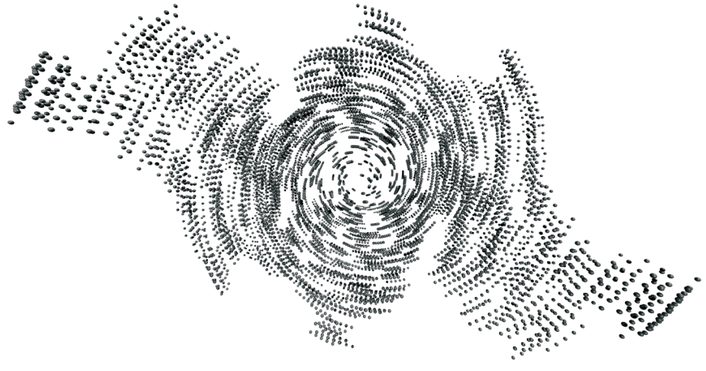


Protein conformational plasticity and aggregation

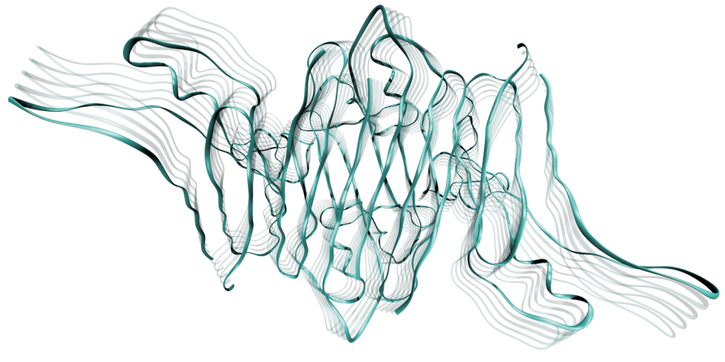
Nuno Loureiro Ferreira

2010

Nuno Ricardo Santos Loureiro da Silva Ferreira

Protein conformational plasticity and aggregation

Transthyretin and δ–toxin as two case studies



Coimbra
2010

Protein conformational plasticity and aggregation

•

Transthyretin and δ-toxin as two case studies

Nuno Ricardo Santos Loureiro da Silva Ferreira

Universidade de Coimbra
2010

Dissertation for the obtention of Ph.D. degree in Chemistry, Biological Chemistry specialty, submitted to the Faculty of Sciences and Technology, University of Coimbra.

Supervisor: Dr. Rui Manuel Pontes Meireles Ferreira de Brito

Departamento de Química
Universidade de Coimbra
3004-535 Coimbra
Portugal

Financial support for the execution and publication of this thesis was provided by Fundação para a Ciência e Tecnologia through grant SFRH/BD/1354/2000.

Aos meus avós, pais e irmãos,
À minha nina e ao nosso filho.

Contents

Abstract

All living organisms possess common attributes that distinguishes them from the surrounding inorganic matter. A complex network of interacting organic molecules are responsible for the existence and perpetual of life. Among the most extraordinary biomolecules are proteins. Proteins are the workhorses of life, they actively participate in all biological processes. This is accomplished through an intrinsic conformational plasticity, a requisite in all biochemical reactions inherently dynamic processes. For most proteins a correct three dimensional structure is essential for its biological function. It is a fact that many diseases can be traced back to malfunction of proteins. Understanding the underlying connection between the biomolecule structure and its dynamics, is thus essential for current day medicine.

Throughout this thesis, two proteins were the subject of study using both experimental and computational techniques. Different biophysical aspects were focused stressing the importance of the usage of different approaches to get complementary views.

δ-lysin is a small peptide excreted by the bacteria *Staphylococcus aureus*, responsible for the disruption of eukaryotic membranes in infection conditions. On its way to the membrane, the peptide experience subtle chemical environment changes with consequences on peptide conformation. The amphyphatic helical structure of this peptide was studied using nuclear magnetic resonance (NMR), and its conformational plasticity and propensity for aggregation in solution was assessed through molecular dynamics (MD) simulations and NMR. The first stages of the lytic process were also modeled by analyzing MD trajectories describing the interaction of the peptide with a zwiterionic membrane. The results show that peptide aggregation may play a key role in the process.

Transthyretin (TTR) is one of the many proteins known to be involved in human amyloid diseases. Depending on the protein variant amyloidogenicity, the native tetrameric complex dissociates into monomeric species with high tendency for aggregation. Soluble aggregates and insoluble amyloid fibrils are responsible for the morphological alterations and cell observed in these pathologies. We developed a computational method using protein docking driven by experimental data, to build a high resolution molecular model of the elementary units that constitute the fibrils. We obtained a polymorphic set of protofilament models, further characterized in terms of helical periodicities and interface match. The overall stereochemical quality of the structures was assessed. This models may be a valuable instrument in the rational design of compounds with therapeutic potential to inhibit amyloid fibril formation.

Sumário

Os organismos vivos possuem atributos comuns que os distinguem do mundo inorgânico. Uma rede complexa de interacções entre moléculas orgânicas propicia a existência e a perpetuação da vida. As proteínas são uma das mais extraordinárias classes de biomoléculas, participando activamente em todos os processos biológicos. Esta ubiquidade das proteínas deve-se à sua *relativa* plasticidade conformacional, um requisito importante nas reacções bioquímicas, processos dinâmicos por definição. A maioria das proteínas tem uma estrutura tridimensional *bem* definida, essencial para a sua função biológica. Muitas das doenças hoje conhecidas têm a sua origem num incorrecto funcionamento das proteínas. O estudo da relação existente entre a estrutura de uma dada biomolécula e a sua dinâmica, é um dos tópicos mais importantes na compreensão dos mecanismos moleculares de múltiplas patologias.

Relativamente a esta temática, duas proteínas foram o sujeito de estudo na elaboração desta tese. Foram utilizadas técnicas experimentais bem como protocolos computacionais, onde diferentes aspectos biofísicos foram abordados.

A bactéria *Staphylococcus aureus* é responsável por inúmeros casos de septicémia. Umas das toxinas excretadas pela bactéria para o plasma, a δ-toxina, é responsável pela desintegração membranar em organismos eucarióticos. Durante o mecanismo infeccioso, o peptídeo experiencia a influência de diferentes ambientes químicos, e por conseguinte com impacto na sua conformação tridimensional. A estrutura anfipática helical do peptídeo foi estudada por ressonância magnética nuclear (NMR). Simulações moleculares foram também produzidas no estudo da plasticidade conformacional e propensão à agregação do peptídeo em solução. A interacção do peptídeo com membranas foi também estudado por dinâmica molecular. Os resultados evidenciam que o estado de

agregação do peptídeo pode ser relevante para o processo lítico.

A transtirretina (TTR) humana é uma das muitas proteínas envolvidas em doenças amilóides. Dependendo da amiloidogeneceidade da variante proteica, a estabilidade do complexo tetramérico nativo é alterado, ocorrendo dissociação e formação de espécies monoméricas com tendência para agregar. Os agregados solúveis e as fibras amilóides insolúveis são responsáveis pelas alterações morfológicas e morte celular observadas nestas patologias. Um método computacional baseado em acoplagem molecular ("docking") utilizando dados experimentais foi desenvolvido nesta tese. O objectivo do protocolo centrou-se na modelação molecular de um modelo protofilamentar, a unidade elementar das fibras amilóides. Os modelos polimórficos obtidos foram agrupados de acordo com a sua periodicidade helical, para posterior análise das interfaces proteicas e qualidade estereoquímica das estruturas. Os modelos produzidos poderão ter um impacto significativo em estudos terapêuticos com o objectivo de produzir fármacos que inibam a formação das fibras amilóides.

Outline of the thesis

The thesis consists of seven chapters, including a general introduction and a final chapter containing some overall conclusions, and perspectives. A more detailed description of the main chapters follows.

Chapter 1 aims to give a short glimpse on how proteins are truly complex and important biological machines. The study of protein biophysics relies on both experimental and computational techniques, and a general background about the methodologies employed throughout the thesis is given. The goal of the author is to introduce the reader to the main concepts used throughout the next chapters, without going into too much detail.

Chapter 2 introduces the importance of the understanding and characterization of protein interactions, as one of the major goals on modern biochemistry. This chapter will focus the attention on the application of pulsed field gradient nuclear magnetic resonance (PFG NMR) techniques to measure the translational diffusion of proteins, which can be correlated with the aggregation state of the protein in solution. In addition, the chapter aims to give the theoretical background of PFG NMR experiments, along with relevant methodological issues on how to setup, perform and analyze PFG experiments. Two protein systems were used to introduce these concepts, namely hen egg white lysozyme and δ-toxin.

Chapter 3 is a classical NMR work on the determination of a peptide structure, namely δ-toxin, in membrane-mimetic environments. The mechanisms of interaction between cytolytic peptides and target membranes are still under strong scientific debate. The three-dimensional structure of these peptides in different chemical environments is essential for the prosecution of such studies.

Chapter 4 introduces the reader to molecular dynamics (MD). It describes

the performance of variants of the GROMOS96 force field on simulations of a lipid bilayer system in water, namely dimyristoylphosphatidylcholine (DMPC), for a set of generally used parameters. The major outcome was an equilibrated system well characterized, ready to be used on more complex simulations.

Chapter 5 assesses the conformational stability of δ-toxin and its propensity for aggregation, through molecular dynamics simulations at room temperature in different chemical environments. A brief comparison between experimental results and MD simulations show that they agree fairly well, and provide a complementary view of the structure and dynamics behavior of the peptide in solution. The initial steps of interaction between the peptide and a DMPC lipid bilayer are also abridged by MD.

Chapter 6 is a protein docking study on the protein transthyretin (TTR), one of the many proteins known to be involved in human amyloid diseases. The chapter focuses on the modeling of amyloid protofilaments, based on available experimental biochemical data for TTR. The models were built using a docking-alignment procedure developed in the group.

Chapter 7 highlights the main conclusions from all previous chapters. It also gives a brief overview about my perspectives on how grid enabled computational sciences are a promising approach to delve further into the intricacies of the protein world.

Estrutura da tese

Esta tese é composta por sete capítulos, incluindo uma introdução geral e um capítulo final de conclusões e perspectivas futuras. Uma descrição mais detalhada do conteúdo de cada capítulo segue abaixo.

O **capítulo 1** pretende dar uma panorâmica muito geral e concisa da importância das proteínas no mundo biológico. No estudo da biofísica de proteínas, diversas técnicas experimentais e computacionais são geralmente empregues. Uma descrição breve e sumária das técnicas utilizadas nesta tese é apresentada. O objectivo do autor neste capítulo é o de introduzir o leitor aos principais conceitos relevantes para a compreensão dos capítulos que se seguem, sem contudo entrar em grandes detalhes.

O **capítulo 2** introduz um dos aspectos mais relevantes na bioquímica moderna, a importância da compreensão e caracterização das interacções entre proteínas. Estas interações podem ser estudadas por diversas técnicas, nomeadamente por ressonância magnética nuclear ("NMR"), recorrendo a sequências de impulsos com gradientes de campo. Esta técnica permite a determinação da difusão translacional das moléculas em solução, sendo que os resultados podem ser correlacionados com o estado de agregação das moléculas em estudo. Este capítulo pretende fornecer alguns dos aspectos teóricos da técnica, bem como aspectos a ter em conta aquando da preparação e execução das experiências. Duas proteínas foram utilizadas como exemplos para introduzir os principais conceitos da técnica, nomeadamente a lisozima da clara de ovo de galinha e a δ-toxina.

O **capítulo 3** é um trabalho clássico de "NMR" na determinação da estrutura de um peptídeo (δ-toxina) em diferentes ambientes químicos. Os mecanismos de interacção entre peptídeos citolíticos e membranas celulares, continua a

ser um dos tópicos em debate pela comunidade científica. A estrutura tridimensional destes peptídeos em diferentes ambientes químicos é de extrema importância para a compreensão dos mecanismos infecciosos.

O **capítulo 4** introduz o leitor ao método de dinâmica molecular ("MD") e descreve o desempenho de diferentes campos de força baseados em GROMOS96. O sistema molecular simulado é constituído por uma bicamada lipídica de dimiristoil fosfatidil colina ("DMPC") em solução aquosa. O principal resultado foi a obtenção de um sistema em equilíbrio bem caracterizado, podendo ser utilizado como ponto de partida para simulações moleculares mais complexas.

No **capítulo 5**, a estabilidade conformacional e a propensão para agregação da δ-toxina em diversos ambientes químicos, são estudadas por dinâmica molecular. Uma breve comparação entre os resultados obtidos anteriormente por "NMR" e os resultados de modelação molecular, mostram que as técnicas são complementares na interpretação do comportamento molecular. As etapas iniciais da interacção entre o peptídeo e uma membrana de "DMPC" também foram alvo de estudo.

O **capítulo 6** aborda a temática das doenças amilóides. A transtirretina é uma entre várias proteínas humanas, responsável pela etiologia desta classe de patologias. A modelação molecular de protofilamentos, a unidade estrutural das fibras amilóides, foi o objectivo principal deste capítulo. Auxilidados por dados experimentais, um protocol computacional baseado em "docking" proteico foi estabelecido.

E por último, o **capítulo 7** é uma resenha das conclusões dos capítulos anteriores. Adicionalmente, um comentário pessoal acerca das novas técnicas computacionais baseadas em tecnologia grid, serve de mote para terminar esta tese. Ao mesmo tempo, pretende lançar novas perspectivas para uma melhor compreensão do mundo complexo das proteínas.

List of abbreviations

n-D	n-Dimensional, where n $= 1, 2$ or 3
AG	AGgregation state
AIR	Ambiguous Interaction Restraint
BPPLED	BiPolar Pulse with Longitudinal Encode-Decode
BSA	Buried Surface Area
CD	Circular Dichroism
COM	Center Of Mass
COSY	COrrelated SpectroscopY
DAC	Digital-to-Analog Converter
DQF	Double Quantum Filtered
DMPC	DiMyristoylPhosphatidylCholine
DMSO	DiMethyl SulfOxide
EPR	Electron Paramagnetic Resonance
FID	Free Induction Decay
HADDOCK	High Ambiguity Driven biomolecular DOCking
HEWL	Hen Egg White Lysozyme
MD	Molecular Dynamics
MM	Molecular Mechanics
MRI	Magnetic Resonance Imaging
MSD	Mean Squared Displacement
NearNI	Near-Native Interface
NonNI	Non-Native Interface
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Enhancement SpectroscopY

PFG	Pulsed Field Gradient
PFGSE	Pulsed Field Gradient Spin-Echo
PME	Particle Mesh Ewald
RF	Radio Frequency
RMSD	Root Mean Square Deviation
S/N	Signal to Noise ratio
SPC	Simple Point Charge water model
SS	Secondary Structure
TOCSY	TOtal Correlated SpectroscopY
TTR	TransThyRetin
VT	Variable Temperature
WT	Wild Type

Chapter 1

General introduction

"Beware of the idols of the mind, the fallacies into which undisciplined thinkers most easily fall. They are the real distorting prisms of human nature [...] Stay clear of these idols, observe the world around you as it truly is, and reflect on the best means of transmitting reality as you have experienced it; put into it every fiber of your being."

Wilson (1998)

1.1 The wonder of life

The universe, as we now it today, was formed after a cataclysmic explosion from a primordial hot and dense initial condition at some finite time in the past, the "primeval atom" (Lemaître, 1931). Though the Big Bang theory does not provide any explanation for such an initial condition, it describes the general evolution of the universe since that instant, 14 billion years ago. Within seconds, the simplest chemical elements were formed and as the universe expanded and cooled, material condensed under the influence of gravity to form stars. Elements heavier than hydrogen and helium were produced in the cores of ancient and exploding stars, and spread throughout the universe. About four billion years ago the planet Earth was formed, and as favorable conditions arose a new journey was to begin.

There is no scientific consensus as to how life originated and all proposed theories are highly speculative. However, most scientific models agree that the availability of different chemical elements (trapped on Earth) under certain abiotic conditions, was of primordial importance for life. As the basic molecules of life were formed, a diverse array of living organisms started to populate Earth's biosphere.

All living organisms possess common attributes that distinguishes them from the surrounding inorganic matter (Stryer, 1995). A complex network of interacting molecules is *all* that is needed for the existence and perpetual of life. These molecules are responsible for the microscopic organization; regulation of the interaction between organism's components; extraction, transformation and storage of energy from the environment; adaptability to the surroundings; fidelity of self-replication; and capacity to change over time by gradual evolution. Astonishingly, all living organisms have the same chemical background thus sharing a common evolutionary origin.

1.2 Proteins - a biomolecular surprise

Life is built from the same set of biomolecules, which is a wonder by itself. Nonetheless, one of the most striking surprises on living organisms is a class of biomolecules named proteins. Proteins are the workhorses of life, and have

evolved through selective pressure to perform specific functions. They regulate a variety of activities in all known organisms, from replication of the genetic code to oxygen transport, and are generally responsible for regulating the cellular machinery and consequently, the phenotype of an organism (Branden and Tooze, 1999).

Proteins are mainly built from a small basic repertoire of 20 building blocks, the amino acids. The blueprint for each amino-acid is encoded on base triplets found in the coding regions of genes. These base triplets are recognized by protein building factories on the cell, which create and successively join the amino acids together into a linear chain. The unique sequence of these building units on the polypeptide chain gives each protein its individual characteristics. The interaction between the amino acids on the polypeptide and their surroundings in the cell, produces a well-defined compact three dimensional shape, the folded protein. Thus, the resulting three-dimensional structure is determined by the sequence of the amino acids (Anfinsen, 1972).

For most proteins the correct three dimensional structure is essential to function. Failure to fold into a functional intended shape usually produces misfolded proteins with different properties. In recent years, with the identification of several neurodegenerative and other diseases as protein folding disorders, the importance of the protein folding problem was highlighted (Dobson, 2004).

The answer to the puzzling question of how proteins are able to actively participate in all biological processes, resides in their intrinsic conformational plasticity. The conformational dynamics of proteins is encoded in their three-dimensional structures, and the early view of such biomolecules as relatively rigid bodies (Tanford, 1968) has been replaced by a dynamic model. The internal motions and resulting conformational changes play an essential role in their physiological function, since all biochemical reactions are inherently dynamic processes. A critical assessment of how these molecules work requires an understanding of the connection between the biomolecule structure, obtained mainly from X-ray diffraction studies and nuclear magnetic resonance (NMR), and the dynamics, which is much more difficult to probe experimentally (Karplus and Kuriyan, 2005). Molecular simulations are becoming an extremely useful tool for scientists, providing links between structure and dynamics. These

computational techniques enable the exploration of the conformational energy landscape which is accessible to protein molecules.

1.3 Consilience

The exciting field of modeling molecular systems has been steadily drawing increasing attention from scientists in a variety of disciplines. In particular, modeling large biological polymers is a truly multidisciplinary enterprise (Schlick, 2002). Each branch of knowledge studies a subset of reality that depends on factors studied in other branches. Atomic physics underlies the workings of chemistry, which studies emergent properties that in turn are the basis of biology.

An emerging belief on a unifying perspective for understanding our universe was revived in the 20^{th} century by Wilson (1998). Wilson uses the term consilience[1] to describe the synthesis of knowledge from different specialized fields of human endeavor. The sociobiologist advocates that the world is orderly and can be explained by a set of natural laws that are fundamentally rooted in biology.

If Wilson's vision is correct, the interlocking web of scientific fields will succeed ultimately in explaining the connections between our biological makeup and human behavior. Undoubtedly, an universal effort to integrate the knowledge is needed to decode life as we know it.

1.4 Molecular modeling

A model is a simplified mathematical description of a system or process, used to assist calculations and predictions. Molecular modeling is therefore concerned with ways to mimic the behavior of molecules and molecular systems (Leach, 1998). The ultimate goal is to study molecular structure and

[1] Consilience, or the unity of knowledge, has its roots in the ancient Greek concept of an intrinsic orderliness that governs our cosmos. With the rise of the modern sciences, the sense of unity gradually was lost in the increasing fragmentation and specialization of knowledge observed in the last two centuries.

function through model building and computation. The model building can be as simple as plastic beads and rods, or as sophisticated as three-dimensional interactive space-filling representations.

The roots of molecular modeling began with the notion that various molecular properties, such as molecular geometry and energy, can be calculated from mechanical-like models when subjected to basic physical forces. Typically, molecular models consist of spherical particles (atoms or group of atoms) connected by springs which represent bonds. The molecule then rotates, vibrates, and translates to assume favored conformations in space as a collective response to the inter and intra-molecular forces acting upon it (Adcock and McCammon, 2006).

Since its first steps in the 1960's, molecular modeling has witnessed a major development with the advent of supercomputers. Several different well established computer modeling procedures are nowadays being used as a partner to experimental work. The computations encompass *ab initio* quantum mechanics, molecular mechanics (MM), and many other established procedures. Often forgotten by experimentalists, is the fact that refinement of experimental data, such as from NMR, is also a component of molecular modeling. The main difference between all these computation techniques resides on the mathematical treatment of the molecular model. Small systems can be studied using quantum mechanics theory, while molecular mechanics is used for bigger systems such as biomolecules (Schlick, 2002). The best computational approach is thus determined by the biological problem under study and the computational power available.

The key in modeling is to develop and apply models that are appropriate for the questions being examined by them. A close connection between theory and experiment is essential. Computational models evolve as experimental data become available, and biological theories and new experiments are performed as a result of computational insights. The questions being addressed by computational approaches are becoming more and more complex. They range from understanding the equilibrium structure of small molecules, to the kinetics of protein folding, and the complex functioning of supramolecular aggregates.

In this thesis, both experimental work using NMR and molecular model-

ing techniques were applied to study different aspects of protein and peptide conformational plasticity and aggregation. The next sections will introduce the reader, in a rather general way, both to the biological systems studied and to the techniques employed.

1.5 Case studies

1.5.1 δ-toxin

A large number of naturally occurring toxins are short peptides, of about 20 to 35 residues, generally with a significant number of polar and charged residues. The distribution of charged and hydrophobic residues along the chain is such that the peptide can adopt the conformation of an amphipathic α-helix, having a hydrophobic and a hydrophilic face. Although length, composition and structure of these peptides are similar in a broad sense, they have vastly different specificities in their toxicity (Andreu and Rivas, 1998).

δ-toxin is a peptide secreted by *Staphylococcus aureus* (Kreger et al., 1971) that efficiently lyses eukaryotic cells and does not require binding to cell surface receptors to exert its action. It is soluble in water and it seems to associate with phospholipid bilayers as an amphipathic α-helix (Lee et al., 1987; Thiaudière et al., 1991).

On its way to the membrane, the peptide must experience subtle chemical environment changes with consequences on peptide conformation. The study of the conformational plasticity and aggregation behavior of the peptide should give some clues for the understanding of the first stages of the lytic mechanism.

1.5.2 Transthyretin

A significant number of human diseases, including many neurodegenerative disorders, originate from the formation and deposition of stable, ordered, filamentous protein aggregates, known as amyloid fibrils. In these pathological states, a specific protein (or protein fragment) changes from its natural soluble form into insoluble fibrils, which accumulate in a variety of organs and tissues (Chiti and Dobson, 2006). More than 40 distinct human proteins have been

associated with amyloidoses, with the amyloidogenic potential being enhanced by specific mutations (Pepys, 2006). The most frequent hereditary amyloidoses are caused by the genetic variants of human transthyretin (TTR).

Transthyretin is a homo-tetrameric protein mostly found in the plasma and the cerebral spinal fluid, which has been identified as the causative agent of such diseases as familial amyloidotic polyneuropathy, familial amyloidotic cardiomyopathy, and senile systemic amyloidosis. It is believed that, in the process of amyloid formation, TTR dissociates to non-native monomeric units, which may act as the building blocks of the amyloid fibrils (Brito et al., 2003). The structural characterization of these fibrils and the identification of the entities involved in fibril assembly are crucial for understanding of the mechanisms of pathogenesis in amyloid diseases, and for the development of appropriate therapeutic strategies.

1.6 NMR spectroscopy

Spin is a fundamental property of nature like electrical charge or mass, and particles like protons, electrons, and neutrons do possess it. The world of the nuclear spins is considered to be a true paradise both for theoretical and experimental physicists, and several of them have been honored with Nobel prizes (Ernst, 1997). Nuclear spin systems possess unique properties that predestinate them for molecular studies. For instance, the nucleus can be considered to be a natural sensor, well localized in the atom center, thus providing information about their immediate vicinity. This characteristic fascinated biological scientists, since it allows a non-invasive analysis of biomolecules in their natural milieu.

Nuclear magnetic resonance spectroscopy is a manifestation of the nuclear spin angular momentum, a quantum mechanical property[2]. Associated with the spin angular momentum there is a nuclear spin magnetic moment μ, generating a small magnetic dipole on each nuclei (Cavanagh et al., 1996). When the nucleus

[2]In order to fully understand the magnetic resonance concept and its applications, a quantum mechanical description should be used, beyond the scope of this text.

is placed in a external magnetic field of strength B_0, there is an interaction between μ and the applied field. The energy of the interaction depends on the relative orientation between μ and B_0, being the lowest energy arrangement when both vector quantities are parallel (α-state). The highest energy is achieved when the magnetic moment is anti-parallel to the field (β-state).

The number of nuclei in α and β-states is described by a Boltzmann distribution. The thermal motion of the molecules randomizes the orientation of the individual nuclear magnetic moments. However, there is a small preference for alignment with the external field, giving rise to a bulk equilibrium magnetization of the sample, parallel to the field direction. This magnetization can be represented by a vector called the bulk magnetization vector. By convention in the NMR coordinate system, the external magnetic field and the net magnetization vector at equilibrium are both along the z-axis. The vector model is a classical way of describing how the bulk magnetization interacts with radio frequency (RF) pulses and its time evolution (Keeler, 2006).

When a nucleus is irradiated with RF energy, the lower energy nuclei spin-flip to a higher state through absorption of a photon having a specific frequency. The equilibrium magnetization is thus disrupted and is tilted away from the z-axis. It turns out that the magnetization vector rotates about the direction of B_0 at a frequency equal to the energy difference between the spin states. The rotation of the magnetization about the applied field is called Larmor precession, and its angular frequency is described by the following equation

$$\omega_0 = -\gamma \frac{B_0}{2\pi} \quad \text{, in Hz} \tag{1.1}$$

where ω_0 is the Larmor frequency, and γ is the gyromagnetic ratio.

The β-state is unfavorable, and the relaxation back to the Boltzmann distribution will occur after the initial RF pulse has disturbed the system. In a pulsed NMR experiment, the precession of the magnetization vector is detected in the transverse xy-plane of the spectrometer as a time-domain signal, the free induction decay (FID). The FID is a sinusoidal function encoding all the Larmor precession frequencies present in the sample under study. Usually, a Fourier transform is applied to deconvolute the signal into a spectrum.

A general feature of NMR spectroscopy is that the observed resonance

frequencies depend on the local environment of individual nuclei. The differences in resonance frequencies are referred to as chemical shifts, and are of paramount importance on the process of assigning the spectrum.

The development of NMR as a powerful method in structural biology has involved a number of critical steps: identification of NMR parameters useful for determining molecular structure; development of methods for reliable assignment of the many resonances in the spectrum, to their respective nuclei in the macromolecule; development of methods to measure a sufficient number of structure-related parameters to give the information needed for structure determination; and finally, development of computational techniques that could translate the structural parameters into the *unique* three-dimensional structure (Wüthrich, 1986).

^{1}H nuclei are particular important in NMR experiments on biomolecules because of their natural abundance. On the other hand, even for small proteins the number of ^{1}H atoms can reach a few hundreds, typically resulting in a complex NMR spectrum. Because many resonances are degenerated, the signals are impossible to assign by using one-dimensional pulse-sequences. To effectively utilize the information available from NMR spectroscopy of biological macromolecules, multi-dimensional methods were developed to improve resolution in the frequency domain (Jeener et al., 1979). The introduction of multidimensional NMR spectroscopy allowed the measurement of two important pair interactions in nuclear spin systems, namely the scalar through-bond electron mediated spin-spin interaction (J coupling), and the through-space magnetic dipolar interaction (the nuclear Overhauser effect, NOE) (Ernst, 1997).

The process of translating a multi-dimensional NMR spectrum into a three-dimensional structure is a laborious one. The NOE's provide information about the distance between individual nuclei, but for these constraints to be of any use, the nuclei giving rise to each signal must be identified. For large proteins, it is common to enrich these biomolecules with ^{13}C and ^{15}N isotopes. The coupling of the signals produced by these NMR active nuclei with ^{1}H signals, helps in the assignment process of the spectra. After having the distance constraints between different atoms of the protein under study, this information is coupled with other geometrical constraints such as chirality, van der Waals radii and bond lengths.

Robust algorithms are then applied to minimize an energy function describing the mechanical model. The final result is a family of closely related structures representing the range of conformations consistent with the experimental data acquired from the NMR spectra (Wüthrich, 2003).

High-resolution NMR has become an important tool for studying the structure, dynamics and molecular interactions of biological macromolecules in aqueous solution. At present, NMR stands besides single-crystal X-ray diffraction as the major method for determining the three-dimensional structures of proteins and nucleic acids. Of the total 57133 atomic coordinate sets deposited in the Protein Data Bank (Berman et al., 2000) by April 2009, about 12% had been determined by NMR. Very significant developments have also emerged in the past few decades in the techniques of solid-state NMR and magnetic resonance imaging.

1.7 Molecular dynamics simulations

The various dynamic processes intrinsic to proteins have time scales ranging from femtoseconds to hours, such as bond-length vibrations and protein aggregation, respectively. They also cover an extensive range of amplitudes and energies. Most of these motions have critical roles in biochemical functions, such as enzymatic reactions and folding events (Berendsen and Hayward, 2000).

Molecular simulations are a very powerful toolbox in modern molecular modeling, enabling us to follow and understand structure and dynamics. Molecular dynamics (MD) is one of the methods that can be employed to simulate the natural motion of biological macromolecules, by keeping track of the individual motion of the atoms on the simulated system (Lindahl, 2008). As already pointed out in section 1.4, simulations can provide a way to test whether theoretical models predict experimental observations. Furthermore, simulations provide a level of detail not accessible through experiments, as well as, tools to refine biomolecular structures.

The molecular dynamics simulations involve the calculation of instantaneous forces present in a molecular mechanics system and the consequential movements on that system. The MM system consists of a set of particles that

move in response to their interactions according to the classical equations of motion (Adcock and McCammon, 2006). The position of the n particles as a function of time, a trajectory, is calculated by numerically integrating Newton's equation of motion

$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2} \quad i = 1,2,\ldots,n \tag{1.2}$$

where r_i is the position of a particle with mass m_i at a specific time t. F_i is the force acting on the particle i and depends on the potential energy of the system $V(r_1, r_2, \ldots, r_n)$.

Given a set of initial coordinates and velocities for the particles in the system, a trajectory can be computed based on the intermolecular interactions of all the particles. From this trajectory many dynamic and thermodynamic properties of the system can be evaluated assuming that the trajectory is an effective representation of the phase space of the system.

The accuracy of any MD simulation will be primarily determined by the accuracy of the potential energy function of the system. This energy function is comprised of a collection of simple functions to represent a minimal set of forces that describe the molecular structures, namely non-bonded (Van der Waals and electrostatic) and bonded (stretching, bending and torsions) interaction terms. These terms are obtained both from experimental and quantum mechanical studies of model compounds. The set of functions and parameters used to describe the interactions of the atoms in a system is called a force field (Mackerell, 2004).

The increasing quality of empirical force field based MD simulations on biomolecules allowed a general acceptance of the method in the scientific community. This acceptance enhanced the atomic description of systems that are currently not accessible to experiment. A wide range of simulations of proteins in solution have been performed providing insights into structure function relationships as well as helping to understand intrinsic dynamic processes such as the mechanism of protein aggregation (Kelley et al., 2009). Simulations of lipid molecules have also been the subject of extensive research over the last decade, both in pure lipid bilayers as well as in more realistic systems such as lipid rafts.

1.8 High ambiguity driven biomolecular docking

Protein-protein interactions play a central role in biochemistry. This can be seen in cell-signaling cascades, enzyme catalysis, the immune response by means of antibody-antigen interactions, and misfolding diseases, just to mention a few examples.

The structure of the macromolecular complex of two interacting proteins provides insights on their functional mechanisms and roles in the cell. Information about which specific residues are involved in the interaction and the degree of conformational change undergone by proteins upon binding can be readily retrieved (Wiehe et al., 2008). Although X-ray crystallography and NMR are the experimental techniques of choice to determine the three-dimensional structure of proteins and protein complexes, in many instances, due to the size of the macromolecular complexes and other experimental limitations, computational methods are required to characterize protein-protein interactions.

In recent years, several methods based on docking emerged as new tools to model biomolecular complexes, namely protein-protein complexes. Protein-protein docking is defined as the prediction of the structure of two (or more) proteins in a complex, given only the structure of the interacting proteins. Most of the available docking methods are based on a two stage protocol. The initial stage treats the proteins as rigid-bodies, while allowing for an efficient search of the six-dimensional space. One protein is fixed in space and the second one is rotated and translated around the first one. For each protein-protein configuration, a score is calculated based on a combination of energetics and shape complementarity terms. A final stage with refinement and re-ranking of the complexes, increases the rank of near-native structures while decreasing the rank of false-positives. This involves making small changes to the highest-scoring predictions obtained on the rigid-body stage, like increasing the flexibility of docking partners (Andrusier et al., 2008).

Biochemical and biophysical experiments are widely used to gain insight into biomolecular interactions (fluorescence studies, resonance energy transfer, etc.). Hence, there is a wealth of experimental sources that can provide information about interfaces of biomolecular complexes. However, these data are generally not used in the computational docking protocols. HADDOCK

(*H*igh *A*mbiguity *D*riven biomolecular *DOCK*ing) is a docking approach that makes use of such interaction data to drive the docking (Dominguez et al., 2003). The information is encoded as ambiguous interaction restraints (AIRs), similar to the ambiguous restraints commonly used in NMR structure calculation (Wüthrich, 1986). These restraints are defined between any residue which, based on experimental data, is believed to be at the interface (active residue), and neighbor residues (passive residues) on the partner molecule.

The HADDOCK protocol consists of three stages during which increasing amounts of flexibility are introduced. The AIRs are incorporated as an additional energy term to the energy function minimized in all docking steps. In the first stage, randomization of orientations (sampling) and rigid body energy minimization is accomplished. In the next stage, a limited amount of flexibility is introduced in docking partners using a semiflexible simulated annealing schema. Flexibility is first introduced into the side chains and subsequently into both side chains and backbones of predefined flexible segments encompassing the active and passive residues. Finally, the solutions are refined in explicit solvent, clustered, and scored using a combination of energy terms, mainly intermolecular van der Waals and electrostatics energies, as well as restraints energies (van Dijk et al., 2005).

The modeling of protein-protein complexes by means of docking has become increasingly popular, as witnessed by the CAPRI (*C*ritical *A*ssessemnt of *PR*edicted *I*nteractions) experiment (Méndez et al., 2005). Stimulation of research in biomolecular docking provides ground to evaluate prediction methods and assess their reliability. Docking is a powerful tool which scientists can use to predict interactions, and thus to assist in drug design protocols.

Chapter 2

Probing protein aggregation in solution by PFG NMR

Anjo da guarda,
Minha companhia,
Guardai a minha alma
De noite e de dia.

2.1 Introduction

Proteins play key roles in virtually all biological processes, namely enzymatic catalysis, transport and storage, coordinated motions, mechanical support, immune protection, excitability, integration of metabolism, and the control of growth and differentiation (Stryer, 1995). All these actions are achieved through an amazing ability proteins have to recognize and interact with highly diverse molecules. This ability is ultimately determined by their conformational plasticity allowing them to regulate complex cellular processes *in vivo*.

Protein association, i.e. protein-protein interactions, is of central importance either under normal physiological conditions as well as in disease states. Recent developments showed that protein aggregation into assemblies rich in β-sheet structure is linked to a series of severe disorders, including Alzheimer's and Parkinson's diseases. Additionally, an increasing interest in this phenomenon comes from the possibility of using highly ordered cross-β protein aggregates, known as amyloid fibrils[1], as novel high-performance and versatile nanomaterials (Tartaglia and Vendruscolo, 2008). Hence, the understanding and characterization of protein interactions are one of the major goals on modern biochemistry, with an enormous importance in industry, medicine and nanobiotechnology.

The first step toward protein structure determination using nuclear magnetic resonance (NMR) is elucidation of appropriate solution conditions that obviate any self-aggregation. Knowledge of the oligomerization state of a protein prior to data collection is rather important, due to the complexity associated with assigning and identifying long-range nuclear Overhauser contacts in the process of determining the three-dimensional structure (Krishnan, 1997).

Biomolecular interactions are generally non-covalent and thus weak, thereby limiting the techniques that can be used to probe molecule association. In general, the oligomerization state of a protein of interest is characterized using biophysical methods like equilibrium ultracentrifugation or dynamic light scattering measurements. Even though the obtained results are reliable, these measurements are done at a protein concentration about three orders of mag-

[1] This subject will be further discussed on chapter 6.

nitude lower than the concentration used for the NMR experiments. Thus, despite a long history of study, the extent, kinetics and mechanisms of protein association remain poorly understood at NMR sample conditions (Price, 2006).

The meteoric rise in sophistication of NMR techniques in the last decade and the availability of commercially good performance spectrometers, and their concomitant ability to provide detailed information on binding of biomolecules, augur well for rapid advances in our understanding of this area of science. NMR, due to its non-invasive nature (most biological molecules are composed of NMR sensitive nuclei), has special significance for measuring diffusion[2] as it does not perturb the thermodynamics of the system or require labeled probe molecules (Price, 1998a). In theory any NMR observable that is affected by the association process can be used as a probe of the association kinetics. Amongst the NMR observables, translational diffusion rates are quite sensitive to structural changes and to binding and association phenomena (Stilbs, 1987).

Translational diffusion measurements by NMR have been utilized in numerous studies ever since the discovery of spin echoes by Hahn (1950). In that pioneering study, the author noticed that the formation of echoes could be disrupted by diffusion in an inhomogeneous magnetic field. Normally we take a great deal of effort to make sure that the applied magnetic field is as homogeneous as possible across the sample. However, the ability nowadays to make the magnetic field inhomogeneous through the usage of pulsed (magnetic) field gradients (PFG), opened alternative ways to make science. Some applications include the determination of the aggregation state of proteins (Altieri et al., 1995), the degree of ligand binding to proteins (Derrick et al., 2002), the characterization of protein folding (Jones et al., 1997), the investigation of polymer mixtures (Antalek et al., 2002), among others. As is often the case in NMR, phenomena that at first sight seem to be only the cause of artifacts are later found to have practical applications (Price, 1996).

This chapter will focus on the application of PFG NMR techniques to

[2]To avoid misunderstandings, whenever the word *diffusion* is employed (e.g. translational diffusion), we are talking about self-diffusion, a umbrella term describing all of the diffusion processes associated to a particular molecule (Lobo, unpublished).

measure translational diffusion of proteins. A in depth explanation of the theoretical background will be given, though far from complete, along with relevant methodological issues on how to setup, perform and analyze PFG experiments.

2.2 Theoretical synopsis

2.2.1 Diffusion

Molecules in solution are in constant motion and experience both rotational and translational movements. The process of translational motion in solution is commonly referred to as self-diffusion and is characterized by the self-diffusion coefficient D. This random motion is driven by the internal kinetic energy of the system and is the most fundamental form of transport responsible for all biochemical reactions (Price, 1997). The molecular interpretation of the diffusion coefficient was "first" put forward by Einstein (1905). He found the simple relationship

$$D_0 = K_B T / f \tag{2.1}$$

where D_0 is the diffusion constant at infinite dilution of the diffusing species, K_B is the Boltzmann constant, T is the temperature, and f is the friction constant or frictional coefficient. For the simple case of a spherical particle with an effective hydrodynamic radius r, in a solution of viscosity η, the friction factor is given by the Stokes's law

$$f = b\pi\eta r \tag{2.2}$$

The dimensionless parameter b reflects the boundary condition between the moving particle and the solvent. However, most macromolecules of biological relevance have more complicated shapes and may include contributions from factors such as hydration. The friction factor thus becomes a more complicated mathematical function (Cantor and Schimmel, 1980).

The Stokes-Einstein equation[3], which results from the direct substitution of equation 2.2 into the Einstein relation, does not include the effects of interparticle interactions. Nevertheless it provides a strongly intuitive framework for analyzing diffusion data. In particular, it shows that the translational diffusion relates directly to a molecule's size, shape and environment. As a consequence, measuring diffusion is a straightforward mean for probing association.

The relation between the diffusion coefficient D and the displacement z of an ensemble of particles undergoing free isotropic diffusion is given by (Crank, 1975)

$$D = \lim_{t\to\infty} \frac{1}{n\mathrm{d}t} \left\langle z^2(t) \right\rangle \tag{2.3}$$

hence, for long enough times t

$$\left\langle z^2(t) \right\rangle = nDt \tag{2.4}$$

where $n = 2$, 4 or 6 for diffusion in one, two or three dimensions, respectively.

In order to measure the translational motion, enough time must be allowed to pass in the experiment for z to be several times larger than the hydrodynamic radius r. This point is particularly important for large molecules such as proteins. If the interval of time used is too short, we may actually measure the translation of a chain segment or the rotational diffusion rather than the translational diffusion (Antalek, 2002).

Because of its non-invasive nature, NMR spectroscopy is a unique tool for studying molecular dynamics in biological systems. There are two main ways in which NMR may be used to study the system dynamics, namely by analysis of relaxation data and by pulsed field gradient NMR (Price, 1997). The relaxation method is sensitive to motions occurring in the picosecond to nanosecond time scale, thus it probes motions on the time scale of re-orientational correlation of the nucleus (rotational diffusion). In the PFG measurements, motion

[3]It is not known why the Stokes-Einstein equation, $D = \frac{K_B T}{6\pi\eta r}$, is not known today as the Stokes-Einstein-Sutherland equation instead. In fact, Sutherland (1905) published a similar relationship for the interpretation of the Brownian motion some months prior to Einstein.

is measured over the millisecond to second time scale, characteristic of the translational motion. Though a diffusion coefficient can be derived from both methods, in the relaxation method a number of serious assumptions must be taken into account, which makes the determination of D not straightforward when compared with the PFG method.

2.2.2 Magnetic gradients

In a PFG NMR enabled machine, along with the static field B_0 produced by the superconducting magnet, a set of coils in the probe will be responsible for generating a field gradient across the sample. The simplest commonly used geometry for producing gradients along a given direction, is the Maxwell pair of coils (Price, 1997). These coils are positioned coaxially outside the radio frequency (RF) coil carrying current in opposite directions. If these coils are z oriented with respect to B_0, they will create a situation where the magnetic field strength is added to the top of the sample volume and subtracted from the bottom, or *vice versa* (see figure 2.1). In other words, along the length of the sample during a gradient pulse there is a uniformly changing magnetic field B_z. This magnetic field, due to the combination of the gradient and the applied field B_0, can be written as

$$B_z = B_0 + G_z z \tag{2.5}$$

where G_z is the magnetic field gradient strength, and z is the coordinate along the field direction. Conceptually, the NMR sample can be envisaged as a column of thin slices, called isochromats, along the z direction. The spins in a given slice experience the same B_z magnetic field (Cavanagh et al., 1996).

Magnetic resonance imaging (MRI) usually makes use of magnetic gradients created on all three x, y, and z directions of the imaging volume (Callaghan, 1991). However, in normal NMR spectrometers it is more common to do one-dimensional (1-D) studies, where the gradient is aligned along the z axis, i.e. parallel to the main field B_0.

There are several technical features designed into the gradient coils, namely shielding and field linearity, which makes them suitable for PFG NMR ex-

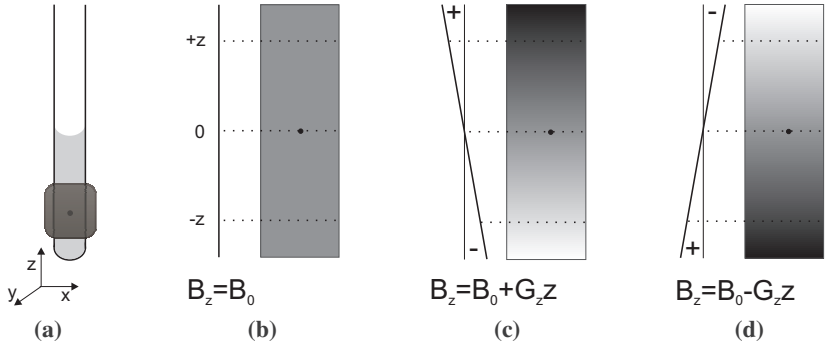


Figure 2.1: Diagrammatic representation of the magnetic field gradient felt by spin isochromats in a standard NMR tube. **(a)** By convention, the NMR tube is oriented along the spectrometer main field B_0, i.e. parallel to the z-axis. Typically, only a small region of the sample is actually excited and detected by the RF coil. This sensitive volume (1–2 cm long), is shown in the diagram by a highlighted gray rectangle. Depending on the probe being used, extra coils can create a magnetic field which varies linearly (i.e., a gradient) along the direction of B_0. The magnetic field B_z due to the combination of the static field B_0 and gradient field, can be written as $B_z = B_0 + G_z z$. **(b)** When the magnetic field applied to the sample sensitive volume is homogeneous, all spin isochromats experience the same field $B_z = B_0$ (for representation purposes, isochromats at $+z$ and $-z$ have the same grayish color to indicate that $B_z^{+z} = B_z^{-z} = B_0$). If current is allowed to flow through the z-aligned gradient coils with a power strength G_z, then B_z varies linearly throughout the sample, i.e. along z-axis. Depending on the polarity of the current flow, **(c)** B_z increases as we go along the positive z-direction (darker spin isochromats) or **(d)** along the negative z-direction. Usually the extra field due to the gradient is zero in the middle of the sample, $z = 0$.

periments. Whenever an electrically conducting material, such as the probe, experiences a changing magnetic field, an electrical current is set up in nearby conducting surfaces (i.e. the probe body), creating a secondary field that opposes the change (Antalek, 2002). This is problematic because the secondary field interferes with the acquisition of the free induction decay (FID). The other design criterion is gradient field uniformity. The applied gradient must be constant (uniform) along the z direction, that is, a constant change in the magnetic field exists at all points in the z direction. The magnetic field changes linearly, but the gradient is constant as it can be seen from equation 2.5. A more detailed analysis of these problems will be discussed in section 2.4.

2.2.3 Spatial-spin labeling through magnetic gradients

All of the NMR theory needed for understanding the effects of B_0 gradients on nuclear spins has the Larmor equation as the origin (see equation 1.1, page 9). Since B_0 is spatially homogeneous, the Larmor frequency ω is the same through out the sample. However, if in addition to B_0 there is a spatially dependent magnetic field gradient G_z, and accounting for the possibility of more than single quantum coherence transitions, ω becomes spatially dependent,

$$\omega_z(n) = n(\omega_0 + \gamma G_z z) \tag{2.6}$$

where z is the position of the spin from the observed nucleus, and γ is the gyromagnetic ratio. The important point in this equation is that if a homogeneous gradient of known magnitude is imposed throughout the sample, the Larmor frequency becomes a spatial label with respect to the direction of the gradient.

In the case of a single quantum coherence ($n = 1$), we can see from equation 2.6 that for a single spin the cumulative phase shift is given by (Price, 1997)

$$\phi_i(t) = \gamma B_0 t + \gamma \int_0^t G_z(t') z_i(t') dt' \tag{2.7}$$

where the first term corresponds to the phase shift due to the static field, and the second term represents the phase shift due to the effects of the gradient. Thus, from the second term of equation 2.7 we can see that the degree of dephasing due to the gradient is proportional to the type of nucleus (γ), the strength G_z of the applied gradient, the duration t of the gradient, and the displacement z of the spin along the direction of the gradient.

2.2.4 Measuring diffusion with magnetic field gradients

From equation 2.6 is apparent that a magnetic field can be used to label the position of a spin, through the Larmor frequency. This is in fact the basis for measuring diffusion.

Pulsed field gradient spin echo (PFGSE) NMR is a method devised by Stejskal and Tanner (1965), and it is derived from the nuclear spin echo concept

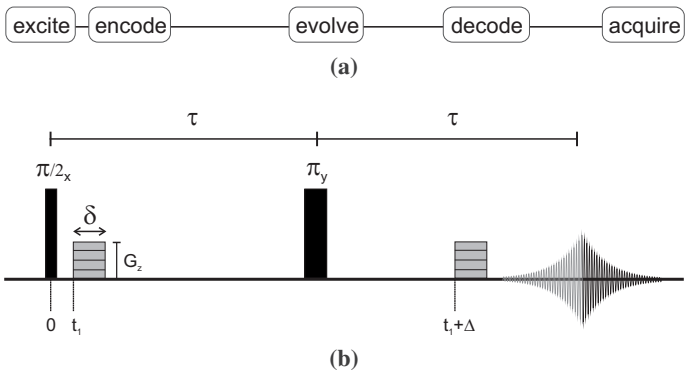


Figure 2.2: (a) The essential components of a PFGSE NMR diffusion experiment. **(b)** The Stejskal and Tanner PFGSE pulse sequence.

of Hahn (1950) and Carr and Purcell (1954). This is the most common approach to measure particle diffusion by NMR, and it is depicted in figure 2.2. Two equal rectangular gradient pulses of duration δ are inserted into each τ period. These gradient pulses, as explained before, are essential for interrogating the effects of translational motion on the signal intensity.

