 3.myoglobin(1A6N) 4.MyoD DNA binding domain (1MDY).

for describing effects such as phosphate-binding and anti-parallel helix motif stabilisation. However, a knowledge of the magnitude of the helix macrodipole is of fundamental interest in understanding protein biophysics and macromolecular electrostatics [134].

The strong vacuum macrodipole is counteracted in aqueous solvent by an electrostatic reaction field due to solvent reorganisation. The charge-fitting method presented here has allowed this reaction field and the resulting effective dipole moment μ_{eff} to be estimated for helices in various geometries in proteins and membranes. The calculated effective helix dipole is found to vary considerably in the different geometries and environments considered. For example, in contrast to the common assumption that the helix dipole increases with length, a decrease in helix dipole strength with increasing number of residues is seen in membrane-spanning peptides. Therefore, it is of prime importance to consider μ_{eff} before assigning any structural or functional role to the helix dipole.

The results indicate that the following three rules of thumb can be used to estimate a helix dipole moment. The rules derived are partly counterintuitive and do not always

correspond to common assumptions. Firstly, the dipole strength is determined principally by the positions of the helix termini relative to the aqueous phase. The amount of solvent-exposed surface area does not determine μ_{eff}. In cases where both the termini are buried (at least 7.5Å from the protein or membrane surface) then the dipole moment is close to that of a helix in a non-polar medium (*e.g.* $\epsilon = 2$) and can thus be strong. This rule holds even for helices that, apart from the termini, are solvent exposed, such as helices linking domains as in calmodulin and troponin. μ_{eff} for helices buried in homogeneous apolar media are high. Thus, the corresponding helix-helix interactions are likely to be strong, favouring the antiparallel helix arrangement. An example of this is the anti-parallel four-helix bundle motif in cytochrome b562 core [141].

The second rule of thumb is that if both termini are solvent-exposed then μ_{eff} will be small, and effectively the same as if the helix were fully solvated in aqueous solution ($\epsilon = 80$). This leads to the observation that μ_{eff} decreases with helix length for transmembrane helices, the opposite length dependence to that in vacuum. This has consequences for de-novo protein design, in that elongating an α-helix by adding residues with high helix propensity will not necessarily increase the helix dipole. Helices with terminal charges in the aqueous layer and the rest buried in a protein or bilayer core have dipole moments as if they were positioned in only the aqueous phase. Therefore, membrane spanning helices or helices spanning soluble proteins have small μ_{eff}. Exposure of termini, and therefore low μ_{eff}, may favour parallel arrangement of helix bundles. One example of this is the SNARE complex [142] which is a parallel four-helix bundle with the helices exposed to solvent. Similarly, transmembrane peptides with the termini extending into the aqueous layer can associate in parallel arrangements.

Thirdly, in cases where one helix terminus is solvent exposed and the other buried, asymmetric reaction field shielding can lead to a relatively high μ_{eff}. An interesting example of this is the relatively high dipole moment of the "thumb" of the DNA binding proteins which might contribute to binding to the DNA phosphate groups. Membrane helices with only one terminus exposed to the solvent will have relatively large values of μ_{eff}. This may be of functional importance in preventing proton transfer in aquaporin [123] and for providing ion affinity in the KcsA K^+ channel [89, 124] and the EcClc Cl^- channel [125].

Helices with large dipole moments which are not part of helix bundles would be unstable in the absence of other stabilising charged groups. Thus, it is energetically favourable to expose the helix termini to the aqueous layer and shield the helix dipole. For proteins such as myoglobin where the helices have negligibly small dipole moments it can therefore be postulated that in the folded state there is no contribution of the helix dipole to structural stability.

Helix dipoles are fundamental to protein electrostatics and may contribute to stability and function. Whether a particular helix has a significant dipole, depends on the solvent exposure of the helix termini. The rules of thumb established here allow qualitative estimation of helix dipole strength from graphical inspection of experimental structures.

Chapter 6

Transmembrane Helix-Helix Association

"The mystery of life begins with the intricate web of interactions...", Linked, Albert-László Barabási.

6.1 Association of Membrane Helices

Association of helices within a membrane is important for many biological processes such as signal transduction [143, 144], protein internalisation [145] and folding of polytopic membrane proteins [2, 27]. Dimerisation motifs such as the GxxxG motif, as described in Section 1.2.6, have been identified to drive association of transmembrane helices. Further, the dimerisation propensity and contribution of certain amino acids such as Gly, and Leu to dimerisation have been determined. However, the energetics of this process is poorly understood. The first step in understanding the energetics of the process is to determine the structures of the transmembrane dimers and oligomers and decompose the contributions of individual amino-acid residues to the association process.

Experimental data studying helix-helix association strongly focusses on glycophorin A since it provides a simple system for understanding the structural basis of helix association. Solution NMR studies in detergent micelles [42] as well as solid state NMR studies [70, 145] of the glycophorin A transmembrane dimer showed that it is a right-handed helix pair with a crossing angle of -35°. The two studies however differ on the positioning of the side chains at the interface and thereby disagreed on the ability of an interfacial Thr to form an inter-helical hydrogen bond. Extensive work on glycophorin A mutants [59, 71] has been done to map the residues important for dimerisation and to characterise the GxxxG dimerisation motif. Similar mutagenesis studies coupled with ultracentrifugation techniques have helped characterise the trend of dimerisation propensity of the glycophorin A transmem-

brane dimer [53]. Mutations of residues involved in the helix-helix interface have been shown to largely decrease dimerisation propensity. In contrast, mutation of certain residues on the opposite sides of the helix interface have been shown to even increase the dimerisation propensity in detergent micelles.

The helix-helix interface has also been mapped for proteins such as synaptobrevin [43], the erythropoietin receptor [44], the M2 channel [146] and phospholamban [147, 148]. Both left-handed (*e.g*, erythropoietin receptor) and right-handed (*e.g*, synaptobrevin) helix pairs have been identified. The difference between the right- and left-handed helix pairs arises from differences in their sequence since the crossing angle determines the residues interacting at the helix interface. For instance, based on the characterisation of the glycophorin A transmembrane dimer it had been postulated that the GxxxG motif facilitates right-handed packing but recently left-handed packing has also been reported for this motif [149]. At present, the general principles and factors governing the energetics of helix crossing angles are not clearly understood.

In this chapter, we propose a method for modelling transmembrane dimers using the five-slab model introduced in this thesis. The modelling methodology is based on sampling a large number of biologically-relevant dimer structures and screening against solvation free energy. We apply this method to two systems: glycophorin A and the erythropoietin receptor transmembrane dimer.

6.1.1 Molecular Modelling of Transmembrane Dimers

Determining the structure of associated membrane helices by molecular modelling has been attempted for various transmembrane dimers and oligomers such as glycophorin A [31, 71], synaptobrevin [61], the CD3 transmembrane domain [150], the erythropoietin receptor [151], the epidermal growth factor receptor [152, 153] and phospholamban [147, 153]. The modelling has been usually carried out in vacuum [61, 71, 152] without considering the membrane environment or in the three-slab membrane model explained in Section 1.3.2 [31]. Mutagenesis data has also been extensively used to refine the modelled structures and to distinguish between probable models. The modelled structures were found to compare well with subsequent experimental mutagenesis data. An analytical residue-scale scoring function based on atomistic van der Waals interaction, has been also proposed for modelling membrane association of helices [154]. The method was used to model glycophorin A, the epidermal growth factor receptor and phospholamban and has been shown to be compatible with existing experimental data.

The *in vacuo* models of the transmembrane dimers do not account for effects such as

anisotropic charge screening. The solvent effects are introduced in the current modelling methodology by using the five-slab membrane model. Introduction of the head-group energetics as well as charge screening by the aqueous medium is expected to improve the accuracy of the modelling method. Further, the exhaustive search in conformational space coupled with scoring based on the solvation free energy of the modelled dimer structures increases computational cost, but paves the way towards an understanding of the energetics of the association process.

6.1.2 Glycophorin A

Glycophorin A is a bitopic integral membrane sialoglycoprotein that is present in the red blood cells. The presence of experimentally-determined structures [42, 70] makes it a model transmembrane dimer to validate new modelling methods [35, 153]. Many site-specific mutagenesis studies measuring the propensity to dimerise have been carried out to probe the glycophorin A transmembrane dimerisation [53, 59, 71, 155].

In this chapter, the glycophorin A transmembrane dimer is modelled using the methodology described in Section 6.2. The modelled structure is compared to existing experimental structures to validate the modelling methodology. Further, the contribution of each amino-acid residue to the free energy of association is calculated. The calculations show that a negative, favourable contribution is attributed to the residues implicated to contribute to dimerisation in the mutagenesis studies.

6.1.3 Erythropoietin Receptor

The erythropoietin receptor, EpoR is a single membrane spanning protein belonging to the hematopoietic cytokine receptor family that has been implicated in maturation of immature red blood cells. Structural evidence suggests that the transmembrane domain has a strong potential to self-interact and therefore, exists as an associated homodimer in membranes [156]. These pre-associated erythropoietin receptor oligomers are not active and binding of erythropoietin is required to induce signalling [156, 157]

Asparagine-scanning mutagenesis studies have proposed that the EpoR transmembrane dimer exhibits a positive crossing angle [44]. These studies have identified mutants such as T242N and L241N which exhibit increased self-interaction compared to the EpoR dimer [44]. However, one should be cautious in assigning these residues to the helix-helix interface in the wild-type EpoR dimer since the introduced bulky amino acid residue may also introduce structural changes.

In vacuo modelling of the EpoR transmembrane dimer has pointed towards a zero crossing angle [151]. Two low-energy structures, differing in the residues occupying the helix-helix interface were reported. The helix-helix interface of the first structure was lined by polar Ser and Thr residues and that of the other by hydrophobic Leu and Ile residues.

The cell surface expression of EpoR is low, but it is enriched in lipid rafts. It has been proposed that localisation of the EpoR in lipid rafts is crucial for amplifying specific signal transduction cascades. Experimental evidence shows that the mutant T242N is impaired in lipid raft localisation thus suggesting a critical influence of Thr^{242} on lipid raft localisation. Initial analysis suggested that the residue Thr^{242} of the wild-type EpoR influences raft sorting by interacting with cholesterol *via* its OH group. However, the cellular distribution of the mutant HA-T242A-GFP5 lacking the OH group at position 242 was comparable to the HA-EpoR-GFP5, thus requiring the hypothesis to be altered. Other possible molecular mechanisms accounting for the impaired raft localisation may be attributed to any structural changes in the dimer conformation introduced by the mutation T242N. Since no experimentally-determined structures exist for the erythropoietin receptor and its mutants, molecular modelling is required.

In the second part of this chapter, the transmembrane dimer of the erythropoietin receptor is modelled. In addition, the structure of the transmembrane dimer of the mutant T242N is modelled to elucidate possible mechanisms that effect lipid raft localisation in the erythropoietin receptor. The mutants A245N, T242A and L241N that localise in lipid rafts are also modelled as control sets.

6.2 Modelling Methodology

The structures of the transmembrane dimers of glycophorin A and erythropoietin receptor are modelled in the five-slab membrane environment. Dimer structures are generated by sampling the relevant conformational space. The ensemble of structures is scored against the solvation free energy, ΔG_{solv}, defined as the free energy change for transferring the molecule from vacuum to its position in the membrane [97]. The equilibrium structure is then determined by the minimum of ΔG_{solv}.