An ensemble of diffusing spins at thermal equilibrium has a net magnetization oriented along the z-axis. After exciting the sample with a $\pi/2$ RF pulse the macroscopic magnetization is instantaneously rotated from the z-axis to the xy plane, and the phase of the spins is coherent. During the first τ period of time, at $t = t_1$, a gradient pulse of duration δ and magnitude G_z is applied, labeling the position of the spins by producing a spatially dependent phase angle. That is to say, the Larmor frequency varies uniformly along the z direction during the gradient pulse, and each plane of the sample perpendicular to the z direction contain spins that will be affected by the gradient pulse in exactly the same way. At the end of the first τ period, spin i experiences the following phase shift (Price, 1997)

$$\phi_i(\tau) = \gamma B_0 \tau + \gamma G_z \int_{t_1}^{t_1+\delta} z_i(t)dt \tag{2.8}$$

During the evolution period and because the spins are always undergoing random translational motion in solution, the extent of which is defined by the self-diffusion coefficient, some will change position along the z direction. At the

end of the first τ period, a π RF pulse is applied and has the effect of reversing the sign of the phase angle. At time $t_1 + \Delta$, the spin position is decoded with an identical gradient pulse. The total phase shift of spin i relative to the $z = 0$ position, at the end of the echo sequence, is given by

$$\phi_i(2\tau) = \gamma G_z \left\{ \int_{t_1}^{t_1+\delta} z_i(t)dt - \int_{t_1+\Delta}^{t_1+\Delta+\delta} z_i(t')dt' \right\} \tag{2.9}$$

If the spins have not undergone any translational motion with respect to the z-axis, the effects of the two gradient pulses cancel and all spins refocus, $\phi_i(2\tau) = 0$. A maximum NMR signal is obtained at $t = 2\tau$. However, if diffusion takes place, some spins are not in the same position along the z-axis during the second gradient pulse as during the first. Therefore, their phase component imposed by the first gradient is not cancelled by the second gradient and the signal diminishes. The degree of dephasing due to the applied gradient is proportional to the displacement in the direction of the gradient in the period Δ, and thus $\phi_i(2\tau) \neq 0$.

From this qualitative description of the Stejskal and Tanner (1965) PFGSE NMR sequence, it is clear that the gradient should be constant for the entire excited region so that the same gradient strength is experienced no matter where the spin is. Additionally, the gradient pulses must have exactly the same strength and length for the sequence to correctly probe the translational motion of spin i.

The equation that is derived to represent the signal intensity generated by the pulse sequence in figure 2.2b is

$$S(2\tau) = S(2\tau)_{G_z=0} \exp\left[-D\gamma^2 G_z^2 \delta^2 (\Delta - \delta/3) \right] \tag{2.10}$$

where $S(2\tau)_{G_z=0}$ represents the attenuation of the echo signal due to relaxation in the absence of a magnetic field gradient, which in the present case corresponds to $S(2\tau)_{G_z=0} = S(0)\exp[-2\tau/T_2]$. S(0) is the signal without attenuation due to relaxation[4], that is, the signal that would be observed immediately after

[4] $S(0) \propto M_w n$ where M_w is the molar mass of the diffusing specie and n is the number of such molecules present (Price, 2000).

the $\pi/2$ pulse, 2τ is the total echo time, and T_2 is the spin-spin relaxation time of the species. It is assumed here that the observed signal results from one population with a single relaxation time T_2. To remove the effects of the signal attenuation due to spin-spin relaxation, equation 2.10 is normalized with respect to the signal obtained in the absence of the applied gradient and thereby the echo attenuation is defined as

$$E(2\tau) = \exp\left[-D\gamma^2 G_z^2 \delta^2 (\Delta - \delta/3)\right] \tag{2.11}$$

By inspection of equation 2.11, it can be seen that to measure diffusion, a series of experiments are performed in which either G_z, δ, or Δ is varied while τ is generally kept constant so that bulk relaxation effects may be factored out.

An obvious way to analyze the data is simply to plot $\ln\left[E(2\tau)\right]$ *versus* $\gamma^2 G_z^2 \delta^2 (\Delta - \delta/3)$, in which case the diffusion coefficient can be obtained from the slope $-D$.

2.2.5 PFGSE: considerations for proteins

Polydisperse systems are commonly encountered when studying both natural (e.g. proteins) and synthetic polymers (Zhou, 1995). For a discrete protein system with i different aggregation species (i.e. monomer, dimer, ..., i-mer) undergoing isotropic free diffusion, at infinite dilution, the observed echo intensity from the Stejskal and Tanner sequence would be given by the following equation instead of equation 2.10 :

$$S(2\tau) = \sum_i \left(S(0)_i \exp\left[-2\tau/T_{2,i}\right] \exp\left[-D_i\gamma^2 G_z^2 \delta^2 (\Delta - \delta/3)\right] \right) \tag{2.12}$$

where the subscript i refers to the aggregation state of the contributing species. The identification of the other variables in this mathematical relation is the same as the one introduced for equation 2.10. This equation is valid if the exchange between the different species is slow on the PFGSE NMR timescale.

A closer inspection of equation 2.12 shows that the observed echo signal $S(2\tau)$ is not weighted by the respective concentrations alone (i.e. $S(0)_i$; see footnote 4) (Price, 1996). The PFG signal attenuation is modulated by three

contributions, namely a distribution of diffusion coefficients D_i, the population of each diffusing species $S(0)_i$, and different relaxation rates $T_{2,i}$.

A problem when performing diffusion measurements of macromolecular systems is that of cross-relaxation (Peschier et al., 1996) . Severe problems with the Stejskal and Tanner sequence arise from the fact that the magnetization evolves entirely in the xy plane (see figure 2.2b). The longest value of diffusion time Δ allowed for the study of macromolecular diffusion is determined by the rate of signal loss due to spin-spin relaxation T_2 of the species over the echo time. Generally, for macromolecules the spin-lattice relaxation time T_1 is much longer than T_2 (Cavanagh et al., 1996), thus pulse sequences that minimize the time spent by spins on the xy plane are preferred for macromolecules. Furthermore, J-modulation effects can degrade the quality of the final spectrum. Because the signal for coupled spins is modulated in the *xy* plane, the final spectrum may contain dispersion peaks (i.e. with positive and negative components), thus strongly affecting the $S(2\tau)$ measured (Keeler, 2006). These problems can be minimized by prudent choice of the pulse sequence. A pulse sequence much more reliable for the study of diffusing protein species will be discussed in section 2.4.1.

If there is a negligible relaxation time difference between the different i-mers in equation 2.12, then relaxation can be normalized out as discussed in section 2.2.4:

$$E(2\tau) = \frac{\sum_i \left(M_{w,i} n_i \exp\left[-D_i \gamma^2 G_z^2 \delta^2 (\Delta - \delta/3) \right] \right)}{\sum_i M_{w,i} n_i} \tag{2.13}$$

where $M_{w,i}$ is the molar mass of the i-th aggregate species and n_i is the number of such molecules in solution. However, if the relaxation properties are significantly different, the difference can be used to detect that the system is not homogeneous (van Dusschoten et al., 1995). On the other hand, if exchange is fast on the NMR timescale the multi-exponential equation 2.13 becomes a single exponential (cf. with equation 2.11),

$$E(2\tau) = \exp\left[-\langle D \rangle_w \gamma^2 G_z^2 \delta^2 (\Delta - \delta/3) \right] \tag{2.14}$$

where $\langle D \rangle_w$ is the mass averaged diffusion coefficient defined by

$$\langle D \rangle_w = \frac{\sum_i M_{w,i} n_i D_i}{\sum_i M_{w,i} n_i} \tag{2.15}$$

Due to the large size of the molecules and thus a large number of nuclear spins, the NMR spectra of proteins are densely crowded with resonance lines (e.g., see figure 2.14). Additionally, peaks from several species present in solution may also overlap, thus making peak assignment a conundrum. For well resolved spectra or with partial signal frequency overlapping, peak deconvolution can be readily made (Antalek, 2007), but such strategy is not feasible for proteins. Protein diffusion measurements are typically performed by integrating over the aliphatic region (i.e. $\sim [0,4]$ ppm). The very small population of exchangeable resonances in this region results in a nearly imperceptible difference in the measured diffusion coefficient. Nevertheless, the PFGSE signal attenuation will have contributions for every i-mer species.

NMR protein solutions are also crowded systems, thus during the diffusion period in a PFGSE measurement, protein molecules do collide with each other often (see section 2.5.6). Obviously, even in the absence of aggregation, the measured diffusion coefficient of any oligomeric species will decrease due to self-obstruction as the protein concentration increases.

2.3 Experimental details

Sample preparation

δ-toxin was kindly provided by H. Birkbeck (University of Glasgow, UK) and its purification was described previously (Birkbeck and Freer, 1988). An NMR sample was prepared by dissolving lyophilized δ-toxin in methyl-d_3 alcohol-d (CD_3OD, 99.8 + atom % in D) to a final concentration of 1 mM.

A suitable amount of hen egg white lysozyme (HEWL; 95% ww pure, balance primarily sodium acetate) was dissolved into dilute HCl (pH 3.0) and dialyzed extensively against dilute HCl (pH 3), at 277 K, using 6000 molecular weight cutoff dialysis tubing from Spectrum Medical Industries Inc., USA. Dialyzed HEWL was quickly frozen at liquid N_2 temperature (72 K) and

lyophilized. Exchangeable hydrogens in HEWL were deuterated by dissolving the lyophilised protein in deuterium oxide (D_2O, 99.9 atom % in D) and the pH was adjusted to 3.1 by addition of small amounts of DCl, without correction for isotope effects. The protein solution was heated in a thermostat bath at 353 K for 45 minutes. Deuterated lysozyme was lyophilized as before and then stored at 253 K until used. A NMR sample was prepared by dissolving the stored HEWL powder in D_2O to a final concentration of 1.5 mM, pH 3.1, at 298 K.

Sample concentrations were measured by absorbance spectroscopy at 280 nm. An estimation for the molar absorption coefficient ε of δ-toxin was taken to be 5500 $M^{-1}cm^{-1}$ (see Pace and Schmid, 1997, equation 2) whereas $\varepsilon_{HEWL} =$ 37646 $M^{-1}cm^{-1}$.

Some of the tests described in this chapter, while setting up the spectrometer, used D_2O (99.9 atom % in D) or doped D_2O (99.9 atom % in D + $GdCl_3$) samples. All reagents were purchased from Sigma Chemical Company, St. Louis, USA.

Protein samples of 150 μL for the PFG experiments were transfered to 5-mm susceptibility-matched microtubes (Cat. BMS-005 from Shigemi, Tokyo), with 12-mm bottom matched glass against the respective solvent. Preparative D_2O samples of 500 μL were placed either in a standard 5-mm NMR tube (Wilmad, USA) or Shigemi microtubes. The position of the tubes in the NMR spinner turbines was adjusted so that the sample was in the most linear region of the magnetic field gradient.

PFG NMR spectroscopy

^{1}H NMR measurements were obtained at 298 K on a Varian ^{UNITY}INOVA 500 MHz spectrometer equipped with a 20 A Highland Model L700 gradient amplifier (Performa II) and a dual-broadband 5-mm direct detection PFG probe. The combination provides a z gradient strength G_z of up to 0.6 Tm^{-1}. Either a Stejskal and Tanner (1965) or a PFG longitudinal encode-decode pulse sequence incorporating bipolar gradient pulse pairs (BPPLED) (Wu et al., 1995) were used. Typically, a value of 12–100 ms is used for Δ, 3 ms is used for δ and τ, and 1–500 ms is used for τ_e. Diffusion coefficients were measured by linearly incrementing G_z, typically with an initial value of 0.019 Tm^{-1} up to 0.6 Tm^{-1}

in 20–32 steps. The combination of G_z, Δ, and δ was chosen to obtain 90% total signal attenuation throughout the experiment.

The original Varian bipolar gradient pulse pairs sequence (bppste) was edited in order to add a pre-saturation scheme. Low-power pre-saturation in the last 1.5 s of the recycle delay was used in first trial acquisitions of HEWL on 90% H_2O / 10% D_2O samples, to suppress the water resonance, but baseline distortion was evident. We also wrote a pulse sequence to determine a safe τ_e delay prior to FID acquisition in order to avoid eddy currents, based on Price (1998b, figure 5; see appendix A, page 195).

All FIDs were acquired using a spectral width of 4 or 8 KHz, digitized into 16 or 32 K data points, and 64–256 transients were co-added to obtain adequate signal-to-noise (S/N) ratios in the final spectra. A recycle delay of 10 s sufficient to allow for full relaxation, i.e. $> 5 \times T_1$, was allowed between each transient.

The FIDs were transfered to a Silicon Graphics Octane2 workstation and processed using the Varian VNMR 6.1B software package. All FIDs were zero filled to 132 K data points, apodized by multiplication with a 1 Hz line broadening function, and Fourier transformed. The spectra were baseline corrected by fitting selected baseline points to a 2^{nd} order polynomial function.

All PFG NMR experiments were done on non-spinning conditions. Deuterium lock was used in every experiment. The temperature in the NMR probe was calibrated using a 100% methanol sample (Raiford et al., 1979).

Calibration of B_0 field gradient strength

The gradient coil constant, 1.890×10^{-5} $Tm^{-1}DAC^{-1}$, was firstly calibrated generating a spatial profile of a doped D_2O sample with the field gradient on during acquisition ($G_z = 0.15$ Tm^{-1}), according to the manufacturer's instructions. A more accurate B_0 field gradient strength was obtained by measuring the self-diffusion coefficient of the residual HDO in a D_2O (99.9%) sample at 298 K. A diffusion coefficient of 1.90×10^{-9} m^2s^{-1} (Longsworth, 1960) for the HDO signal was used for back calculation of the gradient strength from a single exponential decay fitting procedure.

Self-diffusion calculation

Diffusion coefficients were obtained by fitting peak volumes to a single exponential decay described by equations 2.11 (Stejskal and Tanner experiments) and 2.16 (BPPLED experiments). To obtain a sufficient signal-to-noise ratio, the diffusion coefficient of δ-toxin and HEWL were determined using an integral of the aliphatic region, $\approx$ 0.5–1.8 ppm and -1.4–2.4 ppm respectively. No significant differences on calculated diffusion coefficients were observed by fitting peak volumes over aliphatic regions with slightly different limits.

Non-linear regression analysis was performed using the nonlinear least-squares fitting routine of OriginPro v.7.0 (OriginLab Co., USA) which is based on the Levenberg-Marquardt algorithm (Press et al., 1992).

The diffusion coefficients were measured as soon as possible after NMR sample preparation to minimize errors associated with protein aggregation and precipitation.

2.4 Magnetic gradients and associated problems

To achieve a reliable analysis from PFG NMR data, we must reduce or eliminate any experimental artifacts. The PFGSE NMR experiment is susceptible to several artifacts, all of which are manageable with proper care and consideration. There are essentially three problems that give rise to artifacts in the acquired spectra (Antalek, 2002): eddy currents, gradient field non-uniformity, and convection currents.

A brief discussion of these topics will be given in the next sections, along with some general notes on how to minimize their effects. Their significance should be perfectly understood by the NMR spectroscopist.

2.4.1 Eddy currents

An eddy current (also known as Foucault current) is an electrical phenomenon caused when a conductor is exposed to a changing magnetic field. This can cause a circulating flow of electrons within the conductor. These circulating eddies of current create electromagnets, with magnetic fields that

opposes the change of the magnetic field. The stronger the applied magnetic field the greater the currents developed and the greater the opposing field. The fast switching of the gradient pulses in a PFGSE NMR experiment is thus prone to produce such eddy currents within any closely located conductor. The effect produces a magnetic field that can be experienced by the sample and therefore distortion may occur in the spectra[5].

Modern coils are designed in a way to minimize this problem. Actively shielded coils are manufactured placing several coils around and in series with the primary Maxwell pair that produce the magnetic gradient. This way, an equal but opposite field outside the main coils is produced. These two fields cancel each other and the eddy currents that would normally form in the probe body are minimized. Unfortunately no design is perfect and eddy currents, although greatly reduced, are still present.

In addition to directly affecting the spectral quality, eddy currents can also have an effect on the locking mechanism. The lock circuitry is designed to compensate for small changes in the main B_0 magnetic field. Depending upon the time constants built into the circuitry, it may have a response to the applied gradient pulse and even produce shifts in the resonances.

To measure the time that is required for the eddy currents to fully decay, a simple pulse sequence can be devised such as the one shown in figure 2.3a. In this sequence, a gradient pulse is first applied. After a delay τ_e, an RF pulse is applied and a spectrum is acquired. Spectra are acquired with successively shorter τ_e delays to determine the minimum time required for the eddy current effects to decay. The result of this experiment is shown in figure 2.3b for a doped water sample. Each spectrum in figure 2.3b is acquired with a different delay time τ_e. All of the spectra were phased using the phasing parameters of the last spectrum in the series. It is evident from the results that the impact of the eddy currents for the specific probe used vanishes for $\tau_e \geq 200$ ms, with recovery of the full signal strength. This means that one must wait a period of

[5]Though eddy currents are deleterious for PFGSE NMR experiments, they are not always undesirable, as there are some practical applications. Among the most common applications are the magnetic levitation, electrical generators and electromagnet brakes.

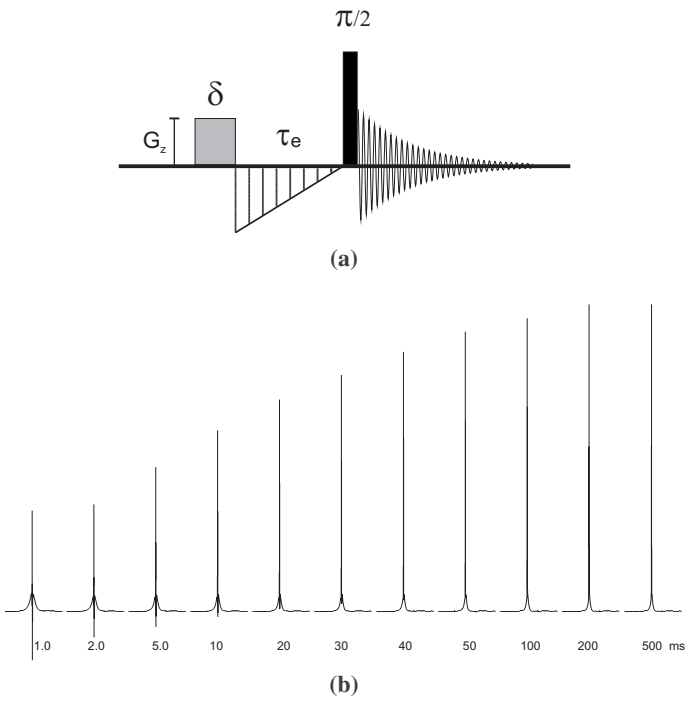


Figure 2.3: **(a)** The `checkeddy` pulse sequence to determine the minimum time necessary for the effects of eddy currents to dissipate. **(b)** Experimental spectra for a sample of doped D_2O at 298 k acquired using the `checkeddy` pulse sequence for t_e values ranging from 1 to 500 ms. The gradient pulse used had a duration δ of 1.5 ms and a strength G_z of 0.6 Tm^{-1}.

time $\geq$ 200 ms after the application of the final gradient in the pulse sequence, and before data acquisition. One must not forget that the results vary widely depending on the particular probe used, since probes have different shielding effectiveness.

Many articles have been published that discuss ways to minimize this ring down time. All of them deal with either changing the shape of the applied gradient pulse or using a composite RF gradient pulse. The advantage of changing the gradient shape is that a gentler rise and fall of the gradient magnitude will lessen the intensity of the eddy currents. The disadvantage is that such a gradient may be less effective than a square gradient pulse.

A better alternative for high resolution work is the use of composite pulses. Figure 2.4 shows a sequence where a simple bipolar gradient composite pulse

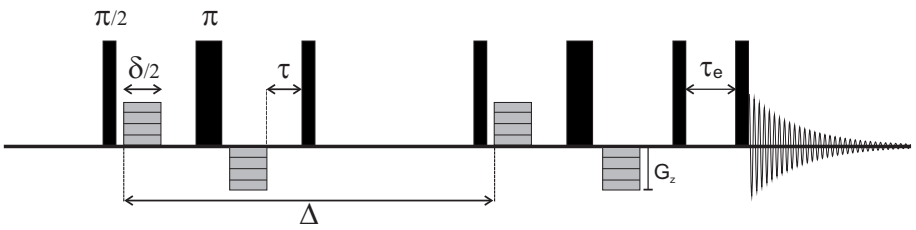


Figure 2.4: The bipolar pulse pair with a longitudinal eddy current delay (BPPLED) pulse sequence.

is used. Placing two gradient pulses with opposite polarity, and a π RF pulse between them, creates a self-compensated composite pulse. In general, the disturbance created by the second gradient pulse in the pair offsets that of the first gradient. The π RF pulse is present so that the magnetization continues to dephase in the same direction during both gradients.

Similar to the Stejskal and Tanner (1965) sequence, the signal attenuation due to diffusion is given by the following equation

$$E = \exp\left[-D\gamma^2 G_z^2 \delta^2 (\Delta - \delta/3 - \tau/2)\right] \tag{2.16}$$

All the experimental determinations of D obtained in this chapter make use of the BPPLED pulse sequence, unless stated otherwise.

2.4.2 Gradient field non-uniformity

For measuring accurate diffusion constants the quality of the gradient coil plays a very important role. The gradients used for diffusion measurements have to be linear across the sample volume, i.e. the field change has to be linear and therefore the field-gradient has to be constant. This is a question of gradient coil design (Kerssebaum and Salnikov, 2006).

The basic derivation of the signal decay function described in equation 2.11 only assumes that the gradient is uniform in the z direction across the entire sample region. Deviations from a uniform gradient will cause systematic deviations from the ideal decay behavior. Once again, this effect is probe dependent and may vary widely. The problem is illustrated in figure 2.5 using a doped water sample at 298 K. The signal attenuation is recorded and a

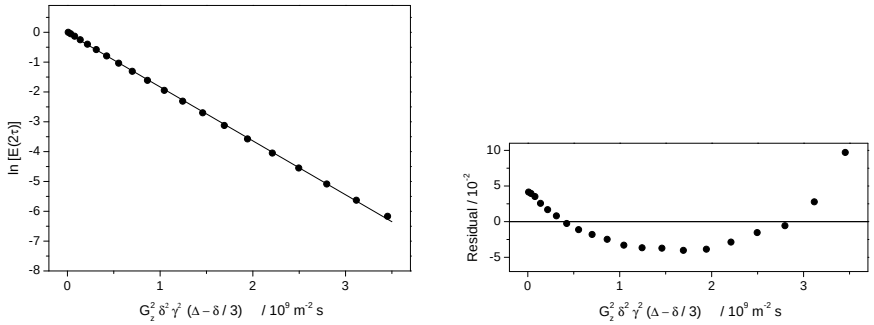


Figure 2.5: (a) A plot of the logarithm of the signal attenuation of the HDO resonance of a D_2O (99.9%) doped sample at 298 K *versus* $(G_z^2\gamma^2\delta^2(\Delta-\delta/3)$. The acquisition parameters for the Stejskal and Tanner experiment were : $\delta = 5$ ms, $\Delta = 15$ ms, and G_z was varied linearly for 20 values up to 0.37 Tm^{-1}. The correlation coefficient for the experimental attenuation curve is $r^2 = 0.9997$. **(b)** The residuals plot (the difference between the experimental data and the fitted data for each point).

linear least-squares fit is applied to the integrated signal of the HDO resonance in figure 2.5a. The residuals obtained from the fitted curve are plotted in figure 2.5b. Although at first glance the data appear to fit well, a close inspection of the residuals reveals a systematic error in the data.

Most gradient coil designs have a "sweet spot" in the very middle of the coil that provides the most uniform gradient (Antalek, 2002). As a consequence the gradient strength may vary substantially as a function of the distance from the coil center. Therefore, only a small portion of the sample actually experiences a strong, uniform gradient. This explains why figure 2.5b presents a systematic error, since some signal from the sample was acquired from outside the "sweet spot". Moreover, most RF coil designs generally excite a region beyond the coil's physical dimensions. This combination produces non-ideal signal decay behavior.

Spin-echo 1-D image profile

One way to visualize the gradient field non-uniformity is to perform a shape analysis on a spin-echo pulse sequence coupled with a 1-D imaging profile experiment (Price, 1998b), as exemplified in figure 2.6d . Accordingly to subsection 2.2.3, if a field gradient is absent then there is no spatial dependence of

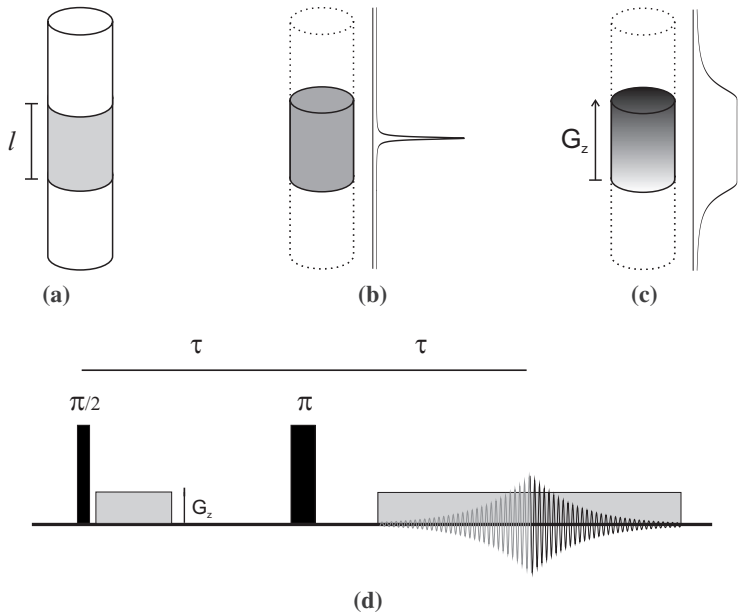


Figure 2.6: Schematic representation of the effect of a magnetic field gradient on the NMR spectrum. **(a)** The sample sensitive volume (gray highlighted) has a length l and ideally a cylindrical shape. **(b)** When the magnetic field applied to the sample sensitive volume is homogeneous, all spin isochromats have the same Larmor resonance frequency (same grayish color). Therefore, the spectrum expected will have the usual Lorentzian narrow line shape. **(c)** When a field gradient is applied, the spins in each sample slice experience different magnetic fields and thus different Larmor resonance frequencies. The resulting spectrum is a very broad line which results from the inhomogeneous field felt by the sample sensitive volume. This spectrum is known as a 1-D image of the sensitive volume region. **(d)** Hahn spin-echo sequence incorporating a read gradient of strength G_z used for obtaining a 1-D image of the sample sensitive volume in the NMR tube.

the resonance frequency. On the other hand, in the presence of a field gradient the observed frequency is position-dependent, according to equation 2.6. These considerations suggest that the observed FID and resulting spectrum will reflect both the gradient and the shape of the sample.

Before getting into the interpretation of the spin-echo 1-D sample imaging experiment, as an aside, some brief considerations about NMR tube quality will be given. The purpose of an NMR tube is to confine a liquid sample in a perfectly cylindrical volume, in a magnetic field. This premise should be

attained either for spinning and non-spinning tubes. The "perfectly" cylindrical shape of the tube can be characterized by the following measures: camber, inside/outside tube diameter, and concentricity (Glass, 1992). While the first two measures can produce a diversity of undesirable performance characteristics in NMR spectroscopy, such as sideband increasing modulation, a failure to conform concentricity requirements can cause major artifacts on diffusion experiments. Concentricity is a measure of the uniformity of the tubing wall thickness. If the thickness of the wall is not maintained throughout the sample, then portions of the sample fall outside the perfect cylindrical volume where the magnetic field may not be homogeneous. That is to say, to rule out tube concentricity problems from any possible non-homogeneity of field felt by the sample, one must carefully choose good quality tubes.

Most NMR tubes have an internal cylindrical shape and by consequence the sample also adopts that same shape (see figure 2.6a). If we arbitrarily take the magnetic field at one end of the cylinder to be B_0, the lowest resonance frequency will be γB_0 accordingly to equation 2.6. At a distance l in the direction of the gradient, the magnetic field is $B_0 + G_z l$, and thus the highest resonance frequency will be $\gamma(B_0 + G_z l)$. Since the number of spins is constant along the cylinder axis, the absolute value of the Fourier transform of the signal must be rectangular with a line width $\Delta\nu$[6], given by (Price, 1998b)

$$\Delta\nu = \gamma G_z l / 2\pi \tag{2.17}$$

Though it is possible to determine the strength of the gradient by analyzing the FID shape, this method is prone to a number of systematic errors as pointed out by other authors (Lamb et al., 1987; Murday, 1973). A more informative and intuitive method is to analyze the Fourier transformed spectrum and apply equation 2.17. One-dimensional profiles of a D_2O doped sample are shown in

[6] A modern high-resolution spectrometer can deliver a field gradient up to 0.6 Tm^{-1}. If we assume that the sensitive volume extends for about $l = 10$ mm, then from the top to the bottom of the sample the magnetic field due to such gradient will vary by 0.006 T. From equation 2.17 and taking the observed nucleus to be proton, we find that the obtained profile has a width of about 255 KHz.

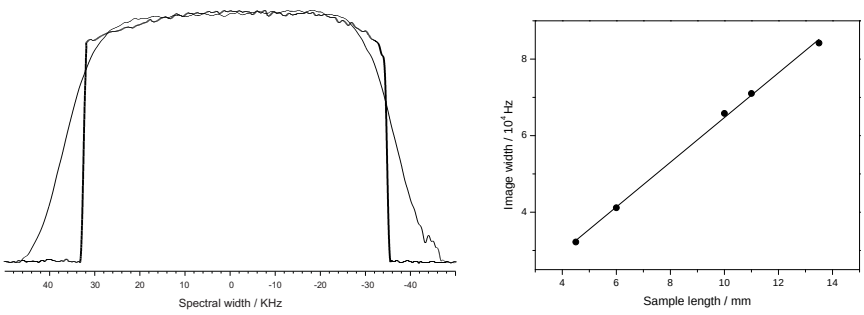


Figure 2.7: The 1-D profile experiment showing the effect of *slice* selection on a NMR tube containing doped D_2O at 298 K. **(a)** ^1H images of the residual water along the z-axis on a D_2O susceptibility matched tube, with a sample height l of 10 mm (solid line), and on a standard NMR tube containing a sample height of about 37 mm (grey line). The data were all acquired with the same read gradient (0.154 Tm^{-1}) and with the same polarity. Both tubes were positioned in the spinner at about the same position, in order to obtain centered profiles; **(b)** A plot of the profile image width *versus* sample length for the D_2O doped sample in the susceptibility matched microtube. The image width was determined at 20% of the peak height. The correlation coefficient is $r^2 = 0.998$.

figure 2.7.

From the theoretical considerations previously presented, the expected 1-D image profile should have a rectangle shape. However, from the profiles given in figure 2.7a that shape is not observed (compare with figure 2.6c). We can interpret the results considering the height of the NMR signal, as well as, the shape of the profiles.

The uneven signal height along the z-axis of the sample is due to a combination of RF and applied gradient inhomogeneity as well as background gradients (Price et al., 2001). Background gradients are due to the magnetic susceptibility differences at sample interfaces, the so called “meniscus effect”. There is a large difference in magnetic susceptibility across the interface of the sample in a standard NMR tube due to the large difference in susceptibility between the air and the sample. However, the glass in a susceptibility matched tube is approximately matched to the magnetic susceptibility of the sample (see figure 2.8). The results in figure 2.7a show that a plateau in signal height is reached for the standard and susceptibility matched tubes. The active volume

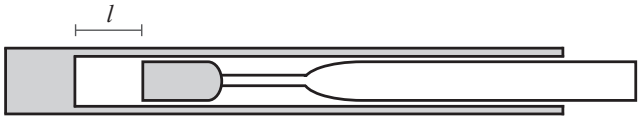


Figure 2.8: Schematic diagram of a susceptibility matched NMR tube. The sample is contained in a cylinder of length l. The bottom of the tube and the glass plug at the top of the sample have their magnetic susceptibility approximately matched to the solvent used.

of the sample in the standard tube is such that the meniscus is out of the RF coil active region. If the z position of the standard tube is changed in the probe, it is readily seen that a plateau is not attained (not shown). This does not happen as easily for a susceptibility matched tube. Even if a susceptibility matched tube is not centered in the probe, and thus with a greater probability for the RF coil to see the sample insert glass interface, the intrinsic characteristics of the tube reduce the probability of having background gradients. This is still true for relatively short samples. Nevertheless, even for these special NMR tubes, it is advisable to center the sample in the RF uniform region, since most of these tubes do allow a film of solvent between the plug and tube wall.

The applied gradient is more or less constant up to a sample length of 14 mm, since the image width increases linearly with sample length as shown in figure 2.7b. Although the sample in the susceptibility matched tube has a very precise cylindrical shape, the images of samples longer than 14 mm do not have sharp cutoff frequencies (not shown), resembling the image of the sample in the standard NMR tube given in figure 2.7a. The loss of vertical sides is probably largely due to increasing RF inhomogeneity as the sample moves away from the RF coil center (Price et al., 2001).

Typically, the physical extent of the RF sensitivity region is 15 mm (Welch, 1998a), whereas the region of constant gradient is at most 10 mm long. Antalek (2002) showed how easily the gradient strength can vary by 30% across a 15 mm sample region, corresponding to a variation of about 50% in the diffusion coefficient along the same direction. Clearly, to obtain the most accurate diffusion coefficients, one must find a way to obtain a signal from only a small region in the center of the sample, or to compensate for the non-homogeneity of the gradient field.

To deal with the non-uniformity of the z-gradient, several methods have been suggested, namely constraining the sample, post-processing of data by fitting a gradient strength distribution function, and sample slice selection. Physically constraining the sample, as already explained above by the usage of susceptibility matched tubes, is probably the most simple approach. Figure 2.7b shows that we were able to constrict the sample to a length l of about 4.5 mm, without loosing gradient constancy for the probe used. Although this is a good approach, constraining the sample to such small l sample lengths brings another problem: extreme difficulties in shimming the sample, i.e., to turn the B_0 field homogeneous around the sample. Thus a compromise must be met between using the smallest possible sample length in the active B_1 homogeneity region, and at the same time to have a good set of shimming parameters. Damberg et al. (2001) showed that the accuracy and precision of measured diffusion coefficients by the Stejskal-Tanner spin-echo pulsed field gradient experiment can be significantly improved using a parameterized gradient distribution function. A better approach involves the selection of the sample region to be excited through a slice selection, widely used in MRI. The idea is to use the gradient in conjunction with a selective RF pulse to excite only those spins in the middle of the sample region. As an example, a slice-selection technique was applied to the BPPLED pulse sequence with good results for chemical mixtures (Park and Lee, 2006).

2.4.3 Bénard convection

Most modern spectrometers have a variable temperature (VT) unit to control the actual sample temperature. A thermocouple near the sample senses the temperature, which the VT controller continuously compares with the user-requested temperature, changing the heater current accordingly (Welch, 1998b). The VT gas is introduced through the bottom of the sample region, travels around the side of the tube and exits through a port near the top of the sample region. Thus, it is quite possible that convection cells are produced along the

long axis of the sample[7], not excluding the possibility of transverse temperature gradients be formed too.

If the temperature gradient is large enough, convective flow will be induced along the z-axis. As convection results in the physical movement of the spins, it interferes with the formation of gradient spin echoes. This adds a velocity dependent phase term to the diffusion and will perturb the ideal decay in a PFGSE NMR experiment by superimposing an oscillating behavior in the signal decay (Mao and Kohlmann, 2001). In practice, the presence of convection appears as an increase on the measured diffusion coefficient (Stilbs, 1987).

Convective flow only takes place when the temperature gradient exceeds a critical value. Several factors contribute to it, namely, the shape of the sample, the viscosity and thermal expansivity of the liquid, the acceleration due to gravity, VT gas flow rate, among other things (Loening and Keeler, 1999). Several formal methods were presented to measure, visualize and compensate for thermal convection (Jerschow and Müller, 1998; Nilsson and Morris, 2005). A simple way to observe if a given NMR sample is being affected by convective flows, is to allow the signal in a PFGSE NMR experiment at the proper temperature to attenuate at least 95% and examine the decay behavior (Antalek, 2002). If there is no effect from the temperature and effects from gradient field non-uniformity are eliminated, the decay should be linear on a semilogarithmic plot. Figure 2.9 demonstrates an experiment with a D_2O sample at 298 K.

If critical convection problems were present in the sample, resonances should begin to exhibit phase distortion for the higher gradients, which was not observed. The resulting signal decay plot for the HDO resonance in figure 2.9c shows a very good correlation. If we compare the residuals plot in figure 2.9d with the experiment described earlier on figure 2.5, it is readily seen that the

[7] A convection cell is a phenomenon of fluid dynamics that occurs in situations where there are temperature differences within a fluid body. When a volume of fluid is heated, it expands and becomes less dense and thus more buoyant than the surrounding fluid. The colder, denser fluid settles underneath the warmer, less dense fluid and forces it to rise. Such movement is called convection, and the moving body of liquid is referred to as a convection cell or Bénard cell (Drazin, 2002).

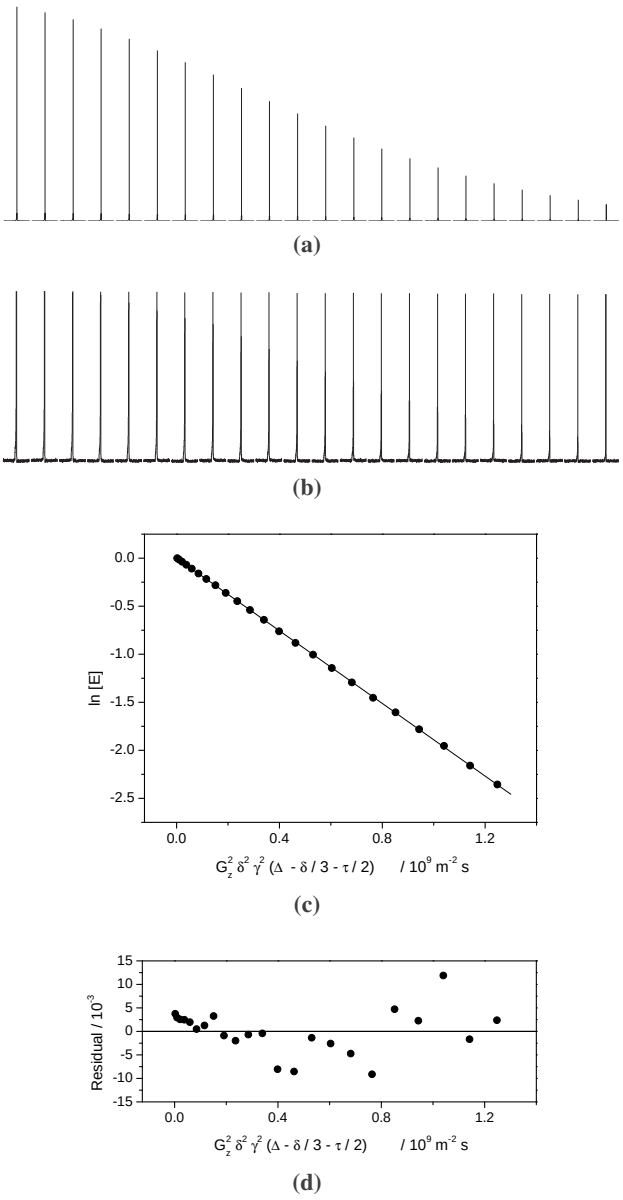


Figure 2.9: Evaluation of temperature gradients using a BPPLED experiment on a 10 mm length D_2O sample at 298 K, using a D_2O susceptibility matched tube. **(a)** Signal attenuation of the HDO resonance. The acquisition parameters used were: $\frac{\delta}{2} = 1.5$ ms, $\tau = 1.5$ ms, $\tau_e = 300$ ms and $\Delta = 12$ ms. Each spectrum is the average of 64 scans. The gradient strength was linearly varied up to a maximum of 0.44 Tm^{-1} with values increasing from left to right of the figure. **(b)** Corresponding normalized plots to the spectrum at highest gradient strength. **(c)** A semilogarithmic plot of the signal attenuation of the HDO resonance. The correlation coefficient for the experimental attenuation curve is $r^2 = 0.99996$. **(d)** The residuals plot.

correlation coefficient between the signal attenuation and the applied gradient has substantially improved. Therefore, evidence of convection currents caused by temperature gradients is not found in our setup.

A common NMR technique to improve sample homogeneity, either regarding the magnetic field or temperature, is to spin the sample tube. Theoretically, it should be allowed to spin the sample in PFGSE NMR experiments since the introduced velocity term is orthogonal to the z direction. In fact, some studies showed the suppression of effects from convection currents using sample spinning (e.g., see Esturau et al., 2001). Nevertheless, there is always the possibility of introducing other problems such as vibrations in the system when spinning the tube. It should also be emphasized that for a proper temperature control, the VT gas flow rate should be high enough to allow for thorough heat transfer along the tube but not so high as to introduce vibrations. Two other approaches could also be used to minimize temperature induced convection: reduce sample diameter and/or reduce sample height. As already pointed out earlier, reducing too much the sample height is problematic since it compromises the line shape (shimming problems). If instead we reduce the sample diameter by using a smaller size tube such as 3 mm, the S/N ratio will decrease. Nevertheless, the effect of reducing the sample diameter is much more important than reducing the height (Antalek, 2002). Goux et al. (1990) describes the effects of all these parameters on PFGSE NMR experiments.

For completeness, another solution is to use convection current compensated pulse sequences (Sørland et al., 2000).

2.5 Setup, acquisition and PFGSE analysis

The PFGSE experiment is a powerful technique to obtain diffusion coefficients. Though D is calculated in a straightforward way from a curve fit to the experimental data, the NMR spectroscopist should be extremely careful and diligent concerning several sources of potential artifacts. Some clues on how to minimize the most problematic ones were discussed in section 2.4.

Before proceeding to spectra acquisition, temperature and gradient fine tuning should be done. Also, determination of the minimum time required for

the eddy current effects to dissipate should be taken into consideration. Only after this prerequisites are met, the spectrometer can be considered ready for proper NMR acquisition of PFGSE data. The present section aims to be a general guide to setup the spectrometer, such that meaningful D values can be computed. Data acquisition and processing will be also discussed, and particular attention is paid to the application of diffusion measurements to the study of protein diffusion.

2.5.1 Temperature calibration

Modern spectrometers control the probe temperature through built-in hardware, providing a coarse estimate of the actual sample temperature. More accurate schemes have been developed making use of the temperature-dependence of the chemical shifts of methanol and ethylene glycol to calibrate the actual temperature of the sample in the probe (Raiford et al., 1979), or piezoelectric thermometers (Wang and Leigh, 1994).

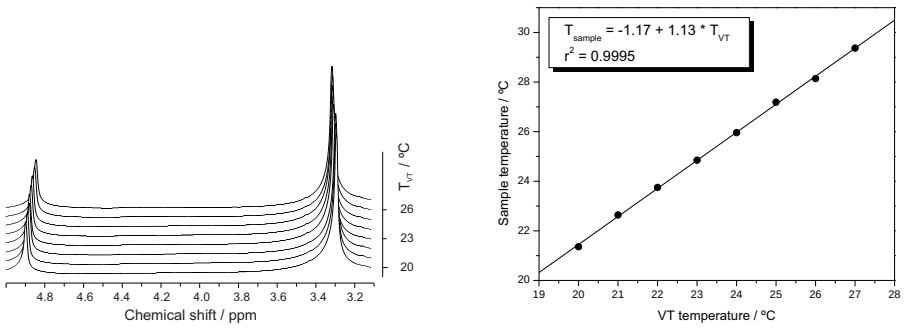
The most common method is the measurement of temperature-dependent chemical shifts of standard compounds. Over the range 250–320 K, the difference in chemical shift $\Delta\delta$ between the methyl and the hydroxyl resonance of a 100% methanol sample, is given by equation 2.18.

$$T = 403.0 - 29.53\Delta\delta - 23.87\Delta\delta^2 \quad , \Delta\delta \text{ in ppm} \tag{2.18}$$

Samples of ethylene glycol are used for high-temperature calibration (300–370 K), and the difference in chemical shift ($\Delta\delta$ in ppm) between the methylene and hydroxyl resonances of 100% ethylene glycol is given by equation 2.19 (Cavanagh et al., 1996).

$$T = 466.0 - 101.6\Delta\delta \quad , \Delta\delta \text{ in ppm} \tag{2.19}$$

An example of temperature calibration close to room temperature is given in figure 2.10. The obtained temperature calibration curve shown in figure 2.10b indicates that there is an offset of *circa* 1.84 degrees between the real sample temperature and the VT temperature in the conditions of spectrometer operation used.



(a) **(b)**

Figure 2.10: Temperature calibration close to room temperature for a 5-mm probe on a Varian ^{Unity}INOVA 500 spectrometer, with a 100% methanol sample. **(a)** Spectra were obtained as a function of VT temperature. The temperature-dependent chemical shift change of methyl and hydroxyl resonances is clearly observed; **(b)** Temperature calibration fitting curve. The fitting correlation parameter obtained expresses a good confidence for experimental work performed within the temperature range studied.

Typically, when the probe temperature is changed one must allow at least 10 minutes for the probe and sample temperature to reach equilibrium. The correlation between the actual and VT temperature can be improved, if the flow rate of the cooling/heating gas is increased, or if the thermocouple is moved closer to the sample. It should be noted however, that spectrometer systems are designed and tested with a limited range of VT gas flow rates[8]. Sizable deviations from this flow rate range may result in significant inaccuracy on temperature calibration.

If, as is often the case, intramolecular or intermolecular exchange processes are taking place in the sample, such as conformational equilibria, protein aggregation, or ligand binding, the appearance of the spectrum may be particularly sensitive to temperature (Lian and Roberts, 1993).

[8]The Varian VT unit is optimized for a gas flow rate of about 10 $Lmin^{-1}$ (Welch, 1998a)

2.5.2 Dissipation of eddy current effects

After setting up the correct VT temperature for your experiment, allow the sample to reach thermal equilibrium. Usually 10 to 15 minutes is sufficient if the temperature of the spectrometer room is controlled.

The next step should be the determination of the time required for the complete dissipation of eddy current effects for the maximum gradient strength to be used. From figure 2.3b it can be seen that a delay $\tau_e \geq 200$ ms is enough for the eddy currents to dissipate for the probe used in the present study. A $\tau_e = 300$ ms was used in all subsequent PFGSE experiments.

2.5.3 Gradient calibration

Acquiring a PFGSE experiment to measure diffusion coefficients involves the build-up of a gradient ramp (see section 2.2.4). The effective gradient output delivered to the probe is a linear function of the input parameter controlling the current amplitude. Every gradient probe will have its own characteristic gradient strength for a given current value, thus the need for a proper calibration. If the amount of current delivered to the probe is known, the gradient strength is also known through a calibration constant *gcal*, in units of $\mathrm{Tm^{-1}A^{-1}}$, given by the following relation[9]

$$G_z = gzlvl1 \times gcal \tag{2.20}$$

Usually, we do not concern ourselves with current values because the level of current is controlled by a DAC (Digital-to-Analog Converter) value in the software (*gzlvl*1 in VARIAN systems). Thus, *gcal* can be written in units of $\mathrm{Tm^{-1}DAC^{-1}}$.

There are a number of papers describing the most common gradient cali-

[9] On VARIAN systems, the pulsed field gradient strength is specified by the *gzlvl*1 parameter. It specifies the pulsed field gradient DAC value and ranges from -32767 to $+32768$ for a 16-bit gradient module like the PerformaII. To convert this value to an absolute gradient strength, a gradient calibration constant *gcal* must be calculated, such that $G_z = gzlvl1 \times gcal$. Therefore, *gcal* stores the proportionality constant between the parameter values (DAC units) controlling the desired gradient and the intensity of the gradient expressed in $\mathrm{Tm^{-1}}$ (Welch, 1998c)

bration procedures (Price, 1998b; Yadav et al., 2008). Traditionally, gradient calibration in PFGSE experiments is made either through one-dimensional image analysis of the spin echo in a steady field gradient, or indirectly through analysis of the echo decay of a PFGSE experiment on a single-component solution with a known molecular self-diffusion rate (Stilbs, 1987). These two methods were used in this work and will be further discussed below.

As an aside, it must be emphasized that most articles published using PFGSE experiments, do calculate diffusion constants by a linear regression of the attenuation data curves obtained. However, this approach gives unequal weighting to the noise, particularly as the NMR signal intensity approaches zero (Price, 1998b). When accuracy is needed, it is better to use nonlinear regression of the attenuation equation onto the experimental data[10].

1-D image analysis for gradient calibration

The line width of the Fourier transformed spectrum of a sample of known geometry can be used to calculate the gradient strength (Lamb et al., 1987; Murday, 1973). The methodology was described in sub-section 2.4.2. Briefly, the presence of a read gradient during echo acquisition will result in a spatial dependence of the resonance frequency. If the sample length l is known, the line width from the 1-D sample profile can be measured and used to determine the gradient strength.

In the case illustrated in figure 2.7a, a line width of 65.8 KHz was attained for a sample length of 10 mm in the susceptibility matched tube. A DAC value of 8000 out of a total of 32768 was used (see footnote 9). Using equation 2.17 we can estimate a gradient strength, $G_z = 65.8 \text{ KHz}/(4.26 \times 10^4 \text{ KHzT}^{-1} \times 0.010 \text{ m}) = 0.154 \text{ Tm}^{-1}$. The gradient calibration constant is then $gcal = 0.154/8000 = 1.925 \times 10^{-5} \text{ Tm}^{-1}\text{DAC}^{-1}$.

The main virtue of the method is that calibration can be performed without

[10]Throughout this chapter, PFGSE experiments were acquired by linearly changing the gradient strength G_z. All PFGSE attenuation curves are described by an exponential function with the independent variable G_z being a squared parameter. As such, for a proper sampling of the decaying curve, G_z should had been linearly incremented as G_z^2.

Table 2.1: Self-diffusion coefficients of reference compounds for gradient calibration at 298 K

Compound	Diffusion coefficient $/10^{-9}$ m^{2}s^{-1}	Reference
H_2O	2.30	(Mills, 1973)
D_2O (99.9%)	1.902	(Longsworth, 1960)
C_6H_6	2.207	(Collings and Mills, 1970)

the knowledge of the sample diffusion coefficient. Major problems are that large receiver bandwidths are required at even modest gradient strengths as shown by the above G_z estimation. Also the gradient broadened spectrum makes it difficult to obtain adequate S/N. Restricting the length of the sample or using slice selective pulses, one can alleviate problems with spectrometer bandwidth however restrictions due to S/N become more significant (Price, 1998b). *A priori* this methodology gives an estimate of the gradient calibration constant with an error $\leq 5\%$, if the sample dimensions are precisely controlled and not temperature dependent (Holz and Weingärtner, 1991).

This 1-D imaging "projection" analysis can additionally be used for checking the constancy of the gradient over a range of temperatures and over the length of the sample (see figure 2.7b).

Reference sample for gradient calibration

Whereas the gradient calibration constant obtained from a 1-D image method should be considered only as an estimate, one needs to go a step further to obtain the proper gradient calibration value. The best way to fine tune *gcal* is to use a suitable reference compound having a known diffusion coefficient D_{known}. The temperature, solvent conditions, and concentration need to be controlled. Some commonly used standard samples and their diffusion coefficients are listed in table 2.1. Systematic investigations of isotope effects on water diffusion have yielded accurate values for the limiting self-diffusion coefficient of HDO in D_2O (Holz and Weingärtner, 1991; Longsworth, 1960). A suitable primary standard sample to work at 298 K is thus residual water (HDO).

The prime advantages of the H_2O/D_2O method of calibration are (Stilbs, 1987) :

- accurately known and tabulated diffusion coefficients
- relatively low temperature dependence of D
- long T_2 value and not strongly temperature dependent
- deuterons provide an excellent field lock signal

If the sample experimental conditions used for gradient calibration (i.e. sample shape, delays, pulse lengths, gradient strengths, VT gas flow, etc.) are also used in subsequent experiments, then non-ideal gradient behavior is automatically canceled.

There are at least two approaches of getting an absolute calibration constant for a specific probe using a sample compound as reference. We can simply acquire a PFGSE NMR experiment with a D_2O (99.9%) sample, and instead of determining D (which is a known variable in this case) from the attenuation curve fitting, the calibration constant is calculated. A gradient calibration curve is shown in figure 2.11.

The obtained calibration constant value ($gcal = 1.890 \times 10^{-5}$ $Tm^{-1}DAC^{-1}$) is within 2% of the initial value calculated using the 1-D imaging technique. This new value should be more accurate due to the intrinsic problems of the 1-D image calibration technique.

A different mathematical approach to obtain the calibration constant was proposed by Antalek (2002), taking into account an estimative of the gradient constant, $gcal_{est}$. Making use of the same HDO attenuation curve values in figure 2.11, a diffusion coefficient D_{meas} is calculated using equation 2.16. From this equation it is readily seen that DG_z^2 is linearly proportional to $ln(E)$. Thus, for a single ln(E) value the following must be true

$$D_{meas}(gzlvl1 \times gcal_{est})^2 = D_{known}(gzlvl1 \times gcal)^2 \tag{2.21}$$

After rearranging equation 2.21, the actual gradient calibration constant $gcal$ can be calculated from the following (Antalek, 2002)

$$gcal = gcal_{est}\sqrt{\frac{D_{meas}}{D_{known}}} \tag{2.22}$$

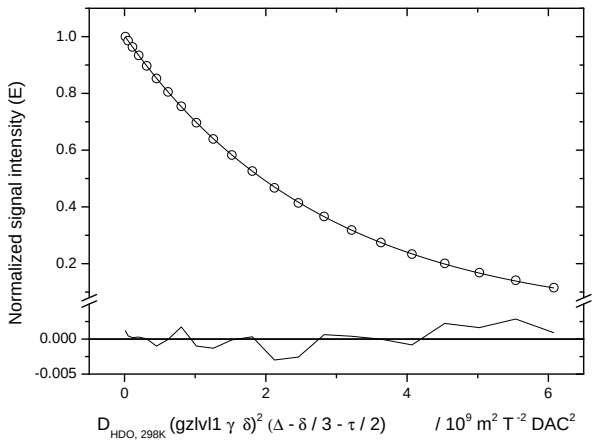


Figure 2.11: Gradient calibration using a D_2O susceptibility matched tube with a 10 mm of D_2O (99.9%) at 298 K. A nonlinear squares fit of the signal attenuation of the HDO resonance *versus* $(gzlvl1\gamma\delta)^2(\Delta-\delta/3-\tau/2)$ is shown. The acquisition parameters for the BPPLED experiment were : $\frac{\delta}{2}=1.5$ ms, $\Delta=12$ ms, $\tau=1.5$ ms, $\tau_e=300$ ms and *gzlvl*1 was varied linearly for 22 values up to 22000 DAC. 64 scans per spectrum were collected. The correlation coefficient for the experimental attenuation curve is $\chi^2=0.02$, with a value for the calibration constant equal to $gcal=(1.890\pm0.001)\times10^{-5}$ $Tm^{-1}DAC^{-1}$. The residuals plot is also shown below the fitting curve.

Considering $gcal_{est}=1.925\times10^{-5}$ $Tm^{-1}DAC^{-1}$ previously estimated from 1-D image profile, a nonlinear fit of equation 2.16 to the experimental data plotted in figure 2.11, $D_{meas}=1.83\times10^{-9}$ m^2s^{-1} is obtained. Making use of equation 2.22 the same value for *gcal*, i.e. 1.890×10^{-5} $Tm^{-1}DAC^{-1}$, is obtained.

2.5.4 Parameter optimization

Before going to the laboratory, it is extremely useful to simulate a diffusion decaying curve for the compound under study, if the value of D is known or estimated. Once the eddy current delay is determined for a specific probe, it is necessary to optimize the parameters that determine the decay function described by equation 2.16. To keep the timing constant throughout the whole experiment and to factor out spin-spin relaxation (see sub-section 2.2.4), the gradient strength G_z is chosen as the variable parameter, the τ delay is kept constant as well as the diffusion time Δ, and the diffusion gradient length δ.

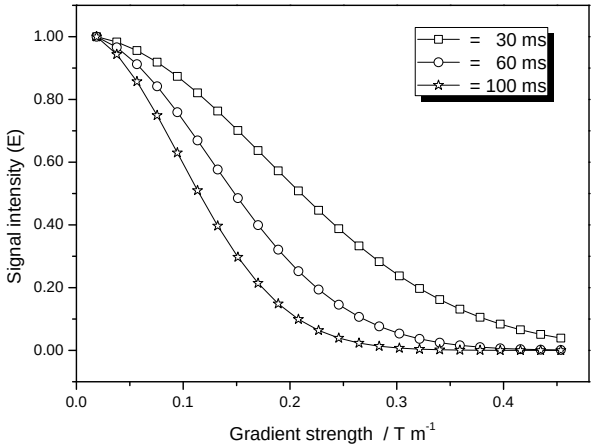
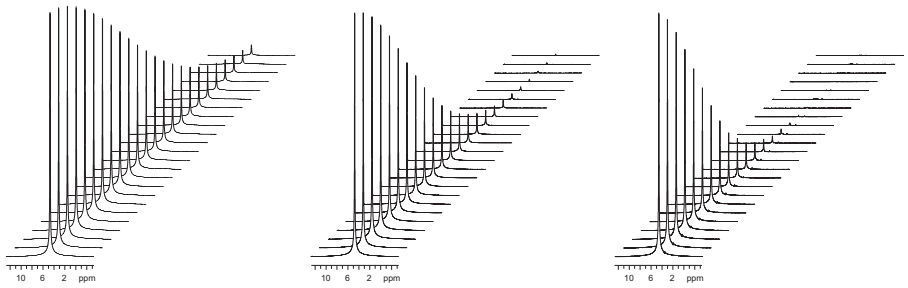


Figure 2.12: Simulated diffusion decay curves for a D_2O sample at 298 K, using the BPPLED pulse sequence. The gradient strength G_z was varied in 24 steps, while maintaining a constant diffusion gradient length $\frac{\delta}{2} = 1.5$ ms, a delay $\tau = 1.5$ ms along with an eddy delay $\tau_e = 300$ ms. The sample molecules were allowed to diffuse for $\Delta = 30, 60$ and 100 ms. The diffusion coefficient for residual HDO was considered to be 1.90×10^{-9} $m^2 s^{-1}$ (Longsworth, 1960).

The simulation curves for three different Δ values are shown in figure 2.12.

The aim is to obtain a curve where data points are properly distributed along the whole decay curve, such that proper data fitting can be made. If we consider an attenuation of 90% on the signal intensity as the threshold to avoid complications due to low S/N, it can be seen from figure 2.12 that for $\Delta = 30$ ms the decay provides a good data sampling to be adjusted to equation 2.16. A set of BPPLED spectra for D_2O (99.9%) at 298 K acquired with the same acquisition parameters as in the above simulated decay curves is shown in figure 2.13.

Depending on the accuracy with which the diffusion constant is to be determined, the signal decay should be allowed to attenuate to the lowest possible value above the noise level. An *apriori* diffusion decay simulation for a given NMR signal should always be stressed out, such that the right acquisition parameter values, or at least a rough estimation, are readily known for the setup of a real experiment.



(a) **(b)** **(c)**

Figure 2.13: ^{1}H-NMR spectra of a D_2O sample, at 298 K, using the BPPLED sequence. The acquisition parameters used were $\frac{\delta}{2} = 1.5$ ms, $\tau = 1.5$ ms, $\tau_e = 300$ ms and **(a)** $\Delta = 30$, **(b)** 60, **(c)** and 100 ms. Each spectrum is the average of 64 scans. The gradient strength was varied linearly in 24 increments up to a maximum of 0.45 Tm^{-1}, with values increasing from front to back in the spectra of the figure.

2.5.5 Solvent signal suppression

Limited sample availability, often the case in protein NMR experiments, limited solubility or other problems such as aggregation, strongly limit the solute concentration to be used in diffusion measurements. Additionally, the presence of large solvent resonances, such as the water resonance in biological samples, present a big challenge to proper acquisition of NMR spectra. Water molecules are typically present at concentrations more than three-orders of magnitude greater than solute concentration. Such strong NMR signals prevent optimal use of the analogue-to-digital (ADC) converter of the spectrometer[11], and obscure peaks near the solvent resonance. Efforts have been directed to ensure that the solvent signal is greatly reduced in amplitude while the rest of the spectrum remains, as far as possible, unchanged. Several methods of solvent suppression have recently been reviewed by Price (1999).

Traditionally, selective irradiation of the solvent peak has been the most common method since it is both simple and easily incorporated into multidimensional experiments (Price, 1996). This method, often called pre-saturation, was

[11] Fourier spectroscopy requires a digitized free induction decay and hence an ADC converter. There is has a practical limitation on the dynamic range handled by the ADC and no part of the FID can be allowed to overflow that range (Freeman, 1988).

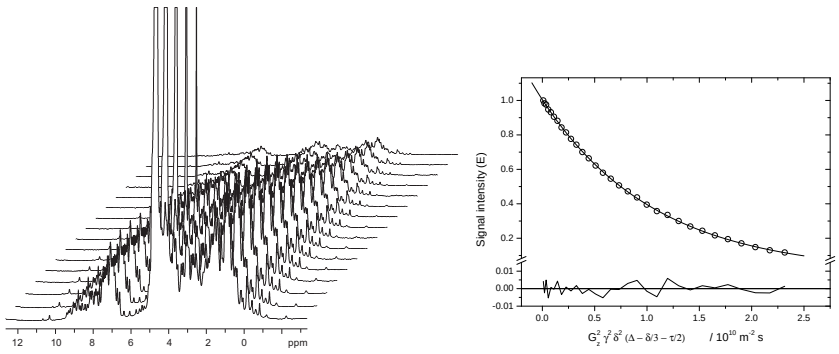
incorporated in the BPPLED pulse sequence. Albeit its general usage, the first try outs to suppress solvent peaks from protein solutions were not satisfactory to pursue with this solvent suppression scheme. It suffers from the disadvantage that signals close to the solvent signal are difficult to observe.

In addition to measure diffusion, gradients also provide the foundations of highly efficient means for water suppression, particularly suited to the difficult requirements set in protein NMR measurements (Price, 2000). A suppression sequence may be designed to take advantage of an inherent difference between the solute and solvent molecules, such as translational diffusion. The difference in the translational diffusion coefficients between rapidly diffusing solvent molecules and the generally much more slowly moving solute molecules, can be used in combination with a PFGSE sequence to achieve solvent suppression (Stilbs, 1987). Accordingly to equation 2.16, a faster diffusion species will be more strongly attenuated than a slower moving species. An example of a diffusion measurement on hen-egg white lysozyme in D_2O is given in figure 2.14.

The water and the lysozyme peaks have quite different signal attenuation by virtue of their very different diffusion coefficients. For example, at 298 K, residual water has a diffusion coefficient of about 1.9×10^{-9} m^2s^{-1} (table 2.1) and a large protein such as lysozyme has a diffusion coefficient nearly two orders of magnitude smaller. Thus, the solvent resonance will be greatly attenuated compared to the protein resonances if a pulse sequence is used that incorporates some form of gradient spin echo sequence, like the BPPLED.

2.5.6 Protein diffusion case studies

The relationship between protein structure and protein function is well known and NMR has long been used to determine the three-dimensional structures of such biomolecules (Wüthrich, 1986). Due to the intrinsically low sensibility of NMR to collect experimental data, often saturated protein solutions must be used in the millimolar concentration range to produce meaningful results. Such sample conditions may bring to the NMR spectra interpretation, an extra difficulty while assigning it since self-aggregation is promoted by such crowded solutions. Nevertheless, such macromolecular solution compli-



(a) **(b)**

Figure 2.14: **(a)** Typical set of BPPLED ^{1}H NMR spectra for hen egg white lysozyme 1.5 mM in 99.9% D_2O, at pH 3.1 and 298K. The experimental parameters used were $\Delta = 100$ ms, $\frac{\delta}{2} = 1.5$ ms, and $\tau = 1.5$ ms. Each spectrum is the average of 256 scans. In this experiment, a longitudinal eddy current delay τ_e of 300 ms is used to minimize spectral artifacts resulting from residual eddy currents. The gradient strength G_z was varied up to a maximum of 0.60 Tm^{-1} with values increasing from front to back in the spectra of the figure. A sample length of 10 mm was used in D_2O susceptibility-matched micro tubes. A recycle delay of 10 s was allowed between scans. **(b)** Plot of the normalized signal attenuation of the aliphatic region (~ -1.4 to 2.4 ppm) of the lysozyme spectra *versus* $G_z^2\gamma^2\delta^2(\Delta - \frac{\delta}{3} - \frac{\tau}{2})$. The solid line represents a nonlinear least squares fit to the experimental data giving a diffusion coefficient for lysozyme of $D = (9.32 \pm 0.02) \times 10^{-11}$ $\mathrm{m}^2\mathrm{s}^{-1}$. The correlation coefficient for the fitting curve is $\chi^2 = 0.09$. The residuals plot is also shown below the fitting curve.

cations are also present in biological systems. Because crowding influences macromolecular association and conformation, it will play a role in all biological processes that depend on noncovalent associations and/or conformational changes, such as protein synthesis and protein oligomerization. Proteins in the biological milieu are generally in some equilibria between different aggregation states. Protein aggregation plays an important role under normal physiological conditions, but it is also implicated in disease conditions, such as amyloid related diseases (see chapter 6). Thus, a biochemist can not neglect crowding effects when trying to mimic biological conditions in a *in vitro* study, since living systems are ubiquitously crowded at the biochemical level (Ellis, 2001).

The advent of pulsed field gradients increased the range of available NMR experiments. The usage of well-designed gradient sequences decreases the experimental acquisition time and generally perform better than those using

extensive phase cycling[12], especially when suppressing extremely strong solvent signals (Thomas et al., 1999). One of the most promising NMR experiments making extensive usage of PFG are the PFGSE related pulse sequences. These are being used to elucidate the complicated solution behavior of proteins such as self-association and aggregation, nevertheless the technique is still in its infancy (Price, 2000). The following sub-sections will show the application of PFGSE to two protein systems.

Hen egg white lysozyme

After having calibrated the spectrometer with a sample of D_2O it is convenient to benchmark the current spectrometer setup with a well characterized protein sample. Lysozyme serves as a widespread model system in various fields of biochemical research, since it is one of the most studied proteins.

The lysozyme solution used in the experiment depicted in figure 2.14 is a crowded system. In other words, the average spacing between protein molecules is much less than the mean-square displacement of the particles over the time scale of the PFGSE diffusion experiment. For example, the average spacing between lysozyme molecules in a 1.5 mM solution is about 13 nm[13]. In the BPPLED experiment in this study the time interval allowed for molecule diffusion was 100 ms, thus the mean-square displacement calculated from the Einstein equation 2.4 is about 4.5 μm. Hence, during the diffusion measurement there is a high probability for different lysozyme molecules to collide numerous

[12] A high-resolution NMR pulse sequence is defined by a sequence of radiofrequency pulses and a procedure for selecting a required set of coherence transfer pathways. When a RF pulse is applied it excites an enormous number of coherence transfer pathways, and not all give signals of interest. Phase cycling is the most commonly applied approach to suppress unwanted coherence transfer pathways.

[13] A rough estimation of the average distance between protein molecules can be calculated using the excluded volume concept, which is determined by the distance between the center of mass of a reference particle in relation to other macromolecules in solution upon closest approach. The HEWL protein volume can be determined knowing that monomeric lysozyme is a nearly spherical prolate ellipsoid with major and minor semi-axes ratio of 24 : 13 Å for the non-hydrated molecule (Dubin et al., 1971); or, it can be determined knowing the molecular weight of lysozyme (14.32 KDa) and from the tabulated partial specific volume of ~ 0.7 cm^{3}g^{-1} (Cantor and Schimmel, 1980)

times. Although there is considerable disagreement in the literature regarding the limits of lysozyme solubility, a sample containing 1.5 mM lysozyme should be unsaturated at pH 3 and it is likely that lysozyme remains monomeric (Price et al., 1999). Consistent with this, the NMR spectra of lysozyme under these conditions have sharp resonance lines.

The diffusion constant of a molecule will be a function of both the temperature at which the measurement was performed and the solvent viscosity (see equations 2.1 and 2.2). Solvent viscosities are themselves a function of temperature. Rather than deal with all these variables, it is convenient to convert the experimentally measured diffusion coefficients to what would be observed if the measurements were carried out for example at 293 K in pure water (Cantor and Schimmel, 1980) :

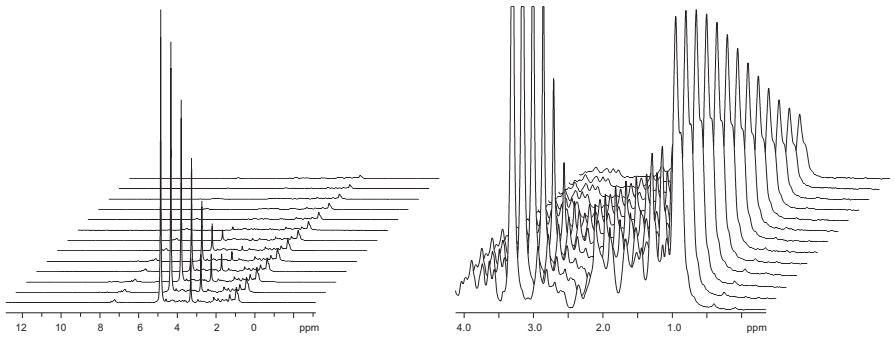
$$D_{w,293} = D_{obs,T}\left(\frac{293}{T}\right)\left(\frac{\eta_{w,T}}{\eta_{w,293}}\right)\left(\frac{\eta_s}{\eta_w}\right) \tag{2.23}$$

where $D_{w,293}$ is the diffusion constant in pure water at 293 K, $D_{obs,T}$ is the measured diffusion coefficient in the actual solvent at the experimental absolute temperature T, $\eta_{w,T}$ and $\eta_{w,293}$ are the viscosities of water at absolute temperatures T and 293 K, respectively. η_s and η_w are the viscosities of the actual solvent used and water at a common temperature, respectively.

Using viscosity tabulated values (Cho et al., 1999) and the experimental data (figure 2.14), a diffusion coefficient of $D_{w,293} = 1.01 \times 10^{-10}$ m^{2}s^{-1} was obtained for lyzozyme using equation 2.23. The result is in line with published results at similar experimental conditions (see Price et al., 1999, table 2). However, there is a wide range of values published. No systematic errors were observed in the attenuation curve when fitted to equation 2.16.

δ-toxin

δ-toxin is a 26 residue peptide produced by *Staphylococcus aureus* which lyses eukariotic cells in preference to bacterial cells *in vivo* (Dhople and Nagaraj, 1993; Tappin et al., 1988). The peptide has a high propensity to aggregate in solution due to its amphipatic character and zero net charge at neutral pH (Fitton, 1981). Knowing the aggregation state of biomolecules, or at least not



(a) **(b)**

Figure 2.15: (a) Typical set of BPPLED ^{1}H NMR spectra 1mM δ-toxin, at 298 K, in CD_3OD. The acquisition parameters used were $\frac{\delta}{2} = 1.5$ ms, $\Delta = 40$ ms and $\tau = 1.5$ ms. Each spectrum is the average of 256 scans. In this experiment, a longitudinal eddy current delay τ_e of 300 ms is used to minimize spectral artifacts resulting from residual eddy currents. The gradient strength G_z was varied linearly up to a maximum of 0.47 Tm^{-1} with values increasing from front to back in the spectra of the figure. Samples of ≈ 150 μL in CD_3OD susceptibility matched micro tubes were used. **(b)** Spectral expansion of the aliphatic region used to calculate the diffusion coefficient. The solvent peaks are at ≈ 4.8 and 3.3 ppm.

underestimate the possibility of aggregation, is essential for NMR structure determination. The wrong assumption that a protein is in a monomeric state may lead to erroneous interpretation of the NOE connectivities which would result in wrong assignments, and thus in non-meaningful 3-D structures. In spite of the several studies previously performed on this peptide system, still little is known about its aggregation state.

Figure 2.15 shows the signal attenuation in a PFGSE NMR experiment of δ-toxin in CD_3OD. As shown previously for HEWL (figure 2.14), it is readily seen that due to a large difference in diffusion coefficient between the solvent and the peptide, very different rates of peak attenuation are observed.

The integration of the signal intensity was done using the aliphatic region $[0.5 - 1.8]$ ppm to avoid the large contributions of the solvent peak resonances and to discard the low upfield resonances with a small S/R. The signal attenuation data is plotted in figure 2.16 along with the curve fitting to a single exponential equation.

Except for some cases of restricted diffusion, a plot of the natural logarithm

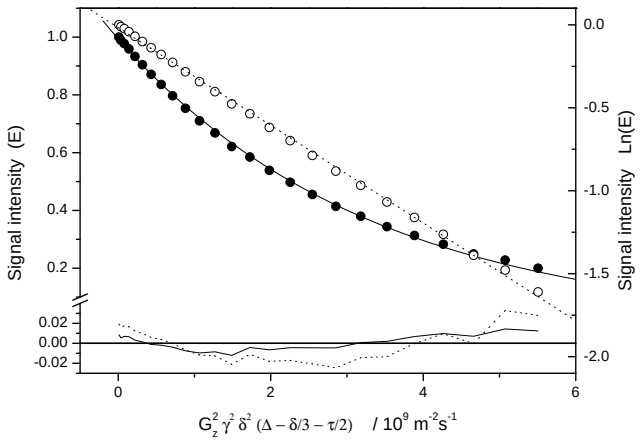


Figure 2.16: Attenuation plots of the normalized (closed circles) and logarithm (open circles) signal intensity of the aliphatic region ($\sim$ 0.5 to 1.8 ppm) of the δ-toxin BP-PLED spectra *versus* $G_z^2\gamma^2\delta^2(\Delta-\frac{\delta}{3}-\frac{\tau}{2})$. The solid and dotted lines represent nonlinear and linear least squares fits to the experimental data, giving diffusion coefficients of $(3.03\pm0.02)\times10^{-10}$ and $(2.94\pm0.02)\times10^{-10}$ m^2s^{-1}, respectively, for the peptide. The correlation coefficient for the nonlinear fitting curve is $\chi^2=0.09$ ($r^2=0.9993$), whereas for the linear fit is $r^2=0.9989$. The residuals are also shown at the bottom of the graph.

of the signal intensity versus the square of the gradient amplitude should be a straight line for the diffusion of a single component in the case of free diffusion (Price, 1997). Nevertheless, the straight line depicted in figure 2.16 only approximates the signal decay behavior. A closer look of the data, in particular the analysis of the residuals plot, reveals that the experimental data decay shows an upward curve trend, indicating polydispersity (Antalek et al., 2002).

As briefly pointed out in sub-section 2.2.5, protein solutions are intrinsically polydisperse systems. Nonetheless, any aggregate built of a large number of monomers will be too large to give any significant contribution to a NMR spectrum, owing to its short transverse relaxation time T_2 (Keeler, 2006). Hence, we can reasonable assume that low-weight molecular species give the major contributions to the decay of the resonance intensity in the diffusion experiment. Another important aspect is that chemical shifts of such low-weight aggregates are usually overlapped, thus the resulting attenuation curve is the sum of the contributing species. δ-toxin due to its physical-chemical properties is extremely

prone to aggregation even at concentrations on the μM range (Thiaudière et al., 1991). In the present experimental conditions (mM range) the peptide is aggregated, albeit two-dimensional NMR spectra of δ-toxin do not show segregation of peaks belonging to different oligomeric species in solution (see chapter 3).

A rather obvious aspect to be taken in consideration when analyzing PFGSE attenuation profiles is the lifetime of the oligomeric species. Though it is likely to be very protein specific, the kinetics of formation/disruption of oligomeric species is expected to be slow on the time-scale of the PFGSE experiment ($\Delta = 100$ ms) (Price, 2000). A monodisperse solution should give a single exponential attenuation curve. If aggregation occurs, as it is the case, the attenuation curve should have systematic errors relative to a single exponential, if the exchange is slow (see equation 2.13). Danielsson et al. (2002) showed that the trend of residuals obtained from the fitting curves to one-exponential function may give information on the presence of slowly exchanging species in the sample. The results shown in figure 2.16 have a systematic error on the residuals, either in the linear and nonlinear curve fitting, consistent with the presence of aggregates in a slow exchange regime.

To ensure that the sample was in equilibrium with respect to sample temperature and state of self-association, diffusion measurements were taken spaced in time. The results are shown in figure 2.17.

A time-dependent variation on the peptide diffusion coefficient is evident, with a 5% increase on the D value on the first 6 days (144 hours) of sample incubation. Though this variation is small, the results seem to suggest that the solution is not in equilibrium and that slow exchange processes are occurring. On the other hand, we noticed a deposition of peptide at the bottom of the NMR tube at the end of the 6^{th} day. Thus, though we can say that the oligomers are in slow exchange on a timescale of hours to days, another factor is coming into play. As time proceeds, more of the δ-toxin peptide is converted into very high molecular weight oligomers and ultimately precipitates, which are undetectable by liquid NMR. As more of the peptide comes out of solution, the remaining pool of dissolved peptide will be effectively at lower concentration. As an immediate consequence, such pool will be less obstructed and more likely to

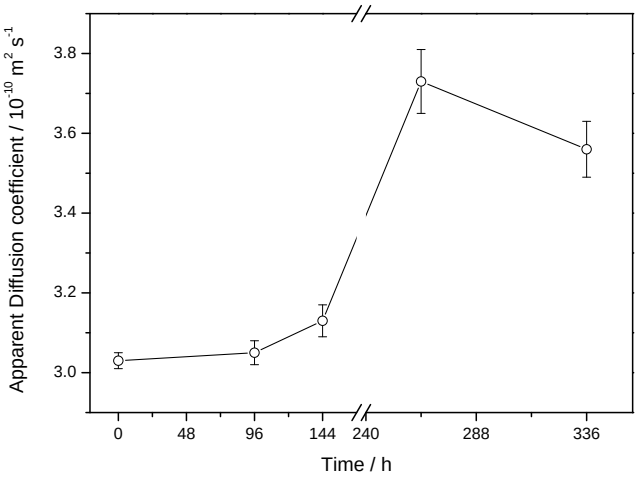


Figure 2.17: Apparent diffusion coefficient for δ-toxin in CD_3OD as a function of incubation time determined from PFGSE NMR experiments. The BPPLED parameters used are the same as in figure 2.16. The acquisition of the first measurement started about 1 hr after the sample was inserted into the probe. During the experiment, the sample temperature was controlled at 298 K. The last 2 data points were obtained after dilution of the sample to $\approx$ 0.4 mM with CD_3OD. Diffusion coefficients were calculated from nonlinear fits to the respective signal attenuation decays, with all residual plots showing the same systematic error trend reported in figure 2.16 (not shown). The error bars correspond to error associated with the calculation of *D*.

be monomeric (or in a lower molecular weight aggregate) and hence faster diffusing.

Dilution of the sample was performed on the 6^{th} day of incubation to a final concentration of about 0.4 mM, to dissolve the peptide deposits. After 5 days of incubation, no peptide precipitates were observed at the naked eye, and we measured again the apparent diffusion coefficient of δ-toxin. We observed an increase of *D* of about 18%, which could be explained by the less crowded surrounding environment. On the other hand, the variation trend on the diffusion coefficient between the 11^{th} day (264 hours) and the 14^{th} day (336 hours) is now different from the one observed in the first 6 days. Again, a slow exchange regime is observed but the equilibrium conditions have changed. The decrease on the apparent diffusion coefficient indicates that high-molecular weight species are being formed in solution. Hence we are in the presence of a complex system, in which the high propensity of δ-toxin to self-assemble is

evident.

2.6 Conclusions

This chapter does not claim to be exhaustive in its coverage of the theoretical background of PFGSE NMR, nor on the experimental difficulties encountered by a spectroscopist and available technical solutions to surpass them. It is nevertheless an attempt to give a general overview of how the technique can be used to infer information from NMR active diffusion species in solution, giving a special emphasis to proteins and peptides. A number of good reviews can be consulted for an in depth study of some tips and tricks such technique poses, as well as, to the extensive applicability to bio-molecular systems (Antalek, 2002; Brand et al., 2005; Johnson Jr., 1999; Price, 1997, 1998b, 2000; Söderman and Stilbs, 1994; Stilbs, 1987; Waldeck et al., 1997).

Almost 60 years after Erwin Hahn's paper (Hahn, 1950) about the spin echo phenomenon, it is spectacular how an *artifact* on Hahn's observations, led to the development of techniques to map molecular displacements, to study the motion of fluids in human bodies or to the development of magnetic resonance imaging. The study of diffusion by NMR has made many contributions over the years, establishing a new set of pulse field gradient based sequences. The recognition that PFG can greatly improve speed and spectral quality of multidimensional NMR is becoming a common place on modern routine high resolution spectrometers (Morris, 2002).

PFG NMR is a very powerful alternative to traditional methods, such as dynamic light scattering and analytical ultracentrifugation, for studying the self-diffusion of proteins. It provides a convenient and quick way for studying the translational motion of molecules non-invasively. For instance, the concentration range used is exactly that used in NMR structure determination, whereas the traditional methods of studying protein aggregation are generally carried out at much lower concentrations.

To extract maximum information from PFGSE experiments, the effects of obstruction and relaxation-weighting need to be better understood to enable greater molecular weight selective information to be obtained (Price, 2006).

Importantly, the reliability of the diffusion measurements depends on the spectrometer and gradient system being well characterized and calibrated. The development of new multiple-quantum PFGSE sequences, more sophisticated obstruction theories, theoretical modeling of association processes, as well as hydrodynamic models (Aragon and Hahn, 2006), are needed for a proper analysis of the resulting NMR data.

Acknowledgments

The author would like to thank the following persons for reprints, and prompt replies to his questions: Professor Joseph Hornak, Leslie Schwartz, Professor Gareth Morris, William Price, and Brian Antalek. I am indebted to Professor Sergio Aragon for stimulating the computation of hydrodynamic properties from peptide 3-D models.

Chapter 3

NMR solution structure of δ-toxin in membrane-mimetic environments

O rosto
das coisas
lavado
e não há
temor

o rosto
do homem
chovendo
e as coisas
que o reconhecem

e o ósculo:
as coisas
como canto-chão
da alma

Paulo Carvalho

3.1 Introduction

Despite their structural and functional diversity, living entities do possess common features in their defense-and-surveillance systems against competing or attacking organisms (Andreu and Rivas, 1998). Cytolytic peptides comprise an important subset of the molecular armament protecting the host organism (Saberwal and Nagaraj, 1994), in both eukaryote and prokaryote kingdoms. These naturally occurring toxins, usually short linear peptides not exceeding 40 residues, have received considerable attention in the last 30 years, as possible alternatives for the conventional antibiotics. As resistance to commonly used antibiotics increases, the scientific community is devoting an enormous effort on research and development of new compounds. Characterization of structure-function relationships in these toxins may allow the design of economically viable and more efficient peptide analogues (Bechinger and Lohner, 2006).

The primary sequence of the peptide toxins vary widely, but generally have a significant number of polar and charged residues. The distribution of charged and hydrophobic residues along the polypeptide chain is such that the toxin can adopt the conformation of an amphipatic α-helix, having a hydrophobic and a hydrophilic face (Thiaudière et al., 1991). Although length, composition and structure of these peptides are similar in a broad sense, they have vastly different specificities and target the membranes of different cells.

The mechanisms of binding to membranes remain still elusive (Gregory et al., 2008), nevertheless it seems relatively certain that no cell-surface receptors are involved in the recognition mechanism (Wade et al., 1990). Hence, some kind of molecular recognition mechanism appears to reside at the level of the lipid bilayer (Nizet et al., 2001). Recent kinetic studies of peptide-membrane interactions seem to point to a general mechanism in which membrane raft-domains play a key major role (Pokorny and Almeida, 2005).

Lytic peptides can be classified into three major groups: those which are selectively active against eukaryotic cells, those which are active against prokaryotic cells and those active against both eukaryotic and prokaryotic cells. Some

peptide toxins are secreted by bacteria, such as δ-toxin[1] secreted by pathogenic strains of *Staphilococcus aureus*, which lyse eukaryotic cells in preference to bacterial cells (Kreger et al., 1971). Others, such as magainins (Zasloff, 1987) and dermaseptins (Mor et al., 1991), secreted by the African clawed frog *Xenopus lævis*, and the cecropins (Steiner et al., 1981), isolated from insect hemolymph and pig intestine, are targeted at prokaryotes and lower eukaryotes, such as fungi and protozoa. A third category of toxins, which include melittin (Habermann, 1972) and the mastoparans (Argiolas and Pisano, 1983), are secreted as part of the venom of bees and wasps and affect prokaryotic and eukaryotic cells alike.

In particular, Staphylococcal δ-toxin is a 26 amino acid residue peptide with four basic and three acidic residues, a negative C-terminus, and a formylated N-terminus, giving it a zero net charge at neutral pH. Such physicochemical properties facilitates its aggregation in aqueous solution (Fitton, 1981). Due to its amphipathicity, δ-toxin is soluble not only in aqueous solutions but also in organic solvents such as chloroform/methanol (Fitton et al., 1980). The helical wheel projection of the peptide is shown in figure 3.1, where it becomes evident that the helix formed contains a hydrophobic face and a charge-rich hydrophilic face. The available evidence, mainly from circular dichroism and nuclear magnetic resonance (NMR), indicates that the peptide is largely unstructured at low concentrations in aqueous solutions while adopting an amphipatic α-helical conformation upon aggregation at higher concentrations in aqueous solution, in organic solvents, and when bound to lipid membranes (Lee et al., 1987; Tappin et al., 1988; Thiaudière et al., 1991). Bladon et al. (1992) argued that the extent of helical character in δ-toxin is influenced by the solvent composition as well as the polarity of the N-terminus.

Before binding to membranes most lytic peptides are present in an aqueous solution as monomers in an unfolded state, or in equilibrium with oligomers in which the peptides exist in an α-helical conformation (White and Wimley,

[1] Several names appear in the literature describing this peptide, such as δ-toxin (Bernheimer and Rudy, 1986), δ-lysin (Freer and Birkbeck, 1982), and δ haemolysin (Fitton et al., 1980). Through out this chapter we will use the term δ-toxin with no particular reason.

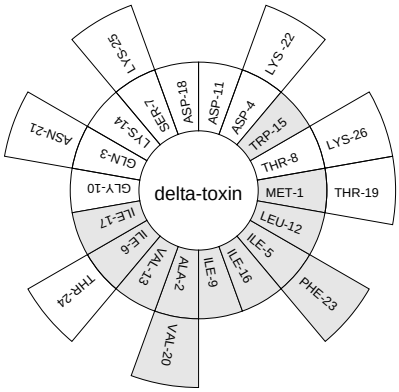


Figure 3.1: Edmundson helical wheel projection of δ-toxin. Hydrophobic and hydrophilic residues are represented by shaded and open areas, respectively, according to the whole-residue free-energy of transfer from water to *n*-octanol (White and Wimley, 1999).

1998). Lytic peptides are soluble in aqueous solution, but when interacting with the target membrane conformational changes occur which allow them to associate with and ultimately lyse the membrane (Shai, 1999). The conformational changes are due to the different chemical environment that the peptide encounters upon interaction with the cell membrane. Using NMR we will evaluate the conformation of δ-toxin, in aqueous solution and membrane mimetic environments, such as methanol, methanol / H_2O mixtures, and dimethyl sulfoxide.

3.2 Materials and methods

3.2.1 Peptide synthesis and sample preparation

δ-toxin (formyl-MAGDIISTIGDLVKWIIDTVNKFTKK) was a gift from Dr. Birkbeck (University of Glasgow, Scotland) and its purification protocol was described previously (Birkbeck and Freer, 1988).

Deuterium oxide (D_2O, 99.9 atom % in D), methyl-d_3 alcohol (CD_3OH, 99 atom % in D), methyl-d_3 alcohol-*d* (CD_3OD, 99.8 + atom % in D), chloroform-*d* ($CDCl_3$, 99.8 atom % in D), and dimethyl-d_6 sulfoxide (DMSO, 99.9 atom

% in D) were purchased from Sigma-Aldrich Chemical Company, St. Louis, USA. DCl and NaOD used to adjust the solution pH were obtained from Merck, Germany.

Samples for NMR were prepared by dissolving lyophilized δ-toxin in 500 μL of the chosen solvent to a final concentration of $\sim$ 3 mM. Preliminary one-dimensional (1-D) ^{1}H spectra were recorded in each of the solvents mentioned above to qualitatively assess the peptide aggregation state. Final results were taken from spectra in the following solvents or solvent mixtures (v/v) : CD_3OH; 75%/25% CD_3OH/H_2O; 65%/35% CD_3OH/H_2O; and DMSO. The pH* of protic solvents was measured with a glass electrode at room temperature, without isotope effect correction (Glasoe and Long, 1960), and adjusted to ≈ 3 with the addition of a few microliters of NaOH/NaOD or HCl/DCl solutions. The low pH imparts the peptide with a net positive charge, which minimizes its aggregation. After peptide solubilization, samples were transfered to 5-mm NMR tubes with Pasteur glass pipettes.

3.2.2 NMR spectroscopy

NMR spectra were acquired at a ^{1}H resonance of 499.825 MHz on a Varian ^{Unity}INOVA 500 spectrometer at a temperature of 298 K or 303 K. The spectrometer is equipped with an Oxford Instruments Co. super-conducting magnet and coupled to a Sun Ultra 30 workstation running VNMR 6.1B. A dual-broadband 5-mm direct detection PFG probe, as well as, a 5-mm indirect detection probe, both from VARIAN Inc., were used. The sample temperature was calibrated against an ethylene glycol standard. Spectra were referenced to the central component of the quintet due to the CD_2H resonance of methanol at 3.315 ppm and 2.49 ppm for DMSO; D_2O and $CDCl_3$ spectra were referenced to the singlet resonance at 4.75 and 7.24 ppm, respectively.

1-D NMR spectra and two-dimensional (2-D) double quantum filtered (DQF) correlated spectroscopy (COSY) (Rance et al., 1983), total correlated spectroscopy (TOCSY) (Bax and Davis, 1985; Griesinger et al., 1988) and nuclear Overhauser enhancement spectroscopy (NOESY) (Jeener et al., 1979) spectra were recorded. The sign of the frequency in the indirect dimension was discriminated using the hyper-complex method (States et al., 1982).

Typically, 1-D spectra were collected with 32 scans and 32k data points in the time domain, with a spectral width of 8000 Hz, unless stated otherwise. 2-D spectra were collected with 256 increments (32 to 96 scans each) in t_1 and 2k data points in t_2, using a spectral width of 7000 Hz in both dimensions. High resolution DQF COSY spectra were also recorded, typically with 1024 increments in t_1 and 4k data points in t_2. The water signal was suppressed by low-power pre-saturation during the relaxation delay. TOCSY experiments were performed with a mixing time of 40 and 80 ms using a clean MLEV-17 pulse sequence (Bax and Davis, 1985; Griesinger et al., 1988). NOESY experiments were collected with mixing times of 150 and 250 ms.

1-D free induction decays (FIDs) were apodized with a 3 Hz line broadening function and zero filled to 132k data points prior to Fourier transformation. TOCSY and NOESY FID's were apodized by a 80º phase shifted squared sine-bell window function in both dimensions, whereas DQF COSY were apodized by a 0º phase shifted squared sine-bell window function in t_2 and a 0º sine-bell function in the second dimension. Drift correction (DC) offset correction was obtained using the last 20% of the FID prior to zero filling the time-domain data. A Gibbs filter switch was applied in the case of TOCSY and NOESY experiments before Fourier transform the data to yield data matrices of 2k x 2k for TOCSY and NOESY spectra, and 4k x 4k for high-resolution DQF-COSY spectra. Baselines were typically corrected using a fourth order polynomial function in both dimensions. A FLATT baseline correction algorithm was further applied to the t_2 dimension in TOCSY experiments (Guntert and Wüthrich, 1992). The NMR data was processed using home-built Felix 95.0 macros on a Silicon Graphics Octane workstation running the program Felix (Biosym Molecular Simulations, USA). The assignment procedure and NMR restraints generation was also accomplished with Felix.

3.2.3 Structure calculations

NOE-derived interproton distance restraints were classified into four ranges: strong (1.8–2.8 Å), medium (1.8–3.8 Å), weak (1.8–5.0 Å), and very weak (1.8–6.0 Å). $^3J_{HNH\alpha}$ coupling constants were measured from a high-resolution DQF COSY and converted to dihedral angle restraints ($^3J_{HNH\alpha} < 6$ Hz, $\phi =$

-60 ± 30). χ_1 torsion angle restraints for Val residues were assessed based on analysis of $^3J_{H\alpha H\beta}$ coupling constants and NOE patterns (Barsukov and Lian, 1993). Structures were calculated using the Xplor-NIH NMR molecular structure determination package, version 2.16.0 (Schwieters et al., 2006). The library files supplied with the Xplor-NIH program were modified to include a formyl-methionine residue.

For structure calculation, the covalent structure of δ-toxin with random ϕ, ψ, and χ angles, with *trans* planar peptide bonds, was used as starting structure. A total of 100 structures were calculated using simulated annealing in torsion angle space (Schwieters and Clore, 2001a) based on NMR data recorded on δ-toxin in CD_3OH at $pH^\star$ 3 and 298 K. Structure calculation included three main stages: high temperature dynamics (8000 steps at 3500 K), followed by a slow cooling simulated annealing protocol for convergence (279 cooling steps of 0.2 ps each, from 3500 K to 25 K), and a final torsion angle minimization. The target energy function comprises square well potential terms for the NOE-derived interproton distance restraints and torsion angle restraints; harmonic potentials for covalent geometry; quartic van der Waals repulsion potentials (Nilges et al., 1988); and a torsion angle data base potential of mean force (Clore and Kuszewski, 2002). A final set of 20 structures with the lowest overall energy were selected for further analysis, based on stereochemistry and energy considerations, using the programs PROCHECK-NMR (Laskowski et al., 1996) and VMD-XPLOR (Schwieters and Clore, 2001b). Figures of the structures were prepared using VMD (Humphrey et al., 1996).

3.2.4 Data deposition

^{1}H NMR chemical shifts for δ-toxin in CD_3OH have been deposited in BioMagResBank (Ulrich et al., 2008). The three-dimensional (3-D) cartesian coordinates for the ensemble of the 20 best structures have been deposited in the Protein Data Bank (Berman et al., 2000) under accession number 2*kam*.

3.3 Results and discussion

The lengthy process of solution structure determination of a protein by NMR, justifies a prior assessment of the spectra quality (Cavanagh et al., 1996). If basic attributes of the ^{1}H NMR spectrum, such as signal-to-noise ratio (S/N), linewidth, and chemical shift dispersion are not ideal for proper protein characterization, then sample conditions should be changed to improve spectra quality or alternative methods should be considered. Once the experimental conditions are defined and the spectra acquired, each NMR signal must be associated (assigned) to a nucleus in the molecule under study. The assignment procedure is based on information gathered from the NMR spectra, such as the chemical shift (reporting the chemical environment), scalar couplings (reporting through-bond interactions), and dipolar couplings (reporting through-space interactions).

The use of 2-D NMR techniques, such as DQF COSY, TOCSY, and NOESY, dramatically increases the resolution of protein NMR spectra. Wüthrich et al. (1982) introduced a systematic method for the full assignment of the 2D NMR spectra of proteins which relies only on information of the amino acid sequence of the protein. This sequential assignment procedure consists of two stages. The first stage involves the identification of the spin-spin coupling pattern characteristic of a particular amino acid residue type, the spin system identification. DQF-COSY and TOCSY spectra were used to identify the amino acid residue spin systems in δ-toxin. The second stage involves the assignment of each spin system to a particular residue in the polypeptide primary sequence. This sequence-specific assignment is achieved by correlating one amino acid spin system with the spin systems of its neighboring residues in the polypeptide primary sequence. This stage of assignment relies on the short-range through-space connectivities observed in NOESY spectra (Redfield, 1993).

Once the assignment of the NMR spectra has been completed, structural data such as scalar couplings and dipolar effects, are translated into restraints on dihedral angles and interproton distances, respectively. The pattern of these constraints allows the identification of secondary structure elements (Wüthrich, 1986). A qualitative assessment of the secondary structure can thus be readily made. The final stage of the structure determination process involves the use of

computational methods, such as simulated annealing and molecular dynamics calculations. These methods aim to sample the conformational space available to the peptide or protein whilst satisfying a set of NMR constraints (Sutcliffe, 1993).

3.3.1 Assessment of the 1-D ^{1}H NMR spectra

On its way to the cell membrane, peptide toxins do experience different chemical environments. A simple way to study the conformational plasticity of such polypeptides, is to determine the peptide 3-D structure on different chemical environments with increasing degree of lipophilicity. Log P partition coefficient is a common descriptor of lipophilicity or hydrophobicity (Thompson et al., 2006). We have chosen four different solvents with increasing log P values, namely water, dimethyl sulfoxide (-1.35), methanol (-0.74), and chloroform (+1.97) (Lide, 1997).

Although the majority of spectroscopic analyses will rely on n-D ($n > 1$) spectra, important preparatory work can be performed by using 1-D) ^{1}H NMR spectroscopy. Spectra of δ-toxin in different solvents are shown in figure 3.2.

One of the parameters that can be derived from 1-D NMR experiments is the chemical shift, particularly for the H^N and H^α from the backbone of the peptide. Even without resonance assignments, the resonance dispersion observed in a 1-D spectrum can be used as a crude estimate of the presence of ordered structures (Wang and Morden, 1997). The chemical shifts for protons in the 20 common amino acid residues in random coil polypeptides have been extensively studied and well characterized (Bundi and Wüthrich, 1979; Wishart et al., 1995). If the polypeptide is denatured, resonances from the amide (random-coil shifts 8.5–8.0 ppm) and α-carbon protons (random-coil shifts 4.4–4.1 ppm) fall within a small range of chemical shifts, i.e. they present reduced chemical shift dispersion. However, folded polypeptides will exhibit a broad range of chemical shifts due to anisotropic magnetic fields of proximal aromatic or carbonyl groups (Cavanagh et al., 1996).

Figure 3.2 shows that δ-toxin may present structure in DMSO and in CD_3OH solutions, since conformation-dependent secondary chemical shifts are observed. On the other hand, in D_2O and CD_3Cl_3 solutions, the peptide must

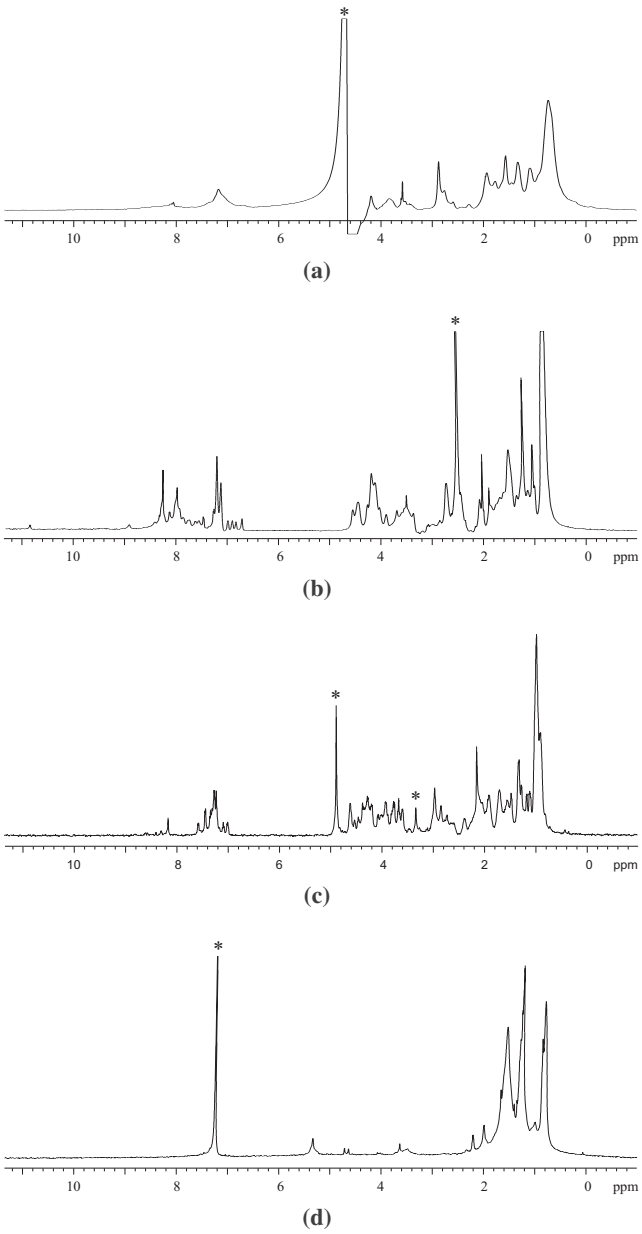


Figure 3.2: 1-D ^{1}H NMR spectra of 3 mM δ-toxin in different chemical environments acquired at 500 MHz and 303 K: **(a)** deuterium oxide; **(b)** dimethyl-d_6 sulfoxide; **(c)** methyl-d_3 alcohol; and **(d)** chloroform-d_3. Spectra were acquired using a standard 1-D two-pulse sequence whereas the spectrum in methanol was recorded using a BPPLED sequence with a gradient pulse of 0.34 Tm^{-1} (see chapter2). All spectra were apodized with a 3 Hz line broadening function prior to Fourier transformation, and referenced to solvent signals (marked with an $*$).

be in a conformationally disordered state (or with very little stable structure), with a characteristic random-coil chemical shift distribution. As expected due to the δ-toxin amphiphatic characteristics, solvents with intermediary log P values (CD_3OH and DMSO) and able to participate both in H-bonding and van der Waals contacts, are the most suited for structure determination. Hence, we collected 2-D NMR spectra for δ-toxin in both of these solvents as well as in CD_3OH/water mixtures, to assess the peptide structure.