ΔG_{solv} is calculated using all-atom models of the transmembrane dimer together with the five-slab implicit membrane model. The electrostatic solvation energy, ΔG_{elec} and cost of cavity formation, ΔG_{np} are summed to compute ΔG_{solv}, as described in detail in Chapter 3. ΔG_{elec} is calculated as the difference between the electrostatic energies in vacuum and in the membrane environment using the Poisson-Boltzmann equation. ΔG_{np} is estimated from the solvent-accessible surface area using Eq. 3.3.

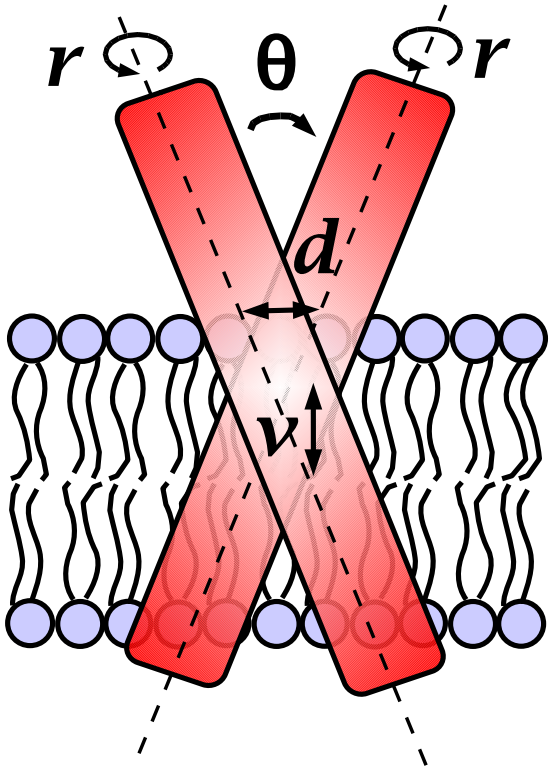

Figure 6.1: Parameters describing the orientation of a helix-helix pair in the membrane: the inter-helical distance, d, the crossing angle, θ, the membrane insertion, v and helix rotation, r_i. The parameters v_i and r are sampled for both helices independently. d is defined as the minimum distance between the two helix axes and θ the angle between them. v is the distance of the centre of mass of each helix from the centre of the membrane. r is the angle of rotation of each helix about its axis.

The relevant conformational space was sampled by a thorough search of structural parameters of a membrane dimer. For a helix-helix pair in a membrane, the structure is described by six parameters: the inter-helical distance, d, the crossing angle, θ, the membrane insertions, v_i and helix rotation angles, r_i (see Fig. 6.1). d is defined as the minimum distance between the two helix axes and θ the angle between them. v_i is the distance of the centre of mass of helix i from the centre of the membrane. r_i is the angle of rotation of helix i about its axis.

6.2.1 Helix-Pair Construction

The back bone structure of the transmembrane segment of glycophorin A was taken from the PDB (1AFO). Values were assigned to d, r_i, θ and v_i systematically so as to sample the relevant conformational space. θ was varied from $+60°$ to $-60°$ to sample both left-handed and right-handed helix-pairs. r_1 and r_2 were varied between $0°$ and $360°$ to sample

all possible helix interfaces. d was varied from 5 Å to 10 Å and v_1 and v_2 between 0 Å and ±5 Å. The side chains were positioned using the software SCWRL [120] (described in Section 4.2) for every generated dimer conformation and assigned after the dimer parameter assignation. The dimers were energy minimised in a dielectric medium of ϵ=2 using the CHARMM potential [99] (version 30b1) with 1000 steps of Steepest Descent followed by and 1000 steps of Newton-Raphson minimisation with 1kcal/mol harmonic constraints. The structures were then incorporated in the five-slab membrane model and scored based on the value of ΔG_{solv}.

The transmembrane segment of EpoR [sequence:LILTLSLILVLISLLLTVLALLS] was modelled as an ideal α-helix with ψ=-47° and ϕ=-57° [2]. Values were then assigned to sampling parameters as above. The side chains were again positioned using the software SCWRL. The dimers were energy minimised as above in a dielectric medium of ϵ=2 using the CHARMM potential. The structures were then incorporated in the five-slab membrane model and scored based on the value of ΔG_{solv}. The T242N, T242A, L241N and A245N mutants were modelled by the same method. The volume of the transmembrane dimers were calculated with a probe radius of 2.2 Å. The probe radius used was larger than 1.4 Å commonly used for the radius of a water molecule used as a probe radius for soluble proteins. For membrane helix-helix pairs, a larger radius is used since it approximates the radius of a methyl group [112] and also fills in the internal cavities in the helix-helix interface.

The modelling approach proposed here is based on an improved representation of the energetics of helix-pair association in membranes. Further improvement could be achieved by the introducing backbone and side-chain flexibility. Such improvements are however computationally expensive and prohibited by present computational resources.

6.2.2 Decomposition of Free Energy of Dimerisation

The free energy of dimerisation, $\Delta\Delta G_{dimerisation}$ is defined as the difference in total energy between the bound (ΔG_{bound}) and the unbound ($\Delta G_{unbound}$) state. Here, we focus only on the electrostatic component on the free energy and we obtain:

$$\Delta\Delta G^{elec}_{dimerisation} = \Delta G^{elec}_{bound} - \Delta G^{elec}_{unbound} \tag{6.1}$$

The bound state refers to the modelled, energy-minimised structure of the transmembrane dimer. The unbound state refers to each helix of the helix-pair, in the same conformation in the absence of the other.

The electrostatic energy of each state is obtained from the electrostatic potential, $\phi(r_i)$ and partial charge, q_i at every atom centre i and is given by $\sum_i q_i \phi(r_i)$. Therefore the electrostatic free energy of dimerisation is given by:

$$\Delta\Delta G^{elec}_{dimerisation} = \sum_{i \in C} \frac{1}{2} q_i \phi^C(r_i) - \left(\sum_{i \in M1} \frac{1}{2} q_i \phi^{M1}(r_i) + \sum_{i \in M2} \frac{1}{2} q_i \phi^{M2}(r_i) \right) \tag{6.2}$$

where, C is the set of atoms in the complex and M1 and M2 the set of atoms in the monomers 1 and 2.

Further, the total potential at any point may be expressed as the sum of contributions from individual charges or groups of charges. Thus, for a group of charges, y in the state, z

$$\phi^{y \in z}(r_i) = \sum_{j \in y} \phi^{j \in z}(r_i) \tag{6.3}$$

The contribution of a group j to the binding free energy was calculated as

$$\Delta\Delta G^{contrib}_{j} = \sum_{i \in C} \frac{1}{2} q_i \phi^{j \in y}(r_i) - \left(\sum_{i \in M1} \frac{1}{2} q_i \phi^{j \in M1}(r_i) + \sum_{i \in M2} \frac{1}{2} q_i \phi^{j \in M2}(r_i) \right) \tag{6.4}$$

If the group j is in monomer 1, $M1$, then the electrostatic potential due to group j at the atoms of monomer 2, $M2$ will be zero.

The above electrostatic potentials are calculated by solving the Poisson-Boltzmann equation on a three-dimensional grid with the standard finite-difference method. For the three systems - the two monomers and the dimer, the same three-dimensional grid and the same three-dimensional coordinates were used. Using the same grid and atomic positions allows the artificial contributions to the electrostatic potential arising from the grid to be subtracted out exactly. The method has been applied before to calculate contributions of residues to ligand binding [158, 159].

The side-chain atoms of each amino-acid residue are considered to be a group j in Eq. 6.4. $\Delta\Delta G^{contrib}_{j}$ was calculated for each residue in the transmembrane segment of glycophorin A as the difference of its contribution in the dimer and in the monomer. Further, to compare with experimental data $\Delta\Delta G^{contrib}_{j}$ of a residue from both helices were added.

The method provides a simple way to estimate the contribution of each amino-acid residue to the dimerisation free energy. The contribution of each group is measured directly and not *via* indirect mutagenesis data.

6.3 Helix Association in Glycophorin A

The glycophorin A transmembrane dimer was modelled as described in Section 6.2. The modelled structure is a right-handed helix pair with a crossing angle of θ = -30°. Fig. 6.2 illustrates the modelled transmembrane dimer of glycophorin A. The helix interface is lined by the residues: Ile^{15}, Gly^{18}, Val^{19}, Gly^{22} and Thr^{26}. Though Thr^{26} is present at the helix-helix interface, it is not involved in hydrogen bonding. The helix-interface mapped by the theoretical study and the crossing angle is similar to that of the experimentally-determined structure. For instance, the residues implicated in the GxxxG motif - Gly^{18} and Gly^{22} are present in the helical interface and mediate interactions between the two helices.

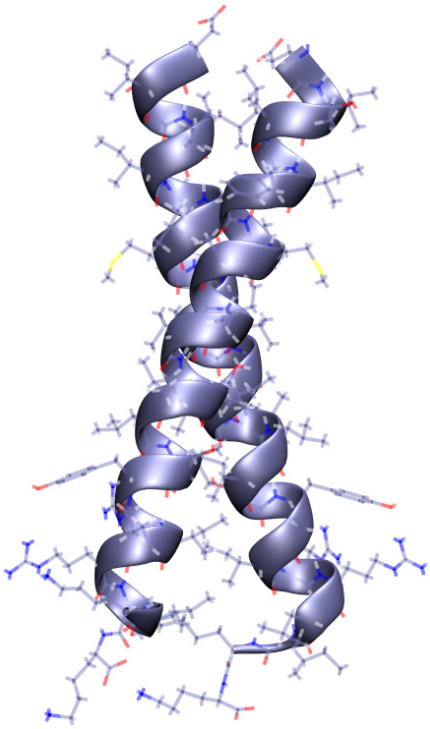

Figure 6.2: The lowest-energy wild-type EpoR transmembrane dimer structure. The dimer has a negative crossing angle of θ = -30°, *i.e.* it is a right-handed helix-helix pair.

The inter-helical distance in the theoretical model is increased by 1 Å compared to the NMR

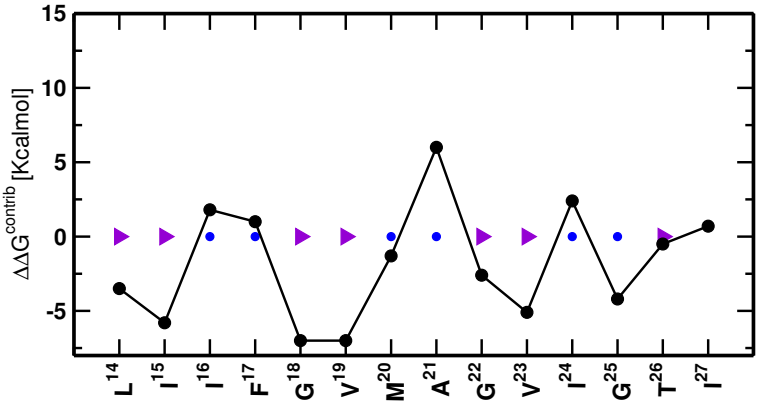

Figure 6.3: $\Delta\Delta G^{contrib}$ of the transmembrane residues of glycophorin A dimer calculated as described in Sec. 6.2.2. The residues are marked with the one-letter code for clarity.

structure in solution. The closer packing of the helices in the detergent NMR structure can be attributed to the fact that the detergent environment is more polar than the five-slab membrane model which better represents the bio-membrane. The modelled structure of glycophorin A captures the essential features of the transmembrane dimer, thus validating the modelling method used here.