3.3.2 Assignment of ^{1}H resonances and secondary structure

δ-toxin in CD_3OH

The structure of δ-toxin in methanol solution has already been studied by Tappin et al. (1988), using high-resolution NMR in combination with molecular dynamics simulations. Though a set of NMR constraints were determined by the authors and a 3-D structure published, no structure deposition was made in the RCSB Protein Data Bank[2].

In the present work, it was possible to identify the complete spin systems of all 26 residues for δ-toxin, with the exception of the Met1 H^{ε}, Leu12 $H^{\delta 1/\delta 2}$ and H^{γ}, and Ile16 $H^{\delta 1}$ protons. The sequence specific resonance assignment is illustrated in figure 3.3 and proton assignments are detailed in appendix table A.1. The atom identifier nomenclature used throughout this chapter, as well as any NMR related nomenclature, follow the guidelines described elsewhere (Markley et al., 1998).

The sequence specific assignment of spin systems relies on the observation of a series of inter-residue NOEs, namely H^N_i–H^N_{i+1}, H^{α}_i–H^N_{i+1}, and H^{β}_i–H^N_{i+1} (Wüthrich, 1986). Using the NOESY spectrum, we were able to trace connectivities between sequential protons along the backbone (H^N or H^{α}) from the N-formyl proton to Asp4, from Ile5 to Lys22, and then from Phe23 to Lys26 H^N. The connectivities were lost due to identical H^N chemical shifts between residues $i, i+1$ ($i = 4, 22$). Figure 3.4 presents a survey of the experimental data used for the secondary structure determination for δ-toxin in CD_3OH. It can

[2]Moreover, the constraints are not available anymore (priv. comm.).

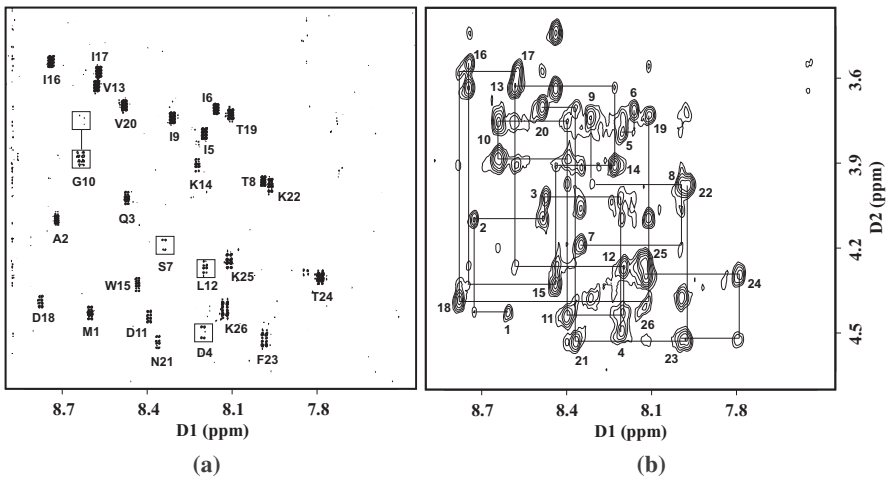


Figure 3.3: $^{1}H^{N}$–$^{1}H^{\alpha}$ fingerprint region of δ-toxin in CD_3OH at pH* 3 and 298 K. The cross-peaks are labeled by the one-letter amino acid code and the residue number. **(a)** High-resolution phase-sensitive DQF COSY spectrum. The boxes indicate cross-correlations with a low signal-to-noise ratio. **(b)** NOESY spectrum acquired with a mixing time of 250 ms, showing the sequential NOE connectivities between spin systems. Connectivities $i, i+1$ are not represented due to spectral overlap in the following cases: I6/S7, I16/I17, T19/V20 and K25/K26.

be seen that, in addition to sequential NOEs extending along all the molecule, there are a considerable number of H^{α}_{i}–H^{N}_{i+3} and H^{α}_{i}–H^{β}_{i+3} NOEs, which are characteristic of a helical structure (Wüthrich, 1986). The number of H^{α}_{i}–H^{N}_{i+4} NOE connectivities are particularly significant, thus suggesting that δ-toxin is an α-helix over most of its length.

$^{1}H^{N}$–$^{1}H^{\alpha}$ vicinal spin-spin coupling constants ($^{3}J_{HNH\alpha}$) can provide further conformational information. These coupling constants are related to the torsional angle ϕ by the Karplus equation (Karplus, 1959). $^{3}J_{HNH\alpha}$ coupling constants were measured from a high-resolution DQF-COSY spectrum. The observed values are indicated in figure 3.4 as well as in appendix table A.4. It can be seen that most of these ^{3}J couplings are less than 6 Hz, consistent with the existence of a helical structure for most of the molecule (Pardi et al., 1984). Hence, the overall analysis of figure 3.4 indicates that δ-toxin in CD_3OH at low pH and 298 K, adopts an α-helical structure between residues Ala2 and Lys24.

A larger helical content was observed in this work when compared with

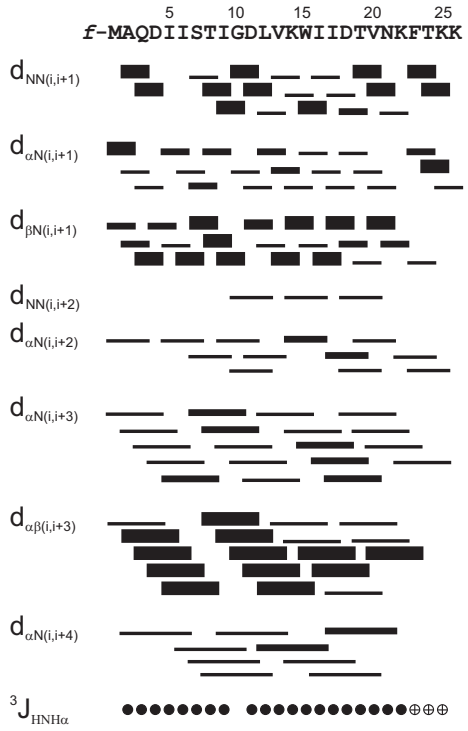


Figure 3.4: Summary of the NMR data used to establish the secondary structure and to calculate the three-dimensional structure of δ-toxin in CD_3OH, at pH* 3 and 298 K. Sequential and medium-range NOE connectivities, as well as, $^3J_{HNH\alpha}$ coupling constants are presented, along with the amino acid sequence. The height of the bars for NOE connectivities indicates the NOE intensity. Values of $^3J_{HNH\alpha}$ are indicated by filled ($^3J_{HNH\alpha} < 6$ Hz) and crossed ($6 \leq {}^3J_{HNH\alpha} \leq 8$ Hz) circles. The absence of non-glycine $^3J_{HNH\alpha}$ coupling constants is due to spectral overlap.

previous NMR studies. Tappin et al. (1988) determined that δ-toxin in methanol adopts a helical structure extending from Ala2 to Val20, whereas Bladon et al. (1992) reported that δ-toxin adopts an α-helical structure between residues Thr8 and Lys22. The main difference between those studies and this one, relies on the initial peptide sample concentration, which was 6 mM on both of them, whereas we used half of that peptide concentration. Our results are in agreement with a recent circular dichroism (CD) study of δ-toxin in methanol at a much lower concentration (20 μM, pH 3, room temperature) (Cruz, 2007). After CD spectra deconvolution, the author reached the conclusion that 93% of the peptide (i.e., about 24 amino acid residues) is in α-helix conformation.

From these observations, we may conclude that on the NMR timescale, the structural content of the peptide in methanolic solution is maintained from the μM to the mM range, but at higher concentrations the peptide may loose some structure. Such difference in peptide concentration among the different studies, undoubtedly has a significant importance on the type of molecule ensembles present in solution, since δ-toxin is prone to aggregation due to its amphipathic character. Thus, we can speculate that different aggregation states in solution may have different regular structure content.

From our data, it was also possible to stereospecific assign the methyl groups of the two Val residues (Barsukov and Lian, 1993). In both cases, large $^{3}J_{H\alpha H\beta}$ coupling constants of about 12 Hz were observed in a high-resolution DQF COSY experiment, together with strong H^{α}–H^{γ} NOE couplings and only one strong H^{N}–H^{γ} NOE. These NMR parameters are characteristic of a *trans* rotamer around the C^{α}–C^{β} bond, and thus stereospecific assignment of the methyl groups is possible (Zuiderweg et al., 1985). We notice that the Val stereospecific assignment determined from our data does not agree with the one previously published (Tappin et al., 1988).

δ-toxin in CD_3OH / D_2O

The structure of δ-toxin was also assessed in two largely methanolic solutions, namely 75% CD_3OH / 25% D_2O and 65% CD_3OH / 35% D_2O, at $pH^{\star}$ 3 and 303 K.

The $^{1}H^{N}$–$^{1}H^{\alpha}$ fingerprint region of a DQF COSY experiment of δ-toxin in 75% CD_3OH / 25% D_2O shows all expected correlations for a 26 residue peptide (see figure 3.5a). However, additional cross correlations are observed in the fingerprint region, indicating that δ-toxin does not exist in a single NMR ensemble of conformers in such experimental conditions. Nevertheless, though the assignment process becomes more difficult to accomplish, it was possible to fully characterize the 26 spin systems, with the exception of some sidechain protons of Ile, Asn, Met and Lys residues. We were also able to stereospecifically assign the methyl groups of the two Val residues in the 75% CD_3OH / 25% D_2O system. The sequence specific assignment is detailed in the appendix table A.2. On the other hand, the $^{1}H^{N}$–$^{1}H^{\alpha}$ fingerprint region

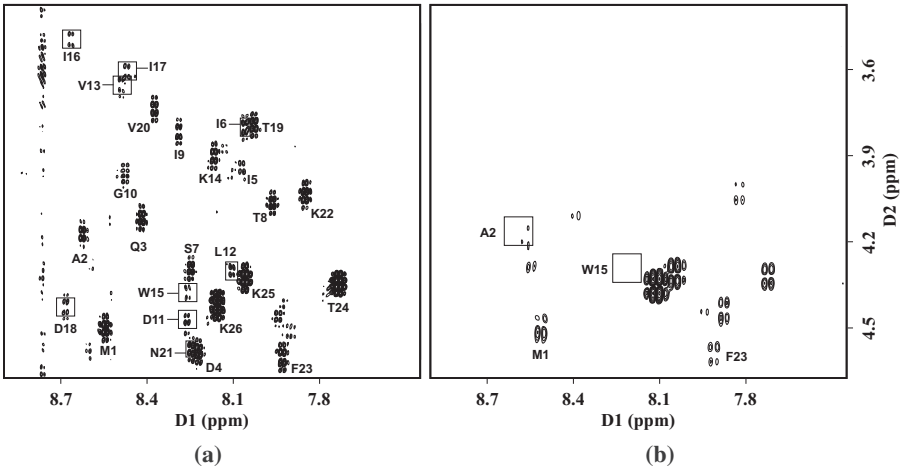


Figure 3.5: $^1H^N$–$^1H^\alpha$ fingerprint region of δ-toxin in CD_3OH / D_2O at pH* 3 and 298 K. **(a)** High-resolution phase-sensitive DQF COSY spectrum of δ-toxin in 75% CD_3OH / 25% D_2O. **(b)** Phase-sensitive DQF COSY spectrum of δ-toxin in 65% CD_3OH / 35% D_2O. Assigned cross-peaks are labeled by the one-letter amino acid code and the residue number. Square boxes indicate cross-correlations with a weak signal-to-noise ratio or partially overlapped.

of δ-toxin in 65% CD_3OH / 35% D_2O shows a very small number of cross correlations (see figure 3.5b), suggesting that δ-toxin is less structured. Even with a severe lack of $^1H^N$–$^1H^\alpha$ correlations, it was possible to identify four spin systems, namely Met1, Ala2, Trp15 and Phe23, mainly based on the unique patterns arising from the aromatic residues, and through identification of the ^{1}H-formyl resonance. The N-terminal residues were sequentially assigned from NOE connectivity analysis. The assigned proton resonances for δ-toxin in 65% CD_3OH / 35% D_2O are detailed on the appendix table A.3.

Figure 3.6 summarizes the inter-residue NOEs observed for δ-toxin in 75% CD_3OH / 25% D_2O. Connectivities between sequential H^N or H^α protons were observed along the backbone from the N-formyl proton to Asp4, from Ile6 to Val13, and then from Trp15 to Lys26. Along with sequential NOEs, there are also a considerable number of H^α_i–H^N_{i+3} and H^α_i–H^β_{i+3} NOEs, characteristic of a helical structure. We also observe some sparse medium and strong H^α_i–H^N_{i+2} NOEs, which are characteristic of a tighter helical structure (3_{10}-helix), as well as H^α_i–H^N_{i+4} NOEs which are only found in α-helices. If we compare the NMR

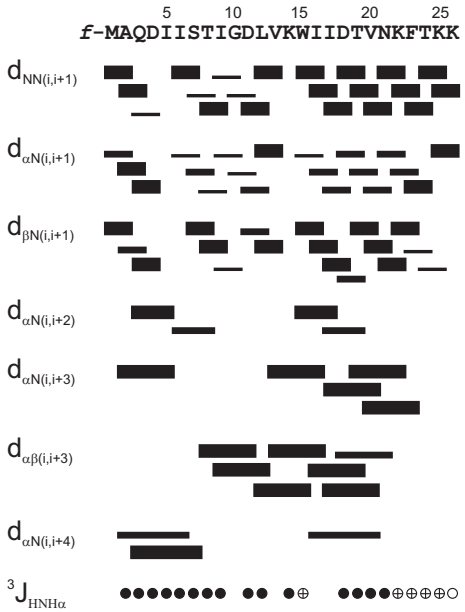


Figure 3.6: Summary of the NMR data used to establish the secondary structure of δ-toxin in 75% CD_3OH / 25% D_2O, at pH* 3 and 298 K. Sequential and medium-range NOE connectivities, as well as $^3J_{HNH\alpha}$ coupling constants are presented, along with the amino acid sequence. The height of the bars for NOE connectivities indicates the NOE intensity. Values of $^3J_{HNH\alpha}$ are indicated by filled ($^3J_{HNH\alpha} < 6$ Hz), crossed ($6 \leq {}^3J_{HNH\alpha} \leq 8$ Hz) and open ($^3J_{HNH\alpha} > 8$ Hz) circles. The absence of non-glycine $^3J_{HNH\alpha}$ coupling constants is due to spectral overlap.

derived structural restraints, qualitatively depicted on figures 3.4 and 3.6, we can conclude that the secondary structure of δ-toxin in both solvent systems is mostly an helix. Furthermore, the helix seems to be more regular and stable on neat methanol (α-helix), since we were able to observe NOE connectivities and $^3J_{HNH\alpha}$ coupling constants consistent with such secondary structure motif, through out the peptide primary sequence. In 75% CD_3OH / 25% D_2O, the NMR data is consistent with an helix from residues Ala2 to Ser7, and then from Thr8 to Asn21. However, the data is characteristic of a less regular α-helix.

The structure of δ-toxin in a CD_3OH/H_2O (2:1) solution has been investigated by Bladon et al. (1992) using a combination of two-dimensional NMR experiments and molecular modeling techniques. Their work demonstrated that δ-toxin assumes a well-ordered helical three-dimensional structure in methanol-water solutions, extending from Ile5 to Thr24. As previously explained, the

experimental conditions are not the same and, as pointed out by the authors, due to the limited NMR data, further studies should be done.

Our data, in particular the presence of medium and strong H_i^{α}–H_{i+2}^{N} NOEs in the regions Gln3–Thr8 and Lys14–Thr19, and the smaller number of H_i^{α}–H_{i+4}^{N} NOEs, suggests an overall more mobile helical structure in 75% CD_3OH, undergoing significant conformational averaging. Moreover, the α-helix seems to be less stable between residues Thr8 and Trp15.

δ-toxin in DMSO

The 1-D ^{1}H-NMR spectrum of δ-toxin in dimethyl-d_6 sulfoxide (DMSO) shows a broad chemical shift distribution of resonances, characteristic of folded proteins (see figure 3.2b). We can also observe that peptide resonance linewidths are generally broader, when compared with the methanol system (figure 3.2c). Nuclear spin relaxation is a consequence of coupling of the spin system to the chemical surrounding environment (Keeler, 2006). DMSO has a high viscosity which in turn increases the rotational time of the polypeptide in solution. Solute molecules in this environment will have shorter relaxation times. This is reflected on the NMR spectrum as line broadening, though other causes such as aggregation or chemical exchange effects may also contribute to it (Cavanagh et al., 1996). 1-D spectra were acquired with increasing sample temperatures to a maximum of 323 K (data not shown). Qualitatively we do not noticed any significant sharpening of the resonance lines.

The $^1H^N$–$^1H^{\alpha}$ fingerprint region of a DQF COSY experiment of δ-toxin in dimethyl-d_6 shows that only 21 spin systems were assigned unambiguously (figure 3.7). Residues Ile6, Ser7, Gly10, Asp11, and Asp18 were not identified. The difficulty to assign these residues can be explained from a large broadening of the respective resonances, with a concomitant peak merging into the spectrum baseline. The identified spin systems are characterized in the appendix table A.5. A brief comparison of the fingerprint region of all DQF COSY spectra acquired, shows that the chemical environment felt by the peptide protons in DMSO is quite different from the chemical environment felt on methanolic solutions. This suggests that the three-dimensional structure of the peptide may be different in the studied solvents. It is to expect then, a quite different pattern

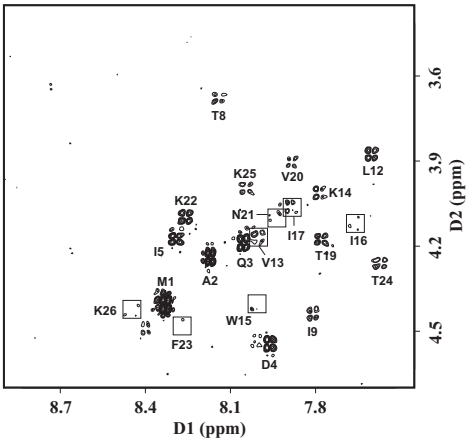


Figure 3.7: High-resolution phase-sensitive DQF COSY spectrum of δ-toxin in DMSO at 303 K. The region shown contains the $^{1}H^{N}$–$^{1}H^{\alpha}$ cross-peaks, labeled by the one-letter amino acid code and the residue number. The boxes indicate cross-correlations with a low signal-to-noise ratio.

of the NMR data used to establish the secondary structure.

Figure 3.8a summarizes the inter-residue NOEs observed for δ-toxin in DMSO. Several gaps in the sequential connectivities are observed, thus impairing a more in depth analysis of the peptide structure under such experimental conditions. A small number of inter-residue NOEs was observed and for only six residues we were able to determine $^{3}J_{HNH\alpha}$ coupling constants (see appendix table A.4). Nonetheless, these NMR restraints seem to indicate a structure more closely related to β-structure than a helical structure, namely due to $^{3}J_{HNH\alpha} > 8$ Hz.

Due to the scarce data available we also assessed the chemical shift index (CSI) for δ-toxin in this DMSO sample, in order to qualitatively check for any secondary patterns. Briefly, since the chemical shift of a nucleus is sensitive to the environment, it should also contain structural information. Wishart et al. (1991) have demonstrated that ^{1}H NMR chemical shifts are strongly dependent on the character and nature of protein secondary structure . Wishart et al. (1992) have produced a simple method for secondary structure determination by analyzing the difference between the $^{1}H^{\alpha}$ chemical shift of each residue in a protein and that reported for the same residue type in a random coil conformation. Helical segments have groupings of $^{1}H^{\alpha}$ whose chemical shifts

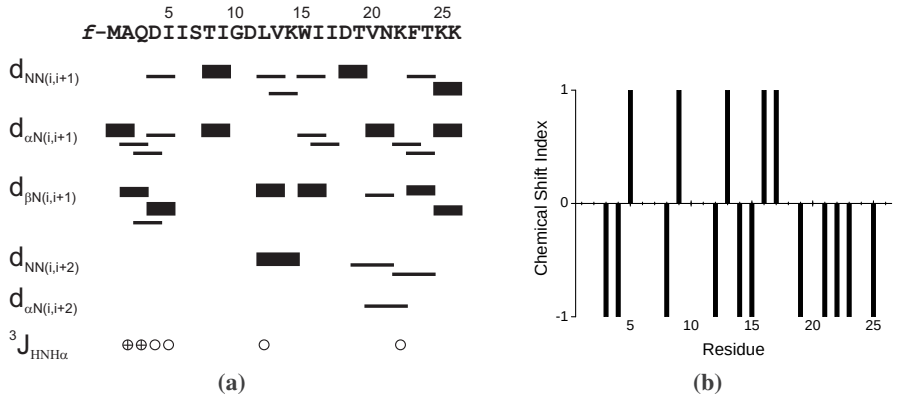


Figure 3.8: (a) Summary of the NMR data used to establish the secondary structure, of δ-toxin in DMSO, at 303 K. The height of the bars for NOE connectivities indicates the NOE intensity. Values of $^3J_{HNH\alpha}$ are indicated by crossed ($6 \leq {}^3J_{HNH\alpha} \leq 8$ Hz) and open ($^3J_{HNH\alpha} > 8$ Hz) circles. The absence of non-glycine $^3J_{HNH\alpha}$ coupling constants is due to spectral overlap and ambiguous spin system assignment. **(b)** Chemical shift index of $^1H^{\alpha}$ *versus* δ-toxin primary sequence in DMSO.

are consistently smaller than the random coil values ($CSI = -1$) whereas β-strands have chemical shift values consistently larger ($CSI = 1$). In this way, the location of helix and strand segments may be identified. Figure 3.8a shows the CSI for δ-toxin in DMSO, and there is no strong indication of the presence neither of α-helix or β-strand secondary motifs.

3.3.3 Three-dimensional solution structure

The quality of the final structure of a protein or peptide is determined to a large extent by the number of constraints available for structure calculation. With exception for the sample δ-toxin in CD_3OH, it was not possible to gather enough experimental constraints from the other solvent systems. The 3-D structure of δ-toxin was thus only calculated from the spectral data collected for δ-toxin in CD_3OH, at pH* 3 and 298 K.

The solution structure of δ-toxin was calculated on the basis of 395 experimental NMR restraints, including NOE derived inter-proton distance restraints and torsional dihedral angle restraints. A slow cooling protocol in torsion angle space was applied to a randomly generated extended structure with correct covalent geometry. A final set of 20 structures with the lowest overall energy

Table 3.1: Structural statistics for δ-toxin in CD_3OH

NOE distance restraints		
All	373	
Intraresidue (i=j)	225	
Sequential ($\|i-j\|=1$)	79	
Short-range ($2 \leq \|i-j\| \leq 4$)	69	
Diehdral angle restraints		
ϕ	20	
χ	2	
Average RMSD from ideal geometry*		
Bonds (Å)	0.0036	± 0.0005
Angles (°)	0.5340	± 0.0083
Impropers (°)	0.3313	± 0.0116
Average RMSD of atomic coordinates (Å)†		
Residues 2 – 22		
Backbone‡	0.25	± 0.07
Heavy atoms	0.67	± 0.09
All residues		
Backbone‡	1.00	± 0.41
Heavy atoms	1.82	± 0.50
Ramachandran analysis¶		
Residues in most favored region (%)	95.0	
Residues in additional allowed regions (%)	4.6	
Residues in generously allowed regions (%)	0.4	
Residues in disallowed regions (%)	0.0	

There was no NOE violations > 0.5 Å or dihedral angle violations > 5°; * The values for the bonds, angles, and impropers show the deviations from ideal values based on perfect stereochemistry; † RMSD of the 20 accepted structures, with respect to the structure with lowest energy; ‡ N, C^{α} and C atoms; ¶ As determined by the program PROCHECK-NMR (Laskowski et al., 1996) for all residues except Gly10.

was selected, with no inter-proton distance restraint violations greater than 0.5 Å, and no angle violations greater than 5°. Statistical parameters for the ensemble of 20 calculated structures are presented in table 3.1.

The ensemble of structures has a low root-mean square deviation (RMSD) across the whole sequence. As expected, the lowest RMSD values are found in the regions containing regular secondary structure; the backbone RMSD in such regions is 0.25 ± 0.07 Å, while that calculated over all residues is 1.00 ± 0.41 Å. This observation of higher RMSD when fitting is done on all residues, agrees well with the lack of secondary structure restraints on the C-terminus region of the peptide. An overlay of the 20 lowest energy solution structures of δ-toxin is shown in figure 3.9a. A Ramachandran analysis indicates that more than 95% of the residues are in the most favored regions of the conformational space available for proteins.

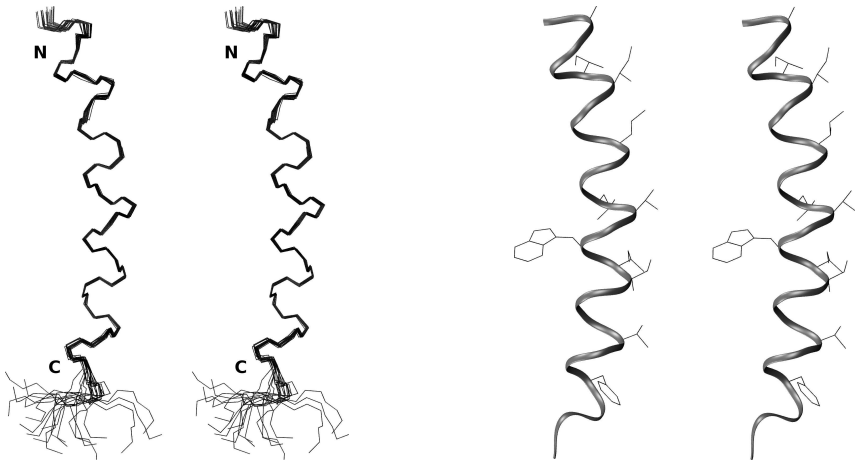


Figure 3.9: Three-dimensional structure of δ-toxin in CD_3OH. **(a)** Superimposed backbone stereo-view of an ensemble of 20 structures. The structures were calculated with Xplor-NIH (Schwieters et al., 2006) using NOE and dihedral derived restraints, at $pH^\star$ 3 and 298 K. The backbone heavy atoms (N,C^α,C) of residues Ala2 to Lys22 have been superimposed using VMD (Humphrey et al., 1996). The letters N and C refer to the amino and carboxy termini, respectively. **(b)** Stereo view of the lowest-energy structure in the family shown in (a), with backbone atoms represented by a ribbon, and side-chains of hydrophobic residues by thin solid lines.

A cartoon of the lowest-energy structure obtained from the final ensemble of structures is shown in figure 3.9b. The side-chains of hydrophobic residues are depicted, showing clear segregation on one side of the helical structure, with the exception of Trp15 on the hydrophilic side. The structure is a well formed α-helix between residues 2 and 23/24, while the first residue of N-terminus and residues 25–26 in the C-terminus are in random coil.

3.4 Concluding remarks

We assessed the secondary structure content of δ-toxin in membrane-mimetic environments by NMR. In neat methanol the peptide has an α-helical content of about 93%, forming a well-ordered α-helical structure between residues 2 and 24, at low pH. An increase of the water content in the methanolic solution leads to a decrease in the peptide helical content. In methanol/water

(75%:25%), at low pH*, the helical content of δ-toxin is about 77%, whereas in methanol/water (65%:35%) resonances with random coil chemical shifts start to predominate in the NMR spectra. In neat water, the ^{1}H 1-D NMR spectrum shows a poor resonance dispersion, indicating the absence of regular structure, probably due to aggregation. The secondary structure of δ-toxin in dimethyl sulfoxide was also assessed, but in spite of a broad resonance dispersion in ^{1}H 1-D NMR spectra, the available NMR data is not enough to tell if there is a well-formed secondary structure motifs. Nonetheless, the data seem to suggest that no helical motifs are present in such experimental conditions. These observations demonstrate that the peptide conformation is strongly dependent on the experimental conditions.

Once δ-toxin is excreted by *Staphylococcus aureus* to the plasma, on its way to eukaryotic membranes a wealth of chemical environments are available to the toxin. The conformational plasticity of such toxins ultimately depend on the neighboring molecules. Thus, though the study of peptide structure in membrane-mimetic environments is rather simplistic, it has interesting implications for the study of which conformational structures are available to toxins when getting closer to a lipophilic environment. In that regard, we were able to improve the NMR data available for δ-toxin in methanol, and to deposit in the RCSB Protein data bank an ensemble of structures. Circular dichroism studies are underway to further study how the δ-toxin structure, stability and dynamics vary as a function of the solvent environment, and in association with lipid vesicles. Though peptide-membrane interaction mechanisms are still under strong scientific debate, it seems to be a consensus nowadays that membrane permeation and/or disruption depends on a number of parameters such as the nature of the toxin and the membrane lipids, peptide concentration and environmental conditions (Bechinger and Lohner, 2006).

NMR and CD studies on peptide conformation in aqueous, membrane-mimetic environments, and associated to lipid vesicles will provide important structural information and help the interpretation of kinetic and equilibrium studies (Pokorny and Almeida, 2004). These structural studies can also provide guidance and a check for computer simulations (see chapter 5).

Acknowledgments

We thank Carla Barreira for her collaboration on the spectral assignment.

Chapter 4

A brief approach to lipid bilayer simulations

一、人格完成に努むること
一、誠の道を守ること
一、努力の精神を養うこと
一、礼儀を重んずること
一、血気の勇を戒むること

Dojo Kun, Gichin Funakoshi

4.1 Introduction

A fundamental precondition for life is the compartmentalization of cells and organelles from the surrounding environment that is provided by biological membranes. Beyond this basic role as a barrier, biomembranes facilitate a number of other functions that are mainly determined by the type of proteins associated with or embedded in the bilayer. Membranes are therefore of high biological relevance.

Molecular dynamics (MD) simulations is a powerful technique for evaluation of interactions on a system of discrete particles. The interactions of the particles are empirically described by a potential energy function from which the forces that act on each particle are derived. Thus the dynamic behavior of the system may be studied using Newton's classical equations of motion, for all atoms in the system (Kandt et al., 2007).

Since their introduction in the late 1950s, MD simulations have become a common tool to investigate structure-activity relationships in biological membranes. Providing the element of dynamics at an atomic level of detail, these simulations facilitate the interpretation of experimental data and give access to information not directly accessible by experiments.

Force fields could be considered the primary assumption in MD simulations, they characterize the interactions between particles. There are many force fields described in the literature, but currently only four are widely used for simulating biological macromolecules: Amber (Cornell et al., 1995), CHARMM (MacKerell et al., 1998), GROMOS (van Gunsteren et al., 1996) and OPLS (Jorgensen and Tirado-Rives, 1988). All four force fields have different strengths and weaknesses, but most importantly they share a number of the same limitations due to the simplifying assumptions necessary in large scale MD simulations (Kandt et al., 2007). Though all of these force fields reproduce fairly well many protein characteristics, the same is not true when simulating biomembranes.

The force field parameters used in biomolecular simulations are generally obtained by a process of refinement against high-resolution experimental data and *ab initio* quantum-mechanical data. Though significant progress has been made in developing modern protein force fields, it has been quite challenging to obtain accurate experimental data to validate and improve lipid models. In fact,

only two phospholipid force fields are in common use today, namely CHARMM and an older version which was built by combining both OPLS and AMBER forcefields with some additional parameters from Berger et al. (1997). It is even more complex to obtain a force field which is able to describe lipid-protein interactions because structures of membrane proteins and peptides typically have a significantly lower resolution than in soluble proteins. Nowadays, the most straightforward approach to simulate membrane/protein systems involves directly combining mathematically compatible protein and lipid force fields. Though this approach has proven to yield useful insights into membrane protein behavior, a re-parameterization of these mixed force fields is clearly needed to provide a consistent force field.

A key difficulty with classical MD simulations of molecules that move between two very different environments is that parameters used to describe them have to be accurate in both environments, and typical biomolecular force fields were parameterized for aqueous solution. Significant improvements are being made in force field re-parameterization taking into account the free-energies of transfer between water and hydrophobic environments as well as the interfacial partitioning of model peptides. Thus, it is expected that a better description of the thermodynamics of membrane proteins at the water lipid interface will be accomplished. Moreover, the parameter sets are starting to accommodate electronic polarizability effects, with the aim of improving protein and lipid force fields (Davis et al., 2009).

The GROMOS force field has been successfully used over the years in biomolecular simulations of a variety of molecules, such as peptides, proteins, sugars and nucleic acids. Recently, this force field was re-parameterized for a correct treatment of aliphatic chains (Schuler et al., 2001), such as the ones composing the phospholipids. The relevance of the new force field is that, unlike many other force fields for lipid simulations, it is consistent with the parameters used for simulating other biomolecules such as peptides and nucleotides.

In the current chapter the force field GROMOS96 parameter set 45A3 is tested on a bilayer of dimyristoylphosphatidylcholine (DMPC; see figure 4.1) in the liquid-crystalline L_α phase in water. The performance of variants of the force field parameters is evaluated in this system, for a set of generally

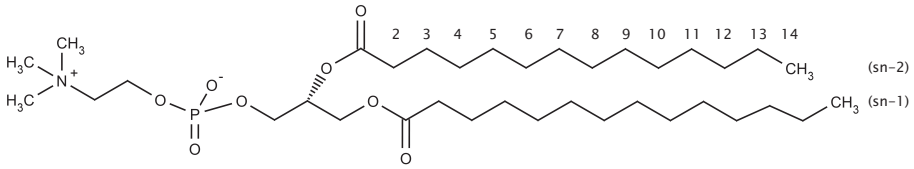


Figure 4.1: DMPC molecule. The aliphatic carbon numbering used in the text is shown. The structure data file used to produce this image was built with Lipid MAPS (Fahy et al., 2007) and manipulated with Marvin (Marvin, 2007).

used parameters in protein simulations. This way, direct comparisons between free protein dynamics in solution and lipid-protein interaction dynamics can be readily made. Though a similar work was already done to benchmark the consistency between a bilayer of dipalmitoylphosphatidylcholine (DPPC) against this same force field (Chandrasekhar et al., 2003), the timescale used by the authors in their simulations was merely demonstrative of the capabilities of the new parameter set.

4.2 Methods

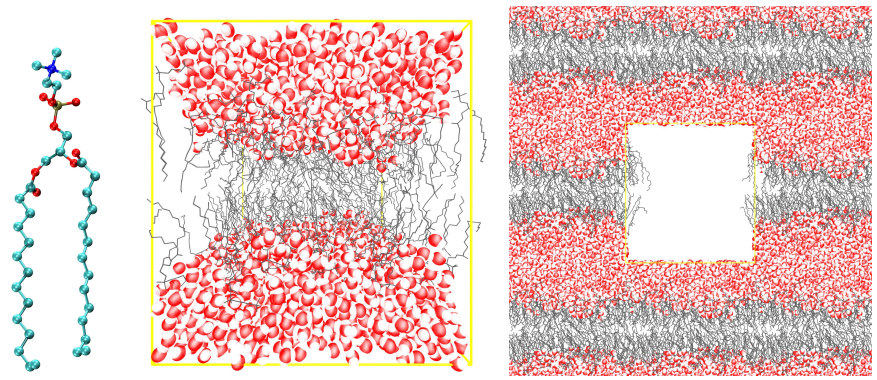
4.2.1 Simulated lipid model

We have simulated a DMPC lipid bilayer in the liquid-crystalline L_α phase, consisting of two layers of 64 lipids each, separated in the z-direction by a layer of 3655 water molecules[1]. The starting structure for the fully hydrated lipid bilayer was downloaded from the web site of the Biocomputing Group at the University of Calgary (`http://moose.bio.ucalgary.ca`), and is depicted in figure 4.2.

4.2.2 Forcefield parameters

The force field parameters for bonded and nonbonded interactions were taken from the recent GROMOS96 parameter set termed 45A3. This is a

[1] By default lipid molecules are modeled with the aliphatic chains towards the z-axis of the simulation box, whereas the lipid "carpet" lies along the xy-plane.



(a) **(b)** **(c)**

Figure 4.2: **(a)** CPK representation of a single DMPC molecule. The carbon atoms are represented by gray spheres, choline nitrogen in blue, phosphorus in green, and oxygen atoms in red. **(b)** Perspective snapshot of the starting configuration of a system containing 128 DMPC lipids (in gray lines) and 3655 water molecules (VDW representation, with oxygens in red, and hydrogens in shadowed white). The boundaries of the simulated box are shown in yellow. **(c)** Under periodic boundary conditions, the system is "replicated" in space, allowing a continuous bilayer phase to be simulated. In the center of the orthographic image, the replication box shown in (b) is omitted to further stress out the periodicity of the system.

united-atom parameter set where the hydrogen atoms are included implicitly by treating the carbons and their attached hydrogens as a single group centered at the carbon. The re-parameterization of the aliphatic hydrocarbons yielded excellent agreement with experimental values for free energy of hydration, liquid densities and heat of vaporization (Schuler et al., 2001). Hence, the new parameter set should be suitable for application to lipid aggregates and bio-polymers.

Four variants of the 45A3 parameter set were tested (see table 4.1), and differ in the van der Waals parameters for the planar ester carbon, which is known to have a significant effect on the cross-section area per lipid (Chandrasekhar and van Gunsteren, 2001). These force field variants have been previously presented (Chandrasekhar et al., 2003). Briefly, the atom type describing the phospholipid ester carbon was modeled as a GROMOS96 atom type 11 (sp2 carbon in planar group), type 45 (tetrahedral sp3 carbon bound to four non-hydrogen atoms), type 45x11 or type 45x12. In these last two cases, the geometric mean of the

Table 4.1: The van der Waals parameters* of the phospholipid planar ester carbon

IAC†	ε / KJ mol^{-1}	σ / nm
11	0.4059	0.3361
45	0.006995	0.6639
45x11‡	0.05328	0.4724
45x12¶	0.02577	0.5773

* The van der Waals interaction is defined as $4\varepsilon\left[(\sigma/r)^{12}-(\sigma/r)^{6}\right]$ (see Appendix B); † van der Waals Integer Atom Code from the GROMOS96 force field to describe the ester carbon;
‡ The geometric mean of type number 45 and type number 11;
¶ The geometric mean of type number 45 and type number 12.

bare tetrahedral carbon (type number 45), and respectively, the planar carbon (type number 11) or the united tetrahedral carbon (type number 12) are used. For water, the simple point-charge (SPC) model was chosen (Berendsen et al., 1981).

The topological description of the DMPC molecule, i.e. the combination of the set of all DMPC atoms with their non-bonded and bonded interaction parameters as well as the connectivity of those atoms in the molecule was built *de novo* . Briefly, atom charges were taken from Chiu et al. (1995) work, charge group boundaries from Chandrasekhar et al. (2003), and the planar ester carbon modeled as "CH0" atom type (Schuler et al., 2001).

4.2.3 Simulation details

The simulations were performed under periodic boundary conditions with the GROMACS package[2], version 3.3 (Lindahl et al., 2001). Simulations were carried out at a constant pressure of 1 bar and a constant temperature of 303 K. A weak pressure-coupling scheme was adopted using the Berendsen barostat (Berendsen et al., 1984). The pressure was controlled semi-isotropically ($Np_{xy}p_zT$ ensemble), where the height of the box (z direction) and the cross-

[2]At the time the simulations were done, there was no compatible GROMACS force field. See footnote on page 212, Appendix B.

section area (x,y plane) were allowed to vary independently of each other. The lipids and water molecules were separately coupled to the heat bath using the Berendsen thermostat (Berendsen et al., 1984). The coupling times for temperature and pressure were set at 0.1 and 1 ps, respectively. All bond lengths were constrained to their equilibrium values by the LINCS algorithm (Hess et al., 1997), and the water geometry was maintained with the SETTLE routine (Miyamoto and Kollman, 1992). The efficiency and stability of both algorithms allow the usage of a leap-frog integration scheme with a time step of 5 fs, as examined and validated in previous publications (Anézo et al., 2003). A bilayer system was simulated with a time step of 2 fs (data not shown) without any change in the equilibrium properties, further demonstrating that a time step of 5 fs provides enough accuracy. Non bonded interactions were evaluated using a twin range cut off of 1.0 and 1.4 nm. The long-range electrostatic forces were treated by the particle-mesh Ewald (PME) summation algorithm (Darden et al., 1993). Interactions within the shorter range cut off were evaluated at every step whereas interactions within the longer range cut off were evaluated every 4 steps.

The initial fully hydrated bilayer of DMPC molecules was firstly minimized using the force field parameter set 45A3-11 (system label B_11) with a quasi-Newtonian algorithm according to the low-memory Broyden-Fletcher-Goldfarb-Shanno approach (Byrd et al., 1995). This configuration was then used as the starting point for the bilayer simulations depicted on table 4.2, using the same Maxwellian velocity distribution. The only difference between these four lipid simulated systems is the force field used. Each of the systems was equilibrated for 10 ps before starting production runs of 50 ns. Coordinates of the full system were saved every 1 ps for further analysis.

A box of neat water (simulation label SPC, on table 4.2) was also run, in order to calculate the water specific volume. The system consisted of a dodecahedron box containing 3227 SPC molecules, simulated during 2 ns with the same parameters described above, where applicable.

The simulations took $\sim$ 2000 CPU hours on a cluster of Sun Fire X4100 computing nodes.

Table 4.2: Overview of MD simulations performed

Label	System	Water molecules	Number of atoms
B_11	water + DMPC	3655	16853
B_45	water + DMPC	3655	16853
B_45x11	water + DMPC	3655	16853
B_45x12	water + DMPC	3655	16853
SPC	water	3227	9681

4.2.4 Data analysis

In a molecular dynamics simulation, a description of the simulated system is normally stored frequently at regular τ time intervals. Thus, the data points for a given property $P(\tau)$ are usually not independent but highly correlated with each other (Allen and Tildesley, 1996). The standard error in the average of a correlated fluctuating quantity, such as $P(\tau)$, can be obtained using a block average procedure (Hess, 2002). Briefly, the data acquired during the simulation time τ_{run} is divided in a number of n blocks of length τ_b, such that $n * \tau_b = \tau_{run}$, and averages are calculated for each block b. The error for the total average for all the blocks is calculated from the variance between averages of the n blocks $\bar{P}_b$ as follows

$$s_e(\bar{P}_b) = \sqrt{\frac{1}{n(n-1)} \sum_{b=1}^{n} (\bar{P}_b - \bar{P}_{run})^2} \tag{4.1}$$

The standard error $s_e(\bar{P}_b)$ as a function of data block size τ_b, assuming that data blocks are independent, is then fitted to an exponential function. The best estimate of the standard error of the average for correlated data is then given by the limit at large block size of the fitted curve. All error estimates in this chapter were calculated using block averaging unless stated otherwise.

Analysis of trajectories was done either with built-in Gromacs tools or home built VMD scripts (Humphrey et al., 1996).

4.3 Results and discussion

The lipid bilayer was characterized by analysis of complementary information, collected from the description of both the spatial direction of the bilayer normal (electron/mass density profiles and lipid tail order parameters) as well as the lateral direction along the bilayer (area per lipid and lipid lateral diffusion). Also, volume per lipid and water permeation through the bilayer is assessed.

4.3.1 Basic lipid properties

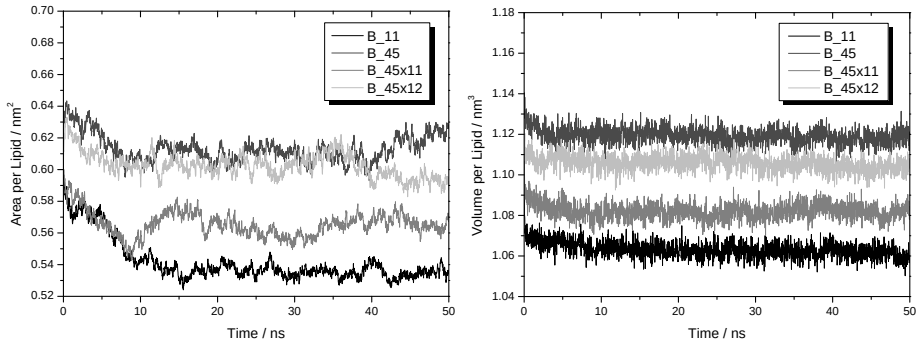
The area per lipid A_L is one of the most important quantities characterizing lipid bilayer systems and one of the most common ways to determine if a simulation of a bilayer system has reached equilibrium. When the A_L reaches a stable value, most other structural properties of the lipid bilayer do not change either. Importantly, A_L can also be compared with experimental values available from X-ray diffraction data (Högberg and Lyubartsev, 2006).

The area per lipid was computed using the following equation

$$A_L = 2\frac{X_{box}\, Y_{box}}{n_L} \tag{4.2}$$

where X_{box} and Y_{box} are the box dimensions on the x-y plane respectively, and n_L is the total number of lipids in the system.

The time evolution of A_L for all simulations is shown in figure 4.3a. After 20 ns, the initial decay of A_L reaches an approximate plateau, allowing an estimate of the average area per lipid $\langle A_L \rangle$ taking into account the last 30 ns of simulation and listed on table 4.3. In the subsequent analysis of all molecular dynamics trajectories, the first 20 ns were discarded unless stated otherwise. The effect of varying the planar carbon van der Waals parameters on the area per lipid is readily seen. Thus, the A_L is acutely sensitive to the change in size of the ester carbon, in agreement with previous results (Chandrasekhar et al., 2003). All variants of the force field parameter set 45A3 result in an A_L within an acceptable range of the experimental value for DMPC of 0.606 nm^2 recently determined by X-ray diffraction studies (Kučerka et al., 2005). Remarkably, bilayer systems B_45 and B_45x12 show an error smaller than 1%.



(a) **(b)**

Figure 4.3: Time evolution of **(a)** the area and **(b)** volume per DMPC molecule for variants of the force field parameter set 45A3.

The volume per lipid V_L is another pivotal quantity, since it correlates the bilayer lateral structure (described by A_L for example) to the bilayer transverse structure (Nagle and Tristram-Nagle, 2000b). V_L was calculated for different trajectories using the following equation

$$V_L = \frac{V_{box} - (\nu_{spc}\, n_{spc})}{n_L} \tag{4.3}$$

where V_{box} is the total volume of the system, ν_{spc} is the water specific volume, and n_{spc} is the total number of water molecules. The water specific volume was calculated from a constant pressure simulation (system SPC; see table 4.2) of a pure water box at the same temperature used for the lipid system. We obtained a value ν_{spc} of 1.02910^{-3} m^3Kg^{-1}, in agreement within 2% of the experimental value (Lide, 1997).

Table 4.3: Average Area and Volume per lipid for variants of the force field parameter set 45A3

System	$\langle A_L \rangle$ / nm^2	$\langle V_L \rangle$ / nm^3
B_11	0.536 ± 0.001	1.0620 ± 0.0003
B_45	0.612 ± 0.003	1.1184 ± 0.0004
B_45x11	0.564 ± 0.002	1.0819 ± 0.0005
B_45x12	0.601 ± 0.003	1.1050 ± 0.0011

A brief inspection of table 4.3 shows that all variants of the force field parameter set 45A3 show also a good agreement of V_L with the experimental value. V_L is within 4% of the published value (1.101 nm^3; Kučerka et al., 2005). Once again, bilayer systems B_45 and B_45x12 show the best agreement with the experimental value, with system B_45x12 having almost a negligible error ($< 1\%$). The remaining results of the chapter aim to better characterize the bilayer system simulated with the variant 45x12 GROMOS 45A3 force field.

The bilayer thickness D_B is also a property related to the bilayer transverse structure and can be defined using a relation like $D_B = 2\frac{V_L}{A_L}$ (Nagle and Tristram-Nagle, 2000b). Using the average results listed on table 4.3 we obtain a value of $D_B = 3.68$ nm for the simulated DMPC trajectory. The result is within 2% of the experimental value (Kučerka et al., 2005). This parameter is of extreme importance for protein/peptide - lipid systems, since the mechanism of interaction between polypeptides and lipids seems to be coupled to D_B.

4.3.2 Density profiles

Electron densities, defined as electronic charge per unit volume, also provide information about the structure of bilayers, along the normal (z-axis) to the bilayer plane. Electron density profiles can be determined from bilayer MD simulations through assignment of the appropriate number of electrons[3] to the atomic nuclei sites, and then binning[4] the transmembrane axis over the entire simulation cell. Due to the inherent fluctuations in fully hydrated fluid phase bilayers during simulations, the trajectory was manipulated through spatial translations in order to center the xy-bilayer plane at $z = 0$ to reduce statistical error, before further analysis.

The electron density is related to the structure factor measurable in X-ray experiments and may serve as one of the basic criteria for validation of molecular dynamics simulations (Benz et al., 2005). Figure 4.4 shows the

[3]The number of electrons per united atom group was assigned to the atomic number of the corresponding united atom group.

[4]At least one z-slice per 0.1 nm of the xy-plane was used.

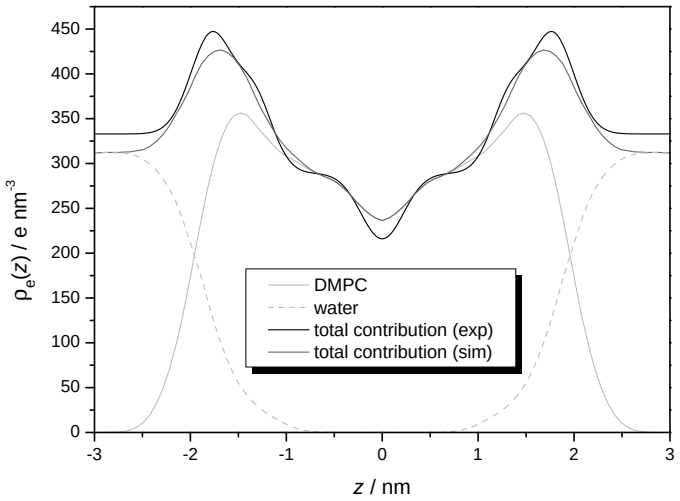


Figure 4.4: Electron density profile of a simulated DMPC bilayer (sim). Contributions from water and DMPC molecules for the total density profile are also shown. The density profile obtained from experimental work (exp) is also depicted (Kučerka et al., 2005).

electron density profile obtained for the DMPC system B_45x12, as well as the individual contributions from lipids and water. The electron density curve show the typical behavior for a lipid bilayer, namely a maximum associated with the electron-dense areas in the headgroups of the phospholipids, and the minimum at the bilayer center due to the lower electron density of methyl groups at the ends of the lipid tails (Nagle and Tristram-Nagle, 2000b). The experimental profile is also shown correlating very well with the DMPC simulation (Kučerka et al., 2005).

From the electron density profile the bilayer head-to-head distance D_{HH}, defined as the distance between the two maxima in the plot, can be estimated. This yields a value of $D_{HH} = 3.4$ nm, which is slightly less than the experimentally observed of 3.5 nm (Kučerka et al., 2005). Notice that if one considers the electron density profile formed by the lipids only, without the water contribution, then the electron density maxima occur at a closer distance to the bilayer midplane than the maxima coming from the total density. In this case, we obtain a $D_{HH} = 3.0$ nm. The difference in maxima positions comes from the strongly varying water density in the headgroup region (Högberg and Lyubartsev, 2006).

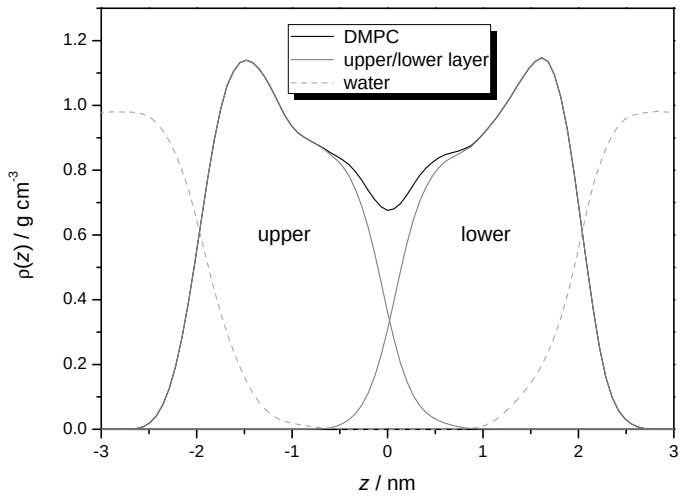


Figure 4.5: Mass density profiles across a DMPC bilayer obtained from the simulation of system B_45x12. The lipid component is divided into the contributions from the two separated leaflets, providing a measure of interdigitation.

Either way, the results from the simulation obtained in this work agree better with experimental results than the results from a recent work (Högberg and Lyubartsev, 2006).

A similar profile can be obtained if instead of considering electron density we consider atomic mass per volume. Mass density profiles across a DMPC bilayer are shown in figure 4.5. The masses of the hydrogen atoms were accounted for in the calculation. Again, we can see that the density of the lipid decreases near the center of the bilayer, a phenomenon known as lipid trough (Nagle and Tristram-Nagle, 2000a). Furthermore, it is easily seen that the tails of DMPC lipids can penetrate up to a value slightly higher than 0.5 nm into the other leaflet. A higher conformational entropy on the terminal methylene groups of the lipid tails explains this trend.

4.3.3 Lipid tail order parameters

Deuterium NMR spectroscopy has provided a wealth of information regarding the organization of membrane components (Petrache et al., 2000). Lipid order parameters are a measure for the orientational mobility of the carbon-

deuterium (C-D) bond and are defined as

$$S_i = \frac{1}{2}\left\langle 3\cos^2(\theta_i) - 1 \right\rangle \tag{4.4}$$

where θ_i is the angle between the *i*th C–D bond vector and the bilayer normal. The angular brackets denote a time and ensemble average. The order parameters of the acyl lipid chains are often used to validate or calibrate molecular simulations (Vermeer et al., 2007). The order parameters S_{CD} for the DMPC acyl chains were calculated from the DMPC simulation of system B_45x12.

Since our simulations employ a united-atom force field, no explicit information about the acyl methylene hydrogen positions is available and they must be reconstructed. It is common practice to assume that C_i–H vectors are perpendicular to the vector connecting C_{i-1}–C_{i+1} and that these carbons and methylene hydrogens attached to C_i adopt a tetrahedral conformation. In practical terms, this means that for the outermost carbon atoms of the chain, no order parameter can be measured. A more detailed explanation about S_{CD} calculation can be found in Douliez et al. (1998) and Patra et al. (2006).

The results for the order parameters are shown in figure 4.6 and provide a qualitative measure of the degree of order along the lipid acyl chains. The agreement between experiment (filled symbols) and simulation (open symbols) is best at the end of the acyl chains and acceptable for most of other carbon positions in the middle of it. Furthermore, the order parameters presented in this chapter are once more in better agreement with experimental data than recent simulation studies on pure DMPC systems at the same temperature (Högberg and Lyubartsev, 2006; Vermeer et al., 2007). The experimental S_{CD} profiles of the acyl chains show a broad plateau for the half of methylene groups closer to the glycerol and headgroup region, followed by a reduction in ordering toward the central region of the bilayer. This apparent plateau in the experimental data is due to the unresolved splittings of ^{2}H resonances near the glycerol. Obviously, such degeneracy is not observed on the molecular simulation. The profiles also show that on average, the S_{CD} from the sn-2 chain is higher when compared with the sn-1 chain. The sn-2 is tethered via the glycerol backbone closer to the aqueous interface, and thus it is more stretched or ordered relatively to sn-1 chain, which starts deeper in the bilayer (personal communication by Michael

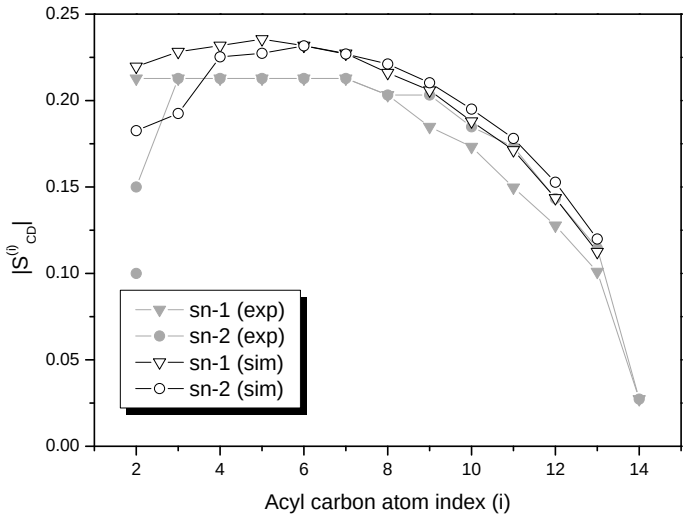


Figure 4.6: Deuterium order parameter $S_{CD}^{(i)}$ profiles as a function of carbon number along the sn-1 and sn-2 acyl chains for DMPC. Experimental (exp) order parameters were derived from the quadrupolar splittings of de-Packed ^{2}H NMR spectra (Petrache et al., 2000). The double resonances for the C_2 carbon of the sn-2 chain are also shown. Theoretical values were derived from our molecular dynamics simulation, using variant 45x12 GROMOS 45A3 force field, at 303 K.

Brown).

4.3.4 Lateral lipid dynamics

To gain information about the diffusion of the lipids in the simulated bilayer, the diffusion coefficient was calculated from the Einstein relation (cf. with equation 2.3)

$$D_{xy} = \lim_{t\to\infty} \frac{1}{2n} \frac{\mathrm{d}}{\mathrm{d}t} \left\langle |\Delta r_{xy}(t)|^2 \right\rangle \tag{4.5}$$

where $\left\langle |\Delta r_{xy}(t)|^2 \right\rangle = \left\langle |r_{xy}(t_0+t) - r_{xy}(t_0)|^2 \right\rangle$ is the mean square displacement (MSD) in the xy-plane during time t starting from initial time t_0, n is the dimensionality of the system, and D_{xy} is the lipid lateral diffusion coefficient. The angular brackets denote an ensemble and time average over all initial time moments t_0 for the given time t. The MSD was calculated for the center of mass (COM) of each DMPC molecule and averaged over time and over all the lipid

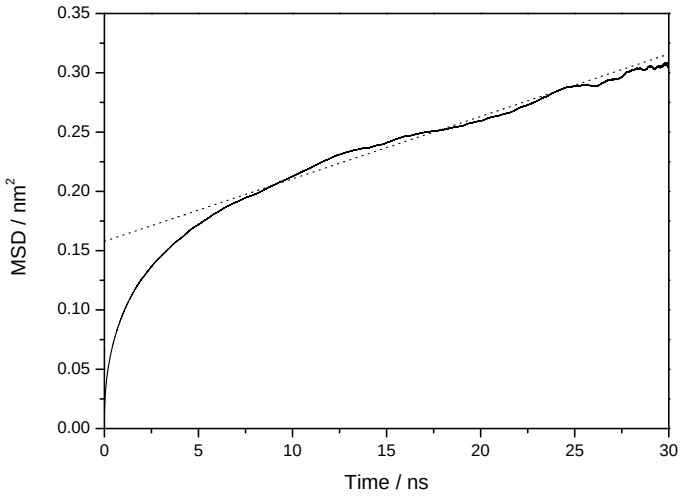


Figure 4.7: DMPC mean square displacement (MSD) profile. The simulations was performed at 303 K using the variant 45x12 GROMOS 45A3 force field. The diffusion coefficient was calculated after fitting the MSD profile equation 4.5, and discarding the first 5 ns of diffusion free-flight.

molecules. r represents the COM positions. To improve the statistics, the time origin t_0 was shifted every 10 ps.

Although the center of mass of the whole system is reset after each time step, both lipid layers can acquire some drift velocity and develop an opposite COM motion while the total COM motion for the system is still zero (Anézo et al., 2003). The random relative motions of the two layers give rise to an apparent supra-diffusivity that is purely artificial and needs to be removed (Lindahl and Edholm, 2001). The COM of the upper and lower monolayers were separately removed in all simulations performed, thus not needing further processing.

As can be seen from figure 4.7, the mean square displacement profile can be fairly well fitted to a straight line for $t > 5$ ns yielding a diffusion coefficient of 5.3×10^{-8} cm^{2}s^{-1}. The obtained value is acceptable when compared with the experimental result of 6×10^{-8} cm^{2}s^{-1} obtained by fluorescence recovery after photo-bleaching (Almeida et al., 1992) but somewhat below the experimental NMR result of 9×10^{-8} cm^{2}s^{-1} at the same temperature (Orädd et al., 2002).

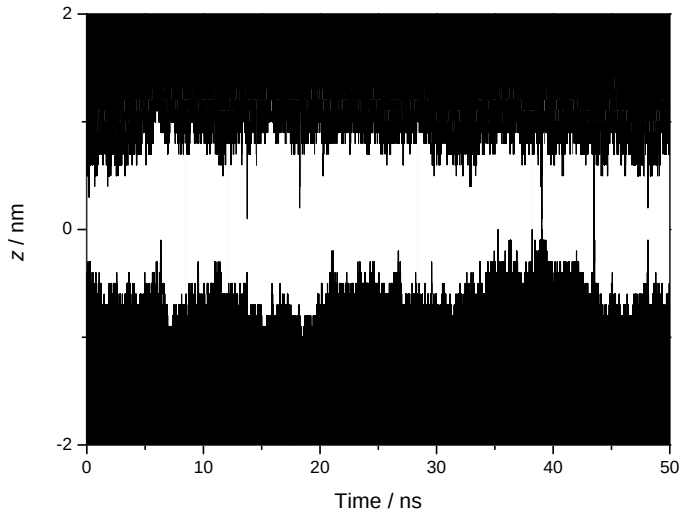


Figure 4.8: The z-coordinates of all water molecules (in black) are plotted as a function of time for system B_45x12. Water is able to penetrate into the DMPC bilayer (mid-plane located at $z = 0$).

.

4.3.5 Bilayer water permeation

We also analyzed the penetration of water molecules into the bilayer. A qualitative overview can be obtained by plotting the z component of the positions of the water molecules on each frame along the trajectory, after centering the bilayer at $z = 0$. The result is shown in figure 4.8. It is easily seen that the density of water molecules is reduced in the center of the bilayer, nonetheless from time to time water molecules are able to penetrate deeper into the bilayer.

Crossing events cannot be inferred directly from figure 4.8, since penetration of water into the bilayer is not *per se* an indication that is going to cross it. To have an estimate on the number of crossing events in the DMPC bilayer, we have first to characterize the solvent molecules regarding their spatial position. For each trajectory frame, we can assess the position of the water molecules in relation to the bilayer molecules. Moreover, we can categorize molecules by their vicinity to the upper leaflet or bottom leaflet. Since it is of little interest if a water molecule is just scratching at the surface of the bilayer, it is important to know if the water molecule reaches the part of the leaflet where it can form a long lived hydrogen bond. This simple ideas give a functional definition to

follow the bilayer crossing of the solvent. In practice, we calculated per each frame which water molecules are near to the glycerol-moieties of the lipids, thus more prone to cross the bilayer, and their z positions were assessed along the trajectory.

Though we did not quantitatively calculated the lifetime of the hydrogen-bonds between water molecules and DMPC lipids, we observed by visual inspection that once a water molecule penetrates deeper in the bilayer, it can stay there for tens to hundreds of ps. On average, 24 water molecules are found between the middle of the bilayer and the average plane formed by the glycerol oxygens. Also, we determined that only 3 water molecules were able to cross the bilayer during the 30 ns of trajectory analyzed.

The initial goal for studying these crossing events was mainly to assess bilayer stability along the simulation. On the other hand, from the literature it is clear that water and solute permeability across lipid membranes are also a major subject of study, nonetheless the mechanisms by which permeants cross bilayers remain unclear. Mathai et al. (2008) proposed a new model for permeability, establishing that the area occupied by the lipids is the major determinant and the hydrocarbon thickness is a secondary determinant. Molecular dynamics simulations provide a wealth of information not accessible from experimental work. Thus, a more detailed study on the crossing events using MD on different bilayers and strategies similar to the one devised above, should give a more comprehensive knowledge about these mechanisms.

4.4 Conclusions

The goal of this chapter is two-fold. Though protein simulations are already becoming a standard protocol to provide further atomistic insight into experimental work, biomembrane simulations are still a new field. This can be understood by the diverse and central roles proteins play in biological systems, and thus more scientific effort was put on a correct parameterization of proteins and peptides. As such, our first goal was to tie together these two classes of biomolecules in a force field able to treat them in a more realistic way than most of the force fields available at the time this work was done. As an immediate

consequence, it will allow us to model more complex bio-systems such as the interaction of proteins with bilayers (see chapter 5). The second goal was to build a simple bilayer system often used by experimentalists, like the DMPC bilayer. This will allow us to compute properties describing such bilayer system, and compare them in a systematic way with experimentally determined properties.

Research on membranes and protein/membrane interactions has been growing steadily. Lipid membranes in the liquid crystalline phase, also known as the L_α phase, are generally considered to be relevant for models of biological membranes. Lipids can influence the structure and function of membrane proteins (intrinsic or not) and in turn membrane proteins have influence on the lipid dynamics (Lee, 2005). Therefore, it is important to be able to characterize the structure and dynamics of a lipid bilayer.

Simulations provide information on an atomic level, at a resolution that still cannot be obtained experimentally. However, putting together parameters into a force field that makes sense is a daunting task. Simulations can then suffer both from uncertainties in the force field, though uncertainty is also found in experimental data (Nagle and Tristram-Nagle, 2000b), and relatively short simulation times due to the computational expensive task of calculating interactions between the Newtonian bodies. As new force fields arise and validation against experimental data shows promising results, new possibilities emerge for computational scientists.

Most of the work published in the last decade on membrane simulations using the GROMOS force field is based on the work of Berger et al. (1997). Though this force field is still valid for lipid systems, it is not the best choice to predict interactions between lipids and proteins. Newer force fields have been published in the last years, which seem much more promising to simulate complex bio-systems according to some authors (Schuler et al., 2001). In this work, we used variants of the GROMOS96 45A3 parameter set to study a DMPC system in the L_α phase, using a set of MD parameters consistent with the ones used in protein simulations.

Analyzing both the area per lipid and volume per lipid on the obtained trajectories, we could assess the best set of force field parameters. System B_45x12

showed the best agreement of A_L and V_L properties with available experimental data. The only difference among the variants of GROMOS96 45A3 employed in this study, is the description of the phospholipid ester carbon. Several results do point out that a careful choice of parameters to run MD simulations should always be followed by a critical assessment of the results (Norberg and Nilsson, 2003).

Not only the parameters describing the molecules are an important issue, but also important are the parameters describing the MD methodology. As an example, nowadays most authors agree upon the use of PME to treat the electrostatics because the properties of bilayers (namely A_L and V_L) are very sensitive to the electrostatic cut-off distance (Anézo et al., 2003). As already mentioned, we used methodological parameters throughout this work consistent with both lipid and protein simulations and also in agreement with recent literature (Vermeer et al., 2007). A much more detailed comparison between different simulations conditions is underway, taking into consideration for example into consideration different electrostatic treatments (PME, cut-off and reaction field), number of charged groups on the topological description of the lipid, time step employed to solve the Newton equations, among others.

Regarding the group of properties evaluated, results show good agreement between the structural and dynamic properties of lipids and the available experimental data. Moreover, our results seem to be in better agreement with experimental published data (at least for the properties computed) than that of other simulations on a similar system (Högberg and Lyubartsev, 2006).

Though only one trajectory simulation per set of parameters studied was analyzed in this chapter, in fact we did several simulations for each set. The comparison between different simulations is not a trivial task (Wassenaar, 2006), nevertheless, the average values of the computed properties are in agreement between the simulation *replicas*.

It was not our goal to make an extensive study of the DMPC system, rather to build and equilibrate a bilayer such that it could be used in forthcoming studies. This brief incursion into bilayer systems showed us that these multimeric systems are not at all simple to simulate. The complexity of cellular membranes and the dynamics of its components poses a difficult task, but rewarding

advances are being made.

Acknowledgments

We acknowledge the computer resources provided by the Advanced Computing Laboratory, Milipeia Cluster, Department of Physics, University of Coimbra, and the support of Fundação para a Ciência e Tecnologia through grant SFRH/BD/1354/2000 (to N.L.-F.). The author wishes also to warmly acknowledge Michael Brown and Horia Petrache for their remarks in lipid order parameters, John Nagle for sharing electron density profile values, Michel Patra and Alexander Lyubartsev for the discussions about calculation of density profiles, and Cláudio Soares who introduced me to the intricacies of the GROMOS96 force-field.

Chapter 5

Plasticity and aggregation of δ-toxin

"Certainly no subject or field is making more progress on so many fronts at the present moment, than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that *all things are made of atoms*, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

Feynman et al. (1963)

5.1 Introduction

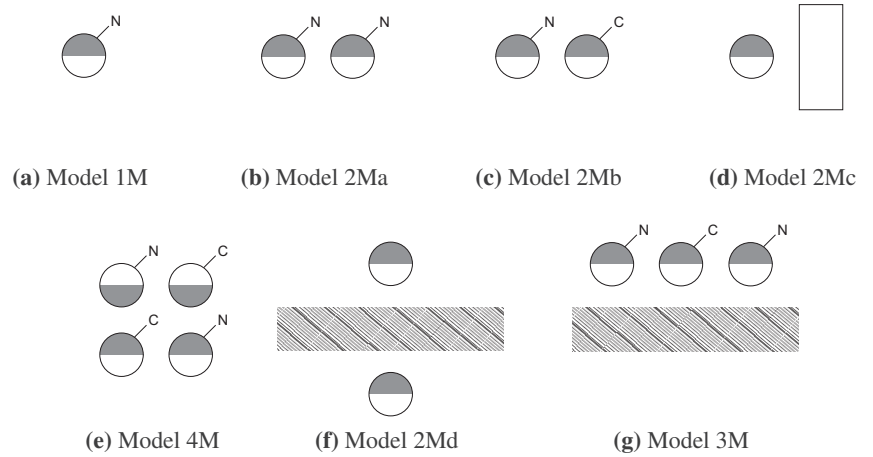
Staphylococcus aureus, a Gram-positive bacterium, produces a number of extra cellular toxins with diverse biological activities. One of such toxins is δ-toxin, a 26 amino acid haemolytic peptide (Fitton et al., 1980). The peptide has an unusually high content of hydrophobic amino acids, four basic and three acidic residues, a negative C-terminus, and a formylated N-terminus, giving it a zero net charge at neutral pH. The polar residues are regularly spaced between hydrophobic patches, forming an amphipatic helix (Fitton, 1981).

δ-toxin lyses eukaryotic cells in preference to bacterial cells in vivo (Dhople and Nagaraj, 1993; Kreger et al., 1971). Specificity of δ-toxin and other cytolytic peptides towards certain cell types is still under investigation. Yet, it seems certain that no cell-surface receptors are involved in the recognition mechanism. Synthetic peptides with all-D-amino acid enantiomers show the same activity as the natural, all-L-amino acid peptides (Wade et al., 1990).

The primary sequence of cytolytic peptides from different organisms vary widely, but many are able to form amphipathic α-helices. The distribution of charged amino acids along the helix axis varies among these cytotoxic peptides, but most frequently they carry a positive net charge at neutral pH. Nevertheless, it was shown that a net charge on these short peptides cannot be the sole cause for their target cell specificity, and other factors may have to be considered, such as the distribution of charge along the peptide axis (Dathe and Wieprecht, 1999).

Due to its charge characteristics, δ-toxin has a high propensity to aggregate in aqueous solution. Furthermore, nuclear magnetic resonance (NMR) data showed that the conformation of δ-toxin is highly solvent dependent. In water the peptide aggregates, probably forming large oligomers, but in methanol these oligomers are small enough to be studied by NMR (Tappin et al., 1988). In methanolic solution, the peptide forms a rather well ordered amphiphatic helical structure between residues 2 and 24 (see chapter 3).

On its way to the membrane, the peptide must experience subtle chemical environment changes with consequences on peptide conformation. The study of the conformational plasticity and aggregation behavior of the peptide should give some clues for the understanding of the first stages of the lytic mechanism.



(a) Model 1M **(b)** Model 2Ma **(c)** Model 2Mb **(d)** Model 2Mc

(e) Model 4M **(f)** Model 2Md **(g)** Model 3M

Figure 5.1: Schematic representation of simulated model systems with the amphipathic δ-toxin in bulk solvents and in a bilayer environment. **(a)** A monomeric peptide, **(b)** an head-to-head peptide dimer, **(c)** a head-to-tail dimer, **(d)** a dimer in a perpendicular arrangement, and **(e)** an anti-parallel tetramer, were simulated in neat solvents, namely water, methanol, and DMSO. **(f)** A DMPC bilayer with one peptide per leaflet, and **(g)** a peptide trimer in anti-parallel arrangement were also simulated. Circles represent a top-view of δ-toxin in a α-helix conformation; the hydrophobic moiety of the peptide is shown as a shaded area; peptide N and C terminus are also depicted. The bilayer is represented by the shaded rectangle in panels (f) and (g).

In order to assess the conformational stability and plasticity of δ-toxin, we performed molecular dynamics (MD) simulations, at room temperature, in different membrane-mimetic solvents. A more realistic system composed by a bilayer of zwitterionic dimyristoylphosphatidylcholine (DMPC) in the liquid-crystalline L_α phase in water was also simulated, in the presence of one to three δ-toxin peptides per leaflet.

5.2 Methods

5.2.1 Simulated systems

To study the propensity of δ-toxin to aggregate in solution, and the adsorption of the peptide to a lipid bilayer, molecular dynamics simulations of a number of peptides in different environments have been performed. Several

Table 5.1: Overview of the molecular dynamics simulations performed

Label*	Solvent	Peptide molecules	Solvent molecules	Number of atoms
W_1M	water	1	7806	23686
W_2Ma	water	2	9533	29135
W_2Mb	water	2	9533	29135
W_2Mc	water	2	9533	29135
W_4M	water	4	11571	35785
M_1M	methanol	1	3503	10777
M_2Ma	methanol	2	4186	13094
M_2Mb	methanol	2	4186	13094
M_2Mc	methanol	2	4186	13094
M_4M	methanol	4	5172	16588
D_1M	DMSO	1	1763	7320
D_2Ma	DMSO	2	2104	8952
D_2Mb	DMSO	2	2104	8952
D_2Mc	DMSO	2	2104	8952
D_4M	DMSO	4	2549	11268
B_2Md	water + DMPC	2	8686	32482
B_3M	water + DMPC	3	6384	25844

* Labeling of simulations throughout this chapter was based in the following rules: the first capital letter denotes the solvent (W for water, M for methanol, D for DMSO, and B for a bilayer in water phase); the following character strings denote the model system depicted in figure 5.1.

molecular systems were built as shown in figure 5.1 and listed in table 5.1.

A canonical right handed α-helical δ-toxin was constructed using InsightII (Biosym/MSI, 1995), and used as a starting structure for all relevant simulations. The starting structure for a fully hydrated DMPC lipid bilayer was produced in chapter 4.

Several models of δ-toxin were centered in a rhombic dodecahedron box unit cell, and solvated with three solvents of different polarity, namely water, methanol and dimethyl sulfoxide (DMSO). Briefly, the number of peptides in the system ranged from one to four peptides (figures 5.1a–e). The peptide multimeric models are arranged such that no van der Waals clashes exist, and peptides can be fully solvated (approximately 1.7 nm separate peptide centers). Moreover, the hydrophilic/hydrophobic interface between adjacent peptides was oriented in a way such that no preferential contacts exist, with the exception of model 4M (figure 5.1e) where the hydrophobic patches of the 4 peptides are facing each other and away from the bulk solvent. The cell dimensions were chosen to allow for at least 1.2 nm between any peptide atom and the nearest

box edge. Thus, at least 2.4 nm separate the simulated model to its periodic images. The final concentration of the peptides in solution is high (mM range) in order to facilitate the aggregation of the monomers.

We also simulated the interaction between δ-toxin and a well-characterized DMPC lipid bilayer in the liquid-crystalline L_α phase. The systems simulated contained one, or three δ-toxin peptides, 128 DMPC lipids, and water molecules (figures 5.1f–g). Because periodic boundary conditions were used in all simulations, we oriented the peptides with its long axis roughly parallel to the membrane surface (xy-plane), such that the system is kept at a reasonable size. Initially, the α-helical peptides are placed in the water phase close to one of the monolayers of an equilibrated bilayer, approximately 1.7 nm separating the peptide centers of mass (COM) and the average position of lipid headgroup nitrogen atoms on the xy-plane. No van der Waals clashes exist between neighbor peptides and lipids, allowing full hydration of each biomolecule. In model 2Md (figure 5.1f) one peptide per monolayer was placed either with the hydrophobic or the hydrophilic moieties facing the lipids. In model 3M (figure 5.1g) the asymetric distribution of the peptides mimics the addition of peptide to a solution containing cells or liposomes, in which peptides adsorb to one monolayer only.

5.2.2 Simulation details

All MD simulations were performed with the Gromacs package version 3.3 (Lindahl et al., 2001). The GROMOS96 45A3 united atom force field (Schuler et al., 2001), variant 45x12 (see chapter 4), was used to describe peptide and lipids. DMSO was described by Geerke et al. (2004), and methanol was described by the B3 three-site model (Walser et al., 2000). Water was modeled explicitly using the simple point charge (SPC) model (Berendsen et al., 1981). The isothermal compressibility was set to $12.06x10^{-5}$, $4.5x10^{-5}$, and $5.25x10^{-5}$ bar^{-1}, for methanol, water, and DMSO, respectively (Easteal and Woolf, 1985; Geerke et al., 2004). The protonation state of ionizable groups was chosen to correspond to pH 7.0. No counter ions were added since all systems are already electro neutral. The time step used for the integration of the equations of motion was 2 fs. Non bonded interactions were evaluated using a

twin range cutoff of 1.0 and 1.4 nm. Interactions within the shorter range cutoff were evaluated at every step whereas interactions within the longer range cutoff were evaluated every 10 steps. The long-range electrostatic forces were treated by the particle-mesh Ewald (PME) summation algorithm (Darden et al., 1993). All bond lengths were constrained to their equilibrium values by the LINCS algorithm (Hess et al., 1997), whereas water geometry was maintained with the SETTLE routine (Miyamoto and Kollman, 1992). Simulations were carried out at a constant pressure of 1 bar and at constant temperature, 298.15 K for peptide/solvent simulations and 303 K for peptide/lipid simulations. A weak pressure-coupling scheme was adopted using the Berendsen barostat (Berendsen et al., 1984), with a coupling time of 1 ps. Temperature was controlled using the Berendsen thermostat (Berendsen et al., 1984), with a coupling time of 0.1 ps.

For the aggregation simulations of δ-toxin in the three different solvents, peptide and solvent molecules were independently coupled to the heat bath. Anisotropic pressure coupling was applied such that the three dimensions of the box (x,y,z) are allowed to vary independently ($Np_xp_yp_zT$ ensemble).

For the peptide/bilayer simulations, lipids, peptide and water molecules were separately coupled to the heat bath. A fine control of the temperature is necessary to keep the DMPC lipid bilayer in the fluid phase. The pressure was controlled semi-isotropically ($Np_{xy}p_zT$ ensemble), where the height of the box (z direction, i.e., along the lipid hydrocarbon chains) and the cross-sectional area (xy plane) were allowed to vary independently of each other.

Starting velocities were randomly assigned from a Maxwellian distribution with different random seeds for each of the simulations. For all simulations energy minimization in vacuum was performed to remove bad contacts within the protein. After solvation, energy minimization was performed again to remove bad contacts between solvent and protein/lipid atoms. Prior to production runs, a short equilibration was performed with position restraints on all heavy atoms. Production runs of 50 ns for solvent/peptide systems and 150 to 250 ns for peptide/lipid systems were made. Coordinates of the full system were saved every 1 ps for further analysis.

Display and analysis of trajectories was done either with built-in Gromacs

tools or home built VMD scripts (Humphrey et al., 1996). The simulations took $\sim$ 37000 CPU hours and were run on two different clusters (Sun Fire X4100 computing nodes; dual Intel Xeon computing nodes) as well as in commodity computers.

5.2.3 Data analysis

The first 20 ns of each simulation were discarded prior final analysis. All error estimates were calculated using block averaging (see chapter 4) unless stated otherwise. Thermodynamic parameters characterizing the MD simulations, and distance between periodic images, were assessed from Gromacs output files. Comparison of peptide structures was done by measuring root mean square distances (RMSD). The secondary structure content of the peptides was determined using the DSSP criteria (Kabsch and Sander, 1983). The aggregation of peptides, both in solution and on the bilayer interface, was monitored by clustering and visual inspection. Penetration of peptides on the bilayer was monitored by measuring the distance to the bilayer midplane.

A more detailed description of the applied methodologies will be introduced whenever necessary along the next sections. Time evolution of properties are depicted through out this chapter with a time interval of 10 ps.

5.3 Results and discussion

We start by making some considerations about quality assurance of the MD simulations, before proceeding to data analysis. The results from MD trajectories of δ-toxin in the different chemical environments will be discussed after. Emphasis on the secondary structure of the peptide and aggregation state, both in solution and in the lipid bilayer interface, will be given. Whenever possible, the results presented in this chapter will be compared with experimental data.

5.3.1 Convergence of molecular dynamics simulations

After a MD simulation has been performed, a quality assessment of the simulation should be performed. This quality assurance step involves the study for convergence of thermodynamic parameters, such as system temperature and

energy, towards equilibrium. We are particularly interested on studying the behavior of a peptide in solution. And so, relaxation of the protein structure should also be inspected, usually in terms of the Cartesian distance from a reference structure. Peptides have an intrinsically high degree of conformational freedom, and thus a great care should be taken when simulating systems with periodic boundary conditions. Solute molecules such as peptides must not interact between adjacent periodic images, otherwise such interactions could lead to unphysical effects. In short, prior to data analysis of molecular dynamics simulations and final interpretation of the results, the overall MD system evolution should be assessed in order to check if a set of properties describing the system are in equilibrium. Nevertheless, it should be mentioned that convergence for a given property is not always attained due to a high relaxation time when compared with the simulated time.

Thermodynamic parameters

Figure 5.2 shows, as an example of the study of convergence, the evolution of several properties during a molecular dynamics run on a system composed of four δ-toxin monomers in water (simulation W_4M). As can be readily seen from figures 5.2a–b, thermodynamic parameters such as potential and kinetic energy, and temperature, are already in equilibrium. In fact, as explained in the methods section, a minimization/equilibration stage on every simulation is performed before the production run, in order to prepare the system to evolve from a system in thermal equilibrium. The density of the W_4M system is very close to the experimental value (1000 Kgm^{-3}), further reinforcing the equilibrium of the system and the quality of the data. A summary of the average values for some thermodynamic properties and other physical properties describing the simulation box for all simulations performed, is presented on the appendix table A.6. Both thermodynamic and physical properties reach equilibrium values almost at the beginning of the simulation. Average system density values are in accordance with experimental data.

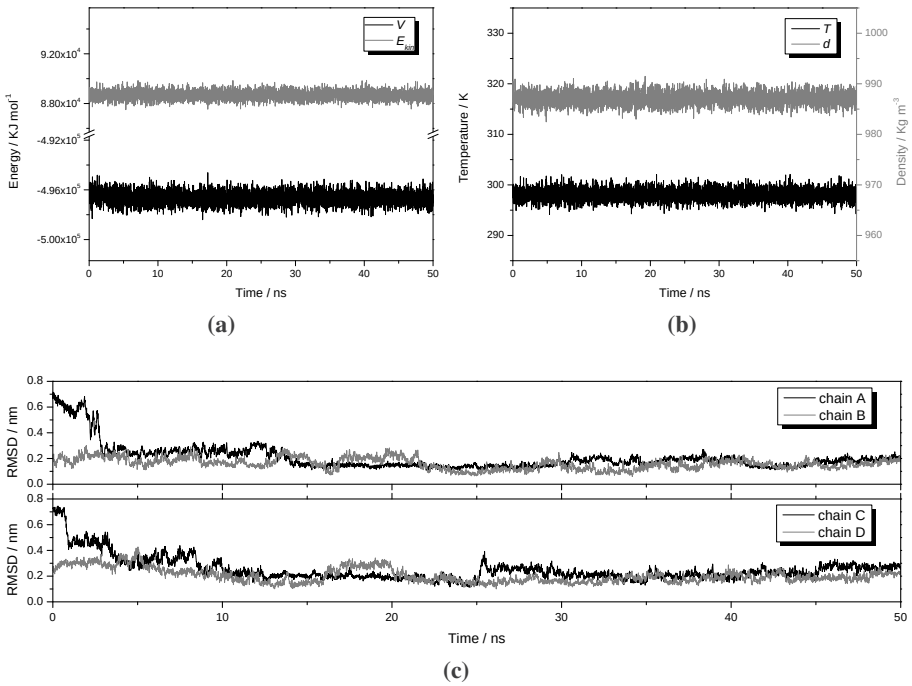


Figure 5.2: Time evolution of properties characterizing the δ-toxin W_4M tetramer model in water. **(a)** Potential energy V, kinetic energy E_{kin}, **(b)** temperature T, and density d variations as a function of simulation time. **(c)** Backbone RMSD from the trajectory average structure, for each peptide chain.

Root mean square distance

A widely used way to compare the structures of biomolecules is to translate and rotate one structure with respect to the other. The most common measure of the fit between the two structures is the RMSD between pairs of atoms, defined as (Leach, 1998)

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \tag{5.1}$$

where n is the number of atoms over which the RMSD is measured and d_i is the distance between the coordinates of atom i in the two structures, after a body superposition.

A molecular dynamics simulation produces a range of structure snapshots.

Convergence of the structure towards an equilibrium state can be computed by calculating the RMSD for each snapshot from a reference structure, usually the starting structure of the production simulation run. For a protein simulation, the RMSD value from the starting structure will eventually level off, indicating that the structure has reached a certain distance from the reference structure and then keeps that distance more or less constant throughout the rest of the simulation. Nonetheless, this does not guarantee that the structure as reached an equilibrium state, since the number of conformations available also increase as the simulation proceeds. A better way to check structure convergence is to check the convergence towards an average structure.

Figure 5.2c shows the backbone RMSD from the average structure for each δ-toxin chain in the W_4M system. While chains B and D have a RMSD value which fluctuates throughout the simulation around a distance value of 0.2 nm from the respective chain average structure, chains A and C do show a different behavior with the RMSD values leveling off at around 15 ns. Thus, though the thermodynamic parameters discussed above are already in equilibrium, the structure of the peptides in solution take much longer time to equilibrate. This is especially true if the starting conformation of the peptide unfolds during the simulation. In this case longer simulation times are needed for the peptide to probe all the conformational space available and reach an equilibrium structure.

Regarding all the other simulations of δ-toxin in neat solvents (water, methanol, and DMSO) as well as in the bilayer system, RMSD from the average structure for each peptide chain shows plateau regions starting at different times on the simulation run, and of varying time length (data not shown). A summary of backbone RMSD values from the respective average structures is given in table 5.2. A low average RMSD value for a given simulation is indication that structural relaxation was attained, whereas a high average RMSD value is indicative of a higher propensity for structural plasticity. A brief analysis of these RMSD values indicates that the δ-toxin structure is better defined in both water and methanol solutions, when compared with DMSO systems. δ-toxin in bilayer systems also show the convergence for an average structure. Based on the obtained RMSD profiles for simulations, all the statistics presented hereafter were made by discarding the first 20 ns of peptide simulations in neat solvents,

whereas half of the trajectory was discarded when analyzing bilayer simulations, unless stated otherwise.

Interactions between periodic images

Molecular dynamics simulations aim to provide information about the properties of a macroscopic sample. Yet, the size of the system is limited by the available storage space on the host computer, and more importantly, by the execution speed of the MD algorithms (Allen and Tildesley, 1996). Thus, a MD system has a finite number of N-particles, usually confined in a semi-regular space-filling polyhedra, such as a rhombic dodecahedron or cubic box. To simulate bulk phases, system boundary conditions should be treated in such a way that mimic the presence of an infinite bulk surrounding the N-particle system. This is usually achieved by employing periodic boundary conditions, in which the motion of the N-particles in the bounding box is reproduced in each of the neighboring boxes. This will create an infinite periodic boundless lattice system. On such a system, we must assure that there is no direct interactions between periodic images of a particle, otherwise artifacts will occur due to the unphysical topology of the simulation. As an example, Weber et al. (2000) showed that for a polyalanine peptide, the α-helical conformation is stabilized by artificial periodicity relative to any other configuration sampled during simulations. The authors demonstrated that this artificial stabilization is larger for smaller unit cells, being responsible for the absence of peptide unfolding.

Another important factor to be taken into account is the box type used during the simulation. The choice of box type can have a statistically significant effect on the outcome of a simulation, acting as a constraint on the dynamics behavior of the solute, restricting the conformational ensemble of the system (Wassenaar, 2006). It was shown that the optimal box, without introducing major artifacts, should have an approximate spherical symmetry such as the rhombic dodecahedron. This box minimizes effects related to the box shape or the resulting distribution of solvent. Nevertheless, Wassenaar also points out that the nature and magnitude of such effects are strongly dependent on the protein studied. δ-toxin in neat solvents was simulated using rhombic dodecahedron boxes, whereas a cubic box was used for simulations of bilayer

Table 5.2: Summary of the RMSD from the average structure, secondary structure, and aggregation state for δ-toxin simulations in different solvents.

			Secondary structure[¶]			Aggregation state[‖]			
label[*]	chain[†]	RMSD[‡]	structure[§]	coil	α-helix	1M	2M	3M	4M
W_1M	A	0.35 ± 0.04	12.8 ± 0.9	8.7 ± 0.5	9.6 ± 0.9	100			
W_2Ma	A	0.23 ± 0.01	14.7 ± 0.4	8.1 ± 0.2	12.2 ± 1.0	0	100		
	B	0.34 ± 0.07	9.1 ± 0.4	12.3 ± 0.8	5.4 ± 0.2				
W_2Mb	A	0.28 ± 0.04	14.7 ± 1.1	7.9 ± 1.0	9.8 ± 1.5	0	100		
	B	0.19 ± 0.01	18.6 ± 0.8	5.5 ± 0.1	15.7 ± 1.0				
W_2Mc	A	0.26 ± 0.06	20.9 ± 1.0	3.8 ± 0.4	17.5 ± 1.6	0	100		
	B	0.27 ± 0.03	15.2 ± 0.3	7.7 ± 0.3	12.9 ± 0.2				
W_4M	A	0.17 ± 0.01	13.5 ± 0.8	8.6 ± 0.2	11.1 ± 3.2	0	0	0	100
	B	0.14 ± 0.01	21.5 ± 0.3	4.2 ± 0.3	18.7 ± 0.5				
	C	0.22 ± 0.01	12.8 ± 0.4	10.0 ± 0.4	10.4 ± 0.7				
	D	0.18 ± 0.01	18.8 ± 0.1	6.0 ± 0.1	15.2 ± 0.4				
M_1M	A	0.26 ± 0.06	17.4 ± 0.4	6.0 ± 0.2	15.6 ± 0.5	100			
M_2Ma	A	0.44 ± 0.03	13.7 ± 0.9	8.8 ± 0.4	12.1 ± 0.9	52.9	47.1		
	B	0.38 ± 0.05	14.8 ± 1.5	9.1 ± 2.0	14.0 ± 1.3				
M_2Mb	A	0.34 ± 0.02	13.5 ± 1.0	9.6 ± 0.5	12.5 ± 1.0	21.4	78.6		
	B	0.16 ± 0.01	20.9 ± 0.4	4.3 ± 0.2	20.0 ± 0.6				
M_2Mc	A	0.31 ± 0.05	15.5 ± 1.2	6.7 ± 0.6	12.8 ± 0.9	63.6	36.4		
	B	0.43 ± 0.11	13.9 ± 1.8	8.5 ± 1.0	12.1 ± 2.5				
M_4M	A	0.28 ± 0.07	19.8 ± 1.3	5.2 ± 2.9	18.1 ± 0.8	20.9	16.2	34.0	28.9
	B	0.31 ± 0.06	17.9 ± 0.5	6.5 ± 0.5	16.3 ± 0.5				
	C	0.30 ± 0.09	16.8 ± 7.1	6.6 ± 1.9	15.2 ± 6.2				
	D	0.37 ± 0.01	13.8 ± 0.8	10.3 ± 0.4	12.1 ± 1.0				
D_1M	A	0.49 ± 0.06	0.7 ± 0.2	16.1 ± 0.5	0.0 ± 0.0	100			
D_2Ma	A	0.62 ± 0.01	0.2 ± 0.1	16.6 ± 0.9	0.0 ± 0.0	83.5	16.5		
	B	0.57 ± 0.04	0.3 ± 0.1	19.3 ± 0.4	0.1 ± 0.1				
D_2Mb	A	0.40 ± 0.05	11.9 ± 1.1	10.4 ± 0.2	9.8 ± 0.8	92.5	7.5		
	B	0.42 ± 0.09	8.7 ± 1.6	10.5 ± 1.2	7.0 ± 2.6				
D_2Mc	A	0.47 ± 0.09	5.9 ± *nan*	13.9 ± 3.9	4.6 ± *nan*	92.0	8.0		
	B	0.48 ± *nan*	7.3 ± *nan*	11.2 ± *nan*	5.1 ± 5.9				
D_4M	A	0.49 ± 0.06	3.3 ± 2.5	14.1 ± 1.8	1.9 ± 1.4	59.0	26.7	14.0	0.3
	B	0.49 ± 0.05	4.5 ± 2.1	14.2 ± 0.9	3.0 ± 1.8				
	C	0.69 ± 0.09	11.7 ± 0.3	16.6 ± 0.4	1.2 ± 0.6				
	D	0.53 ± 0.22	5.3 ± *nan*	13.1 ± *nan*	4.1 ± 9.8				
B_2Md	A	0.18 ± 0.01	12.7 ± 0.2	12.8 ± 0.1	6.4 ± 0.9	100			
	B	0.19 ± 0.02	11.3 ± 0.2	10.6 ± 0.4	8.2 ± 0.3				
B_3M	A	0.19 ± 0.01	11.1 ± 0.1	10.8 ± 0.2	10.4 ± 0.1	0	0	100	
	B	0.21 ± 0.01	13.2 ± 0.3	9.6 ± 0.3	12.4 ± 0.2				
	C	0.14 ± 0.03	10.6 ± 0.7	9.9 ± 0.2	7.8 ± 2.4				

Analysis was performed in the time window 20 to 50 ns on neat solvent simulations whereas on bilayer systems we discarded the first half of trajectory. When an invalid fit is attained during block averaging error estimation, a *nan* symbol is reported. [*] Identification of molecular dynamics simulation. [†] Peptide chain identification. [‡] Root mean square deviation (RMSD) of backbone atoms from average structure. [¶] Secondary structure analysis using the DSSP algorithm (Kabsch and Sander, 1983); number of residues with structure, coil and α-helix patterns are reported. [§] Total number of residues belonging to non-coil patterns used in secondary structure definition. [‖] Aggregation state of peptides in the simulation boxes. Depending on the number of initial peptides, the aggregates can be composed by monomers (1M), dimers (2M), trimers (3M) and/or tetramers (4M). Reported values are population percentages.

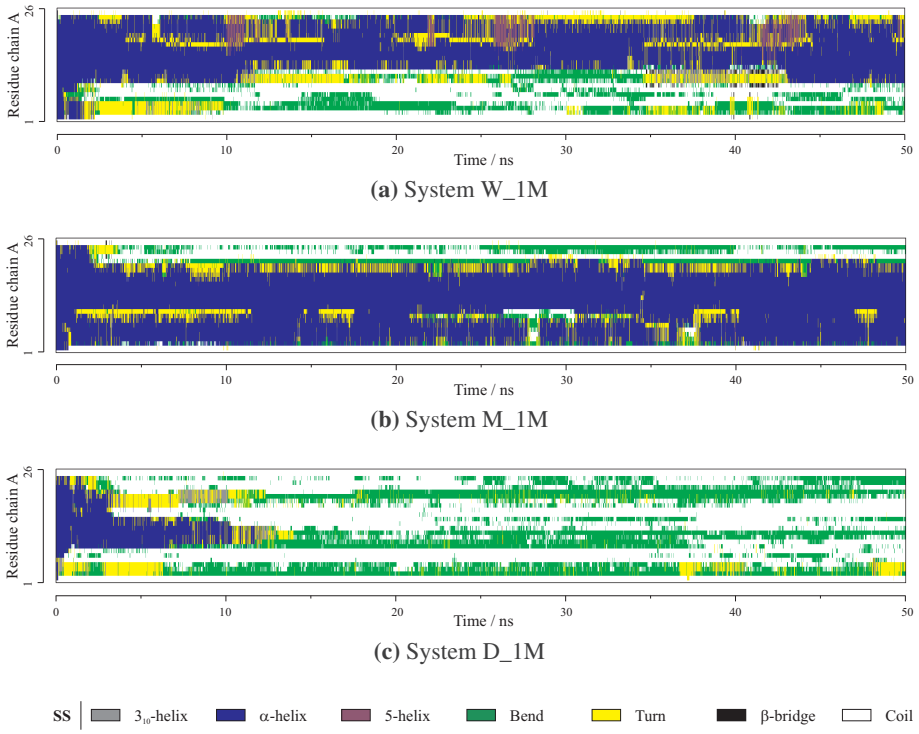
systems.

Appendix table A.7 (see page 201) shows that the shortest minimum distance between periodic images d_{min}^{PBC} of δ-toxin chains in water simulations is 2.8 nm, whereas for methanol, DMSO and bilayer boxes, d_{min}^{PBC} is 1.7, 0.8, and 1.2 nm, respectively. The minimal size of the unit cell necessary to ensure a negligible periodicity-induced perturbation will depend on the nature of the system (i.e., on the range of the intermolecular potential) and on the properties under study (Allen and Tildesley, 1996). Usually, the distance between protein periodic images should be maintained slightly larger than the largest of the cutoff thresholds used for van der Waals or Coulombic interactions. These cutoffs thresholds are a necessary approximation scheme to compute the potential energy of the system, and their definition is force field dependent. The GROMOS force field (Schuler et al., 2001) was parameterized in such a way that the largest cutoff used was 1.4 nm, thus a substantial interaction between a particle and its own images in neighboring boxes will occur if $d_{min}^{PBC} \leq 1.4$ nm. The results in table A.7 indicate that this condition only happened on trajectories D_1M, D_2Ma, D_2Mb, B_2Md, and B_3M. Nonetheless, such short d_{min}^{PBC} distances are transitory in the trajectory, as can be deduced from the average shortest distance $\overline{d_{min}^{PBC}}$ between peptide chain periodic images.

We can also state from d_{min}^{PBC} results, that δ-toxin is structurally more flexible in DMSO than in water or methanol, and that the initial α-helical conformation of the peptide was lost during the simulation, at least for simulations D_1M, D_2Ma, and D_2Mb. All δ-toxin systems in neat solvents were built with a solvent layer of 1.2 nm, thus a minimum distance between peptide images of at least $d_{min}^{PBC} = 2.4$ nm in the first frames of each trajectory, should be expected. If a peptide molecule sees its periodic image during a trajectory and the system volume is constant (see appendix table A.6, page 200), means that a substantial unfolding of the starting conformation must have occurred.

5.3.2 Secondary structure variation in different solvents

δ-toxin in a canonical α-helix conformation was placed at the center of a dodecahedron truncated box and three periodic boundary systems were built, in which the peptide was immersed in water, methanol, or DMSO. After the



(a) System W_1M

(b) System M_1M

(c) System D_1M

Figure 5.3: Secondary structure (SS) of δ-toxin in **(a)** water, **(b)** methanol, and **(c)** DMSO as a function of simulation time.

initial equilibration of the system, trajectory runs of 50 ns were produced. The secondary structure (SS) content of δ-toxin was computed based on a pattern-recognition process of hydrogen bonds and geometrical features (Kabsch and Sander, 1983). Figure 5.3 shows the secondary structure of δ-toxin as a function of time in water (system W_1M), methanol (system M_1M), and DMSO (system D_1M).