The modelled dimer structure is used to calculate the contribution of each residue to the dimerisation free energy. $\Delta\Delta G^{contrib}$ for the transmembrane amino-acid residues, calculated as described in Section 6.2.2, is plotted in Fig. 6.3. It is seen that the electrostatic contribution of each residue to the dimerisation process matches closely the experimental dimerisation propensities. The residues implicated in the dimerisation motif are seen to have a negative $\Delta\Delta G^{contrib}$. The other residues have a zero or positive contribution. The residues Gly^{18} and Gly^{22} which are implicated in the GxxxG motif have a negative contribution. It is obvious that a larger side chain at this position would destabilise the dimer not just electrostatically but also sterically. In the modelled structure, Thr^{26} is not involved in hydrogen bonding and has a contribution close to zero.

Four residues - Ile^{16}, Phe^{17}, Ala^{21} and Ile^{24} have a positive $\Delta\Delta G^{contrib}$ and hence contribute unfavourably to the dimerisation process. This behaviour is reflected in mutagenesis studies, where substitution by Ala at positions 16, 17 and 24 resulted in mutant dimers having more favourable free energies of dimerisation compared to the wild-type dimer [53]. The implication of these destabilising effects, which is indicated to be of electrostatic nature in this calculation, is not clearly understood and may be due to the presence of long range interactions as has been recently proposed [155, 160].

A highly destabilising contribution was calculated for Ala^{21}. The destabilising nature of

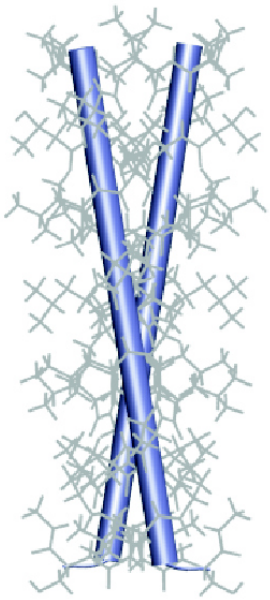

Figure 6.4: The structure of the EpoR transmembrane dimer shows a positive crossing angle $\sim 20°$.

Ala^{21} can not of course be revealed by experimental alanine-scanning mutagenesis studies. The preference of Ala^{21} for the monomeric state could be due to the fact that it lies at the helix-helix hinge directly below Phe^{17}. In the monomeric state, an energetically favourable interaction of the methyl hydrogens with the phenyl ring of Phe^{17} is possible, thereby increasing the preference for the monomeric state.

6.4 Molecular Modelling of the Erythropoietin Receptor

The structure of the wild-type EpoR transmembrane dimer is modelled in atomic details in the five-slab implicit membrane model. The mutant T242N which is impaired in lipid raft localisation is also modelled. In addition, three mutants T242A, A245N and L241N are modelled as control sets.

The modelled EpoR dimer (Fig. 6.4) is a left-handed helix-helix pair, *i.e.* it has a positive crossing angle of $\theta \sim 20°$. A positive crossing angle has been earlier predicted by mutagenesis studies [44].

The models of EpoR generated by the modelling methodology indicated that two low-energy interfaces are possible which are shown in in Fig. 6.5. Both the modelled interfaces are indi-

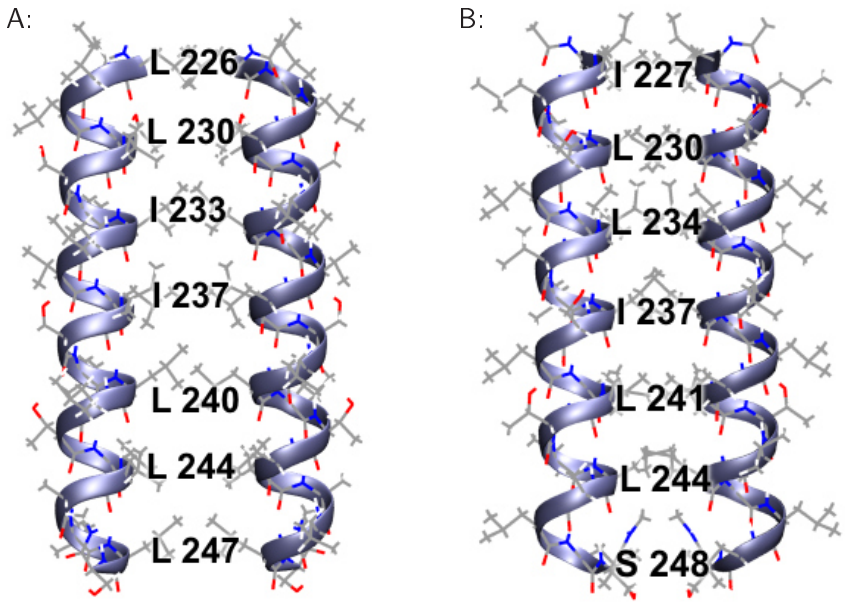

Figure 6.5: The two lowest-energy wild-type EpoR transmembrane dimer structures. Both interface **A** and **B** are lined with non-polar long chain residues. In **A**, the inter helical distance, $d = 12$ Å and in **B** $d = 10$ Å. The interfacial residues in interface **A** are Leu^{226}, Leu^{230}, Ile^{233}, Ile^{237}, Leu^{240} and Leu^{244}. The helix-helix interface in **B** consists of Ile^{227}, Leu^{230}, Leu^{234}, Ile^{237}, Leu^{241}, Leu^{244} and Ser^{248}.

cated to be functionally active by experimental evidence [151]. Previous *in vacuo* modelling studies [151] have also identified interface 2 (Fig. 6.5B) but not interface 1 (Fig. 6.5A). A different interface lined with Ser and Thr residues was identified as being energetically favourable in those studies.

The first low-energy interface is shown in Fig. 6.5. The inter helical distance is $d = 12$ Å. The residues Leu^{226}, Leu^{230}, Ile^{233}, Ile^{237}, Leu^{240}, Leu^{244} occupy the interfacial positions and Thr^{242} points outwards. In the second structure, the inter helical distance is $d = 10$ Å. The helix-helix interface is occupied by Ile^{227}, Leu^{230}, Leu^{234}, Ile^{237}, Leu^{241} and Leu^{244}. Surprisingly, the four polar residues, Thr^{229}, Ser^{231}, Ser^{238} and Thr^{242} are not hydrogen bonded in both structures. However, except for Ser^{238}, these residues are situated in the polar head-group region or the core head-group interface and thus do not evoke large energy costs.

The water-accessible surface of the structures modelled is shown in Fig. 6.6. Both the structures are well packed and have a volume of 160 $Å^3$ and 155 $Å^3$ respectively.

The modelled T242N mutant dimer also exhibited two low-energy interfaces. The two structures are shown in Fig. 6.8. The mutant is also left-handed with a crossing angle of $\theta \sim 20°$. In both the structures, the helices are further apart, $d = 14$ Å compared to the

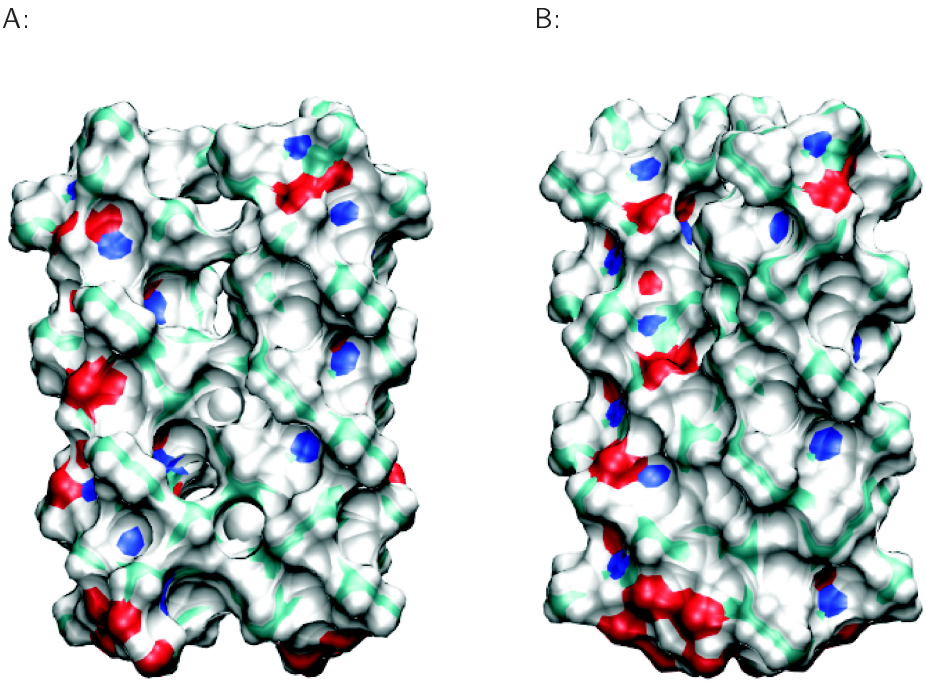

Figure 6.6: The water-accessible surface area of the two low-energy structures of the EpoR transmembrane dimer. Both the structures are well packed but differ in inter-helical distances of $d = 12$ Å and $d = 10$ Å respectively.

wild-type dimer. The first structure (Fig. 6.8 A) represents an unusual orientation with just three amino-acid residues Leu^{228}, Val^{235} and Asn^{242} lining the helix-helix interface. Asn^{242} is involved in hydrogen bonding and stabilises the conformer in the absence of favourable side-chain packing. In the second structure (Fig. 6.8 B) the helix interface consists of a larger number of residues: Ile^{227}, Ser^{231}, Leu^{234}, Ser^{238}, Leu^{241} and Ala^{245}. Though Asn^{242} is not directly involved in the helix-helix interface it bends back into the interface. The packing of the bulky side chain group in the helical interface accounts for the increased inter-helical distance.

The other mutants modelled, T242A, L241N and A245N were similar to the wild-type, especially with respect to the inter-helical distance. This is not surprising since point mutations do not usually introduce large conformational changes. The modelled T242A mutant has a crossing angle $\theta \sim 10^{\circ}$ and an inter-helical distance of $d = 10$Å. The lowest-energy conformers of the other Asn mutants modelled also had small inter-helical distances of 10Å. The lowest energy structures of A245N does not involve hydrogen bonding *via* the Asn^{245} and the Asn points away from the helical interface in all structures. The L241N mutant represents a case where two dissimilar conformers have similar energies. In the first structure, $d = 10$Å and the Asn^{241} points away from the helical interface. The inter-

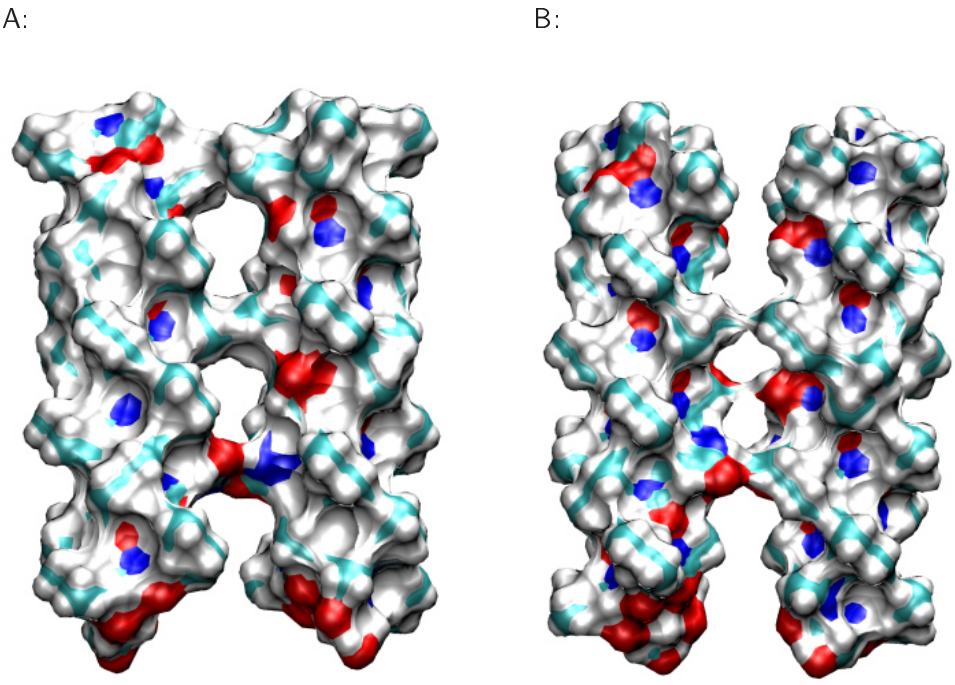

Figure 6.7: The water-accessible surface area of the two low-energy structures of the T242N mutant transmembrane dimer. Both structures are poorly packed and have large inter-helical distance of $d = 14$ Å.

helical distance in the second conformer is 14Å and involves Asn^{241} mediated inter-helical hydrogen bonding. These mutants were modelled as control sets and close packing of these mutant dimers confirmed the importance of close packing in lipid raft localisation.

Lipid Raft Sorting

The erythropoietin transmembrane dimer is seen to be located in lipid rafts, whereas a single mutation impairs lipid raft localisation of the T242N mutant. The inefficient receptor sorting of the T242N mutant into lipid rafts may stem from the structural differences of the wild-type EpoR transmembrane dimer and the T242N transmembrane dimer. The main difference between the two modelled transmembrane dimer structures is that the T242N dimer exhibits a bulkier structure that prevents close packing of the transmembrane helices in contrast to the EpoR dimer. Since lipid raft sorting is thought to be mediated by self assembly and clustering of lipid molecules around a given peptide [13], the bulky T242N dimer may disrupt tight packing of the neighbouring lipid molecules, thereby inhibiting lipid raft sorting.

The epidermal growth factor receptor family member ErbB2 that has also been observed

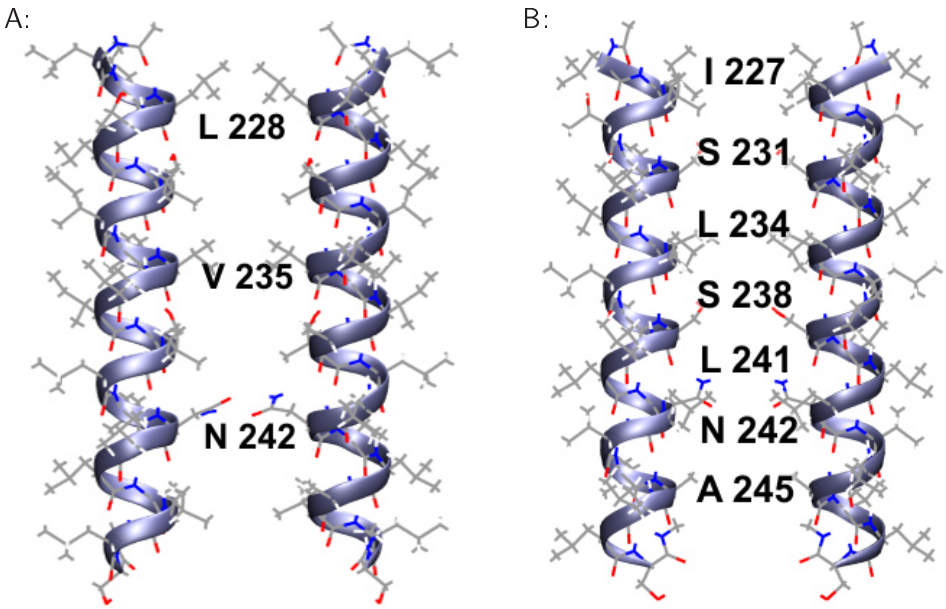

Figure 6.8: The two low-energy structures of the T241N transmembrane dimer. Both A and B are left-handed and have a crossing angle of $\sim 20°$. The structures differ in the helix interface but $d = 14$ Å in both cases.

to partition into lipid rafts [161] exhibits close packing of the transmembrane dimer with an inter-helical distance of approximately 8 Å [152]. The Gly residues present at the ErbB2 transmembrane interface imply a shorter inter-helical distance compared to the EpoR transmembrane interface that is dominated by bulkier residues such as Leu (resulting in a larger inter-helical distance). Therefore, close packing of both the EpoR and ErbB2 transmembrane dimers may correlate to their ability to be sorted into lipid rafts. At present, it is not possible to dissect if close packing of the EpoR transmembrane dimer is a targeting signal or a means of the receptor to stabilise lipid raft association, as there might be additional motifs in the EpoR sequence contributing to lipid raft sorting. However, the observation that an increase in inter-helical distance abrogated lipid raft localisation of the EpoR supports the importance of close packing of the transmembrane dimer. Close packing of the transmembrane segments could therefore be a crucial mediator for lipid raft localisation of receptor proteins.

6.5 A Critical View of the Modelling Method

The modelling methodology developed as a part of this thesis allows an atomic-detailed modelling of transmembrane dimer structures. In this method, an exhaustive conformational search is performed followed by a scoring based on the solvation free energy of the

dimer. The modelling method is validated by comparing to known structures of the glycophorin A transmembrane dimer. The modelled structure shows a favourable contribution to the free energy of dimerisation for the residues implicated in dimerisation by alanine-scanning experiments. Since only the electrostatic component of the dimerisation free energy was considered, it is implied that the dimerisation process is largely electrostatics driven.

All-atom models of the erythropoietin receptor and its mutant transmembrane dimers have shown that crucial conformational changes may accompany point mutations and affect the function of the membrane proteins. In the present case, the structural changes introduced by the mutant T242N are not quantitatively large since the inter-helical distance increases by just 2 Å. However, the increased bulkiness might play a crucial role in altering the biological properties of the erythropoietin receptor. The inefficient receptor sorting of the HA-T242N-GFP5 into lipid rafts could arise from the increased inter-helical distance and volume of the mutant T242N dimer compared to the wild-type EpoR dimer. Since lipid raft sorting is mediated by self assembly and clustering of lipid molecules around a given peptide [13], the bulky and loosely-packed T242N dimer may disrupt tight packing of the neighbouring lipid molecules, thereby inhibiting lipid raft sorting.

The modelling method proposed here provides a first step towards membrane protein structure prediction and understanding the energetics of helix-helix association. The current method does not consider the van der Waals type interactions of the lipid molecules with the peptides. However, it is assumed that within the membrane core, acyl-chain interactions between peptide-peptide and lipid-peptide are likely to be energetically similar. Since the peptide-peptide van der Waals type interactions are not included in the scoring function of the modelling methodology, the lipid-peptide interactions have also been neglected. At present, due to computational costs, the method is limited to smaller systems such as those investigated in the present chapter. However, coupling the method to efficient search algorithms to improve the sampling step, promises to allow the investigation of larger oligomeric complexes such as phospholamban. Such detailed, energy-based approaches are necessary to understand the energetics of complex formation and to model high-resolution structures.

Chapter 7

Concluding Remarks

"This mystery is considered insoluble, for the very reason which should cause it to be regarded as easy of solution", The Murders in Rue Morgue, Edgar A. Poe.

Several aspects of membrane protein architecture have been studied in this thesis, providing a detailed insight into the structure and energetics of membrane-bound peptides. In particular, the orientation, electrostatics and association of membrane peptides have been analysed. A large number of peptides were investigated, among them glycophorin A, melittin and WALP that have become prototypes for studies on membrane peptides. This chapter summarises the key findings of the thesis. The limitations of the methods used are also discussed. The results presented here are a step towards a better understanding of the structural organisation of membrane proteins and its relation to function. New questions raised during this work and methods to address these problems are discussed in the last part of this chapter.

7.1 A Critical Look at the Five-Slab Membrane Model

A five-slab continuum membrane model that distinguishes between the membrane core, head-group region and the aqueous layer is introduced in the thesis. The membrane model is found to be adequate to describe the energetics of peptide analogues, membrane peptides as well as dimers of transmembrane peptides. The inclusion of the head-group region is seen to influence the energetics of peptides. For instance, indole, a tryptophan side-chain analogue, is seen to partition in the head-group region of the model in accordance with experimental results. Furthermore, including head-group energetics is demonstrated to be necessary for accurate predictions of membrane-peptide orientations. The tilt angle of the

20-mer WALP is significantly overestimated by the three-slab models, but is well reproduced in the five-slab model due to presence of the head-group region. The polar head-group region region allows the tryptophan residues to anchor the peptide, thus reducing its tilt angle. Therefore, accounting for the head-group region provides an important step towards a physically realistic, yet computationally efficient membrane representation.

In addition to the head-group energetics, the five-slab model includes the non-polar interactions between the peptide and the solute molecule implicitly. The non-polar effects in the aqueous layer are introduced by an empirical two parameter description of the non-polar solvation energy, ΔG_{np} derived from the partitioning of alkanes between liquid alkane and water [112]. The empirical model used here calculates ΔG_{np} from the solvent accessible surface area rather than the excluded solvent volume, which would be more accurate from a theoretical point of view. In general, however, such methods have been shown to be accurate at a coarse-grain level and parametrisation on the system under consideration (as in the present case) increases accuracy of the calculation [162].

The membrane model makes certain simplifications and assumptions which one needs to bear in mind while considering the results. For instance, the model neglects the van der Waals type interactions of the lipid molecules with the peptides. Such interactions play an important role during helix association, since the environment of the monomer and the associated complex would be different. The model assumes that within the membrane core acyl-chain interactions between peptide-peptide and lipid-peptide are energetically similar. Since the peptide-peptide van der Waals type interactions are not included in the scoring function of the modelling methodology, the lipid-peptide interactions have also been neglected. These interactions may be incorporated by including dummy atoms or using an all-atom representation of the lipid bilayer and could improve the accuracy of the model. For the head-group region, neither the van der Waals type interactions nor the cost of cavity formation were included. No experimental or theoretical estimates of ΔG_{np} for the head-group region are available. However, incorporating a term similar to the aqueous layer but smaller in magnitude does not affect the location of the free energy minimum for control calculations performed on polyalanine peptides.

The five-slab continuum membrane model used in this thesis is demonstrated to provide a computationally efficient and accurate environment to model membrane peptide orientation and association. Though an all-atom representation of the membrane models the energetics of the lipid-peptide systems more accurately, especially for modelling association processes, limited computational resources make it non-tractable at present. The continuum model is of course essential for certain problems such as calculating the electrostatic potentials of the membrane helices in the presence of mobile ions in the bulk aqueous layer.