The initial α-helicity of δ-toxin is not maintained during W_1M simulation, as can be seen in figure 5.3a. The N-terminus starts to unfold until approximately the 10^{th} residue, a few ns after the beginning of the simulation. Nevertheless, the α-helical content stabilizes for the rest of the trajectory. The average peptide secondary structure content for all simulations performed is shown in table 5.2. Half of the residues have a well characterized structure, with 10 residues in the C-terminal region maintaining an α-helix conformation. This relatively small

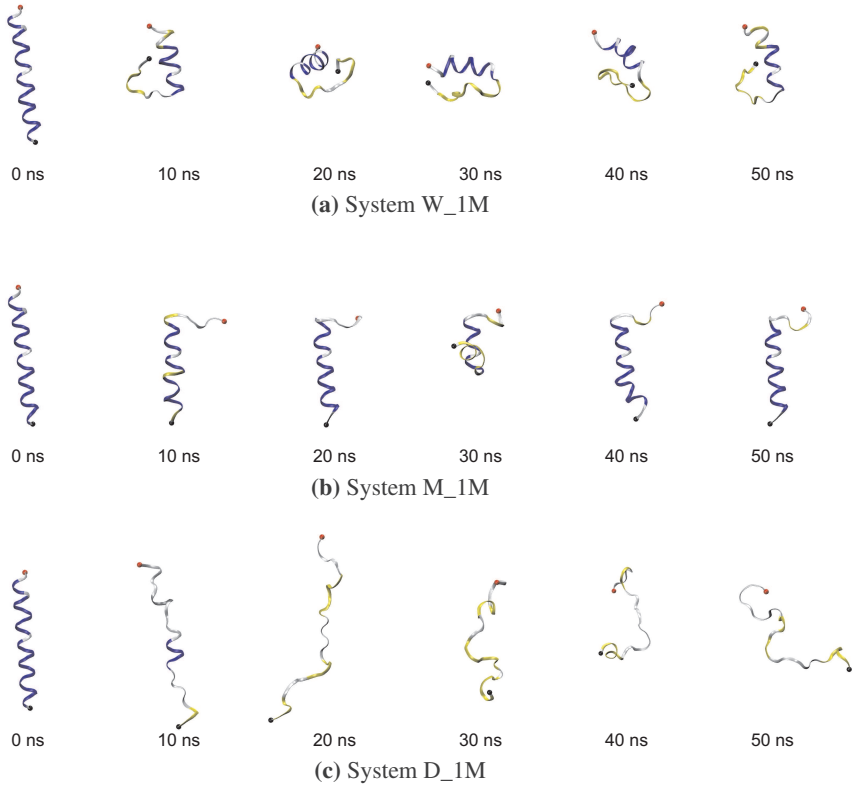


Figure 5.4: Snapshots of δ-toxin taken at 10 ns intervals in **(a)** water (system W_1M), **(b)** methanol (system M_1M), and **(c)** DMSO (system D_1M). The peptides are drawn as ribbons and colored accordingly with secondary structure assigned by the DSSP algorithm (see figure 5.3). The C-terminus is represented by a solid black sphere and the N-terminus by a red sphere. The solvent molecules are not shown for clarity, and all images are at the same scale for direct comparison.

α-helix fluctuates (unfolding-refolding) but is rather stable for the 50 ns of the simulation. Figure 5.4a displays snapshots of the polypeptide backbone during the course of the simulation.

δ-toxin simulated in a methanol box shows a quite different trend. On average, 16 residues are maintained in an α-helix conformation (figures 5.3b, 5.4b). The α-helix starts at about the 3^{rd} residue from the N-terminus of the peptide, and continues until approximately the 20^{th} residue in the sequence. Notice also that the helix is not continuous, and from time to time it occurs a

break on the helix on the 10^{th} residue (e.g., time window between 2 and 12 ns).

On the other hand, the peptide in a DMSO box unfolds very rapidly (figures 5.3c, 5.4c). The unfolding starts again at the N-terminus, followed by a decrease of α-helicity from the C-terminus. A small helix of 7 residues (from residue 9 to 15) is maintained until the 10^{th} ns, where a complete loss of structure occurs for the remaining of the D_1M simulation. On average, 16 residues are in random-coil, thus suggesting low stability of regular secondary structure for δ-toxin in DMSO.

The results suggest that the δ-toxin α-helix is more stable in methanol then in water, and a complete absence of α-helix is observed in DMSO. A comparison with the available experimental results will be given in the next section, where systems closer related to the experimental conditions attained inside a laboratory tube were simulated and analyzed.

5.3.3 Secondary structure *vs* aggregation state

Often, molecular dynamics simulation systems have a high concentration of solute, like NMR samples, due to the small and finite simulation box. δ-toxin monomeric systems discussed in the last subsection are at a protein concentration of about 7-8 mM, reaching concentrations of about 18-21 mM in the modeled systems with four peptide chains discussed in this section. In such crowded systems, peptide chains influence each other and thus the structure of the peptide should be strongly influenced by the presence of other peptides. To get a better description of the secondary structure content of δ-toxin in solution, aggregation systems were modeled in water, methanol, and DMSO. A brief introduction to the systems simulated was already done in the methodological section 5.2, still a more detailed description will be given below.

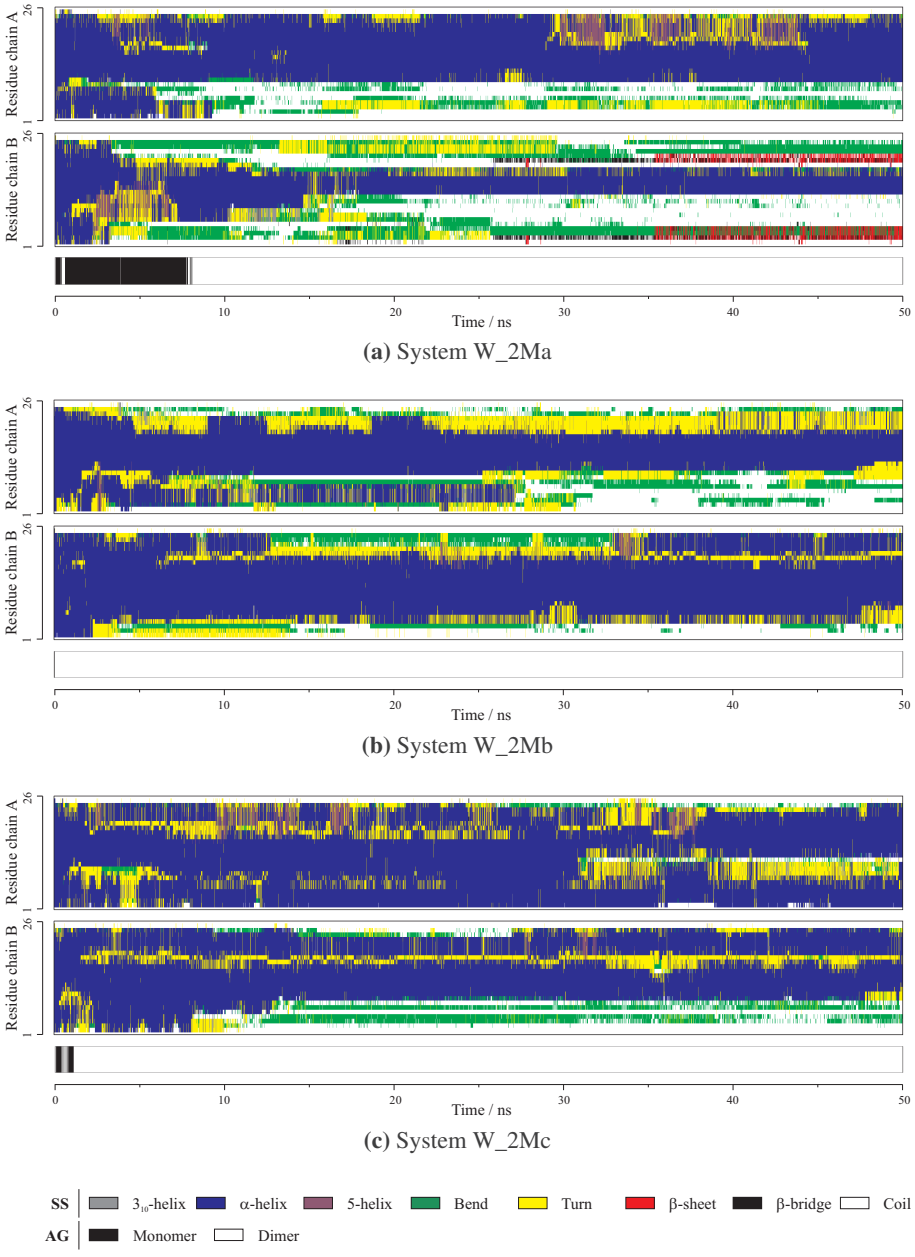
The starting configuration of δ-toxin in all models is an α-helix, the same used for the monomeric systems. Employing rotations and translations in the three-dimensional space four models were built, namely three dimers and one tetrameric system. A parallel head-to-head system with respect to the peptide terminus (system labels X_2Ma; figure 5.1b) will give us insights on the effects electrostatic repulsion has on the secondary structure of both peptide chains. An anti-parallel head-to-tail configuration (system labels X_2Mb; figure 5.1c) is *a*

piori a more stable conformation due to peptide terminus charge attraction, and will be compared directly to the head-to-head model. A perpendicular dimeric model was also built, introducing randomness with respect to the starting orientation between chains (system labels X_2Mc; figure 5.1d). An head-to-tail tetrameric configuration was also build to model a more complex system in solution (system labels X_4M; figure 5.1e). All chains were positioned relative to each other at a distance of 1.7 nm, with no overlapping atoms. No preferred orientation relative to δ-toxin amphipathicity was given between adjacent chains, with exception of the tetrameric model where hydrophobic residues are pointing to the center of the model, as can be seen in figure 5.1. After solvent addition to each model system, at least two solvent layers are present between peptide chains.

Multimeric models in water

Figure 5.5 shows the evolution of the secondary structure motifs on the simulations of dimeric δ-toxin models in water. The head-to-head system model W_2Ma shows that the two chains have different behaviors, as can be seen from figure 5.5a. Whereas chain A shows a relatively well formed α-helix between residues 10 to 24, chain B is much more unfolded throughout the simulation time, maintaining an α-helix pattern between residues 13 to 18. The repulsion between adjacent peptide chain termini leads to both peptides to rotate relative to each other, such that electrostatic repulsion is minimized (data not shown). On the other hand, the head-to-tail model (figure 5.5b) is more stable with respect to α-helix content, due to the electrostatic attraction between N and C termini. Notice that helices macro-dipoles may also play a role on the interaction between the neighboring chains. The perpendicular dimeric model W_2Mc has, on average, the highest content on structure between all dimeric models studied here. At least 50% of the residues are in a α-helical conformation. A more detailed overview of the secondary structure content per peptide chain and studied model is reported in table 5.2.

Though the results on the secondary structure content of δ-toxin dimeric models in water suggest a higher content on overall structure with respect to the monomeric model W_1M, not much can be inferred about how the presence of



(a) System W_2Ma

(b) System W_2Mb

(c) System W_2Mc

Figure 5.5: Secondary structure (SS) of δ-toxin in bulk water for **(a)** a head-to-head model (system W_2Ma), **(b)** a head-to-tail model (system W_2Mb), and **(c)** a perpendicular model (system W_2Mc), as a function of simulation time. The aggregation (AG) state is also depicted underneath the secondary structure stripes.

other peptide chains in solution influence the structure contents of each modeled system. A more detailed analysis can be envisaged if both the aggregation state and the secondary structure are assessed simultaneously.

δ-toxin was considered to be aggregated if there was at least one atom of one peptide chain at less than 0.5 nm from any atom of a neighboring chain. This immediately gives a functional definition on the aggregation state of the peptide in solution. Though this is a *very generous* definition for aggregation, the results were qualitatively the same for different distance cutoffs. At any time during a simulation, the peptide can be in different aggregation states, depending on the initial number of chains on the simulation box. For all simulations performed, we only considered the biggest peptide cluster present. The chosen distance threshold value of 0.5 nm corresponds, on average, to two layers of water molecules between peptides. Though at this distance atoms from neighbor chains are still apart from each other (i.e., no van der Walls contacts), electrostatic interactions between peptide charged groups are quite strong in solution and felt several nanometers way. In fact, this is the main reason why the computation of long range electrostatic forces in MD simulations using cutoff schemes is being left aside.

The aggregation state for δ-toxin dimeric models in water is shown in figure 5.5, depicted as gray colored stripes beneath the secondary structure matrix plots. In the first nanoseconds of simulation W_2Ma, the peptide has a tendency to be in a monomeric state, which it can be understood due to the electrostatic repulsion. As soon has peptide *termini* have rearranged themselves (as well as the other charged residues) to a more stable conformation, δ-toxin forms a dimer throughout the rest of the simulation. On the other hand, system W_2Mb has a great tendency to collapse into a dimeric aggregate right from the first instances, which is also maintained during the simulation. System W_2Mc also shows a high tendency for aggregation, though in the first nanosecond monomeric species are found. This is explained by the randomness of the starting structure, where no peptide *termini* alignment preference was given. Once peptide chains *termini* repulsion forces are minimized, a stable dimeric aggregate is formed.

The simulation of a δ-toxin tetrameric model also presents a high content

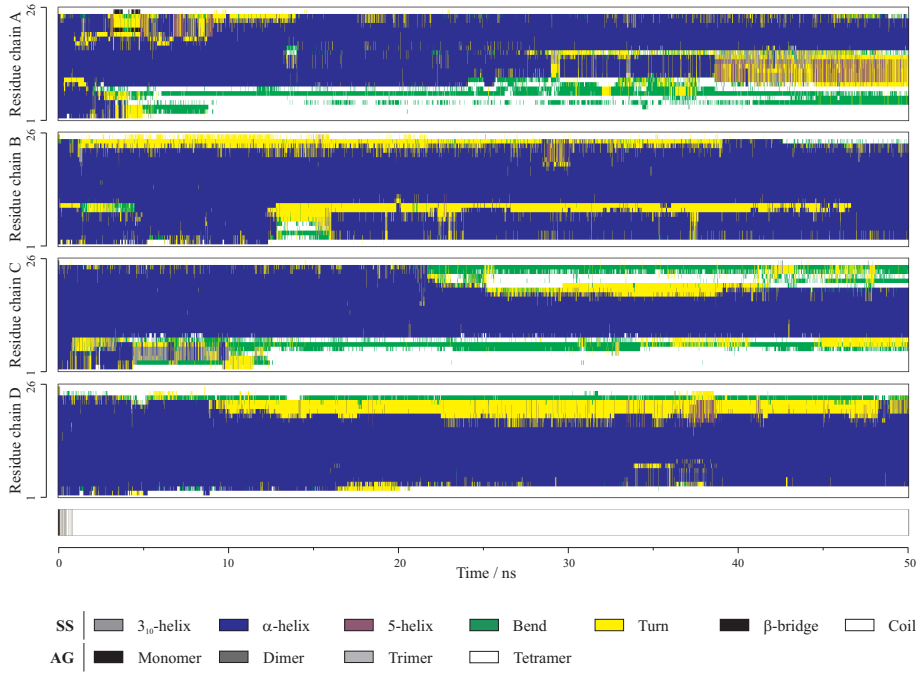


Figure 5.6: Secondary structure (SS) of δ-toxin in bulk water for a tetrameric model (system W_4M) as a function of simulation time. The aggregation (AG) state is also depicted underneath the secondary structure stripes.

of regular structure (figure 5.6), with chains B and D with more than 50% of the residues in an α-helix conformation. The peptide tetramer is formed rather rapidly in the first nanosecond, and is stable throughout the simulation period. Overall, these results seem to suggest that irrespective of the starting model, δ-toxin has a high tendency for aggregation in water solutions (see table 5.2). Nevertheless, structure content seems to be dependent on the starting model structure.

Water is not at all a good solvent for amphiphatic peptides, since it lacks the ability to interact favorably with hydrophobic patches on the peptide surface. This mismatch leads to an inevitable aggregation of the peptides where these hydrophobic regions are clustered together away from the solvent. Remarkably, all simulations of δ-toxin discussed above, collapse quickly into stable aggregates which are maintained throughout trajectories. Regarding the structural content present on the aggregates seen in solution, the way peptides aggregate should

certainly play a major role. A δ-toxin peptide in solution is more prone to structural rearrangements since energetically is a non-favorable system. Water molecules will try to bend the peptide structure until an energetical minimum compromise between enthalpic and entropic factors is attained. We have shown from the simulation of the peptide monomer (system W_1M; figure 5.4a) that most of the initial structural content is lost, namely only 10 residues (approx. 38% of all residues) maintain an α-helical structure on average. Whereas, in the dimer and tetramer models, the α-helix content per peptide chain increases on average, ranging from 20 to 70% of the total number of residues (table 5.2).

Experimental data regarding how δ-toxin behaves in simple solvents is rather scarce. Cruz (2007) showed by circular dichroism that δ-toxin in water at 20 μM (ambient temperature) has a residual content of α-helix of about 30%. Analytical ultracentrifuge and quasi-elastic light-scattering experiments suggest aggregation of the peptide above 2 μM with formation of very large asymmetrical species (Thiaudière et al., 1991). The self-associated species are essentially in a alpha-helical conformation (70%), but once dissociation is promoted in more diluted samples, the α-helix content severely decreases down to 35%. Though the experimental conditions were not exactly the same as the ones employed on the MD simulations performed in this chapter, the results point to a fair agreement between experimental and theoretical techniques. Furthermore, NMR results reported on chapter 3 (see figure 3.2a) also indicate the presence of aggregation, and a low dispersion of peak resonances was indicative of a random-coil structure in solution. The apparent discrepancy of the results when comparing the structural secondary content from NMR with MD can be due to several reasons, namely the solvent chemical-physical characteristics as well as the time scale of the techniques.

Heavy water (D_2O) is often used as a solvent (or co-solvent) in protein NMR experiments instead of light water (H_2O). The NMR experiment cited above (figure 3.2a) was performed in D_2O, and though we can state that aggregation is visible from the spectra, we cannot rule out the presence of residual structure on δ-toxin. It is known that deuterium oxide establishes stronger hydrogen bonds with proteins thus stabilizing (in general) the overall protein structure (Loureiro-Ferreira, 1997). On the other hand, deuterium oxide is a poorer solvent for

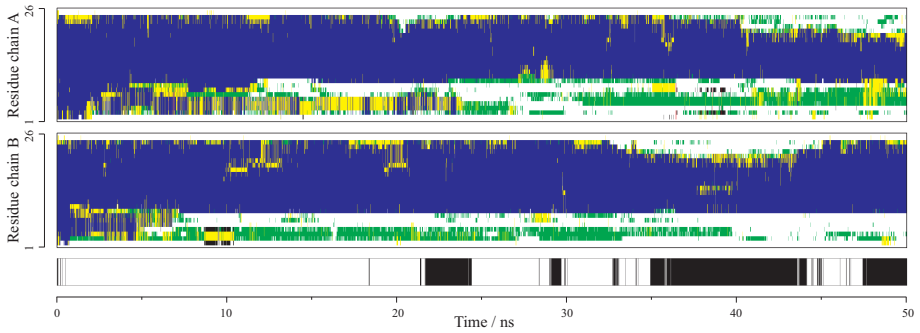
nonpolar amino acids than water (Efimova et al., 2007) thus promoting a stronger aggregation effect on proteins with exposed hydrophobic patches. It is to expect that a more dramatic conformational change would happen on an amphipathic α-helix in D_2O, since hydrophobic patches will be even less stable in this solvent than in normal water. A more pronounced unfolding of the peptide leading to a high content of random coil is thus possible to occur in such environment.

Another aspect to be taken into consideration is the time scale of experimental and MD techniques. MD is a computational expensive technique, both in data collection and CPU time, hence the difficulty in producing long simulations. Albeit *extremely* long simulations are already being produced (Freddolino et al., 2008), current force fields still are inefficient in describing folding-unfolding processes (Freddolino et al., 2009). In short, we cannot assert that the average secondary structural content of δ-toxin peptide chain(s) after 50 ns of simulation will change to a more random-coil like final structure/aggregate.

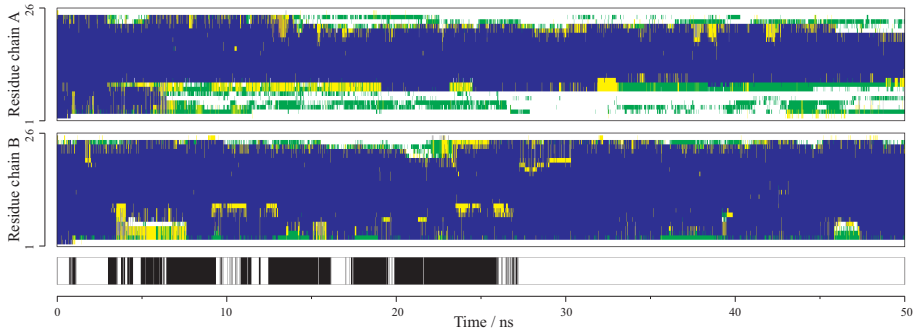
Multimeric models in methanol

The secondary structure content of δ-toxin dimer models in methanol is shown in figure 5.7. Residues in the head-to-head system model M_2Ma (figure 5.7a) show a higher content in structure, particularly in α-helix conformation, when compared with the same model in water. A relatively stable helix between residues 10 and 20 is maintained in chain A, whereas in chain B the α-helix conformation spans from residue 9 until residue 24. The head-to-tail model M_2Mb (figure 5.7b) shows the highest α-helix content per chain from all dimer simulations in methanol. Chain A has 50% of the residues in helix conformations, with chain B having more than 90% of the residues in the same conformation category. System model M_2Mc (figure 5.7c) has around 50% of its residues in α-helix.

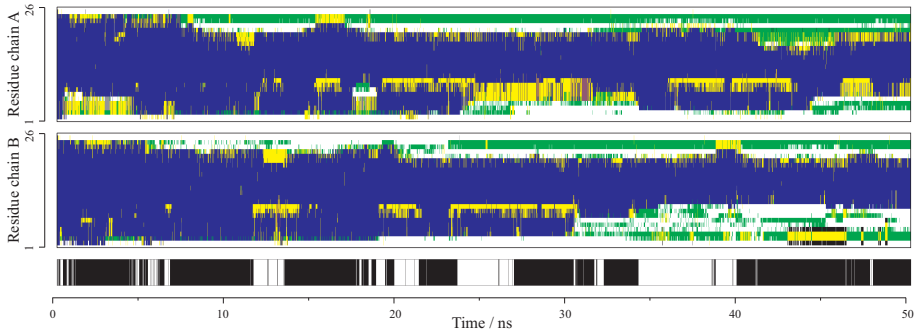
A rather different profile on the aggregation state is observed in the methanol simulations of δ-toxin dimer models, when compared with the previous simulation results in water. Depending on the starting model, most of the δ-toxin population is found in a monomeric form (models M_2Ma and M_2Mc), whereas in the head-to-tail model M_2Mb, dimeric entities populate almost 80% of the



(a) System M_2Ma



(b) System M_2Mb



(c) System M_2Mc

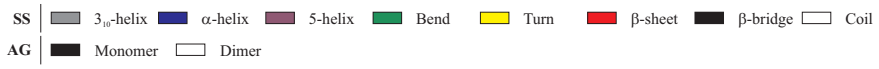


Figure 5.7: Secondary structure (SS) of δ-toxin in bulk methanol for **(a)** a head-to-head model (system M_2Ma), **(b)** a head-to-tail model (system M_2Mb), and **(c)** a perpendicular model (system M_2Mc), as a function of simulation time. The aggregation (AG) state is also depicted underneath the secondary structure stripes.

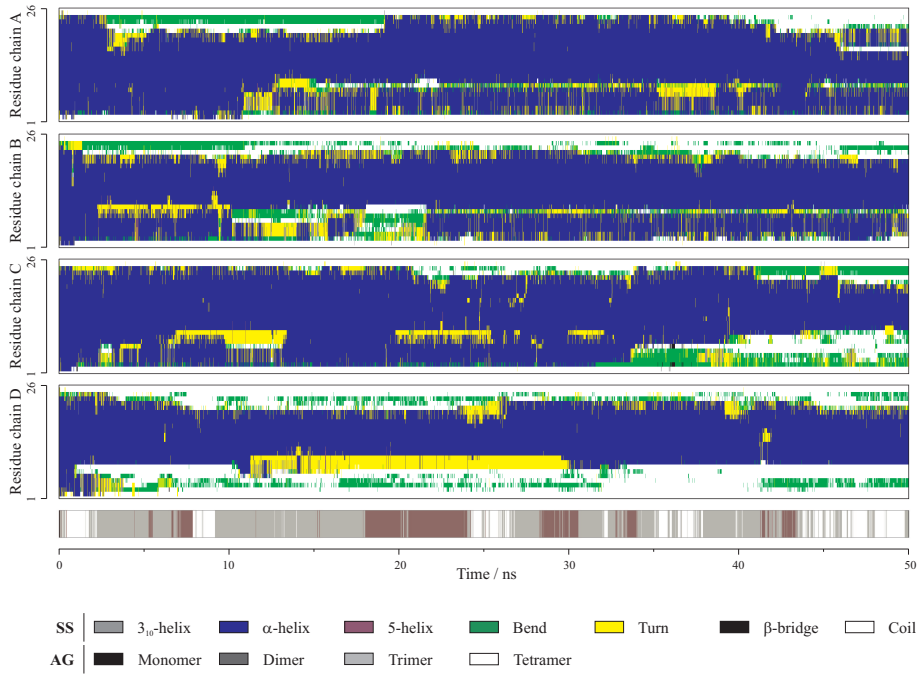


Figure 5.8: Secondary structure (SS) of δ-toxin in bulk methanol for a tetrameric model (system M_4M) as a function of simulation time. The aggregation (AG) state is also depicted underneath the secondary structure stripes.

statistically relevant simulation time considered during analysis (table 5.2).

Secondary structure along simulated time for the tetrameric model of δ-toxin in methanol is depicted on figure 5.8. All chains have more than 50% of their residues in an α-helical conformation, spanning different regions of the peptide primary sequence. Nevertheless, around residue 10 there is always a break in the helix. Notice that residue 10 is a Gly, which may confer a hinge point on the α-helix due to the lack of a bulky side-chain. During the simulation there is no predominance of a specific aggregation state, though higher aggregates of three and four peptide chains populate more than 60% of simulation time.

Overall, the results for δ-toxin in methanolic solution, irrespective of the aggregation model, show that the peptide is more prone to acquire an α-helix conformation than in water. The chemical properties of methanol, namely the fact that its molecule has both a hydrophilic and a hydrophobic moiety, *a*

priori will be a good candidate solvent either for polar or less polar molecules. Though we did not analyse in detail how the solvent is arranged around the peptide during simulations, it is expected that the amphipathic characteristic of this solvent will reveal itself. On average it should be expected to find the methyl group of the solvent near the hydrophobic patches of the peptide, and conversely the hydroxyl group in the first layer of hydration of peptide charged/polar residues.

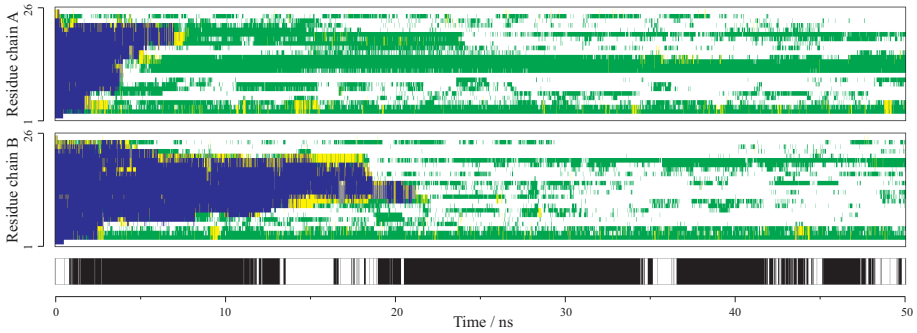
The experimental tertiary structure of δ-toxin in a methanolic solution at 298K was calculated in chapter 3, and is congruent with an α-helix extending from the 2^{nd} to the 24^{th} residue. Deconvolution of CD spectra of δ-toxin also showed that more than 90% of the residues are in a helix conformation (Cruz, 2007).

Multimeric models in DMSO

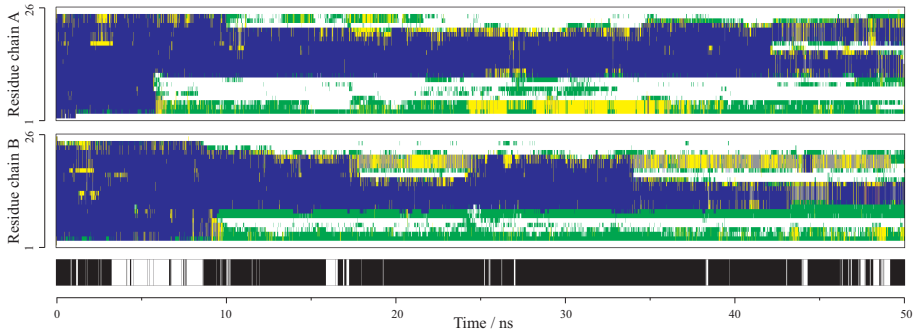
The secondary structure evolution for the three dimeric models of δ-toxin in DMSO are shown in figure 5.9. Clearly, the head-to-head model D_2Ma is the least stable whereas the head-to-tail model D_2Mb is the most stable, in terms of structure content. After 20 ns of simulation, model D_2Ma has lost all the initial secondary structure, and most of the residues are in random coil (table 5.2). Model D_2Mb shows a higher content in structure, but folding and unfolding fluctuations occur during all the trajectory. Model D_2Mc starts to loose structure form the N-terminus, and eventually all α-helix content is lost after 45 ns for both chains. In these three simulations, δ-toxin is almost always found in a monomeric state, thus exposing all the residues to the surrounding environment. The monomeric simulation discussed earlier in sub-section 5.3.2 gave already indications about a very low stability of the starting α-helical model, and the same trend is seen on these multimeric system models.

The tetrameric model also shows the same characteristic loss of secondary structure during the simulation (figure 5.10). The presence of tetrameric molecular entities in solution is residual ($< 1\%$), with the monomeric state populating almost 60% of the simulation time (table 5.2).

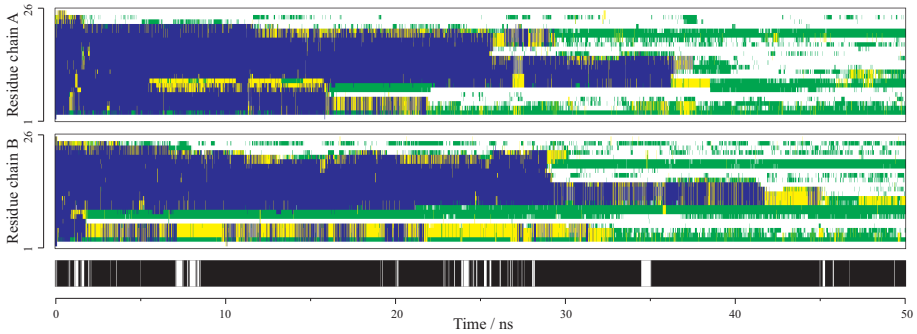
DMSO molecules are stronger competitors to accept available hydrogen bonds than water molecules (Vishnyakov et al., 2001). In practice, the use of



(a) System D_2Ma



(b) System D_2Mb



(c) System D_2Mc

Figure 5.9: Secondary structure (SS) of δ-toxin in bulk DMSO for **(a)** a head-to-head model (system D_2Ma), **(b)** a head-to-tail model (system D_2Mb), and **(c)** a perpendicular model (system D_2Mc), as a function of simulation time. The aggregation (AG) state is also depicted underneath the secondary structure stripes.

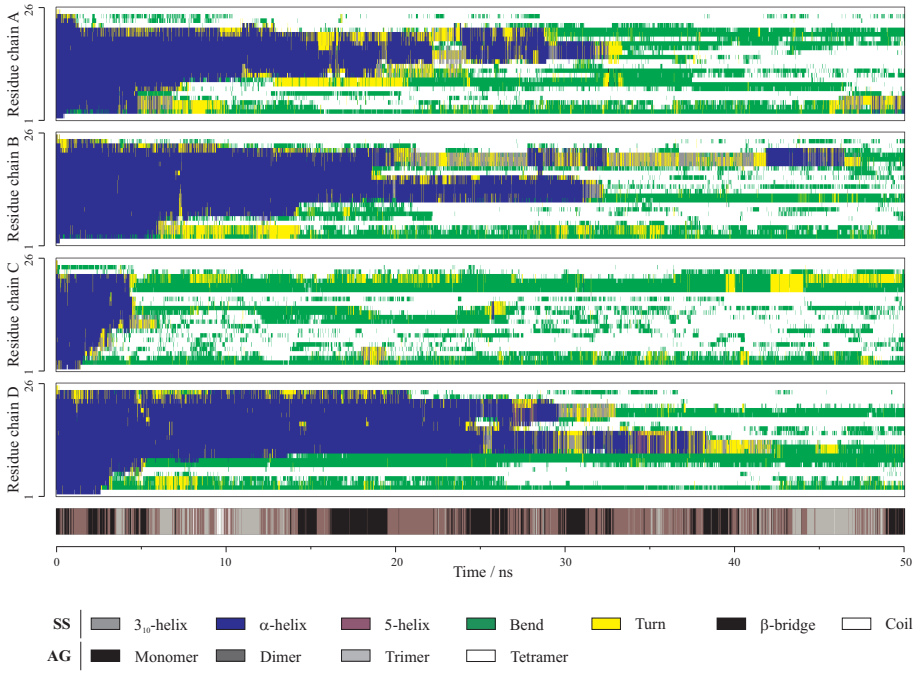


Figure 5.10: Secondary structure (SS) of δ-toxin in bulk DMSO for a tetrameric model (system D_4M) as a function of simulation time. The aggregation (AG) state is also depicted underneath the secondary structure stripes.

DMSO in aqueous protein solutions above a certain concentration can lead to peptide unfolding or protein denaturation, due to a disruption of the intramolecular backbone hydrogen bonds (Jackson and Mantsch, 1991). In the above δ-toxin simulations in DMSO, it is clearly shown that once the solvent disrupts the hydrogen-bonds which maintain the α-helix, the peptide will never recover from a random-coil state. Furthermore, the high content on monomeric species in solution is indicative of the strong solubilizing properties of DMSO molecules.

The NMR data reported on chapter 3 is inconclusive about the secondary structure of δ-toxin, but the obtained restraints do not indicate the presence of helical structure.

In summary, the results presented in this sub-section show that δ-toxin α-helix structure in methanol is both more stable and larger in its extent than in water, while it has a high propensity to unfold in DMSO. In water, the content

in α-helix increases when a more crowded solution is attained. On the other hand, all methanolic systems do show a high content on α-helix, even for the monomeric system.

Though the secondary structure of δ-toxin in water and methanol are similar from the molecular dynamics results reported above, the aggregation dynamics are quite different. Peptide chains in water tend to collapse quickly in very stable aggregates. As soon as the aggregate is formed, it is maintained during the rest of the simulation without disruption. This propensity for aggregation in water was suggested to be a protective measure against proteases (Pokorny et al., 2002). On the other hand, the aggregation state in methanol solution is rather variable, with a constant formation and disruption of aggregated species, though multimeric states have a higher propensity to form. We showed in chapter 2 using pulsed-field NMR studies that δ-toxin in methanol is polydisperse. Though we were not able to discriminate an aggregation state from the NMR data, results of molecular dynamics simulations on the present chapter suggest also a polydisperse solution, with no predominance on the molecular species present in solution.

Figure 5.11 shows ribbon representations of the peptide chains in the tetrameric model in the three solvent systems studied, where the preservation of secondary structure and aggregation state as a function of time can be inferred.

5.3.4 δ-toxin interaction with DMPC

A more complex assemblage composed of δ-toxin and a DMPC bilayer was also simulated. Two system models were simulated, either in a symmetric (system B_2Md, figure 5.1f) or asymmetric (system B_3M, figure 5.1g) arrangement of peptides relatively to the bilayer leaflets.

In system B_2Md, one peptide per bilayer leaflet was simulated, positioned in the bulky water phase in a parallel arrangement to the xy-bilayer plane . The idea beyond this initial system setup was to study the interaction of the peptide with the lipid bilayer from two different initial peptide orientations, in a single simulation box. Due to its amphiphatic character, one of the peptides (chain A) was placed with the hydrophobic patches facing the bilayer surface, whereas the

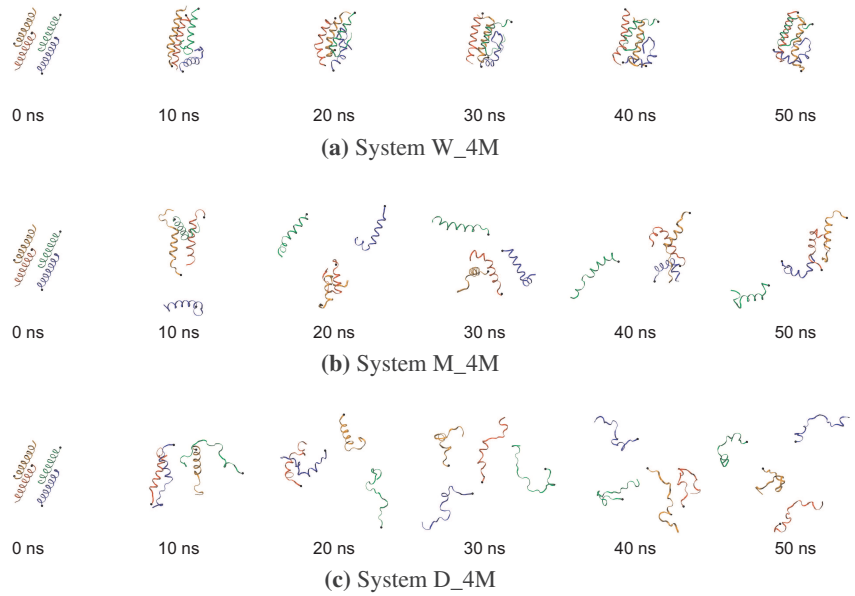
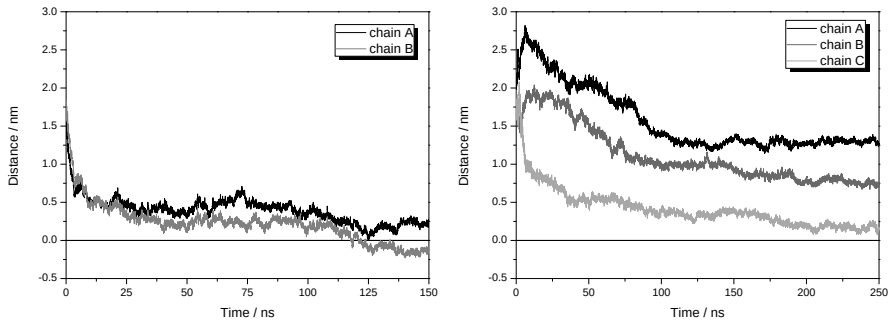


Figure 5.11: Snapshots of four δ-toxin chains taken at 10 ns intervals in **(a)** water (system W_4M), **(b)** methanol (system M_4M), and **(c)** DMSO (system D_4M). The peptides are drawn as ribbons and colored by chain identification (A in blue, B in red, C in green, and D in orange). The C-terminus is represented by a solid black sphere. The solvent molecules are not shown for clarity, and all images are at the same scale for direct comparison.

peptide on the other leaflet (chain B) was placed with the hydrophilic residues facing towards the bilayer surface.

In system B_3M, three δ-toxin peptides were oriented with the long axis parallel to the membrane surface. The peptide starting model was an head-to-tail-to-head configuration, with the hydrophilic patches facing the membrane surface. This model system was built based on several experimental documented data. Most α-helical amphipathic peptides remain largely oriented parallel to the membrane (Bechinger, 1999), and appear to associate side by side (Talbot et al., 2001). Pokorny et al. (2002) investigated binding and insertion of δ-toxin into phospholipid bilayer vesicles. Based on the results obtained, the authors formulated a detailed kinetic mechanism, and concluded that the peptide translocates across the bilayer as a small, transient aggregate, most likely a trimer.



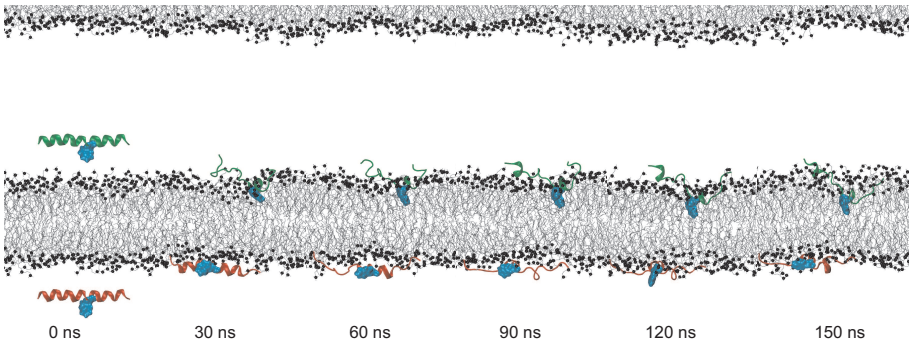
(a) System B_2Md **(b)** System B_3M

Figure 5.12: Interaction of δ-toxin peptide chains and the DMPC bilayer surface, for simulations with **(a)** one peptide (B_2Md) and **(b)** three peptides (B_3M) per leaflet. The distance between peptides and the bilayer was measured between the center-of-mass of each peptide and the average z component of both lipid nitrogen and phosphorus atoms.

In order to assess the interaction of the peptides with the biomembrane, the distance between each peptide in the system to the *xy*-bilayer plane was measured as a function of time (figure 5.12). Briefly, for each time frame on the simulation, the orthogonal distance between the center-of-mass (COM) of the peptide and the bilayer plane formed by both nitrogen (N) and phosphorus (P) lipid atoms, was measured. The lipid N atoms are the first bulky polar atoms at the bilayer interface interacting with the peptide. Furthermore, due to bilayer flexibility and to the bilayer headgroup tilt angle[1], we further considered the P atoms to assess an average polar plane at the bilayer surface per trajectory frame analysed.

Figure 5.12a shows that both δ-toxin peptide chains on system B_2Md gradually approach the bilayer surface in the first 25 ns. A plateau average distance between peptide chains and bilayer surface of about 0.25-0.5 nm is

[1]The angle between the headgroup P-N vector, pointing from the phosphorus to the nitrogen atom, and the bilayer normal was calculated for the DMPC system B_45x12 in chapter 4 (data not shown). The average value taken from the angular distribution is 80.6 ± 0.1 degrees, indicating headgroups lying essentially flat down the bilayer plane. This behavior is in accordance with previous published data on phosphocholine lipids (Bechinger and Seelig, 1991; Högberg and Lyubartsev, 2006).



(a) System B_2Md

Figure 5.13: Snapshots of the binding process of a δ-toxin monomer to a DMPC bilayer, from the trajectory system B_2Md. The nitrogen and phosphorous atoms of the phosphocholine headgroups are shown as black spheres. The remaining lipid molecule atoms are represented by black thin lines. The peptides are in a cartoon representation; chain A in red, chain B in green. The Trp residues molecular surface is shown in cyan. The water molecules soaking the system were removed for clarity, and the periodic box image of the lower leaflet is partially shown.

then maintained for most of the simulation. In the last 50 ns it is clearly seen that both chains do penetrate into the bilayer. Though the general behavior is the same for both peptides, chain B seems to have a higher tendency to interact with the bilayer than chain A. The same information is displayed qualitatively as snapshots in figure 5.13.

Chain B was positioned in the bilayer box with the Trp residue pointing to the bilayer surface, whereas chain A had the same hydrophobic residue pointing away from the lipidic surface. Because both amphiphatic chains were placed in the bulky water, they are free to rotate and translate, still, the presence of a bilayer system is felt right from the beginning of the simulation. Moreover, it seems that the initial orientation of the peptide regarding the bilayer surface is not important, since both chains behave in the same way, i.e., the adsorption kinetics is comparable. A closer look of the snapshot image taken at 30 ns (figure 5.13, paying particular attention to δ-toxin Trp residue orientation), show that chain A rotated along its long axis, exposing the hydrophilic peptide residues towards the bilayer surface. The Coulomb interactions between the lipid charged headgroups and the peptide charged groups, play certainly a major role on this initial step. Since these are long range interactions, diffusion of the

peptide will be conditioned by a pulling effect towards the bilayer, regardless its orientation in solution.

Differences on the behavior of the peptides on the B_2Md system are evident at a later stage of the simulation. Figure 5.12a shows that chain B is closer to the bilayer than chain A after 25 ns of simulation. The orientation of the Trp residue seems to play a major role here. As soon as the Trp residue is able to accommodate itself among the lipid headgroups, it acts as an anchor to the peptide. Moreover, its seems to be a driving force for the peptide to get deeper into the membrane. After 150 ns, δ-toxin Trp in chain B reaches the midplane of the bilayer, and the COM of the peptide is *circa* 0.5 nm below the nitrogen/phosphorus atoms plane (NP-plane) of the bilayer leaflet. On the other hand, the Trp in chain A does not cross this NP-plane, and the peptide remains only adsorbed at the bilayer surface.

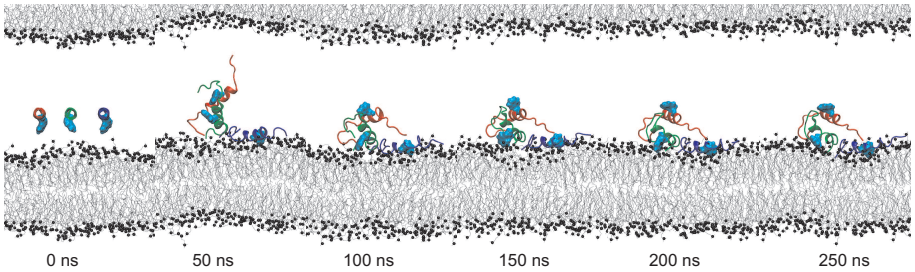
The results obtained for system B_2Md do suggest that the Trp residue is important in the first stages of peptide/bilayer interaction. This is a common trend found in many amphipathic peptides, like cecropins, where a hinge prone region in the middle of the peptide seems to be essential for function (Marassi et al., 1999). Chain A was not able to get deeper into the membrane surface, and two plausible explanations arise. First, considering that the Trp residue of chain A was not in an "ideal" position relative to the bilayer for interaction, it will take time for the chain to rotate exposing the Trp to the bilayer surface. The adsorption of the hydrophobic moiety of chain A to the membrane, which was positioned toward the bilayer surface, would also impose a constraint in the degrees of freedom available to all peptide residues, including the Trp. Slow processes are thus extremely difficult to be studied by molecular dynamics due to its finite simulation time. Secondly, the Trp of chain B was already interacting deeply with the bilayer. The sequestering of lipid molecules by δ-toxin impart a more ordered phase to the bilayer, which *per se* increases the energetic barrier for a solute to permeate it. Furthermore, once both peptides are at the bilayer interface, a mass equilibrium is attained, impairing further peptide bilayer penetration. Pokorny and Almeida (2004) showed that at low P/L (1/50), δ-toxin promotes a graded efflux of the contents of large POPC vesicles. The authors suggested that the bilayer must be under curvature strain, which

arises from an imbalance in the membrane-bound peptide concentration across the bilayer. Once equilibrium is established, membrane curvature disappears, and peptide translocation occurs at negligible rates.

The secondary structure average content of both peptide chains on system B_2Md is mainly random coil (up to 50%; see table 5.2). The simulation results show that the δ-toxin monomer in the presence of a DMPC bilayer is not very stable as an α-helix. Actually, the results are very similar to the ones obtained for the simulation of the monomer in a water box (system W_1M). Several biophysical techniques were used to characterize the structure of δ-toxin in aqueous solution as well as interaction with micelles and lipid membranes (Lee et al., 1987; Thiaudière et al., 1991). Accordingly to the experimental data available, the peptide is largely unstructured at low concentrations, while adopting an α-helical conformation at higher concentrations when bound to micelles or lipid membranes. Nevertheless, such apparent discrepancy between peptide secondary structural content between system B_2Md and published data, is an indication of the intrinsic difficulty in building an initial meaningful model, able to reproduce the experimental data.

It is certain that folding/unfolding are complex biological processes (Dill et al., 2007; Levinthal, 1968). This poses problems on molecular dynamics simulations, such that starting a simulation from a folded protein, or from an extended conformation with the expectation of getting a refolded body, is not at all comparable. The folding/unfolding landscape available to the protein has several energy barriers in between, and molecular dynamics may not be able to cross them under the conditions and on the time scales used. When it comes to unfolding and refolding, the real question is "what causes it"[2]. If it is due to instability of the folded structure given the conditions, then it is quite unlikely that the secondary structure will reform. That may be an issue of force field, but may also be intrinsic to the protein. On the other hand, it may just temporarily unfold, and refold soon after, if it is in an equilibrium (Reif et al., 2009). δ-toxin is not stable as an α-helix in aqueous solutions, unless high aggregates protect exposed hydrophobic residues from the aqueous environment. In the

[2]Private communication from T. Wassenaar.



(a) System B_3M

Figure 5.14: Snapshots of the binding process of a δ-toxin trimer to a DMPC bilayer, from the trajectory system B_3M. The nitrogen and phosphorous atoms of the phosphocholine headgroups are shown as black spheres. The remaining lipid molecule atoms are represented by black thin lines. The peptides are in a cartoon representation; chain A in red, chain B in green, and chain C in blue. The Trp residues molecular surface is shown in cyan. The water molecules soaking the system were remove for clarity, and the periodic box image of the lower leaflet is partially shown.

initial stages of simulation B_2Md, δ-toxin residues are completely exposed to the solvent. The helices unfold as it would be expected for an amphipathic peptide in aqueous solution. When δ-toxin finally interacts more tightly with the membrane, most of the α-helical content was gone. In the simulated time it was not possible to see the re-folding back of δ-toxin.

In the second peptide/bilayer system studied (system B_3M), the goal was to study the interaction of several (three) δ-toxin peptides with a DMPC bilayer in fluid phase. Figure 5.12b shows a quite different trend on the behavior of peptide bilayer interaction when compared to system B_2Md. Though in general terms the COM of all peptides is closer to the NP-plane by the end of the simulation, only one peptide (chain C) strongly interacts with the bilayer surface. The COM of the other two chains (A and B) stays on average between 1.0 to 1.5 nm from the NP-plane. However, we should stress that these results based on the distances calculated using the peptide COM, should be taken only as qualitative measures of the overall process. δ-toxin is prone to major conformational changes, thus the COM of the peptide could be distant from the target membrane, and still a large percentage of its residues could be adsorbed to the bilayer surface.