7.2 Understanding the Energetics of Helical Peptide Orientation in Membranes

The energetics of membrane-peptide orientation is analysed in detail in this thesis. The heterogeneous environment experienced by a single peptide in the membrane causes it to orient and tilt to a characteristic angle. It is seen that a solvent reaction field countering the charges in or close to a region of high polarity favours dipole orientations that allow maximum charge screening. Thus, peptide-dipole groups that approach the dielectric boundaries of the system favour an orientation perpendicular to the membrane plane. Furthermore, the solvent reaction field is shown to determine the membrane orientation of blocked polyalanine helices. The orientation of the shorter polyalanine and WALP peptides is attributed to the screening of the peptide backbone charges. Estimates of the magnitude of this screening by calculating the effective dipole moment of the helices are also reported in the thesis. The screening is shown to depend on the proximity and geometry of the aqueous phase and its relevance to membrane protein topology is discussed.

Another important factor determining peptide tilting is the cost of cavity formation in the solvent region. The longer peptides which are exposed to the solvent tend to tilt to minimise the water-accessible surface area. The calculations show that the cavity term influences the steepness of the tilt-angle potential such that the longer peptides undergo smaller tilt-angle fluctuations than the shorter ones. A further point of interest is the experimental finding that the 19-mer WALP peptide has a non-zero tilt angle but nevertheless orients roughly perpendicularly to the membrane plane [40]. The present analysis suggests that the relatively large fluctuations of the 19-mer WALP peptide tilt angle may have led to this observation. Experimentally estimating fluctuations and tilt angles in synthetic peptides such as WALP peptides will help understand membrane protein architecture.

The decomposition of solvation free energy allows the rationalisation of the dependence of the peptide tilt angle on length of peptides. A systematic study of the length-dependent tilting of the synthetic peptides is carried out in the thesis. Bilayer thinning that has been implicated to compensate the hydrophobic mismatch between the peptide length and the bilayer thickness is shown not to affect the tilt angles of the shorter peptides.

The experimental tilt angle of glycophorin A is well reproduced by the five-slab model. It was shown that side-chain polarity accounts for the tilting since it incorporates a larger number of hydrophobic side chains into the non-polar membrane core region while allowing the flanking polar residues to interact with the head-group region. Further, the surface aligned orientation of melittin in the absence of a membrane potential is reproduced by the model. Two different surface-aligned orientations of melittin are distinguished. The

preference of one is shown to depend on the energetically-favourable interaction of the hydrophobic side chains with the membrane core in one which is not possible in the other.

7.3 Macromolecular Electrostatics of Membrane Helices

In membrane proteins, the α-helix dipole has been implicated in protein stability [73, 129] and function [89, 123, 125]. Therefore, despite being a macroscopic property that neglects near-field effects, the helix dipole provides insights into membrane-protein architecture and function. The effective dipole moment, μ_{eff} of various membrane helices is reported in this thesis and has been found to vary considerably in the different geometries and environments considered. Rules of thumb have been established allowing qualitative estimation of helix dipole strength from graphical inspection of experimental structures.

In contrast to the common assumption that the helix dipole increases with peptide length, a decrease in helix dipole strength with increasing number of residues is seen in transmembrane peptides. For single helices, the helix dipole is seen to increase with tilt angle. Extending the analysis to various geometries and environments in soluble proteins has enabled the establishment of the rules of thumb discussed in Chapter 5. The dipole strength is seen to be determined principally by the positions of the helix termini relative to the aqueous phase, *i.e*, the dipole is substantial if one or both termini are embedded in a non-polar region but is screened out if both termini approach the aqueous phase.

As a consequence of the solvent screening, transmembrane peptides with both termini extending into the aqueous layer are seen to have low dipole moments. An implication of the screening out of the helix dipoles is that the helices can associate in parallel arrangements. The proposed screening out of the helix dipole for longer helices, may explain the finding that the orientation of the wild-type Pf3 coat protein is dependent on the membrane potential, whereas the longer mutant peptides are oriented independent of membrane potential [163]. The observations may be rationalised in terms of solvent screening of the helix dipole in the longer mutant peptides such that both helix termini can be translocated across the membrane bilayer independent of the electrochemical membrane potential.

Solvent screening of the helix dipole is shown to be reduced even if just one of the helix termini is buried in the membrane core. Membrane helices with one terminus exposed to the solvent are shown to have relatively large values of μ_{eff}. The large value of the helix dipole may be of functional importance in preventing proton transfer in aquaporin [123] and for providing ion affinity in the KcsA K^+ channel [89, 124] and the EcClC Cl^- channel [125].

Interestingly, it is seen that the solvent reaction field plays a crucial role in membrane-

protein energetics - both in determining peptide orientations and in modulating the screening of the helix dipole. Thus, although the membrane helices interact primarily with the bilayer, the aqueous layer influences its properties substantially.

7.4 Modelling α-Helical Membrane Dimers

The next step towards understanding the energetics of membrane proteins is to analyse the association of single helices in the membrane. The structures of glycophorin A, the erythropoietin receptor transmembrane dimer and its mutants are modelled and seen to compare well with experimental mutagenesis data. The five-slab membrane model to represent the environment of these transmembrane dimers has led to new insights in the association between helix-pairs.

The transmembrane dimer of glycophorin A is modelled in the five-slab membrane model and shown to reproduce the experimentally-determined helix interface. Decomposition of the free energy of dimerisation shows a favourable contribution of the residues implicated in mutagenesis experiments to contribute to dimerisation . The correspondence of this calculated contribution to the alanine-scanning mutagenesis experiments is only qualitative since conformational changes accompanying point substitutions and entropy changes on dimerisation are neglected. However, the calculations provide a stepping stone towards understanding the energetics of helix-helix association processes.

The transmembrane dimer structures of the erythropoietin receptor and several of its point mutants are also modelled in the five-slab membrane model. Analysis of the modelled structures led to the postulation that close packing may be a pre-requisite for lipid-raft localisation. The modelled structure of the T242N mutant dimer exhibits a bulkier and less well packed structure than the wild-type dimer which may be responsible for the mutant to be impaired in lipid raft localisation.

7.5 Outlook

The questions addressed in this thesis, such as the energetics of membrane proteins and membrane-protein interactions, are of prime biological importance and relevant for membrane protein structure determination. Addressing the further questions raised during this work and improving the models promises to further the understanding of membrane-protein architecture. The first step in this direction is a broader application of the membrane model and the methods presented here. For instance, extension of the methods to other synthetic

peptides and β-barrel membrane proteins will further the understanding of membrane protein topology. Incorporation of other parameters such as membrane potential will increase the biological relevance of the model. An open question that remains in this field is the kinetics of helix-helix association. At a functional level, these associated helix-pairs may localise in lipid rafts and initiate down-stream signalling. Characterising these lipid raft assemblies will lead to an understanding of the partitioning of proteins in and out of these assemblies.

7.5.1 Peptide Orientations in the Membrane

The focus of this thesis is on single α-helical membrane peptides. Extensions to larger membrane proteins are however straightforward. Using the tools presented here, similar analyses can be performed on larger α-helical and β-sheet proteins. For instance, the interaction of the aromatic amino-acid residues with the head-group region in porins has been often discussed [164] and may be studied using the five-slab membrane model. A related question is the identification of the residues interacting with the bilayer interface and the depth of membrane insertion of the transmembrane segment. The membrane model can be used to identify these interfacial residues by sampling the relevant conformational space. An extension of this method would be to model the aggregation of large protein complexes such as the light harvesting complex and the respiratory chain.

In the first part of this thesis, the energetic factors determining helix orientations were quantitatively assessed. The solvent reaction field and cost of cavity formation is seen to determine membrane peptide orientation for simple peptides such as polyalanine and WALPs. However, side-chain polarity is the dominant driving force for the tilting of the more complex peptides such as glycophorin A and melittin. Further research using both experimental and theoretical approaches, is required to understand the energetics of tilting of the more complex peptides and the role of each side chain in determining membrane protein orientation. For instance, the role of non-polar residues not studied in this thesis, such as Ile and Val could be analysed in detail using this model. The 'snorkelling effect' of the Lys residues, which anchors the peptides and has a function similar to that of tryptophan residues is a further interesting topic to be investigated in the context of the five-slab model. Inclusion of side-chain flexibility within the head-group region may prove to be mandatory for studying this anchoring effect since it involves non-standard rotomeric states of the Lys residue. KALP peptides are similar to WALP peptides but contain Lys residues at the peptide termini and provide a simple system to study this effect and may be used as a model system for further studies.

The effect of bilayer thinning has been studied in this thesis for short polyalanine peptides

and was shown to have no effect on the helix orientation of these peptides. However, a more extensive study would be worthwhile especially to investigate the more complex peptides. The five-slab membrane model allows the study of the effect of bilayer perturbations and can be used in a straightforward manner to study the effect of the bilayer thickness on membrane peptide orientation. Whether or not these perturbations actually do occur is outside the scope of the analysis and has to be determined by experiments.

The membrane potential has also been implicated to influence membrane-peptide orientations. For instance, application of a membrane potential has been shown to change the orientation of melittin from surface-aligned to transmembrane [165]. Though the current work has not studied the effect of membrane potential, the five-slab model can be extended to include its effect.

7.5.2 Charge Screening and Effective Dipole Moment

Estimation of the solvent screening of transmembrane peptides provides insight in the membrane protein topology, such as parallel *vs.* anti-parallel helix association. Though the rules of the thumb established here allow estimation of the helix dipole for various geometries and environments, it will be interesting to calculate the accurate value of the helix dipole before assigning any functional or structural role to it. The method developed in this thesis may also be used to estimate effective charges and dipoles of other secondary structural elements such as β-sheets.

Significant dipole moments are often attributed to whole proteins and have been implicated in protein aggregation [166], ion channel function [167] and toxin binding [168]. The present method can be extended to determine the dipole moment and macromolecular electrostatics of larger proteins. The main limitation of the calculations would be the available computational resources. Whole-protein dipoles may also be determined experimentally, especially for globular proteins [169] and allow a direct comparison between experimental and calculated values.

7.5.3 Helix-Helix Association

The theoretical calculation of association processes in membranes is a subject still in development. Accordingly, the modelling of the α-helical membrane dimers raised many questions. The first is a quantitative estimation of dimerisation energies. In this thesis, the entropic contribution to the free energy, ΔG_s has not been calculated. ΔG_s is composed of both favourable and non-favourable terms. Two important unfavourable contribution to $\Delta\ G_s$ is the loss of lipid entropy (solvophobic effect) and the loss of translational entropy

upon dimerisation. The entropic contribution might be estimated by standard equations of simple diffusive systems. As association of membrane-bound helices can be described by a two-dimensional diffusion process, only two degrees of translational freedom are lost upon dimerisation as compared to three in soluble protein association. An estimate of the entropic contributions of internal degrees of freedom can be obtained from normal mode analysis. The major challenge of such analysis will be to consider the energetic contributions of the surrounding water lipid molecules and to distinguish these from the protein contributions.

An even more accurate estimate of the dimerisation free energy of helix-helix pairs may be obtained by umbrella sampling [170]. Molecular dynamics simulations including an all-atom membrane environment at varying inter-helical distances could be used to calculate the potential of mean force using umbrella sampling. The choice of the reaction coordinate in this case is straightforward since helix association may be thought to be a simple two-dimensional diffusion process. Although all-atom simulations are computationally expensive and difficult to set up they provide meaningful insights into such association processes. The role of the bilayer lipids in driving helix associations might be determined by calculating the dimerisation propensities of different transmembrane segments in the membrane and in vacuum. The difference between the two quantities would give an estimate of the solvophobic effect of the membrane environment.

A important question regarding helix-helix association is the definition of the thermodynamics of the association process. Ultracentrifugation techniques such as sedimentation equilibrium analytical ultracentrifugation have been used to measure the free energy of dimerisation of helix-helix pairs [53, 155]. These techniques access a quantity defined as the mole-fraction standard free energy and involve helix-helix pairs in various detergent solutions such as C_{14} betaine and C_8E_5 micellar solutions. In membranes, however, the kinetics of the process is governed by diffusion on a two-dimensional lattice and the free energy change per unit surface area of membrane may be physically more relevant for membrane helices.

An important limiting factor of all presented calculations has been computational resources. Optimisation of the computational requirements will allow a more thorough sampling of dimer parameters. Of course, with improvement in computational resources and better sampling, more detailed models become feasible. Combining a more accurate energy function with an efficient conformational search algorithm would certainly improve the modelling procedure.

7.5.4 Lipid Raft Localisation

Molecular Mechanisms influencing lipid raft sorting remain an open question. Many motifs have been proposed which drive partitioning into membrane rafts. It is seen that a mutant that is not closely-packed is also deficit in lipid raft localisation. It is thus proposed that close packing of helices is a pre-requisite for lipid raft localisation. At present, it is not possible to distinguish whether close packing of the EpoR TM dimer is a targeting signal or a means of the receptor to stabilise lipid raft association. The presence of additional motifs contributing to lipid raft location in the EpoR sequence is definitely possible.

A more detailed study of lipid raft localisation using experimental and modelling studies iteratively is required for a detailed understanding of the principles governing lipid raft localisation. A systematic study of point mutants and double mutants in different transmembrane dimers may provide clues on the factors governing lipid raft localisation. A characterisation of lipid rafts in terms of their thickness, polarity, packing density and electrostatic properties is also required.

Understanding the energetics of the structural organisation of membrane proteins is a fundamental issue in molecular cell biology which can be solved only by an interdisciplinary approach that combines techniques of theoretical computational physics with knowledge gained from bio-chemistry and molecular biology experiments. It is not only has a fundamental scientific significance but also has practical applications in fields ranging from structural biology to drug designing. Studies in these fields usually focus on experimentally characterising membrane-protein architecture and its interaction with the membrane bilayer. There is also a growing body of literature which deals with this problem theoretically. In this thesis, we report theoretical calculations that provide insights into various aspects of membrane peptides such as their orientation, electrostatic properties and association. The work presented here constitutes a small but significant step towards understanding the energetics of transmembrane proteins and its relation to protein structure and function.

References

[1] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter. *Molecular Biology of the Cell.* Garland Science, New York, 2002.

[2] C Branden and J Tooze. *Membrane Protein, Chapter 12 in Introduction to Protein Structure.* Garland Science, New York, 1998.

[3] I Ubarretxena-Belandia and D M Engelman. Helical membrane proteins: diversity of functions in the context of simple architecture. *Curr. Opin. Struct. Biol.*, 11:370–376, 2001.

[4] P L Yeagle and A G Lee. Membrane protein structure. *Biochim. Biophys. Acta*, 1565: 143, 2002.

[5] S J Singer. Some early history of membrane molecular biology. *Annu. Rev. Physiol.*, 66:1–27, 2004.

[6] S J Singer and G L Nicholson. The fluid mosaic model of the structure of cell membranes. *Science*, 175:720–731, 1972.

[7] D L Nelson and M M Cox. *Lehninger Principles of Biochemistry.* W H Freeman and Co., New York, 2004.

[8] S H White and W C Wimley. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.*, 28:319–336, 1999.

[9] D J Volsky and A Loyter. Inhibition of membrane fusion by suppression of lateral movement of membrane proteins. *Biochim. Biophys. Acta*, 514:213–24, 1978.

[10] K Simons and D Toomre. Lipid rafts and signal transduction. *Mol. Cell Biol.*, 1: 31–39, 2000.

[11] J M Berg, J.L. Tymoczko, and L Stryer. *Biochemistry.* W H Freeman and Co., New York, 2002.

[12] H Arthur and K Watson. Thermal adaptation in yeast: growth temperatures, membrane lipid, and cytochrome composition of psychrophilic, mesophilic, and thermophilic yeasts. *J. Bacteriol.*, 128:56–68, 1976.

[13] R G W Anderson and K Jacobson. A role for lipid shells in targeting proteins to caveolae, rafts and other lipid domains. *Science*, 296:1821–1825, 2002.

[14] D A Brown and A E London. Functions of lipid rafts in biological membranes. *Ann. Rev. Cell Dev. Biol.*, 14:111–136, 1998.

[15] R G W Anderson. The caveolae membrane system. *Annu. Rev. Biochem.*, 67:199–225, 1998.

[16] C Thiele, M J Hannah, F Fahrenholz, and W B Huttner. Cholesterol binds to synaptophysin and is required for biogenesis of synaptic vesicles. *Nat. Cell Biol.*, 2:42–49, 2000.

[17] M J Schlesinger. Proteolipids. *Annu. Rev. Biochem.*, 50:193–206, 1981.

[18] P Scheiffele, M G Roth, and K Simons. Interaction of influenza virus haemagglutinin with sphingolipid-cholesterol membrane domains via its transmembrane domain. *EMBO Journal*, 16:5501–5508, 1997.

[19] S A Tatulian and L K Tamm. Secondary structure, orientation, oligomerization and lipid interactions of the transmembrane domain of influenza hemaglutinin. *Biochemistry*, 39:496–507, 2000.

[20] M Gandhavadi, D Allende, A Vidal, S A Simon, and T J McIntosh. Structure, composition and peptide binding properties of detergent soluble bilayers and detergent resistant rafts. *Biophys. J.*, 82:1469–1482, 2002.

[21] H Lodish, A Berk, L S Zipursky, P Matsudaira, D Baltimore, and J Darnell. *Molecular Cell Biology.* W H Freeman and Co., New York, 2000.

[22] M Tomita, H Furthmayr, and V T Marchesi. Primary structure of human erythrocyte glycophorin A. isolation and characterization of peptides and complete amino acid sequence. *Biochemistry*, 17, 1978.

[23] H Luecke, B Schobert, H T Richter, J P Cartailler, and J K Lanyi. Structure of bacteriorhodopsin at 1.55 angstrom resolution. *J. Mol. Biol.*, 291:899–911, 1999.

[24] A Arora, F Abildgaard, J H Bushweller, and L K Tamm. Structure of outer membrane protein a transmembrane domain by NMR spectroscopy. *Nat. Struct. Biol.*, 8, 2001.

[25] J L Popot and D M Engelman. Membrane protein folding and oligomierzation - the 2-stage model. *Biochemistry*, 29:4031–4037, 1990.

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

[27] W F DeGrado, H Gratkowski, and J D Lear. How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundle. *Prot. Sci.*, 12:647–665, 2003.

[28] N Ben-Tal, A Ben-Shaul, A Nicholls, and B Honig. Free-energy determinants of α-helix insertion into bilayers. *Biophys. J.*, 70:1803–1812, 1996.

[29] S Bernéche, M. Nina, and B. Roux. Molecular dynamics simulation of melittin in a dimyristoyl phosphatidylcholine bilayer membrane. *Biophys. J.*, 75:1603–1618, 1998.

[30] K Diraviyam, R V Stahelin, W Cho, and D. Murray. Computer modeling of the membrane interaction of FYVE domains. *J. Mol. Biol.*, 328:721–736, 2003.

[31] W Im, M Feig, and C L Brooks III. An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.*, 85:2900–2918, 2003.

[32] A Kessel, D S Cafiso, and N Ben-Tal. Continuum solvent model calculations of alamethicin-membrane interactions: Thermodynamic aspects. *Biophys. J.*, 78:571–583, 2000.

[33] T Lazardis. Effective energy function for proteins in lipid membranes. *Proteins*, 52: 176–192, 2003.

[34] D Murray, L Hermida-Matsumoto, C A Buser, J Tsang, C T Sigal, and N. Ben-Tal. Electrostatics and the membrane association of Src: Theory and experiment. *Biochemistry*, 37:2145–2159, 1998.

[35] V Z Spassov, L Yan, and S Szalma. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J. Phys. Chem. B*, 106:8726–8738, 2002.

[36] N Sreerama and R W Woody. On the analysis of membrane protein circular dichroism spectra. *Prot. Sci.*, 13:100–112, 2004.

[37] A Watts, I J Burnett, C Glaubitz, G Grobner, D A Middleton, P J R Spooner, J A Watts, and P T F Williamson. Membrane protein structure determination by solid state NMR. *Nat. Prod. Rep.*, 16:419–423, 1999.

[38] M R R de Planque, E Goormaghtigh, D V Greathouse, R E KoeppeII, J A W Kruijtzer, R M M Liskamp, B de Kruiff, and J A Killian. Sensitivity of single membrane-spanning α-helical peptides to hydrophobic mismatch with a lipid bilayer: Effects on backbone structure, orientation and extent of membrane incorporation. *Biochemistry*, 40:5000–5010, 2001.

[39] J A A Demmers, E van Duijn, J Haverkamp, D V Greathouse, R E KoeppeII, A J R Heck, and J A Greathouse. Interfacial positioning and stability of transmembrane peptides in lipid bilayers studies by combining hydrogen/deuterium exchange and mass spectroscopy. *J. Biol. Chem.*, 276:34501–34508, 2001.

[40] P C A van der Wel, E Strandberg, J A Killian, and R E Koeppe. Geometry and intrinsic tilt of a tryptophan-anchored transmembrane α-helix determined by ^{2}H NMR. *Biophys. J.*, 83:1479–1488, 2002.

[41] T M Weiss, P C A van der Wel, J A Killian, R E KoeppeII, and H W Huang. Hydrophobic mismatch between helices and lipid bilayers. *Biophys. J.*, 84:379–385, 2003.

[42] K R MacKenzie, J H Prestegard, and D M Engelman. A transmembrane helix dimer: structure and implications. *Science*, 276:131–133, 1997.

[43] R Roy and D Langosch. Synaptobrevin transmembrane domain dimerization - revisited. *Biochemistry*, 43:4964–4970, 2004.

[44] W Ruan, V Becker, U Klingmüller, and D Langosch. The interface between the self-assembling erythropoietin receptor transmembrane segments corresponds to a heptad repeat pattern. *J. Biol. Chem.*, 279:3273–3279, 2004.

[45] K Hristova, C E Dempsey, and S H White. Structure, location, and lipid perturbations of melittin at the membrane interface. *Biophys. J.*, 80:801–811, 2001.

[46] B A Wallace. Structure of gramicidin A. *Biophys. J.*, 49:295–306, 1986.

[47] B Bechinger. Membrane insertion and orientation of polyalanine peptides: a ^{15}N solid state NMR spectroscopy investigation. *Biophys. J.*, 81:2251–2256, 2001.

[48] P Braun and G H Heijne. The aromatic residues Trp and Phe have different effects on the positioning of a transmembrane helix in the microsomal membrane. *Biochemistry*, 38:9778–9782, 1999.

[49] S Sharpe, K R Barber, C W M Grant, D Goodyear, and M R Morrow. Organisation of model helical peptides in lipid bilayers: insight into the behaviour of single-span protein transmembrane domains. *Biophys. J.*, 83:345–358, 2002.