Figure 5.14 depicts regular snapshots taken from system B_3M. The peptide chains form an entangled aggregate, while interacting with the lipid surface.

The α-helical secondary structure is also lost during simulation as it occurred in system B_2Md, though a considerable number of residues (8 to 12, see table 5.2) are able to maintain this secondary structure pattern. The results from this simulation indicate that the aggregation of the peptides, both in solution and at the DMPC interface, impose a energy barrier for δ-toxin to adsorb strongly to the bilayer. Moreover, without a proper anchoring of the aggregate to the lipid bilayer, penetration of the peptide into deeper regions of the bilayer seems to be impaired.

5.4 Conclusions

In order to assess the conformational stability of δ-toxin, we performed molecular dynamics simulations at room temperature in different chemical environments. To the authors knowledge, this is the first molecular dynamics study on this amphipathic model. The starting structure for all simulations was an ideal right handed α-helical molecular model[3]. Different models were built with the aim of studying the secondary structure content of the peptide as well as its propensity for aggregation in water, methanol, and DMSO solutions. Likewise, the interaction of the peptide with a bilayer composed of DMPC molecules was carried out.

Though reported results in this chapter have considered only a single trajectory simulation per system model, up to five simulations are already produced for a more in-depth comparative description of the overall peptide plasticity in water, methanol, and DMSO solvents. Preliminary analysis show that the multiple MD simulations show identical trends in the same chemical environment.

The simulations carried out in methanol systems showed that the α-helical conformation of the peptide is largely maintained throughout the simulations, whereas in water, larger conformational fluctuations lead to a loss of the secondary structure content. DMSO simulations show that the peptide completely looses its structure. It was also shown that δ-toxin has a high propensity to ag-

[3]The experimental 3-D structure of δ-toxin was not available on the protein data bank when the subject of this chapter was first addressed.

gregate in water, whereas in methanol and DMSO, a complex dynamic behavior of formation and disruption of aggregation species is observed.

When comparing the most recent experimental data obtained with the MD results in the present chapter, we conclude that they agree fairly well and provide a complementary view of the structure and dynamics behavior of δ-toxin in different chemical environments.

Regarding the results from the peptide/bilayer simulations, it was shown in system B_2Md that the Trp residue is very important for δ-toxin to penetrate into the membrane. Furthermore, the aggregation of δ-toxin does not favor the interaction of the peptides with the membrane (system B_3M). In fact, the area per lipid A_L of system B_3M, slightly decreased when compared to a pure DMPC system (see chapter 4) to $A_L = 0.59$ nm^2, supporting the idea that only adsorption to the surface bilayer occurred. A decrease on A_L suggests a closer contact between lipid molecules, in order to counterbalance the interaction of the peptides with the lipid headgroups. On the other hand, A_L increased for system B_2Md ($A_L = 0.63$ nm^2), indicating that the peptide was able to disorganize the lipid surface and insert itself into the bilayer. An increase on A_L suggests that lipid molecules are farther apart, thus lipid molecules were pushed aside by the peptide while insertion occurred.

Several models concerning the molecular mechanism for peptide induced membrane disruption have been proposed, namely the barrel-stave model (Ehrenstein and Lecar, 1977), the toroidal hole model (Ludtke et al., 1996), the carpet model (Shai, 2002), and more recently the sinking-raft model (Pokorny and Almeida, 2004; Pokorny et al., 2002). The number of proposed models is consistent with the present uncertainty on the overall mechanism on how peptides interact with membranes. Furthermore, the crucial question about target cell specificity of these peptides is still under strong scientific debate (Pokorny and Almeida, 2005). Molecular dynamics simulations are a very promising approach to tackle such questions, providing atomistic detail of the process which is virtually impossible to get with experimental techniques alone.

Acknowledgments

We acknowledge the computer resources provided by the Advanced Computing Lab, Milipeia cluster, Department of Physics, University of Coimbra, and the Services and Advanced Research Computing with HTC/HPC clusters (SeARCH), Department of Informatics, University of Minho. Alessandra Villa is acknowledged for sharing the DMSO topology.

Chapter 6

Experimental data-driven modeling of transthyretin amyloid protofilaments

Based on: Nuno Loureiro-Ferreira, J. Rui Rodrigues, Bruno E. Correia, Cândida G. Silva, and Rui M.M. Brito. Experimental Data-Driven Modeling of Transthyretin Amyloid Protofilaments: small changes, different structures, large consequences (*To be submitted*).

6.1 Introduction

Nowadays, the huge interest in amyloid fibrils within the biomedical research community arises from their association with well known amyloid diseases, such as Prion related pathologies (Bovine Spongiform Encephalopathy, Kuru and Creutzfeldt-Jakob), Alzheimer's, Parkinson's and Huntington's diseases, as well as type 2 diabetes mellitus, Familial Amyloidotic Polyneuropathy (FAP) and Senile Systemic Amyloidosis (SSA). Although amyloid diseases are caused by different proteins, they seem to share common molecular mechanisms, which ultimately result in the deposition of ordered protein aggregates mostly in the extracellular environment, though in some pathologies intracellular and intranuclei deposits have been reported.

These supramolecular structures are being a major target of study for the biochemical and biophysical research community, which is trying to unravel the nature of the interactions that make amyloid fibrils a stable structure and the mechanisms by which amyloid fibrils form from monomeric and oligomeric species (Tycko, 2004).

Transthyretin (TTR) is one of the many proteins known to be involved in human amyloid diseases, and over the years was identified as the causative agent of several amyloid pathologies, such as FAP and SSA. TTR is a homotetrameric protein with 127 amino-acids per subunit, 55 KDa of total molecular mass, with a high content of β-sheet, and it is mostly found in the plasma and the cerebral spinal fluid. Each TTR monomer has a β-sandwich fold composed of two four-stranded β-sheets labeled *DAGH* and *CBEF*, as shown in figure 6.1. In the native protein, the β-sheets from two monomers associate edge-to-edge through β-strands *HH'* and *FF'* to produce a dimer composed of two extended β-sheets formed by strands *DAGHH'G'A'D'* and *CBEFF'E'B'C'*. The association of two of these dimers, mainly through hydrophobic interactions mediated by the *AB* and *GH* loops, forms the functional homotetramer. The two known functions of the protein are the transport of thyroxin and when complexed with the retinol binding protein, it helps the carrying of vitamin A.

Current views on amyloid fibril formation by TTR state that, depending on the protein variant or solution conditions, the native tetrameric protein might dissociate to non-native or partially unfolded monomeric species with a high

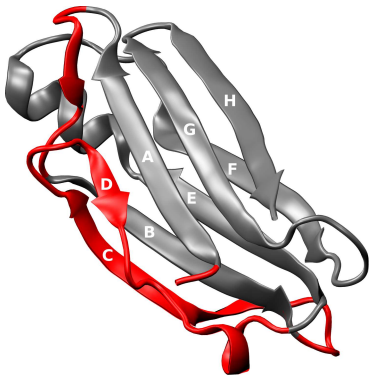


Figure 6.1: Secondary structure of the crystal structure of the wild type TTR subunit (pdb code 1F41; Hörnberg et al., 2000), with the identification of β-strands. A partially disrupted monomer was build by removing strands *C* and *D* and neighbor loops of the native monomer (in red).

tendency for aggregation into soluble oligomers, which grow into insoluble molecular species and eventually mature amyloid fibrils (Brito et al., 2003; Lai et al., 1996; Quintas et al., 1999). Therefore, the structural characterization of these fibrils and the identification of the entities involved in fibril assembly play a crucial role in the understanding of the mechanisms of pathogenesis in amyloid diseases, and in the development of appropriate therapeutic strategies. Though the difficulty of achieving atomic resolution detail for fibrils is well known, considerable progresses are being made toward a more detailed knowledge of these supramolecular structures (Laidman et al., 2006; Nelson et al., 2005).

Proteolysis patterns both *in vivo* (Thylén et al., 1993) and *in vitro* (Goldsteins et al., 1997; Gustavsson et al., 1991) of TTR amyloid fibrils, suggest an increased exposure of strands *C* and *D* in the fibrillar form. Recent X-ray spectroscopy data also demonstrated that residue cysteine 10, which is locate at the start of strand *A* (figure 6.1), is more exposed and oxidized in the fibrils than in the native tetrameric form of TTR (Gales et al., 2003). A conformational change in the *CD* β-strands and *DE* loop regions must occur in order to expose Cys10.

Yeates and collaborators (Serag et al., 2001, 2002) used a series of nitroxide spin-labeled cysteine substituted TTR mutants, for the determination of inter-nitroxide distances by electron paramagnetic resonance (EPR). The authors

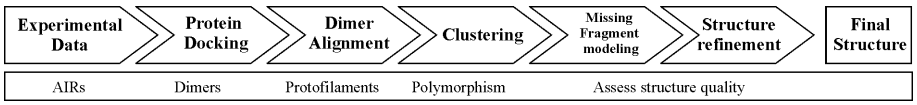


Figure 6.2: Docking-alignment protocol devised to assemble models of amyloid protofilaments of TTR. The protein interfaces of TTR believed to mediate protein aggregation were modeled by protein docking, using experimentally determined data. The resulting complexes were aligned generating multiple protofilaments, and clustered. The structural quality of the chosen structures is then assessed.

proposed an "anti-parallel head-to-head/tail-to-tail" arrangement of the TTR subunits in the fibril, with the native inter-subunit contact between β-strands *F* and *F'* maintained. A new inter-subunit interface is formed between β-strands *B* and *B'*, associated in an anti-parallel arrangement. The formation of this new interface implies the displacement of strand *C* from the β-sheets edge, in order to expose strand *B*. These results are consistent with the formation of a continuous extended cross-β structure characteristic of amyloid and composed by strands $(BEFF'E'B')_n$.

Hydrogen/deuterium exchange studies of TTR in the fibrillar state identified a core region of the TTR subunit with large solvent protection factors . The data indicates that strands *B, E* and *F* and strands *A, G* and *H* of the TTR subunit are, if not totally, partially maintained in the fibrils. The presence of solvent exposed residues in strands *C* and *D*, as well as in the connecting and following loops, clearly shows that this region is not part of the fibrillar core.

Taken together, the experimental results support a protofilament structure formed by a core of two three-stranded β-sheets, *BEF* and *AGH*. Moreover, the interaction between these three-stranded β-sheet monomers in the amyloid fibrils is mediated by two types of interfaces: a near-native interface (NearNI) comprising strands *FF'* and *HH'* and a non-native interface (NonNI) constituted by strands *AA'* and *BB'* of adjacent subunits.

The above experimental findings provided us with impetus to develop and explore a methodology to build a high resolution molecular model of the elementary units that constitute the fibrils (figure 6.2), the so called protofilaments. The methodology, briefly reported in a previous work (Correia et al., 2006), was applied to TTR, but may also be useful for other proteins involved in the formation of amyloid fibrils.

6.2 Computational methods

6.2.1 TTR native monomer

The coordinates for the native monomer were taken from the chain A of the TTR crystal structure deposited at the RCSB Protein Data Bank (Berman et al., 2000) with the accession code 1F41 (Hörnberg et al., 2000). We have opted for structure 1F41, given that it has the largest atomic resolution until the moment, though terminus residues are not present on final structure. Some residues have two well defined conformations with 0.5 of occupancy factor. In these cases, the second conformation was discarded. Crystallization waters were also discarded.

6.2.2 Modeling a TTR non-native monomer

In order to expose strands *A* and *B*, a partially disrupted non-native monomer was built by removing strands *C* and *D* and adjacent loops (residues 36 to 65) from the native monomer. After careful inspection of the resulting structure, loop *AB* (residues 21 to 27) was also removed because it partially occluded the non-native interface that is going to be modeled and it poses steric hindrance problems during the protein-protein docking stage. The N-terminal residues Cys10 and Pro11 form a kink on strand *A*, not allowing a correct approximation of strands *A* and *A'* during the docking procedure, and were also removed.

6.2.3 Protein-protein docking protocol

Docking calculations were performed using a high ambiguity data-driven docking approach, implemented on HADDOCK 1.2 (Dominguez et al., 2003) in combination with CNS (Brünger et al., 1998). The docking protocol consists of three stages: randomization of orientations and rigid body energy minimization; semi-flexible simulated annealing in torsion angle space; and a final refinement in Cartesian space with explicit solvent. For details, see Dominguez et al. 2003.

Two native monomers were docked against each other at their native interface (strands *FF'* and *HH'*), generating near-native dimeric structures. Therein, the interface *FF'–HH'* obtained from the docking procedure will be called

near-native interface, in order to distinguish it from the native dimeric crystallographic interface. A second set of docking runs was performed, where two non-native monomers were docked at their non-native interface (strands *AA'* and *BB'*), generating non-native dimeric structures.

The docking is driven by ambiguous interaction restraints (AIRs), defined between any atom of active residues of the ligand protein and all atoms of all passive residues on the receptor protein (table 6.1). An upper-bound of 2 Å was applied to AIRs, a compromise between hydrogen-hydrogen and heavy-atom minimum van der Waals distances (Dominguez et al., 2003). Active and passive residues were defined based on nuclear magnetic resonance (NMR)) H/D exchange data (Olofsson et al., 2004), and solvent accessibility calculations using the program NACCESS (Hubbard and Argos, 1994). The active residues define the ones experimentally identified to be involved in the interface interaction and having a high solvent accessibility ($> 45\%$ relative accessibility) in the free form of the protein. The passive residues are all solvent accessible neighbors of active residues. Additional sets of restraints for the NearNI and NonNI were defined based on EPR experimental data (Serag et al., 2001, 2002), and were applied between the C atoms of the residues involved in the site-directed spin labeling study. The interface amino-acid residues which constitute the flexible segments in the last two stages of the docking procedure are defined from the active and passive residues used in the definition of AIRs $\pm$ 2 sequential residues.

The modeled non-native monomer is constituted by 3 protein fragments, roughly corresponding to strands *A* (residues 12 to 20) and *B* (residues 28 to 35), and a third fragment with residues 66 to 125. In order to maintain the overall monomer structure, hydrogen bonds were defined as unambiguous restraints[1] , on adjacent strands of each monomer β-sheet, for both partners involved in the

[1] Hydrogen bonds were calculated for TTR native monomer using VMD (Humphrey et al., 1996) builtin routines. A schematic representation of the main-chain hydrogen bonds is depicted in appendix figure A.1, page 205. The unambiguous restraints were defined between adjacent main-chain heavy atoms N and O. Example:

```
assign (segid A and resid 13 and name n) (segid A and resid 105 and name o) 2.88 0.50 0.50
assign (segid A and resid 13 and name o) (segid A and resid 107 and name n) 2.87 0.50 0.50
```

Table 6.1: Restraints used in the docking runs performed with HADDOCK.

			Restraints			
			Ambiguous		Unambiguous	
Interface	Run	Flexible segments	Active residues*	Passive residues†	Inter-residue distance‡	
NearNI	NearNI+epr	84-101 112-125	89,92,94,96 118-124	86,87,99 114-117,125	89(l_{EF})-96(F') 94(F)-94(F') 96(F)-89($l_{E'F'}$) 96(F)-96(F')	12 15 12 25
	NearNI-epr	84-101 112-125	89,92,94,96 118-124	86,87,99 114-117,125		
NonNI	NonNI+epr	12-20 28-35	18,19,28-34	12-17,20,35	29(B)-29(B') 31(B)-31(B') 33(B)-33(B')	12 8 14
	NonNI-epr	12-20 28-35	18,19,28-34	12-17,20,35		

* The active residues correspond to the residues accessible to the solvent in the monomeric form and solvent protected in the fibrillar state (Olofsson et al., 2004); † The passive residues correspond to the residues that are highly solvent accessible and/or neighbors of active residues;
‡ Distances experimentally determined by EPR in Å (Serag et al., 2001, 2002). B, B', F, F', l_{EF} and $l_{E'F'}$, identify the β-strands and loop-EF. Along with these distances, a set of hydrogenbonds between the adjacent strands *AG* and *BE* on the NonNI runs were applied on both runs.

modeling of the NonNI.

A total of 2,000 rigid body docking solutions was generated for each docking run. In addition, for each of the rigid body starting conformations, 10 rigid body trials were performed and only the best solution based on the intermolecular energy (E_{interm}; sum of van der Waals, electrostatic, AIRs and unambiguous energy terms) was kept, bringing the total effective docking trials to 20,000 per docking run. The 200 best solutions according to intermolecular energies were subjected to semi-flexible simulated annealing followed by a final refinement in explicit water. During the simulated annealing and the water refinement stages, the amino-acids at the interface (side chains and backbone) are allowed to move in order to optimize the interface packing. The final refinement in explicit water is performed to improve the energetics of the interface, which is important for a proper scoring of the resulting conformations. The non.bonded energies were calculated with the OPLS parameters (Jorgensen and Tirado-Rives, 1988) using an 8.5 Å cut-off.

The final complexes were clustered using a pairwise backbone root mean squared deviation (RMSD) matrix. Structures were superimposed on backbone atoms, and the RMSD was calculated on backbone atoms of both partners. Then a clustering algorithm described in Daura et al. (1999) was employed . Briefly, for each structure, the number of neighboring structures was counted using a 1 Å cutoff. The structure with the largest number of neighbors constitutes a cluster and is eliminated from the pool of structures. The procedure was repeated for the remaining structures in the pool until the size of the newly extracted cluster was smaller than 4 elements. Clusters were ranked according to their average intermolecular energies.

The buried surface area (BSA) for the obtained docked dimers was calculated using CNS (Brünger et al., 1998), by taking the difference between the sum of the solvent accessible surface area (SASA) for each partner separately and the SASA of the complex. SASA was calculated using a 1.4 Å water probe radius and an accuracy of 0.055 Å. The 10 lowest energy structures of each cluster obtained were selected for further analysis. A broader analysis on the lowest energy cluster was also done using the 100 best structures. The lowest energy structure of the lowest energy cluster is considered to be the

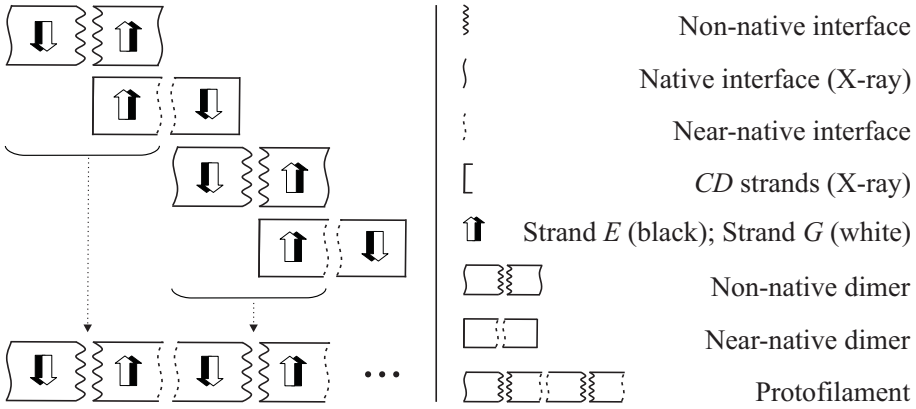


Figure 6.3: Schematic view of the alignment protocol of near-native and non-native dimers, in order to build protofilament structures. The black and white arrows represent the β-strands used to align the dimeric structures.

highest ranking solution, and is assumed to be the best structure generated by HADDOCK.

6.2.4 Analysis of intermolecular contacts

Intermolecular contacts (hydrogen bonds and non-bonded contacts) were analyzed with DIMPLOT, which is part of the LIGPLOT software (Wallace et al., 1995), using the default settings: 3.9 Å heavy-atoms distance cut-off for non-bonded contacts as well as 2.7 Å and 3.35 Å proton-acceptor and donor-acceptor distance cut-offs, respectively, with minimum 90° angles (D-H-A, H-A-AA, D-A-AA) for hydrogen bonds (McDonald and Thornton, 1994).

6.2.5 Protofilament alignment procedure

An alignment protocol was devised to build a protofilament from non-native and near-native dimeric structures obtained in the protein-protein docking scheme (figure 6.3). A Tcl (Ousterhout, 1994) script brings together all the procedures required to build the TTR protofilaments and to perform subsequent analysis. The script runs under VMD's (Humphrey et al., 1996) Tcl interpreter, making use of the major facilities provided for handling molecular structures.

The non-native dimer is used as a building block, and the near-native dimer

as a template model to rebuild the interface between adjacent non-native dimers at the protofilament level. β-strands *E* and *G* (backbone atoms) are used to align the building blocks with the template models, since their relative coordinates were approximately maintained through out the docking procedure. Briefly, two steps are required to build a protofilament composed of two dimeric subunits, for each pair of non-native/near-native docked structures. A superimposition is done between chain *B* of the non-native dimer and chain *A* of the near-native dimer. After this initial alignment, chain *A* of the near-native dimer is discarded and the relative orientation of the resulting 3-mer is saved. This 3-mer will be then subject to a new superimposition, between the left over chain *B* of the near-native structure with chain *A* of a new non-native dimer, producing a 4-mer. Once chain *B* of the last added near-native dimer is discarded, the obtained protofilament is composed solely by four monomeric units coming from non-native dimers. This two-step protocol is repeated until the growing n-mer protofilament has the desired length.

Two sets of protofilaments were built. One of the sets uses non-native and near-native dimers obtained from HADDOCK runs driven only by AIRs (protofil-epr), while the other set of protofilaments was assembled from dimers obtained with AIRs and EPR distance restraints (protofil+epr). The 100 best solutions taken from the cluster with the lowest average intermolecular energy, in each docking run, were used as protofilament building blocks in the alignment scheme. We combined in a pairwise fashion each near-native dimer with each non-native dimer, producing a total of 10,000 protofilament structures for each set of docking conditions.

6.2.6 Protofilament polymorphism

The characterization of each protofilament was accomplished by measuring two geometric properties: the linearity and the helical twist. The linearity of the supramolecular structures was evaluated by measuring the distance between the geometric centers of each dimer composing the protofilament. The twist of the protofilaments was studied by calculating the angle between strand *G* of the first monomer of the first dimer and the same β-strand of the first monomer of the remaining protofilament dimers.

From the two sets of 10,000 protofilament structures, protofil-epr and protofil+epr, about 99.8% of them were selected based on the criteria of having a $r^2 \geq 0.95$ in the linear fitting applied to the distance data. All selected protofilaments present a helical twist. This type of structure implies a periodic factor in the number of dimers. A discrete Fourier transform was applied to the angle data of each protofilament to determine the periodic component. The protofilament structures used for this analysis consisted of 128 dimers. The protofilament structures were built with such a large number of subunits to gain a greater frequency resolution, leading to a better period discrimination. The approach allowed us to cluster the protofilaments based on their helical period.

The representative cluster of protofil-epr and protofil+epr was further characterized. The hydrogen bonding pattern on the protofilaments was studied. We only considered hydrogen bonds involving backbone atoms on both NearNI and NonNI subunit interfaces. One protofilament from each initial set was chosen, so as to maximize the number of hydrogen bonds on both interfaces. These two supramolecular structures will be further refined in order to get a full polypeptide chain for each protofilament subunit. Alongside, the angles between adjacent β-sheet strands were determined as well as between each strand and the protofilament major axis.

6.2.7 Protofilament refinement

In order to rebuild the full polypeptide chain in each protofilament subunit, the previously removed peptide fragments in the non-native monomer and the terminal residues were added (figure 6.4). The approach uses the colony energy concept implemented on the loop prediction program Loopy (Xiang et al., 2002). Though the program was designed to sample the space for correct conformations of a given loop, we used it to predict the conformation of loops and terminal residues missing in the protofilament subunits. For short protein fragments of less than 7 residues, 500 fragment candidates were used, while for longer fragments we used 2000 candidates. Energy minimization was performed on candidate conformations and the best solution, after energy sorting, was chosen.

The added protein fragments were then subjected to a short molecular

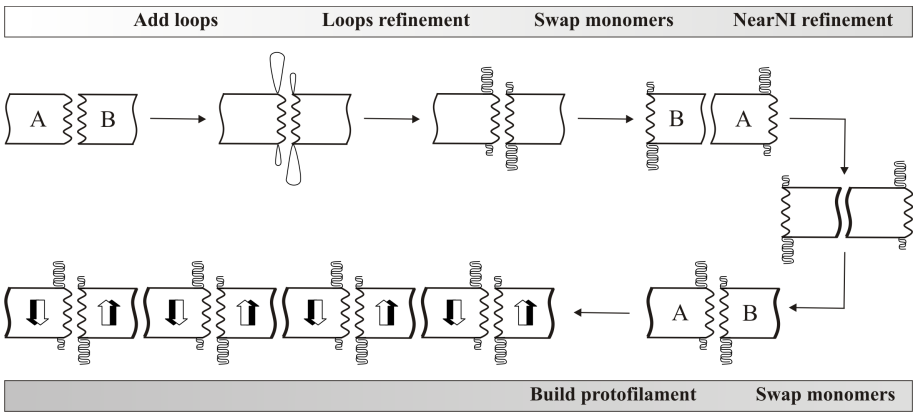


Figure 6.4: Protofilament refinement methodology. The full peptide chain of the partially disrupted monomers was rebuilt and the added fragments subjected to conformational refinement. The alignment protocol described previously was applied to the obtained non-native dimer, and the NearNI modeled was also refined, in order to remove any clashing interactions. The final non-native dimer was then used as the building block in the TTR protofilament assemblage. See computational methods for further details.

dynamics (MD) run in vacuum, while keeping the remaining protein atoms fixed, i.e., without interfering with the NonNI. This non-native dimer with refined loops was then subjected to an alignment scheme with a near-native dimer, in order to reproduce the NearNI obtained with HADDOCK. The NearNI was energy minimized and subjected to a short MD run. The resultant non-native dimer, with refined terminal/loops and a relaxed NearNI, was then used to generate a final protofilament with 20 monomers, using the protofilament alignment procedure described above.

The protofilament was solvated using VMD, and NaCl ions were added to a final ionic strength of 150 mM, resulting in a system of $>$ 350,000 atoms. All atoms were explicitly represented. After energy minimization, the system was first heated for 40 ps under Langevin dynamics at constant volume to the target temperature (298 K), with hard harmonic restraints placed at the backbone atoms, followed by 40 ps at constant pressure. The pressure control was achieved using the Langevin piston Nosé-Hoover method. Equilibration continued for more 40 ps with soft harmonic restraints placed at the backbone atoms and 15 ps without harmonic restraints. The volume was then fixed and

the system was finally equilibrated using classical molecular dynamics in the NVE ensemble for 15 ps. A production run of 100 ps was carried out and coordinates for the whole system were saved every 1 ps.

The simulations were carried out using periodic boundary conditions and a time step of 2 fs, with distances between hydrogen and heavy atoms constrained. Short range non-bonded interactions were calculated with a 12 Å cut-off with the pair list distances evaluated every 10 steps. Long range electrostatic interactions were treated using the particle mesh Ewald summation algorithm and were computed at every step. All energy minimizations and MD procedures were performed with the program NAMD (Phillips et al., 2005), using version 27 of the CHARMM force field (MacKerell et al., 1998).

The overall stereochemical quality of the final structure was assessed with PROCHECK (Laskowski et al., 1993).

The Adaptative Poisson-Boltzmann Solver (Baker et al., 2001) software package was used to describe qualitatively the electrostatic interactions between the final protofilament model and a salty aqueous solution. A linearized Poisson-Boltzmann equation was solved at 150 mM ionic strength with a solute dielectric of 2 and a solvent dielectric of 78.5. The assignment of charges and radius per atom was done with PDB2PQR (Dolinsky et al., 2004) using CHARMM-27 parameters.

All images were produced using VMD, unless stated otherwise.

6.3 Results and discussion

The knowledge of the architecture of amyloid fibrils is essential for understanding the pathological process in amyloidosis and to devise therapeutic strategies. The rational design of drugs that might interfere with or reverse amyloid formation is strongly dependent on the accuracy of target selection. A docking-and-alignment protocol was devised in order to build amyloid protofilament models of TTR (figure 6.2). Briefly, two sets of protein-protein docking computations were performed, to recreate a near-native subunit interface and to build a new non-native subunit interface. The docking procedure is driven by ambiguous/unambiguous interaction restraints, defined accordingly to ex-

perimental data. The protein complexes obtained are then subjected to an alignment/refinement scheme, generating protofilaments with distinct characteristics. A clustering analysis is performed based on the twist periodicity of the protofilaments obtained, and the chosen protofilaments are subjected to a short molecular dynamics run to relax the structures. The overall stereochemical quality of the structures is then assessed.

6.3.1 Defining the protein docking sites

The mechanism by which the native globular fold of TTR may be converted into elongated fibrillar assemblies has been investigated extensively, and numerous studies have shown that dissociation of the native tetrameric structure into partially unfolded monomeric species precedes amyloid formation (Lai et al., 1996; Quintas et al., 2001). The re-association of these partially unfolded species into filamentous protofibrils (Lashuel et al., 1998; Olofsson et al., 2001) accompanied by partial refolding of the subunits, ultimately coalesces into mature rod-like fibrils (Serpell et al., 2000, 1995). Several different models of the TTR amyloid architecture have been proposed (Inouye et al., 1998; Olofsson et al., 2001; Serag et al., 2001, 2002; Serpell et al., 1995). The most recent model was derived from experimentally determined amide proton protection factors (Olofsson et al., 2001). The authors proposed a model of the fibril core, in which the native conformation of the six β-strands *A, B, E, F, G,* and *H* are essentially preserved and able to self-associate into a dimeric (monomer-monomer) native-like conformation. The amide protection factors further suggest an interface region formed by an anti-parallel arrangement of strands *FF'* and *HH'* (NearNI), similar to the native dimeric crystallographic interface. Furthermore, a number of experimental works (Goldsteins et al., 1997; Gustavsson et al., 1991; Olofsson et al., 2004; Serag et al., 2002; Thylén et al., 1993) supports the view that in the fibrils, strands *C* and *D* are displaced from their native position at the edge of the β-sandwich, becoming loosely structured and solvent exposed. Strands *A* and *B* are thus available to participate in intermolecular interactions. Solvent protection of this novel edge region (NonNI), as well as formation of a continuous β-sheet, is then achieved by anti-parallel intermolecular association via strands *AA'* and *BB'*. These observa-

tions are of the utmost importance in the design and validation of our docking approach.

The first goal of this work was to model the two types of interfaces (NearNI and NonNI) suggested as taking part on the TTR protofibril assembly. This was accomplished with two sets of protein-protein docking calculations, each one modeling a different interface. The monomeric subunits that constitute the modeled complexes are treated as semi-rigid bodies retaining a native like fold through out the procedure, with allowed flexibility on the interfaces. Moreover, the experimental data discussed above allows us to pinpoint which residues are involved in the interface interactions, upon monomer stacking in the assemblage of non-branched protofilaments.

In order to generate docked structures using HADDOCK (Dominguez et al., 2003), the identification of the binding interfaces of both interacting proteins is a prerequisite. For this purpose, we used solvent accessibility data determined by solution NMR spectroscopy on TTR fibrils, of the variant TTRY114C (Olofsson et al., 2004). According to the procedure described in the methodological section, for the NearNI modeling the binding interface was defined by 11 active residues (E89, E92, V94, T96 and T118 to N124), since they show a significant solvent protection in the fibril and high solvent accessibility in the monomeric form. Eight surface exposed amino-acids that occur close to the active residues were defined as passive (P86, F87, D99, Y114 to S117 and P125). For the NonNI modeling, 9 active (D18, A19, V28-R34) and 8 passive residues (L12-L17, V20, K35) were assigned in an analogous way. This resulted in a total of 22 and 18 ambiguous interaction restraints, respectively for NearNI and NonNI modeling. AIRs were defined as distance restraints between each active residue of one partner to all active and passive residues of the binding partner. The complete list of restraints as well as the flexible segments for both docking runs are described on Table 6.1.

6.3.2 Modeling TTR near-native dimers

The crystallographic structure of wild type (WT) TTR was used as the starting model in the docking procedure to create the NearNI. A native monomer was docked against another native monomer at their native interfaces (strands

FF' and *HH'*), producing an ensemble of near-native complexes (figure 6.5).

The first approach we took to model the NearNI employed all experimentally available data, namely H/D exchange data and EPR distance restraints. Yeates and collaborators (Serag et al., 2001) determined five inter-spin distances, between spin-labels incorporated at the native interface region using several TTR cysteine mutants, in the fibrillar state. Their study showed that intermolecular residue distances remain largely unchanged from the native tetrameric form for single cysteine mutants 94C, 96C and the double cysteine mutant 89C/96C. On the other hand, two novel interactions are formed upon fibril assembly in the case of mutants 85C and 89C. Several docking trials incorporating different sets of these five EPR distances were run (data not shown). HADDOCK was not able to satisfy all imposed EPR restraints, and the produced interfaces were clearly not satisfactory in geometric terms, since the β-sheets did not show a complementary edge-to-edge connection. The stacking of TTR subunits at the β-sheet edges was previously proposed based on X-ray fiber diffraction studies (Inouye et al., 1998). These preliminary results suggest that if the subunits on the protofilament have a native-like conformation, then in order for residues 85 and 89 to be closer enough between two adjacent subunits, a larger rearrangement than the one allowed by the docking approach should occur. Or, alternatively, the interactions between these two residues could come from neighboring protofilaments, as already proposed by Yeates and collaborators (Serag et al., 2001). In fact, the mechanism of protofilament association into parallel fibrils, capable of a cooperative transformation into twisted super helices, was anticipated to be a generalized scheme for amyloid (Jansen et al., 2005).

Our initial results indicate that care must be taken on the definition of the restraints driving the docking. Distance restraints calculated by EPR have a high ambiguity associated with them, which is inherent to the technique itself, since determined distances come from nitroxide groups attached to flexible side chains of Cys residues. The distance bounds for these inter-nitroxide distances are much larger than the ambiguous distance bounds. In order to account for this data uncertainty, two docking runs were performed in order to check the reliability of the available data, ultimately responsible for the docking driven

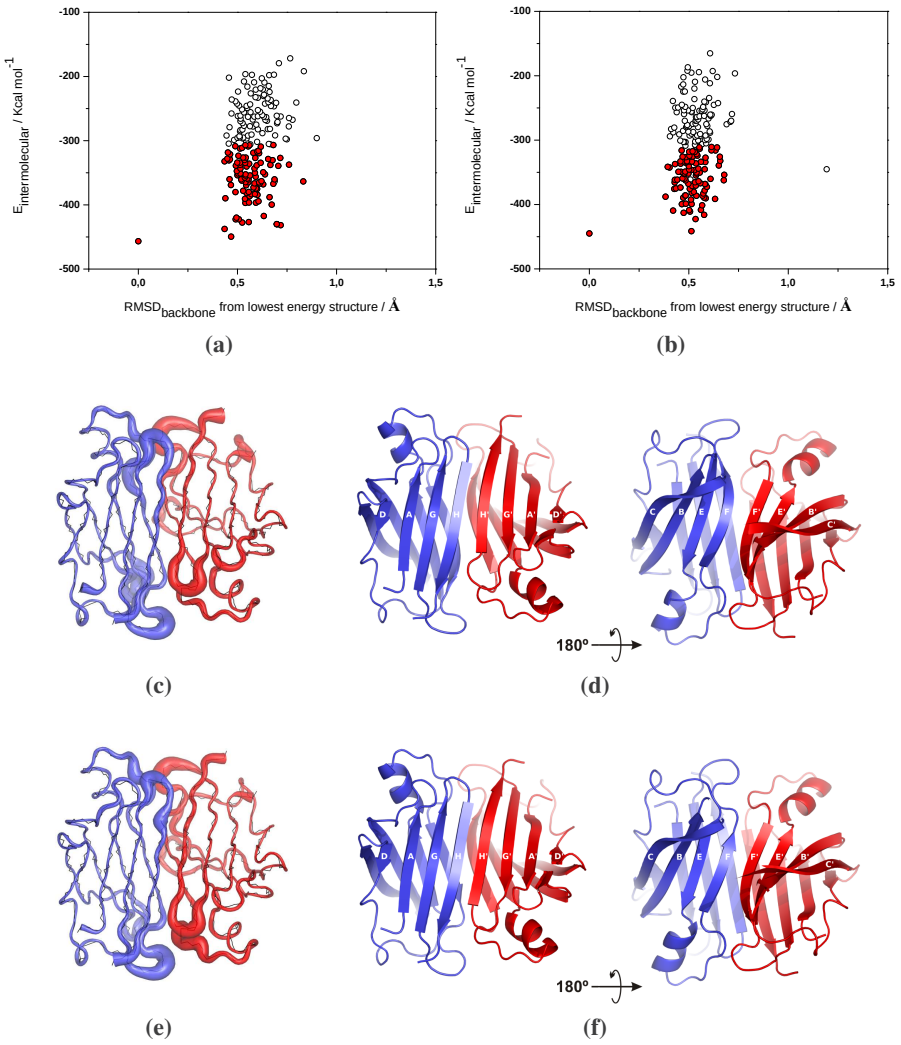


Figure 6.5: Near-native dimers resulting from the docking procedure, using NMR H/D exchange data (b, e and f) and additional EPR data included as unambiguous restraints (a, c and d). **(a)** and **(b)** Intermolecular energies as a function of the backbone RMSD from the lowest energy structure for the 200 docked structures. The 100 best dimers are represented with filled circles. **(c)** and **(e)** Sausage representation of the 10 best dimer complexes. The tube thickness is directly proportional to the RMS fluctuation of the ensemble, after superimposing the structures on the backbone atoms of the flexible interface. The average backbone structure is represented by lines. **(d)** and **(f)** Schematic view of the dimer complex with the lowest intermolecular energy, with the identification of β-strands. Figures c–f were produced with PyMOL (DeLano, 2002).

protocol. We performed a HADDOCK run only with ambiguous interaction restraints, NearNI-epr, and another with the same set of AIRs plus EPR distance restraints, NearNI+epr, having discarded EPR distances from residues 85 and 89 (table 6.1).

For each docking run, 2000 rigid body docking structures were first generated and the best 200 structures in terms of intermolecular energy were subjected to a semi-flexible simulated annealing procedure and final refinement in water. Figures 6.5a and 6.5b show the intermolecular energy E_{interm}as a function of the backbone root mean square deviation from the lowest energy structure for the 200 structures obtained after water refinement. One single cluster with 198 structures could be identified based on pairwise backbone RMSD values using a 1.0 Å cutoff, for both docking runs (with and without EPR distance restraints). The ten lowest energy structures of these clusters were selected for further analysis. Statistics on these structures are summarized in table 6.2.

The docking results between the two runs are similar, though a slightly smaller structural dispersion on the RMSD values of the cluster solutions for NearNI-epr is noticed (table 6.2). The introduction of EPR data should reduce the degrees of freedom of the system, as it occurs when restraints are provided to a NMR structure calculation. From figures 6.5a,b we observe that a faint higher conformational freedom is given by EPR restraints to the NearNI+epr than to NearNI-epr docking run. This is consistent with the fact that EPR distances have, relatively to NMR H/D exchange data, larger distance boundaries. Since the weight given by HADDOCK in this study for both types of restraints was the same, instead of obtaining less structural dispersion we get just the opposite, though is not very relevant when comparing NearNI+epr with NearNI-epr runs. The average energy of the 10 lowest energy structures is -433 and -418 Kcal mol^{-1} for NearNI+epr and NearNI-epr, respectively. For both docking runs, the electrostatic contribution E_{elect} represents the major contribution to E_{interm} being about the double than van der Waals contribution E_{vdw}, as shown on table 6.2. The restraints term $E_{restraints}$ is negligible, consistent with a very low number of restraint violations per structure (see appendix table A.8, page 202). The structure of the complexes obtained are rather well defined, with an overall backbone RMSD from the mean structure of about 0.4 Å (both NearNI+epr

Table 6.2: Statistical analysis of the 10 lowest energy structures of each cluster after clustering of HADDOCK results.

Docking Run	C*	N†	RMSD‡	E_{interm}¶	E_{vdw}§	E_{elec}‖	$E_{restraints}$**	BSA††
NearNI+epr	1	198	0.51 ± 0.20	-433 ± 11	-135 ± 6	-298 ± 13	0.03 ± 0.02	2526 ± 67
NearNI-epr	1	198	0.47 ± 0.16	-418 ± 14	-133 ± 4	-285 ± 13	0.03 ± 0.02	2499 ± 60
NonNI+epr	1	108	0.46 ± 0.17	-560 ± 9	-81 ± 4	-484 ± 8	5.07 ± 0.14	2189 ± 60
	2	78	9.99 ± 0.12	-278 ± 8	-68 ± 6	-218 ± 8	7.57 ± 1.30	1572 ± 31
NonNI-epr	1	121	0.75 ± 0.27	-504 ± 8	-67 ± 4	-442 ± 9	4.65 ± 0.75	1836 ± 44
	2	41	1.11 ± 0.07	-455 ± 30	-67 ± 5	-394 ± 29	5.57 ± 1.97	1832 ± 79
	3	31	0.68 ± 0.05	-459 ± 18	-65 ± 6	-400 ± 19	6.32 ± 1.60	1822 ± 58

* Cluster identification; † Number of structures per cluster;
‡ Average backbone RMSD (Å) from the lowest energy structure of all calculated structures;
¶ Average intermolecular total energy, $E_{interm} = E_{vdw} + E_{elec} + E_{restraints}$. All energy terms are expressed in Kcal mol^{-1}; § Average intermolecular Van der Walls energy; ‖ Average intermolecular electrostatics energy; ** Average distance restraints energy; †† Average buried surface area in $Å^2$.

and NearNI-epr) and a good stereochemical quality with more than 99% of the residues in the most favored and allowed regions of the Ramachandran plot (see table 6.3). The ensemble of the 10 lowest energy structures for the obtained cluster is depicted in figures 6.5c,e.

The average buried surface area of the 10 lowest energy structures is about 2500 $Å^2$ for both clusters, indicating that residues on the dimeric interface are closely packed. The BSA of the native TTR dimer is about 1770 $Å^2$ and among all protein-protein complexes available in the PDB, the average BSA is 1600 $\pm$ 400 $Å^2$ (Conte et al., 1999). A detailed view of the modeled dimer reveals that the interface between two native monomers is predominantly stabilized by an extended network of conserved intermolecular hydrogen bonds, involving either hydrophobic and charged amino-acids, or hydrophobic and polar amino-acids, as expected on the basis of a large electrostatic contribution to the overall intermolecular energy. Intermolecular statistics calculated over the ensemble of the 10 lowest energy structures on the cluster obtained is shown in table 6.4. A salt bridge between Lys76 and Asp99 is formed, though present only on the NearNI+epr run. Conserved non-bonded contacts are also present on the interface as expected, since the E_{vdw} energy term also strongly contributes to the E_{interm} (data not shown).

The dimer topology obtained for both NearNI+epr and NearNI-epr dockings is similar. Figures 6.5d,f illustrate the secondary structure for the dimer of lowest intermolecular energy in each docking run. These near-native structures have their near-native monomers in an anti-parallel edge-to-edge arrangement on strands *FF'* and *HH'* which is in agreement with published data (Serag et al., 2001, 2002). This result was expected since most of the applied restraints were obeyed during the HADDOCK runs. A more careful analysis of the hydrogen-bonding patterns on the NearNI indicates a slip in β-strands register (table 6.4). The results show that the two *F* strands are centered about position 93 instead of 91/92 as in the native structure, and the *H* strands are centered near residue 118, instead of residue 117 (see appendix figure A.1, page 205). Olofsson et al. (Olofsson et al., 2004) pointed out that these results suggest a strengthening in the subunit assembly, since there's a more complementary alignment between subunit β-sheets.

Table 6.3: Statistical analysis of the 10 lowest energy structures of each cluster after clustering of HADDOCK results.

	NearNI		NonNI	
	NearNI+epr	NearNI-epr	NonNI+epr	NonNI-epr
Backbone RMSD (Å) with respect to mean				
Flexible interface backbone*	0.42 ± 0.13	0.43 ± 0.14	0.41 ± 0.18	0.57 ± 0.17
All backbone	0.44 ± 0.13	0.46 ± 0.16	0.54 ± 0.31	0.61 ± 0.29
Number of ambiguous interaction restraints (AIRs)				
From native monomer	11	11		
From non-native monomer			9	9
Total AIRs	22	22	18	18
Number of unambiguous interaction restraints				
Between native monomers	4			
Between non-native monomers			29	26
RMSD from idealized covalent geometry				
Bonds (Å)	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000
Angles (°)	0.389 ± 0.006	0.392 ± 0.005	0.385 ± 0.005	0.389 ± 0.005
Impropers (°)	0.379 ± 0.012	0.379 ± 0.013	0.390 ± 0.005	0.392 ± 0.012
Ramachandran analysis				
Residues in the favoured region (%)	85.0 ± 0.9	84.5 ± 1.7	82.9 ± 3.5	88.2 ± 2.0
Residues in additionally allowed regions (%)	14.8 ± 0.8	15.4 ± 1.7	17.1 ± 3.5	11.8 ± 2.0
Residues in generously allowed regions (%)	0.1 ± 0.3	0.1 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
Residues in disallowed regions (%)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Structural statistics of the 10 lowest energy structures (cluster 1) are based on structures obtained after flexible docking with HADDOCK followed by refinement in explicit water using ambiguous interaction restraints derived from H/D exchange data (NearNI-epr, NonNI-epr runs), and unambiguous interaction restraints derived from EPR data (NearNI+epr, NonNI+epr runs). Hydrogen bonds restraints were also applied on the non-native monomers (see Computational Methods); * The flexible interface is defined from the active and passive residues used in the definition of the ambiguous interactions restraints ± 2 sequential residues.

Table 6.4: Intermolecular contacts statistics calculated over the ensemble of the 10 best docked structures of the cluster with lowest mean intermolecular energy

Interacting Residues				Hydrogen-Bonds		
Donor		Acceptor				
Residue	Chain	Residue	Chain	M-M	M-S	S-S
A. Common contacts between runs NearNI+epr and NearNI-epr*						
Glu 92	A	Val 94	B	10/9		
Val 94	A	Glu 92	B	10/10		
Arg 103	A	Tyr 114	B			9/10
Thr 118	A	Thr 118	B	10/10		
Ala 120	A	Tyr 116	B	10/10		
Val 122	A	Tyr 114	B	6/7		
His 88	B	Thr 96	A		6/7	
Glu 92	B	Val 94	A	10/10		
Val 94	B	Glu 92	A	10/10		
Arg 103	B	Tyr 114	A			10/10
Tyr 116	B	Ala 120	A	6/5		
Thr 118	B	Thr 118	A	10/10		
Ala 120	B	Tyr 116	A	10/10		
B. Specific contacts for NearNI+epr run						
Lys 76	A	Asp 99	B			5
His 88	A	Thr 96	B		6	
Lys 76	B	Asp 99	A			5
Val 122	B	Ser 115	A		5	
C. Specific contacts for NearNI-epr run						
Tyr 116	A	Ala 120	B	5		
Val 122	B	Tyr 114	A	7		
D. Common contacts between runs NonNI+epr and NonNI-epr**						
Arg 104	B	Asp 18	A		9/5	10/6
E. Specific contacts for NonNI+epr run						
Leu 12	A	Val 16	B	7		
Val 14	A	Val 14	B	10		
Val 16	A	Leu 12	B	7		
Val 32	A	Val 30	B	10		
Arg 34	A	Asp 74	B			5
Arg 104	A	Asp 18	B		9	9
Leu 12	B	Val 16	A	6		
Val 14	B	Val 14	A	10		
Lys 15	B	Met 13	A			5
Val 16	B	Leu 12	A	6		
Val 32	B	Val 30	A	10		

Continued on Next Page. . .

Table 6.4: (continued)

Interacting Residues						
Donor		Acceptor		Hydrogen-Bonds		
Residue	Chain	Residue	Chain	M-M	M-S	S-S
F. Specific contacts for NonNI-epr run						
Leu 12	A	Tyr 78	B		7	
Val 16	A	Val 14	B	9		
Val 30	A	Val 30	B	10		
Val 32	A	Val 28	B	7		
Arg 104	A	Ala 19	B		6	
Leu 12	B	Tyr 78	A		10	
Val 16	B	Val 14	A	9		
Val 28	B	Val 32	A	7		
Val 30	B	Val 30	A	9		
Val 32	B	Val 28	A	9		
Arg 104	B	Asp 18	A			5

Intermolecular contacts were analyzed with DIMPLOT (Wallace et al., 1995) and are reported if present in at least 5 of the 10 best structures. The occurrence of main-chain/main-chain (M-M), main-chain/side-chain (M-S), and side-chain/side-chain (S-S) hydrogen bonds is reported, over the ensemble of the 10 lowest energy structures (cluster 1). Chain A and B correspond to the two monomers involved in the docking procedure; *The occurrence is reported as NearNI+epr/NearNI-epr; **The occurrence is reported as NonNI+epr/NonNI-epr.

While there is clearly a dominant hydrogen bond pattern between strands *HH'* on the ensemble of 10 near-native complexes (table 6.4), the interface between strands *FF'* has a reduced amount of hydrogen bonds. Though the propensity for two atoms to form hydrogen bonds is strongly dependent on geometric parameters, namely angles and distances between atoms, the hydrogen bond patterns observed indicate a better alignment between strands *HH'* than strands *FF'*. In native TTR, there are six main-chain to main-chain intermolecular hydrogen bonds between the adjacent *H* strands and neighboring residues, whereas the two *F* strands are separated more widely, and as a result, only two hydrogen bonds (between E89-V94' and V94-E94') are formed directly through main-chain interactions. The remaining hydrogen bonding interactions involve water bridges (figure A.1). Due to the slip in the strands register in the NearNI, the E89-V94 single hydrogen bond is missing, being substituted by a double hydrogen-bond between E92-V94. If we take into

consideration that the stability of neighboring β-strands is mostly accomplished by inter-strand hydrogen bonds, and hydrophobic complementarity on residue side-chains, then the results suggest a more stable interface between *FF'* on the NearNI than the same interface in the native dimer. Furthermore, we cannot rule out the presence of water molecules lying in between the *FF'* strands on the protofilaments, increasing the interface stability, as it happens in the native protein.

6.3.3 Modeling TTR non-native dimers

The NonNI was modeled by docking two partially disrupted monomers, as described in the methods section. These monomers were docked at their NonNI (strands *AA'* and *BB'*) producing an ensemble of non-native complexes (figure 6.6). As in the case of the near-native dimers, a NonNI+epr run (with EPR data) and NonNI-epr run (without EPR data) were performed (table 6.1).