[50] S H White and W C Wimley. Hydrophobic interactions of peptides with membrane interfaces. *Biochim. Biophys. Acta*, 1376:339–352, 1998.

[51] W M Yau, W C Wimley, K Gawrisch, and S H White. The preference of tryptophan for membrane interfaces. *Biochemistry*, 37:14713–14718, 1998.

[52] V Helms. Attraction within the membrane: Forces behind transmembrane protein folding and supramolecular complex assembly. *EMBO reports*, 3:1133–1138, 2002.

[53] K G Fleming and D M Engelman. Specificty in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc. Natl. Acad. Sci. U. S.Å.*, 98:14340–14344, 2001.

[54] W L Ash, T Stockner, J L MacCallum, and D P Tieleman. Computer modeling of polyleucine-based coiled coil dimers in a realistic membrane environment: Insight into helix-helix interactions in membrane proteins. *Biochemistry*, 43:9050–9060, 2004.

[55] A R Curran and D M Engelman. Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.*, 13:412–417, 2003.

[56] M M Javadpour, M Eilers, M Groesbeek, and S O Smith. Helix packing in polytopic membrane proteins: Role of glycine in transmembrane helix association. *Biophys. J.*, 77:1609–1618, 1999.

[57] J Liang. Experimental and computational studies of determinants of membrane-protein folding. *Curr. Opin. Chem. Biol.*, 6:878–884, 2002.

[58] I J Arkin. Structural aspects of oligomeirsation taking place between the transmembrane α-helices of biotopic membrane proteins. *Biochim. Biophys. Acta*, 1565:347–363, 2002.

[59] B Brosig and D Langosch. The dimerization motif of the glycophorin A transmembrane segment in membranes: importance of glycine residues. *Prot. Sci.*, 7:1052–1056, 1998.

[60] R Gurezka and D Langosch. In vitro selection of membrane-spanning leucine zipper protein-protein interaction motifs using possyccat. *J. Biol. Chem.*, 276:45580–45587, 2001.

[61] K G Fleming and D M Engelman. Computation and mutagenesis suggest a right-handed dimer for the synaptobrevin transmembrane domain. *Proteins*, 45:313–317, 2001.

[62] H Gratkowski, J D Lear, and W F DeGrado. Polar side chains drive the association of model transmembrane peptides. *Proc. Natl. Acad. Sci. U. S.Á.*, 98:880–885, 2001.

[63] T Stockner, W L Ash, J L MacCallum, and D P Tieleman. Direct simulation of transmembrane helix association: Role of asparagines. *Biophys. J.*, 87:1650–1656, 2004.

[64] F X Zhou, H J Merianos, A T Brunger, and D M Engelman. Polar residues drive association of polyleucine transmembrane helices. *Proc. Natl. Acad. Sci. U. S.Á.*, 98: 2250–2255, 2001.

[65] J P Dawson, R A Melnyk, C M Deber, and D M Engelman. Sequence context strongly modulates association of polar residues in transmembrane helices. *J. Mol. Biol.*, 331: 255–262, 2003.

[66] D R Fattal and A Ben-Shaul. A molecular model for lipid-protein interaction in membranes: role of hydrophobic mismatch. *Biophys. J.*, 65:1795–1809, 1993.

[67] P Lagüe, M J Zuckermann, and B Roux. Lipid-mediated interactions between intrinsic membrane proteins: A theoretical study based on integral equations. *Biophys. J.*, 79: 2867–2879, 2000.

[68] N Ben-Tal and B Honig. Helix-helix interaction in lipid bilayers. *Biophys. J.*, 71: 3046–3050, 1996.

[69] M K Gilson and B Honig. Destabilization of an α-helix bundle protein by helix dipoles. *Proc. Natl. Acad. Sci. U. S.Á.*, 86:1524–1528, 1989.

[70] S O Smith, M Eilers, D Song, E Crocker, W Ying, M Groesbeek, G Metz, M Ziliox, and S Aimoto. Implications of threonine hydrogen bonding in the glycophorin A transmembrane helix dimer. *Biophys. J.*, 82:2476–2486, 2002.

[71] M A Lemmon, H R Treutlein, P D Adams, A T Brünger, and D M Engelman. A dimerization motif for transmembrane α-helices. *Nat. Struct. Biol.*, 1:157–163, 1994.

[72] W P Russ and D M Engelman. Toxcat: A measure of transmembrane helix association in a biological membrane. *Proc. Natl. Acad. Sci. U. S. Å.*, 96:863–868, 1999.

[73] C M Deber and S C Li. Peptides in membranes: helicity and hydrophobicity. *Biopolymers*, 37:295–318, 1995.

[74] A Grossfield, J Sachs, and T B Woolf. Dipole lattic membrane model for protein calculation. *Proteins*, 41:211–233, 2000.

[75] C E Capener, I H Shrivastava, K M Ranatunga, L R Forrest, G R Smith, and M S P Sansom. Homology modeling and molecular dynamics simulation studies of an inward rectifier potassium channel. *Biophys. J.*, 78:2929–2942, 2000.

[76] D Mihailescu and J C Smith. Atomic detail peptide-membrane interactions: Molecular dynamics of gramicidine S in a DMPC bilayer. *Biophys. J.*, 79:1718–1730, 2000.

[77] D P Tieleman, H J C Berendsen, and M S P Sansom. Surface binding of alamethicin stabilizes its helical structure molecular dynamics simulations. *Biophys. J.*, 76:3186–3191, 1999.

[78] D P Tieleman, J Breed, H J C Berendsen, and M S P Sansom. Alamethicin channels in a membrane: molecular dynamics simulations. *Faraday Disc.*, 111:209–225, 1998.

[79] T B Woolf and B Roux. Structure, energetics and dynamics of lipid-protein interactions: A molecular dynamics study of gramicidine A channel in a dmpc bilayer. *Proteins*, 24:92–114, 1996.

[80] L Forrest, A Kukol, I Arkin, A Tielman, and M Sansom. Exploring the models of *Influenza* M2 channel - MD simulations in a phospholipid bilayer. *Biophys. J.*, 78: 79–92, 2000.

[81] A Kukol, P D Adams, L M Rice, A T Brunger, and I T Arkin. Experimentally based orientational refinement of membrane protein models: A structure for the *Influenza* A M2 H^+ channel. *J. Mol. Biol.*, 286:951–962, 1999.

[82] H I Petrache, A Grossfield, K R MacKenzie, D M Engelman, and T B Woolf. Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: Molecular dynamics calculations. *J. Mol. Biol.*, 302:727–746, 2000.

[83] J Baudry, E Tajkhorshid, F Molnar, J Phillips, and K Schulten. Molecular dynamics study of bacteriorhodopsin and the purple membrane. *J. Phys. Chem. B*, 105:905–918, 2001.

[84] C Kant, J Schlitter, and K Gerwert. Dynamics of water molecules in the bacteriorhodopsin trimer in explicit lipid/water environment. *Biophys. J.*, 86:705–717, 2004.

[85] D P Tielman and H J C Berendsen. A molecular dynamics study of the pores formed by *e. coli* OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.*, 74:2786–2801, 1998.

[86] B L DeGroot and H Grubmuller. Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science*, 294:2353–2357, 2001.

[87] Z Cournia, A C Vaiana, G M Ullmann, and J C Smith. Derivation of a molecular mechanics force field for cholesterol. *Pure Appl. Chem.*, 76:189–196, 2004.

[88] T Woolf, A Grossfield, and D Zuckerman. Electrostatics of membrane systems - complex, heterogeneous environments. In L Pratt and G Hummer, editors, *Simulation and Theory of Electrostatic Interactions in Solution*, volume 492 of *AIP Conference Proceedings*, pages 510–532. American Institute of Physics, 1999.

[89] B Roux and R MacKinnon. The cavity and pore helices in the kcsa k^+ channel: electrostatic stabilization of monovalent cations. *Science*, 285:100–102, 1999.

[90] N Calimet and G M Ullmann. The influence of a transmembrane pH gradient on protonation probabilities of bacteriorhodopsin: The structural basis of the back-pressure effect. *J. Mol. Biol.*, 339:571–589, 2004.

[91] B Roux, S Berneche, and S Im. Ion channels, permeation and electrostatics: Insight into the function of KcsA. *Biochem.*, 39:13295–13306, 2000.

[92] D Beglov and B Roux. Numerical solution of the HNC equation for a solute of arbitrary geometry in three dimensions. *J. Chem. Phys.*, 103:360–364, 1995.

[93] J.D. Jackson. *Classical Electrodynamics*. John Wiley and sons, New York, 1998.

[94] W C Still, A Tempczyk, R C Hawley, and T Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[95] A Onufriev, D A Case, and D Bashford. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comp. Chem.*, 23, 2002.

[96] R G Efremov, D E Nolde, G Vergoten, and A S Arseniev. A solvent model for simulations of peptides in bilayers. i. membrane-promoting α-helix formation. *Biophys. J.*, 76, 1999.

[97] A R Leach, editor. *Molecular Modelling: Principles and applications,*. Pearson Education Limited, 2001.

[98] D A Pearlman, D A Case, J W Caldwell, W R Ross, T E Cheatham-III, S DeBolt, D Ferguson, G Seibel, and P Kollman. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Commun.*, 91:1–41, 1995.

[99] Bernard R. Brooks, Robert E. Bruccoleri, Berry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculation. *J. Comp. Chem.*, 4:187–217, 1983.

[100] W F van Gunsteren and H J C Berendsen. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.*, 29:992–1023, 1990.

[101] B Honig and A Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268:1144–1149, 1995.

[102] G M Ullmann and E-W Knapp. Electrostatic computations of protonation and redox equilibria in proteins. *Eur. Biophys. J.*, 28:533–551, 1999.

[103] F Fogolari, A Brigo, and H Molinari. The Poisson-Boltzmann equation for biomolecular electrostatics: A tool for structural biology. *J. Mol. Recognit.*, 15:377–392, 2002.

[104] W Im, D Beglov, and B Roux. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comp. Phys. Commun.*, 109:1–17, 1998.

[105] H A Stern and S E Feller. Calculation of dielectric permittivity profile for a nonuniform system: application to a lipid bilayer simulation. *J. Chem. Phys.*, 118:3401–3411, 2003.

[106] F Zhou and K Schulten. Molecular dynamics study of a membrane-water interface. *J. Phys. Chem.*, 99:2194–2207, 1995.

[107] H G L Coster and J R Smith. The molecular organization of bimolecular lipid membranes. a study of low frequency maxwell-wagner impedance dispersion. *Biochim. Biophys. Acta*, 373:151–164, 1974.

[108] G Cevc, A Watts, and D Marsh. Titration of the phase transition of phosphatidylserine bilayer membranes. effects of pH, surface electrostatics, ion binding and head group hydration. *Biochemistry*, 20:4955–4965, 1981.

[109] R Cseh and R Benz. Interaction of phloretin with lipid monolayers: Relationship between structural changes and dipole potential change. *Biophys. J.*, 77:1477–1488, 1999.

[110] P I Lelkes and I R Miller. Perturbations of membrane structure by optical probes: location and structural sensitivity of merocyanine 540 bound to phospholipid membranes. *J. Mol. Biol.*, 52:1–15, 1980.

[111] Y Nozaki and C Tanford. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions.establishment of a hydrophobicity scale. *J. Biol. Chem.*, 246:2211–2217, 1971.

[112] D Sitkoff, N. Ben-Tal, and B. Honig. Calculation of alkane to water solvation free energies using continuum solvent models. *J. Phys. Chem.*, 100:2744–2752, 1996.

[113] M Nina, D Beglov, and B Roux. Atomic Born radii for continuum electrostatic calculations based on molecular dynamics. *J. Phys. Chem.*, 101:5239–5248, 1997.

[114] B Lee and F M Richards. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.*, 55:379–400, 1971.

[115] J P Bradshaw, C E Dempsey, and A Watts. A combined x-ray neutron diffraction study of selectively deuterated melittin in phospholipid bilayers: effect of pH. *Mol. Membrane Biol.*, 11:79–86, 1994.

[116] S Frey and L K Tamm. Orientation of melittin in phospholipid bilayers: A polarised ATRI study. *Biophys. J.*, 60:922–930, 1991.

[117] J Torres, A Kukol, and I T Arkin. Use of a single glycine residue to determine the tilt and orientation of a transmembrane helix. a new structural label for infrared spectroscopy. *Biophys. J.*, 79:3139–3143, 1999.

[118] T C Terwilliger and D Eisenberg. The structure of melittin.i.structure determination and partial refinement. *J. Biol. Chem.*, 257:6010–6015, 1982.

[119] H M Berman, J Westbrook, Z Feng, G Gilliard, T N Bhat, H Weissig, I N Shindyalov, and P E Bourne. The protein data bank. *Nucleic Acids Research*, 28:235–242, 2000.

[120] M Bower, F E Cohen, and R L Dunbrack Jr. Side-chain prediction from a backbone-dependent rotamer library: A new tool for homology modeling. *J. Mol. Biol.*, 267: 1268–1282, 1997.

[121] R L Dunbrack Jr and M Karplus. Backbone-dependent rotamer library for proteins: Application to side-chain prediction. *J. Mol. Biol.*, 230:543–574, 1993.

[122] I T Arkin, K R Mackenzie, and A T Bruenger. Site-directed dichroism as a method for obtaining rotational and orientational constraints for oriented polymers. *J. Am. Chem. Soc.*, 119:8973–8980, 1997.

[123] E Tajkhorshid, P Nollert, M Jensen, L J W Miercke, J O'Connel, R M Stroud, and K Schulten. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science*, 296:525–530, 2002.

[124] D A Doyle, J M Cabral, R A Pfuetzner, A Kuo, J M Gulbis, S L Cohen, B T Chait, and R MacKinnon. The structure of the potassium channel: Molecular basis of K^+ conduction and selectivity. *Science*, 280:69–77, 1998.

[125] J Faraldo-Gomez and B Roux. Electrostatics of ion stabilization in a ClC chloride channel homologue from *escherichiacoli*. *J. Mol. Biol.*, 339:981–1000, 2004.

[126] W G J Hol. Effects of the α-helix dipole upon the functioning and structure of proteins and peptides. *Advan. Biophys.*, 19:133–165, 1985.

[127] W G J Hol, L M Halie, and C Sander. Dipoles of the α-helix and β-sheet: Their role in protein folding. *Nature*, 294:532–536, 1981.

[128] R P Sheridan, R M Levy, and F R Salemme. α-helix dipole model and eletrostatic stabilization of 4-α-helical proteins. *Proc. Natl. Acad. Sci. U. S.Ȧ.*, 79:4545–4549, 1982.

[129] T O Yeates, H Komiya, D C Rees, J P Allen, and G Feher. Structure of the reactions center from *Rhodobacter sphaeroides* R-26: Membrane-protein interactions. *Proc. Natl. Acad. Sci. U. S.Ȧ.*, 84:6438–6442, 1987.

[130] H V Joshi and M.S. Meier. The effect of a peptide helix macrodipole on the pK_a of an asp side chain carboxylate. *J. Am. Chem. Soc.*, 118:12038–12044, 1996.

[131] D J Lockhart and P S Kim. Internal stark effect measurement of the electric field at the amino terminus of an α helix. *Science*, 257:947–951, 1992.

[132] E Galoppini and M A Fox. Effect of the electrical field generated by the helix dipole on photoinduced intramolecular electron transfer in dichromorphic α-helical peptides. *J. Am. Chem. Soc.*, 118:2299–2300, 1996.

[133] J S Miller, R J Kennedy, and D S Kemp. Solubilized, spaced polyalanines: A context-free system for determining amino acid α-helix propensities. *J. Am. Chem. Soc.*, 124: 945–962, 2002.

[134] T E Creighton. *Proteins: Structures and Molecular Properties*. W H Freeman and Co., New York, 1993.

[135] W G J Hol, P T van Duijen, and H J C Berendsen. The α-helix dipole and the properties of proteins. *Nature*, 273:443–446, 1978.

[136] A Wada. The α helix as an electric macro dipole. *Advan. Biophys.*, 9:1–63, 1976.

[137] R P Sheridan and L C Allen. The electrostaic potential of the alpha helix. *Biophys. Chem.*, 11:133–136, 1980.

[138] J Warwicker and H C Watson. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.*, 157:671–679, 1982.

[139] R R Gabdoulline and R C Wade. Effective charges for macromolecules in solvent. *J. Phys. Chem.*, 100:3868–3878, 1996.

[140] C Breneman and K J Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials: need for high sampling density in conformational analysis. *J. Comp. Chem.*, 11:361–373, 1990.

[141] C Lange and C Hunte. Crystal structure of the yeast cytochrome bc1 complex with its bound substrate cytochrome C. *Proc. Natl. Acad. Sci. U. S.Á.*, 99:2800–2805, 2002.

[142] J A Ernst and A T Brunger. High resolution structure, stability and synaptotagmin binding of a truncated neuronal SNARE complex. *J. Biol. Chem.*, 278:8630–8636, 2003.

[143] M Mellado, A J Vila-Coro, C Martinez, and J M Rodriguez-Frade. Molecular mechanisms of cytokine receptor activation. *Cell Mol. Biol.*, 47:575–582, 2001.

[144] J Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 103:211–225, 2000.

[145] S O Smith, D Song, S Shekar, M Groesbeek, M Ziliox, and S Aimoto. Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. *Biochemistry*, 40:6553–6558, 2001.

[146] K P Howard, J D Lear, and W F DeGrado. Sequence determinants of the energetics of folding of a transmembrane four-helix-bundle protein. *Proc. Natl. Acad. Sci. U. S. Å.*, 99:8568–8572, 2002.

[147] P D Adams, I T Arkin, D M Engelman, and A T Brunger. Computational searching and mutagenesis suggest a structure for phospholamban, a cardiac ion channel. *Nat. Struct. Biol.*, 2:154–162, 1995.

[148] H K B Simmerman, Y M Kobayashi, J M Autry, and L R Jones. A leucine zipper stabilizes the pentameric membrane domain of phospholamban and forms a coiled-coil pore structure. *J. Biol. Chem.*, 271:5941–5946, 1996.

[149] J D Lear, A L Stouffer, H Gratkowski, V Nanda, and W F DeGrado. Association of a model transmembrane peptide containing Gly in a heptad sequence motif. *Biophys. J.*, 87:3421–3429, 2004.

[150] J A G Briggs, J Torres, and I T Arking. A new method to model membrane protein structure based on silent amino acid substitutions. *Proteins*, 44:370–375, 2001.

[151] N Seubert, Y Royer, J Staerk, K F Kubatzky, V Moucadel, S Krishnakumar, S O Smith, and S N Constantinescu. Active and inactive orientations of the trnasmembrane and cytosolic domains of the erythropoietin receptor domain. *Mol. Cell*, 12: 1239–1250, 2003.

[152] S Fleishman, J Schlessinger, and N Ben-tal. A putative molecular-activation switch in the transmembrane domain of erbB2. *Proc. Natl. Acad. Sci. U. S. Å.*, 99:15937–15940, 2002.

[153] S Kim, A K Chamberlain, and J U. Bowie. A simple method for modeling transmembrane helix oligomers. *J. Mol. Biol.*, 329:831–840, 2003.

[154] Y Park, M Elsner, R Staritzbichler, and V Helms. Novel scoring function for modeling structures of oligomers of transmembrane α-helices. *Proteins*, 57:577–585, 2004.

[155] A K Doura and K G Fleming. Complex interactions at the helix-helix interface stabilise the glycophori A transmembrane dimer. *J. Mol. Biol.*, 343:1487–1497, 2004.

[156] K F Kubatzky, W Ruan, R Gurezka, J Cohen, R Ketteler, S S Watowich, D Neumann, D Langosch, and U Klingmüller. Self assembly of the transmembrane domain promotes signal transduction through the erythropoietin receptor. *Curr. Biol.*, 11: 110–115, 2001.

[157] S N Constantinescu, T Keren T, M Socolovsky M, H Nam, Y I Henis, and H F Lodish. Ligand-independent oligomerization of cell-surface erythropoietin receptor is mediated by the transmembrane domain. *Proc. Natl. Acad. Sci. U.S.A.*, 98:4379–4384, 2001.

[158] G Archontis, T Simonson, and M Karplus. Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Applications to amino acid recognition by Aspartyl-tRNA synthetase. *J. Mol. Biol.*, 306:307–327, 2001.

[159] Z S Hendsch and B Tidor. Electrostatic interactions in the gcn4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding. *Prot. Sc.*, 8:1381–1392, 1999.

[160] R A Melnyk, S Kim, A R Curran, D M Engelman, J U Bowie, and C M Deber. The affinity of gxxxg motifs in transmembrane helix-helix interactions is modulated by long-range communication. *J. Biol. Chem.*, 279:16591–16597, 2004.

[161] C Mineo, G N Gill, and R G Anderson. Regulated migration of epidermal growth factor receptor from caveolae. *J. Biol. Chem.*, 274:30636–30643, 1999.

[162] R M Levy, L Y Zhang, E Gallicchio, and A K Felts. On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.*, 125:9523–9530, 2003.

[163] D Kiefer and A Kuhn. Hydrophobic forces drive spontaneous membrane insertion of the bacteriophage Pf3 coat protein without topological control. *EMBO Journal*, 1999:6299–6306, 1999.

[164] T Schirmer, T A Keller, Y F Wang, and J P Rosenbusch. Structural basis for sugar translocation through maltoporin channels at 3.1 Å. *Science*, 267:512–514, 1995.

[165] C Kempf, R D Klausner, J N Weinstein, J V Renswoude, and M Pincus. Voltage-dependent trans-bilayer orientation of melittin. *J. Biol. Chem.*, 257:2469–2476, 1982.

[166] G G Tartaglia, A Cavalli, R Pellarin, and A Caflisch. The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Prot. Science*, 13:1939–1941, 2004.

[167] S Takashima. The dipole moment of membrane proteins: Potassium channel protein and β-subunit. *Biophys Chem.*, 94:209–218, 2001.

[168] X Huang, H Liu, M Cui, W Fu, K Yu, K Chen, X Luo, J Shen, and H Jiang. Simulating the interactions of toxins with K^+ channels. *Curr. Pharm. Des.*, 10:1057–1067, 2004.

[169] J Antosiewicz and D Porschke. Electrostaics of haemoglobin from measurements of electric dichroism and computer simulations. *Biophys. J.*, 68:655–664, 1995.

[170] G M Torrie and J P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J. Chem. Phys.*, 23:187–199, 1977.