Serag et al. (2002) built a total of 14 TTR single site cysteine mutants, 7 of which could be satisfactorily modified with spin labels and still form amyloid fibrils. The authors focused their attention on the strands near the edge of the native intermolecular β-sheets (strands *B* and *C*), that upon self-assembly give rise to an extended β-sheet. Ultimately, five of the mutants (29C, 31C and 33C on strand *B*; 40C located on the loop connecting strands *B* and *C*; and 46C on strand *C*) showed EPR changes upon fibril formation and, therefore, were informative in the characterization of the NonNI. From these mutants, 5 EPR distances were determined but only 3 of them were used as restraints on the HADDOCK runs. Preparatory HADDOCK runs were not successful in terms of structure convergence when using 40C inter-strand distance restraint, probably due to the highly ambiguity associated to it. In the case of mutant 46C, because strand *C* was removed when building the partially disrupted non-native monomer, we did not consider this distance restraint.

The same docking protocol used on the NearNI modeling was applied to model the NonNI, in which 2000 rigid body docking structures were generated and the best 200 in terms of intermolecular energy were further subjected to a semi-flexible simulated annealing procedure followed by a final refinement in a water shell. Figures 6.6a,b shows the intermolecular energy as a function of

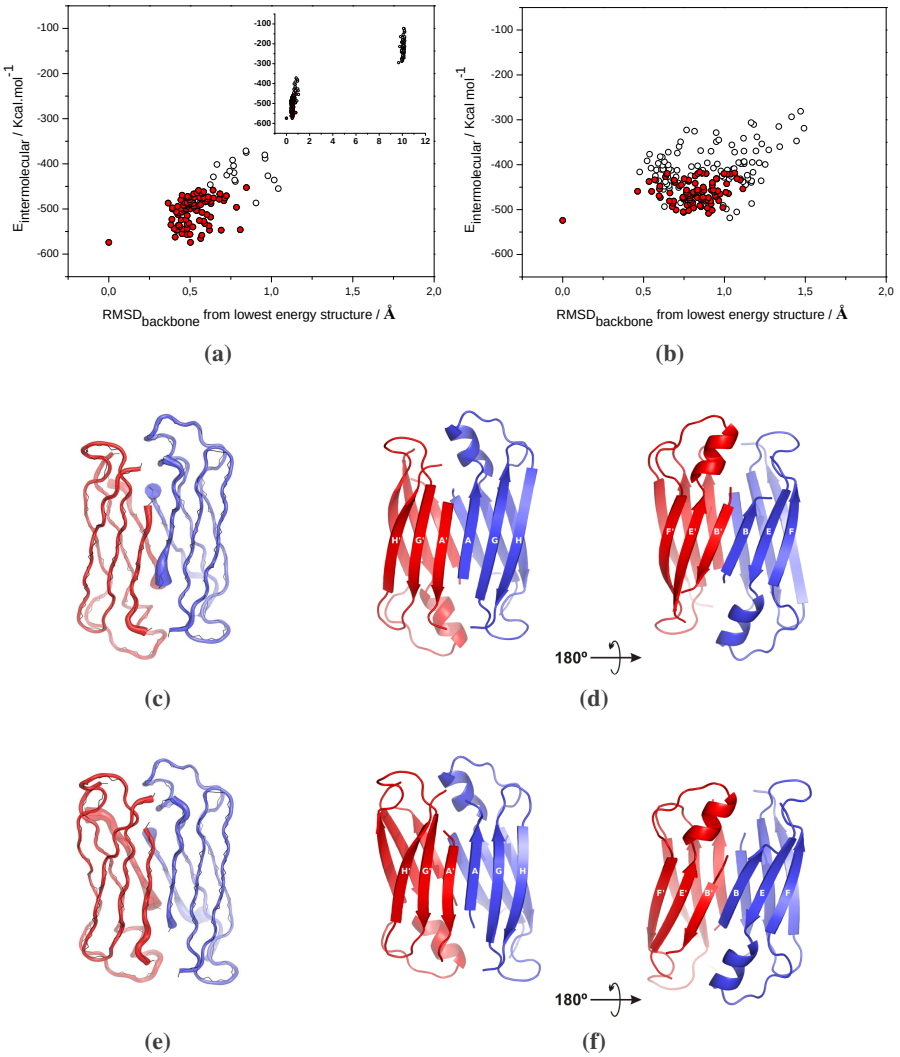


Figure 6.6: Non-native dimers resulting from the docking procedure, using NMR H/D exchange data (b,e and f) and additional EPR data included as unambiguous restraints (a,c and d). **(a)** and **(b)**, Intermolecular energies as a function of the backbone RMSD from the lowest energy structure for the 200 docked structures. The 100 best dimers for the cluster with the lowest intermolecular energy average are represented with filled circles. **(c)** and **(e)**, Sausage representation of the 10 best dimer complexes of the lowest energy cluster. The tube thickness is directly proportional to the RMS fluctuation of the ensemble, after superimposing the structures on the backbone atoms of the flexible interface. The average backbone structure is represented by lines. **(d)** and **(f)**, Schematic view of the dimer complex with the lowest intermolecular energy, with the identification of β-strands. Figures c–f, were produced with PyMOL (DeLano, 2002).

the backbone RMSD from the lowest energy structure for the 200 structures obtained after water refinement.

In the case of the NonNI modeling using the EPR distance restraints, two clusters could be identified based on the pairwise backbone RMSD values, using a 1.0 Å cutoff. The lowest energy cluster ($E_{interm} = -560$ Kcal mol^{-1}) containing 108 structures, best satisfies the ambiguous (AIRs) and unambiguous restraints initially imposed, and shows the largest average BSA at the interface (table 6.2). The average RMSD values of the ten lowest energy structures of clusters 1 and 2 are 0.46 and 9.99 Årespectively, indicating that the structures in cluster 1 are remarkably different from those in cluster 2. Additionally, the difference on the average values of intermolecular energies for clusters 1 and 2, -560 and -278 Kcal mol^{-1}, indicates that the structures in cluster 1 have the most favorable intermolecular contacts. The average values of BSA for these two clusters are 2189 and 1572 Å^2 respectively, indicating that structures on cluster 1 have a better subunit interface. The ensemble of the ten lowest energy structures for cluster 1 is depicted in figure 6.6c. The spatial distribution of representative docking solutions is shown in appendix figure A.2, page 206, where the centers of mass of the docking moving partner (chain B) from the lowest energy structure of each cluster are superimposed on the docking static partner (chain A). It is evident that the interacting surface of non-native monomers is confined to a single patch on strands *A* and *B* of the protein, in spite of the statistical results suggesting a greater conformational change. In fact, the major differences between the non-native interfaces modeled by HADDOCK are in a rotation of one of the non-native monomers in respect to the other in the dimer, along a direction perpendicular to the β-strands.

Regarding the NonNI-epr docking run, 3 clusters of structures were obtained. The values of the intermolecular energies plotted as a function of the backbone RMSD from the lowest energy structure calculated over the 10 best structures of each cluster (figure 6.6b), show that the docking solutions have nearly equal E_{interm} energies, about -500 Kcal mol^{-1} (table 6.2). Furthermore, similar average values for BSA, 1836, 1832 and 1822 Å^2, respectively for cluster 1, 2 and 3, and comparable average RMSD values to the lowest energy structure obtained indicate that the docking solutions display analogous con-

formations. The spatial distribution of the representative docking solutions is shown in figure A.2. There is again indication that there is only one main patch site for the interface interaction between docked non-native monomers. An ensemble of structures for the NonNI dimers modeled is depicted in figure 6.6e. The ten lowest energy structures for the most populated clusters of both NonNI+epr and NonNI-epr runs (lowest average E_{interm}) were selected for further analysis. When the 10 structures are superimposed on the backbone of the mean structure, the RMSD values of all backbone atoms are 0.54 and 0.61 Å, for NonNI+epr and NonNI-epr runs, respectively (table 6.3). The relative small RMSD values indicate that the dimer structure is rather well defined. Moreover, about 99% of the residues are in the most favored and allowed regions of the Ramachandran plot, suggesting a good quality of the structural models (table 6.3). A more careful characterization of the modeled NonNI, indicates that the interface between two non-native monomers is predominantly stabilized by an extended network of intermolecular hydrogen bonds, as observed for the NearNI (table 6.4). There are no conserved specific hydrogen bonds when comparing both NonNI+epr and NonNI-epr runs, with the exception for a salt bridge formed between R104 and D18. In both dockings the E_{elect} contribution is around 6 times the E_{vdw} (table 6.2), which is reflected on the larger number of salt bridges established in the NonNI. The non-bonded contacts are also poorly conserved on both NonNI runs, and contribute in a low extent to the stabilization of the interface (data not shown).

The introduction of EPR data allows for a broader conformational space search during the semi-rigid body docking step, as already explained for the NearNI modeling. This explains why we observe a larger structural dispersion on the NonNI+epr docking compexes relatively to NonNI-epr dimers. The dimer topology obtained for both NonNI+epr and NonNI-epr NonNI docking runs is illustrated in figures 6.6d,f. The non-native structures have their non-native monomers in an anti-parallel edge-to-edge arrangement in strands AA' and BB'. The hydrogen bonding patterns on the NonNI (table 6.4) indicate a center of symmetry on strands A (position 14 for NonNI+epr, and position 15 for NonNI-epr) and B (position 31 for NonNI+epr and position 30 for NonNI-epr) with a slip of strand register of one residue. Serag et al. (2002) proposed a

model where two *B* strands align in an anti-parallel manner centered on position 31, nonetheless the data collected did not allowed the authors to make any assumptions about the alignment of strands *AA'* in the non-native dimer.

6.3.4 Protofilament Polymorphism

Accordingly to HADDOCK philosophy and most of the protein docking software packages, the best produced dimer structure on a docking run typically has the lowest energy and belongs to the lowest energy cluster found. Though this is a general rule, a case by case study should be considered, since the scoring algorithm functions are still evolving (Tame, 2005). Preliminary results did showed us that in fact, the highest ranking solution for each docking run performed in this work, is not *per se* the structure with the best complementarity between the two docking partners (data not shown). The analysis of the 200 lowest energy dimer structures obtained after water refinement for each docking run (figures 6.5a,b and 6.6a,b), do not show a great structural variability within the same cluster (table 6.2). The 100 highest ranking solutions of the lowest energy cluster (cluster 1) in each docking run were chosen to build two TTR protofilaments sets, which allowed us to study the polymorphism of the obtained protofilaments not biasing the results. One of the protofilament sets is produced using NonNI+epr and NearNI+epr dimer structures (protofil+epr), while the other set uses dimers obtained from NonNI-epr and NearNI-epr docking runs (protofil-epr). The structural statistic analysis, the restraints violations and the intermolecular contacts for the 100 best solutions on cluster 1 of each docking run are presented has supplementary material (appendix tables A.9– A.11). The overall conclusions taken from the tables analysis are the same as the ones reported earlier for the analysis of the 10 best solutions.

An alignment scheme briefly reported elsewhere (Correia et al., 2006) was devised to build the protofilaments (figure 6.3). This multi-step protocol superimposes through out the alignment procedure, the same set of two structures: a non-native and a near-native dimer. Each pair of non-native/near-native dimers will produce a unique protofilament, thus 10,000 (100 non-native structures $\times$ 100 near-native structures) supramolecular structures were produced for each protofilament set. The non-native dimers were used as seeds to start building

the protofilaments, since the most recent experimental results support a protofilament structure formed by a core of two three-stranded β-sheets (Olofsson et al., 2004). The near-native dimers were used as template models to grow the resulting protofilaments. The alignment between these two types of dimer structures is done on the backbone of *EG* strands, because they have a lower conformational lability relatively to the interface regions, upon monomer-monomer docking.

The major goal while building the protofilament models was to make use of the gathered information from the docking runs, namely the relative orientation of the monomers in the obtained docked dimeric structures. This way, it is possible to build a supramolecular structure, in which the subunits are arranged in the same way as they were in the complexes obtained during the docking stage. Moreover, the new interfaces between the protofilament subunits will concomitantly maintain the interface chemical features attained on the dimer complexes. Notice that the interface NonNI in the protofilament, is exactly the same as in the non-native dimer. In contrast, protofilament NearNI interface is largely preserved relatively to the original NearNI in the near-native dimer used as a template model (figure 6.3). Though the NonNI in the protofilament is obtained immediately from the non-native dimer, the NearNI is obtained indirectly (see Computational Methods). In addition to that, the structural conformation of *EG* strands is slightly altered during the protein docking, particularly during the semi-flexible and water refinement steps ($\text{RMSD}_{backbone,EG} = 0.6$ Å). In spite of this, the similarity between the NearNI's is high if we compare the interface contacts, namely the hydrogen bonding patterns and the hydrophobic interactions (see appendix figure A.3, page 207). More than 60% of the protofilaments maintain all the main-chain/main-chain hydrogen bonds as in the near-native dimers used to build the protofilaments, whereas the hydrophobic contacts, at the residue level, have a mean similarity value around 75% in both protofil-epr and protofil+epr sets.

The obtained protofilaments have elongated structures with a helical topology, as previously reported (Correia et al., 2006). These structures were then clustered using the helical periodicity, producing characteristic structural families with different geometric properties (figure 6.7). The protofilaments from

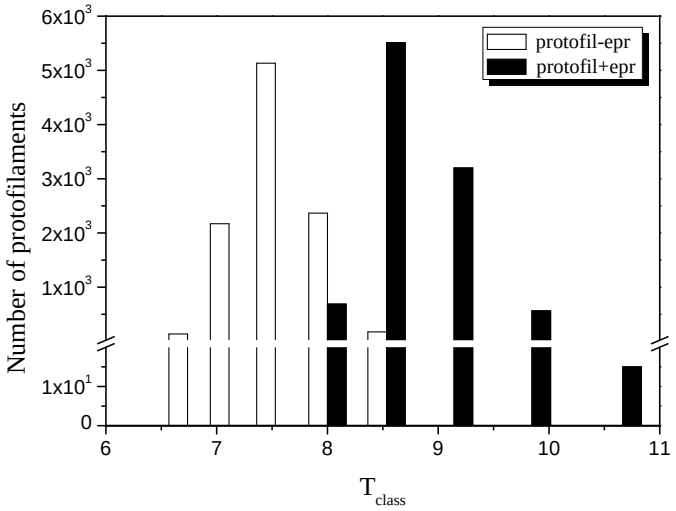


Figure 6.7: Histogram showing the distribution of the protofilaments according to their helical period T_{class}, for the two datasets protofil+epr and protofil-epr studied.

protofil-epr, have helical periods T_{class} in the range of 7 to 9 dimers being the most populated the $T_{class} = 7.5$. Protofilaments belonging to this T_{class} have around 15 monomers (45 β-sheets) per helical turn, with a subunit repeating distance of about 212 Å and a diameter of about 54 Å. The other set of protofilaments assembled with dimers obtained with AIRs and EPR distance restraints (protofil+epr), have helical periods in the range of 8 to 10 dimers, with $T_{class} = 8.5$ being the most populated one. Protofilaments in this T_{class} have around 17 monomers (51 β-sheets) per helical turn, a subunit repeating distance of about 240 Å and an average diameter of about 45 Å. More than 50% of the protofilaments are in these most populated clusters. Both sets of protofilaments show a Gaussian distribution in the helical periodicity clustering, showing that subtle structural differences on the interfaces are of paramount importance for the overall supramolecular geometry. Moreover, the initial set of restraints imposed during HADDOCK runs on the same system, gives slightly different results. This was expected, as the protofilament alignment scheme relies heavily on the subunit interfaces conformation obtained during the protein docking stage. Figure 6.8 depicts examples of protofilaments with distinct helical periods and cross-section diameters.

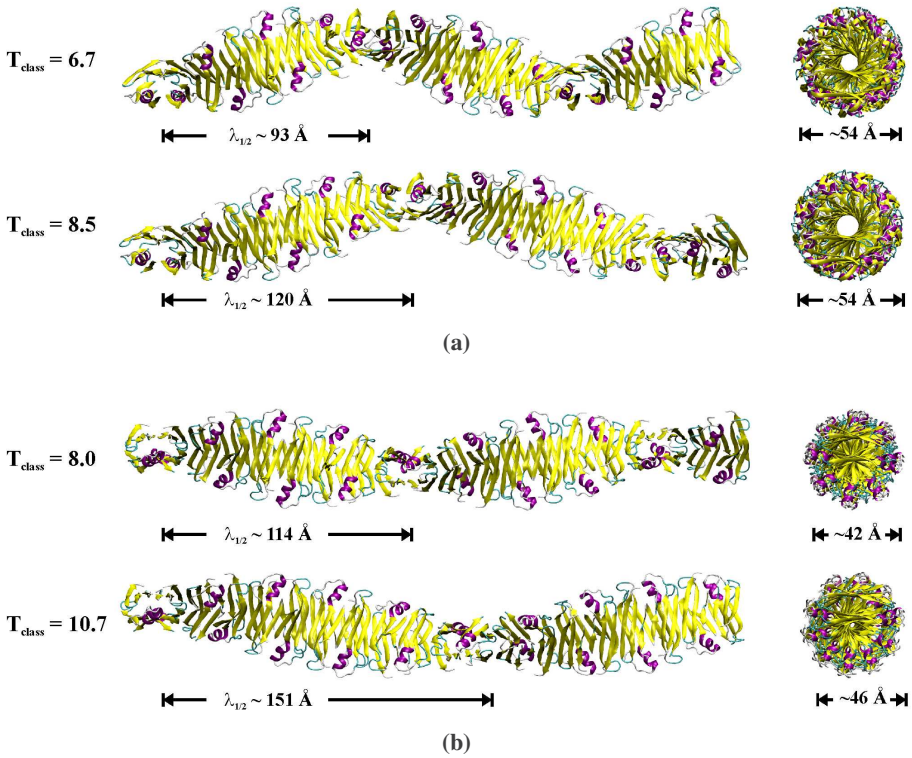


Figure 6.8: Schematic representation of TTR protofilament models obtained from the alignment of HADDOCK dimers produced using **(a)** only AIR restraints or **(b)** with additional EPR data, as a function of their helical periodicity T_{class}. The size of half of the repeating unit ($\lambda_{1/2}$), and the protofilament cross-section dimension including only the core β-strands are also shown. The protofilaments are colored by the secondary structure assignment determined with STRIDE (Frishman and Argos, 1995) (yellow, β-sheet; purple, α-helix; gray, coils), and are all at the same scale.

From the most representative T_{class} cluster for each protofilament set, we set out to choose the final protofilament structures. Since a correct β-sheet pairing between protein interfaces is indicative of good geometrical complementarities, we studied the number of hydrogen bonds in both near-native and non-native interfaces in the protofilaments of the most representative cluster (figure 6.9). The results are similar for both protofil+epr and protofil-epr protofilament groups, with the number of hydrogen bonds on both NearNI and NonNI interfaces ranging from 6 to 20. The Gaussian distribution of the number of hydrogen bonds per interface, show that while the protofilaments belong to the same

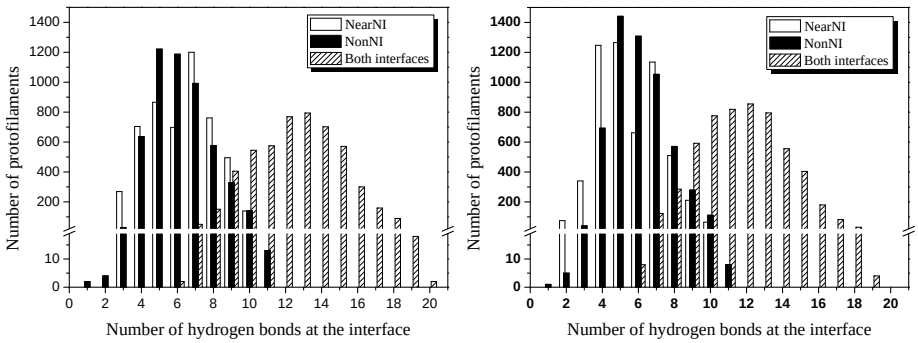


Figure 6.9: Characterization of intermolecular main-chain to main-chain hydrogen bonds on protofilament interfaces NonNI and NearNI. **(a)** Shows the results for the representative cluster of protofil+epr protofilament set ($T_{class} = 8.5$), whereas **(b)** shows the results for the representative cluster of protofil-epr set ($T_{class} = 7.5$).

structural family in terms of helical periodicity, the complementarities between the subunit interfaces are substantially different. The supramolecular structure with the largest number of hydrogen bonds on both interfaces, from the representative clusters characterized above, was chosen for further refinement.

6.3.5 High resolution molecular models of TTR protofilaments

The peptide fragments initially removed to create the TTR partially disrupted monomer (non-native monomer), were added to the two final protofilaments that maximize the number of hydrogen bonds established on protofilament's NearNI and NonNI (figure 6.4). These fragments, comprising β-strands *CD*, loops *AB* and *CD* as well as the terminus residues, were energy minimized and subjected to a short molecular dynamics run, without interfering with the subunit interfaces or the subunit core. A NearNI refinement was applied as soon as the fragments refinement was accomplished. This was necessary since the NearNI on the protofilament is not the same as the NearNI on the near-native dimer that was used to generate that specific protofilament, as explained earlier. Hence, it will ensure that any clashing atoms on the NearNI will be removed prior relaxation of the full protofilament structure. The non-native dimer obtained at the end of the protocol has the same core conformation as in

the beginning of the protofilament refinement scheme, plus the missing residues added and minimized, and the edge strands *AB* optimized. This non-native dimer was then used as the seed to produce a final protofilament with a full polypeptide sequence. Finally, to relax the protofilament structure, we carried out a short molecular dynamics simulation using the program NAMD (Phillips et al., 2005) running on a cluster of 100 commodity computers (Centopeia at the University of Coimbra). The aggregation state and global fold of the protofilament were not affected during the simulation. The overall stereochemical quality of the final structures was assessed with PROCHECK (Laskowski et al., 1993), revealing that $> 75\%$ of the residues are in the most favourable regions of the Ramachandran plot and the very few residues ($< 1\%$) in disallowed regions belong to exposed loops. See the computational methods for a more comprehensive description of the protocol.

In general terms, the representative protofilament of protofil+epr and protofil-epr sets is characterised by a linear fibrillar structure formed by two extended continuous β-sheets, *(BEFF'E'B')*$_n$ and *(AGHH'G'A')*$_n$, with β-strands more or less perpendicular to the main protofilament axis (appendix figure A.4, page 208). This particular trend of the β-strands is generally accepted as a characteristic of amyloid, the so-called cross-β structure (Eanes and Glenner, 1968). The strand edges of each protofilament subunit are also arranged in an anti-parallel manner, as previously reported by Serag et al. (2002) (figure A.4). Both protofilaments show a diameter of about 50 Å, and present an average helical twist of approximately 48 β-strands, i.e. 16 monomeric units per helical turn. A recent fibril modeling on 2-microglobulin, combining sequence and structure conservation analysis as well as protein docking techniques, suggest a helical β-sheet consistent with the cross-β sheet model (Benyamini et al., 2005). A full turn contains 18 monomer units or 52 β-strands, a noticeable similarity between two proteins that have the same native fold.

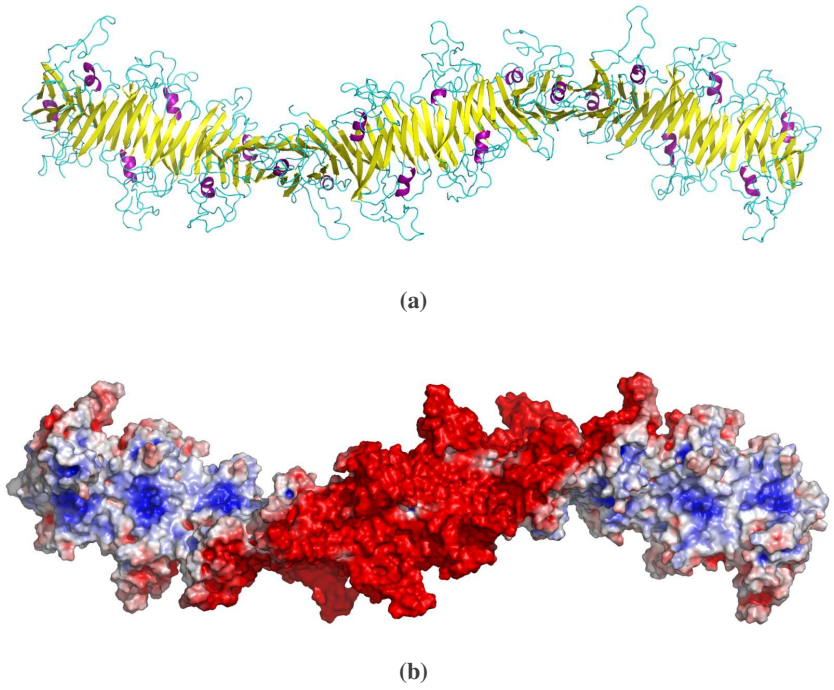
It is interesting to note that, based on X-ray fibre diffraction studies, Blake and Serpell (1996) proposed a model for TTR amyloid consisting of extended β-sheets with a helical twist and a fibre repeating unit of $\approx$ 115 Å, corresponding to 24 β-strands. It must be more than coincidental that our model, based on experimental data as well as energetic and shape complementarity criteria, has a

helical period roughly the double than Blake's model. In fact, it is possible that the X-ray diffraction pattern observed for the fibre could be the result of lateral association of protofilaments shifted by half period. If this is not the case, to build a protofilament with a period of 24 β-strands a much more twisted helical structure is required and a large conformational rearrangement of the β-sheets in the TTR subunit is necessary. Interestingly, our protofilament model shows surface segregation of charged residues in a helical arrangement, which could be responsible for the half-period pairing of protofilaments, in order to avoid electrostatic repulsions. It is clear from figure 6.10 that the outer part of the protofilament, mainly composed by loops of the former *CD* strands in the native TTR monomer, being more readily available to other protofilaments/oligomers present in solution, could have a great significance upon fibril assembly.

6.4 Conclusions

The protein-protein docking methodology employed in this work treats proteins as rigid bodies at the first stage of the docking protocol. Since no flexibility is allowed at this stage, no side-chain or loop rearrangements can occur which might lead to wrong initial orientations of the dimers obtained. Such orientations are generally not corrected during the subsequent HADDOCK semi-flexible molecular dynamic refinement stages (Dominguez et al., 2004). In order to overcome this problem, we could start from an ensemble of structures instead of a single molecule, since an ensemble of structures will reflect flexibility in loops and will show slight differences in side-chains orientation especially for solvent exposed residues. The different side-chains and/or backbone orientations allow a better sampling of all conformational possibilities during the rigid body docking step. An ensemble of structures could be generated applying a short molecular dynamics simulation, and extract the different structures from the trajectory produced. Nevertheless, the docking results obtained, solely based on the available experimental biochemical data in the literature, were promising.

Methodological improvements on the docking and alignment procedure described in this chapter can be envisaged with recent HADDOCK developments,



(a)

(b)

Figure 6.10: Electrostatic properties of a representative protofilament. Potential obtained by solution of the linearized Poisson-Boltzmann equation at 150 mM ionic strength with a solute dielectric of 2 and a solvent dielectric of 78.5. **(a)** Protofilament cartoon representation colored by the secondary structure assignment determined with STRIDE (Frishman and Argos, 1995) (yellow, β-sheet; purple, α-helix; cyan, coils). **(b)** Electrostatic potential mapped on the protofilament molecular surface; a blue color indicates regions of positive potential (> 5 kT/e), whereas red depicts negative potential values (< -5 kT/e). Figures were produced with PyMOL (DeLano, 2002).

namely multi-body docking (de Vries et al., 2007).

Our results show that docking two TTR subunits to recreate a non-native interface involving β-strands *A* and *B* requires full solvent exposure of these strands, otherwise, steric occlusion will prevent an energetically favorable interaction between docking partners. Though this is obvious, the non-native dimers obtained do support Serag's work with the non-native monomers in an anti-parallel edge-to-edge arrangement in strands *AA'* and *BB'*. Moreover, the NonNI interface is energetically favorable with a high buried surface area, stabilized by a network of hydrogen bonds and some salt bridges.

The near-native dimer topology obtained from NearNI modeling, shows also an anti-parallel edge-to-edge arrangement in strands *FF'* and *HH'*. A slip in β-strands register was observed on the NearNI, which suggests a strengthening in the subunit assembly as already pointed out by Olofsson et al. (2004). In fact, the NearNI has a higher average value of BSA than the observed for the NonNI. Cordeiro et al. (2006) also suggested that the alignment of β-sheets was altered upon fibrillization, though the native β-strands are preserved in the amyloid fibril of TTR. On the other hand, there is clearly a dominant hydrogen bonding pattern between strands *HH'*, and in a smaller extent on strands *FF'*. The native TTR also shows this hydrogen bonding profile, where strands *FF'* are stabilized by hydrogen bonding interactions involving water bridges. It is possible that upon protofilament assembly, water bridges could also play an important role on the stabilization of the interface between strand edges. This results could also suggest, that a more pronounced conformational change must occur in the TTR monomer, in order to get a better fitting between strands *FF'*.

The docking and alignment procedure generated a range of protofilament structures, in agreement with the structural polymorphism observed for amyloid fibrils and recently reported (Cardoso et al., 2002; Jansen et al., 2005). To date, to the best of our knowledge, this is the first time an all atom protofilament model is published for TTR. The protofilament structures proposed here have geometric properties in close agreement with the known characteristics of TTR amyloid. Several experimental studies have shown that TTR amyloid fibrils are formed by continuous β−sheet helices (Blake and Serpell, 1996) and protofilaments with diameters in the order of 40–60 Å, as revealed by electron microscopy and Xray fiber diffraction (Serpell et al., 1995; Sunde et al., 1997). Additionally, all the EPR distance restraints and NMR protection factors initially imposed are observed in the final structures, and the generated structures have good stereochemical properties.

Structural models have been built using information from a wide variety of techniques, including X-ray diffraction, electron microscopy, solid state NMR and EPR (Makin and Serpell, 2005). An early model of the protofilament structure arose from analysis of X-ray fiber diffraction patterns from *ex vivo* TTR-Met30 variant amyloid fibrils (Blake and Serpell, 1996), in which

the β-strands were hydrogen bonded to form a continuous β-sheet structure. In this model, four β-sheets are twisted around a central axis in a cross-β structure, though an anti-parallel α-helix model have also been suggested for amyloid formed from TTR . Models retaining largely the structure of the native monomeric form have also been proposed. A double helical arrangement of TTR monomers, where a three-residue β-slip is observed in one of the interface strands, allows the construction of an infinite β-sheet model in which native like β-strands are aligned (Eneqvist et al., 2000; Inouye et al., 1998). Distance restraints from site-directed spin labeling of TTR fibrils permitted the building of a head-to-head/tail-to-tail model, where the edge strands *C* and *D* are displaced (Serag et al., 2002). Although the interfaces in one sheet were revealed, the arrangement of the second sheet could not be studied. More recently, H/D NMR exchange studies do suggest that the protofilament structure is formed by a subunit core of two three-stranded β-sheets (Olofsson et al., 2004) in a native like conformation. Our modeling results do seem to support a two-sheet β-helix in a cross-β fold, based on the premise that the protofilament subunits are in a native like fold.

These models may be refined in the future by the introduction of other experimentally derived constraints, but at this stage may become a valuable instrument in the rational design of compounds with therapeutic potential to inhibit amyloid fibril formation by TTR.

Acknowledgments

We acknowledge the computer resources provided by the Advanced Computing Laboratory, Centopeia and Milipeia clusters, Department of Physics, University of Coimbra, and the support of Fundação para a Ciência e Tecnologia through grant SFRH/BD/1354/2000 (to N.L.-F.). The author wishes also to acknowledge Alexandre Bonvin and the Haddock team for the discussions about protein docking.

Chapter 7

Conclusions and perspectives

Tudo o que faço ou não faço
Outros fizeram assim
Daí este meu cansaço
De sentir que quanto faço
Não é feito só por mim.

Cansaço, Luis de Macedo

The universe is a lonely space, we easily get lost. As far as we yet know, there is only one place, an inconspicuous outpost of the Milky Way called Earth, that sustains life, and even it can be pretty grudging (Bryson, 2004). With a handful of spaced events forged by some mysterious force, atomic particles were bring together giving rise to complex molecules. This molecular cauldron does not have many different basic blocks, still the way they associate and work with each other are responsible for the multitude of organisms populating Earth.

Among such complex molecules, proteins are really unique in the sense that they play key roles in virtually all biological processes. These actions are achieved through an innate ability proteins have to recognize and interact with highly diverse molecules. This thesis is part of my naive journey through the protein world. After all, the understanding of molecular biophysics is no less a daunting task then to know what came out of the primeval atom. We can chronically get lost.

7.1 In conclusion

The study of protein association is nowadays of central importance, since protein-protein interactions are a common trend either under normal physiological conditions as well as in disease states. Consequently, the understanding and characterization of protein interactions are one of the major goals on modern biochemistry. In the last decades, nuclear magnetic resonance (NMR) provided many advances in this field, particularly through its ability to provide detailed information on binding of biomolecules, ranging from self-association to small-ligand binding. Thus, NMR coupled with sophisticated molecular modeling augur rapid advances in our understanding of this area of science. Not only do NMR techniques obviate any exogenous labeling problem, since most biological molecules are composed of NMR sensitive nuclei, but also have the potential to measure association *in vitro* under normal physiological conditions.

Chapter 2 focused on the application of a NMR technique using pulsed-field-gradients (PFG) to measure translational diffusion of proteins. PFG NMR is a powerful technique for studying the self-diffusion of proteins. In order to extract relevant information from PFG NMR experiments (as well as in any

other technique), both the sample and the machine should be well characterized. An introduction to the theoretical background was given, along with the experimental difficulties and available technical solutions to surpass them. Overall, the chapter is meant to be a primer on how the technique can be used to infer information from NMR active diffusion protein species in solution. Also, the idea of protein solutions being a complex polydisperse mixture is stressed out. This fact as strong implications whenever one tries to understand what is going on at the molecular level in a sample, for example when calculating the protein structure or modeling a kinetic model.

We assessed the secondary structure content of a cytolytic toxin excreted to the plasma by *Staphylococcus aureus*, in membrane-mimetic environments using NMR. The results on chapter 3 show that the conformation of δ-toxin is strongly dependent on experimental conditions. In fact, on its way to eukaryotic membranes, a wealth of chemical environments are available to the toxin. It it thus expected large conformational plasticity of such toxins, however the aggregation state should also play a major role, as discussed in a later chapter. The structure of δ-toxin in methanol was re-calculated and deposited in the Protein Data Bank. The toxin forms an amphiphatic α-helix in methanol, which is a characteristic trend of many lytic toxins. The authors do expect that the newly calculated structure will increase the pool of available data to scientists, such that new experiments can be devised to answer relevant questions, such as the mechanism by which toxins interact and lyse membranes.

Among the techniques available to study protein-membrane interactions, molecular dynamic (MD) simulations is becoming a standard protocol. A strong advantage of MD over experimental techniques such as NMR, is its ability to provide atomic detail on the system under study. A correct parameterization of the molecules on the simulated system is a pre-requisite to obtain relevant data. Chapter 4 aims to characterize a simple bilayer system often used by experimentalists, namely dimyristoylphosphatidylcholine (DMPC). We employed a force field parameterized to correctly simulate both protein and lipid molecules (Schuler et al., 2001). An immediate consequence is that it allow us the modeling of complex bio-systems such as the interaction of proteins with bilayers. The properties computed to characterize the DMPC bilayer show good

agreement with the available experimental data, re-enforcing the quality of the force field employed.

The conformational stability of δ-toxin and propensity for aggregation were assessed by MD simulations in different chemical environments, in chapter 5. The interaction of the toxin with a bilayer composed of DMPC molecules was also carried out. The simulations performed in methanol showed that the α-helical conformation of the toxin is largely maintained during simulations, whereas in water and dimethyl-sulfoxide (DMSO), there is loss of the initial secondary structure. It was also shown that δ-toxin has a high propensity to aggregate in water, whereas in methanol and DMSO, a complex dynamic behavior of formation and disruption of aggregation species is observed. Moreover, if the peptide in solution is aggregated, the α-helix is preserved for longer periods in contrast to the monomeric species. Regarding simulations of the toxin interaction with DMPC, it was shown that the bulky residue Trp plays a major role in the process. Also, the amphiphaticity of the helix and aggregation state seems to be important on the kinetics of peptide penetration.

Transthyretin (TTR) is a homotetrameric protein and is one of several proteins known to be involved in human amyloid diseases. It is believed that, in the process of amyloid formation, TTR dissociates to non-native monomeric units, which may act as the building blocks of the amyloid fibrils (Brito et al., 2003). A docking-and-alignment protocol was devised in chapter 6, in order to build amyloid protofilaments of TTR. The docking approach is driven by a combination of shape complementarity and energetic criteria, and uses constraints derived from experimental data obtained for the fibrillar state. Our modeling results do seem to support a two-sheet β-helix in a cross-β fold, based on the premise that the protofilament subunits are in a native like fold. This models may be refined in the future by the introduction of other experimentally derived constraints, but at this stage may become a valuable instrument in the rational design of compounds with therapeutic potential to inhibit amyloid fibril formation by TTR.

7.2 Perspectives

This thesis embraces both experimental and computational techniques used for studying protein biophysics. One approach without the other is *meaningless* since each of them has its own pitfalls and advantages. The next few lines summarize my perspectives regarding the goals of two projects, both with the aim of strengthening cooperation between different scientific fields. Consilience.

•

Driven by increasingly complex problems and propelled by increasingly powerful technology, today's science is as much based on computation, data analysis, and collaboration as on the efforts of individual experimentalists and theorists. But even as computer power, data storage, and communications continue to improve exponentially, computational resources are failing to keep up with what scientists demand on them (Foster, 2004). The Grid concept motivated by distributed computing infrastructures for advanced science is emerging. Coordinated resource sharing and problem solving is becoming truly a multi-institutional task.

NMR is a technique that allows determining three-dimensional (3-D) structures of biomacromolecules, and their complexes at atomic resolution. Knowledge of their 3-D structures is vital for understanding functions and mechanisms of action of macromolecules, and for rationalizing the effect of mutations. 3D structures are also important as guides for the design of new experimental studies and as starting point for rational drug design.

Processing data from NMR to obtain a 3D structure typically involves several steps. Importantly, specialized computer programs are available for each step, each with its own characteristics and often with its own data format. Processing of NMR data has thus become a task for specialists, who understand the data and the data formats, as well as the programs, their installation requirements and their proper usage. Furthermore, NMR data processing requires considerable data storage and computational resources. These factors together have thrown up a barrier for groups in life sciences to employ the full power of NMR. The e-NMR project was started as an European initiative to facilitate data processing, both to allow groups lacking the resources to add NMR to their

toolbox, as well as to allow dedicated NMR groups to shift their attention from standard tasks toward cutting-edge research (Loureiro-Ferreira et al., 2010; Wassenaar et al., 2010).

Protein folding is one of the unsolved paradigms of molecular biology, the understanding of which would provide essential insight into areas as diverse as the therapeutics of neurogenerative diseases or bio-catalysis in organic solvents. Protein folding simulations are time-consuming, data intensive and require access to supercomputing facilities. Once completed, few simulations are made publicly available hindering scientists from accessing data reported in publications, performing detailed comparisons, and developing new analytical approaches. The P-found project aims to create a distributed public repository for storing molecular dynamics simulations, particularly those concerned with protein folding and unfolding (Swain et al., 2010). It aims to provide the tools needed to support the comparison and analysis of the simulations and thus enable new scientific knowledge to be discovered and shared.

•

These perspectives on grid e-science almost conclude the thesis. The topic was mentioned due to my collaboration with both projects. This is where I am, and where I end.

Appendix A

General appendix

Source code

Tables

Figures

Source code A.1: Check-eddy pulse sequence

```
#ifndef LINT
static char SCCSid[] = "@(#)checkeddy.c 1.0 2003";
#endif

/*********************************************************************
checkeddy - Written by NLF & RCR

            Sequence to determine the minimum time necessary for
            eddy currents to dissipate.

            Based on: Price W.S. Concepts Magn. Reson. (1998) 10:197
*********************************************************************\

#include <standard.h>

pulsesequence()
{
   /* declare new variables */
   double g1,d2,gzlvl1;
   extern char gradtype[];

   /*get new variables */
   g1=getval("g1");
   d2=getval("d2");
   gzlvl1=getval("gzlvl1");

   /* equilibrium period */
   status(A);
   hsdelay(d1);

   /* --- gradient delay (d2) --- */
   status(B);
      rgradient('z',gzlvl1);    /* Turn gradient on */
   delay(g1);
      rgradient('z',0.0);       /* Turn gradient off */
   delay(d2);

   /* --- observe period --- */
   status(C);
   pulse(pw,oph);
}

/*********************************************************************

                     pw
                     ..
                     ||
Tx  _________________||FID.acq
          ...
Gz  _____|   |_________________
   |----|----------|------|
      A          B       C

Usage:
1 - choose the biggest gzlvl1 you will use
2 - choose d2 similar to the one you will use
3 - go
4 - choose the best spectrum with the minimum d2 value.

                   ||||||||          |||||||
                   _______          (o) (o)
                   \_/ \_/            O
                     o                \_/
                     \_/

*********************************************************************/
```

Table A.1: ^{1}H NMR assignments* of δ-toxin in CD_3OH at pH 3, 298 K

	H^N	H^α	H^β		others	
N-formyl					HCO	8.182
M1	8.599	4.431	$H^{\beta2/\beta3}$	2.096	$H^{\gamma2/\gamma3}$	2.636/2.581
A2	8.718	4.102	H^β	1.460		
Q3	8.479	4.027	$H^{\beta2/\beta3}$	2.171/2.085	$H^{\gamma2/\gamma3}$	2.451/2.369
D4	8.203	4.510	$H^{\beta2/\beta3}$	3.084/ 2.919		
I5	8.199	3.805	H^β	2.051	$H^{\gamma12/\gamma13}$	1.808/1.203
					$H^{\gamma2}$	0.964
					$H^{\delta1}$	0.911
I6	8.155	3.713	H^β	1.981	$H^{\gamma12/\gamma13}$	1.754/1.251
					$H^{\gamma2}$	0.963
					$H^{\delta1}$	0.879
S7	8.340	4.193	$H^{\beta2/\beta3}$	4.058/3.914		
T8	7.984	3.976	H^β	4.376	$H^{\gamma2}$	1.244
I9	8.308	3.751	H^β	2.037	$H^{\gamma12/\gamma13}$	1.851/1.210
					$H^{\gamma2}$	0.967
					$H^{\delta1}$	0.865
G10	8.634	3.888/3.754†				
D11	8.393	4.444	$H^{\beta2/\beta3}$	3.204/2.759		
L12	8.189	4.264	$H^{\beta2/\beta3}$	1.956/1.907		
V13	8.579	3.632	H^β	2.319	$H^{\gamma1}$ (R)	1.004
					$H^{\gamma2}$ (S)	1.143
K14	8.223	3.913	$H^{\beta2/\beta3}$	2.024/2.001	$H^{\gamma2/\gamma3}$	1.466
					$H^{\delta2/\delta3}$	1.720
					$H^{\varepsilon2/\varepsilon3}$	2.960
					H^ζ	7.764
W15	8.434	4.335	$H^{\beta2/\beta3}$	3.633/3.441	$H^{\delta1}$	7.106
					$H^{\varepsilon1}$	10.233
					$H^{\varepsilon3}$	7.547
					$H^{\zeta2}$	7.342
					$H^{\zeta3}$	6.985
					$H^{\eta2}$	7.099
I16	8.743	3.542	H^β	2.181	$H^{\gamma12/\gamma13}$	1.220
					$H^{\gamma2}$	0.948
I17	8.575	3.590	H^β	1.982	$H^{\gamma12/\gamma13}$	1.856/1.196
					$H^{\gamma2}$	0.950
					$H^{\delta1}$	0.851
D18	8.780	4.394	$H^{\beta2/\beta3}$	3.037/ 2.709		
T19	8.103	3.732	H^β	4.096	$H^{\gamma2}$	1.015
V20	8.486	3.696	H^β	2.210	$H^{\gamma1}$ (R)	0.977
					$H^{\gamma2}$ (S)	1.077
N21	8.367	4.533	$H^{\beta2/\beta3}$	2.897/2.748	$H^{\delta21/\delta22}$	7.638/6.965
K22	7.970	3.990	$H^{\beta2/\beta3}$	1.830/1.700	$H^{\gamma2/\gamma3}$	1.456
					$H^{\delta2/\delta3}$	1.552
					$H^{\varepsilon2/\varepsilon3}$	2.807
					H^ζ	7.760
F23	7.994	4.524	$H^{\beta2/\beta3}$	3.302/3.077	$H^{\delta1/\delta2}$	7.378
					$H^{\varepsilon1/\varepsilon2}$	7.250
					H^ζ	7.220
T24	7.792	4.308	H^β	4.284	$H^{\gamma2}$	1.320
K25	8.108	4.253	$H^{\beta2/\beta3}$	1.902	$H^{\gamma2/\gamma3}$	0.998/0.943
					$H^{\delta2/\delta3}$	1.681/1.511
					$H^{\varepsilon2/\varepsilon3}$	2.960
					H^ζ	7.745
K26	8.124	4.414	$H^{\beta2/\beta3}$	1.948/1.800	$H^{\gamma2/\gamma3}$	1.487
					$H^{\delta2/\delta3}$	1.658
					$H^{\varepsilon2/\varepsilon3}$	2.957
					H^ζ	7.668

* Chemical shift values are in ppm. † $H^{\alpha2/\alpha3}$.

Table A.2: ^{1}H NMR assignments* of δ-toxin in 75% CD_3OH / 25% D_2O at pH 3, 303 K

	H^N	H^α	H^β		others	
N-formyl					HCO	8.184
M1	8.541	4.475	$H^{\beta2/\beta3}$	2.079	$H^{\gamma2/\gamma3}$	2.594
A2	8.623	4.147	H^β	1.452		
Q3	8.418	4.088	$H^{\beta2/\beta3}$	2.108	$H^{\gamma2/\gamma3}$	2.401
D4	8.219	4.551	$H^{\beta2/\beta3}$	3.022/ 2.958		
I5	8.067	3.913	H^β	1.997		
I6	8.055	3.778				
S7	8.248	4.264	$H^{\beta2/\beta3}$	4.030/3.930		
T8	7.961	4.036	H^β	4.358	$H^{\gamma2}$	1.253
I9	8.284	3.790	H^β	2.014	$H^{\gamma2}$	0.960
					$H^{\gamma12/\gamma13}$	1.613
G10	8.477	3.930/3.766†				
D11	8.254	4.469	$H^{\beta2/\beta3}$	3.157/2.782		
L12	8.102	4.276	$H^{\beta2/\beta3}$	1.903	$H^{\delta1/\delta2}$	0.995/0.930
V13	8.483	3.631	H^β	2.272	$H^{\gamma1}$ (R)	0.995
					$H^{\gamma2}$ (S)	1.124
K14	8.166	3.877	$H^{\beta2/\beta3}$	2.026		
W15	8.254	4.346	$H^{\beta2/\beta3}$	3.649/3.415	$H^{\delta1}$	7.147
					$H^{\varepsilon1}$	10.129
					$H^{\varepsilon3}$	7.563
					$H^{\eta2}$	7.118
					$H^{\zeta2}$	7.387
					$H^{\zeta3}$	6.989
I16	8.658	3.467	H^β	2.149		
I17	8.465	3.579	H^β	1.962	$H^{\gamma2}$	0.936
					$H^{\gamma12/\gamma13}$	1.801/1.198
D18	8.682	4.393	$H^{\beta2/\beta3}$	2.969/2.706		
T19	8.026	3.766	H^β	4.001	$H^{\gamma2}$	0.971
V20	8.371	3.713	H^β	2.290	$H^{\gamma1}$ (R)	0.971
					$H^{\gamma2}$ (S)	1.048
N21	8.242	4.540	$H^{\beta2/\beta3}$	2.881/2.764		
K22	7.850	4.018	$H^{\beta2/\beta3}$	1.792/1.669	$H^{\gamma2/\gamma3}$	1.364/1.194
					$H^{\delta2/\delta3}$	1.551
					$H^{\varepsilon2/\varepsilon3}$	2.840
F23	7.926	4.569	$H^{\beta2/\beta3}$	3.309/3.040		
T24	7.739	4.317	H^β	4.276	$H^{\gamma2}$	1.288
K25	8.055	4.299	$H^{\beta2/\beta3}$	1.880	$H^{\gamma2/\gamma3}$	1.499
					$H^{\delta2/\delta3}$	1.692
K26	8.155	4.393	$H^{\beta2/\beta3}$	1.932/1.780	$H^{\gamma2/\gamma3}$	1.481
					$H^{\delta2/\delta3}$	1.674
					$H^{\varepsilon2/\varepsilon3}$	2.975

* Chemical shift values are in ppm. † $H^{\alpha2/\alpha3}$.

Table A.3: ^{1}H NMR assignments* of δ-toxin in 65% CD_3OH / 35% D_2O at pH 3, 303 K

	H^N	H^α		H^β	others	
N-formyl					HCO	8.172
M1	8.511	4.480	$H^{\beta2/\beta3}$	2.053	$H^{\gamma2/\gamma3}$	2.598
A2	8.562	4.153	H^β	1.438		
W15	8.184	4.348	$H^{\beta2/\beta3}$	3.626/3.393	$H^{\delta1}$	7.169
					$H^{\varepsilon1}$	10.106
					$H^{\varepsilon3}$	7.558
					$H^{\eta2}$	7.123
					$H^{\zeta2}$	7.390
					$H^{\zeta3}$	6.981
F23	7.909	4.570	$H^{\beta2/\beta3}$	3.028	$H^{\delta1/\delta2}$	7.341
					$H^{\varepsilon1/\varepsilon2}$	7.271

* Chemical shift values are in ppm.

Table A.4: Three-bond $^{3}J_{HNH\alpha}$ coupling constants for δ-toxin in methanolic and DMSO solutions

	$^{3}J_{HNH\alpha}$ / Hz		
residue	100% CD_3OH	75% CD_3OH	DMSO
M1			
A2	4.2	4.7	6.5
Q3	5.2	5.4	7.0
D4	5.7	5.1	8.2
I5	5.8	5.8	11.4
I6	5.4	5.2	
S7	4.8	5.7	
T8	5.4	6.0	
I9	5.4	5.5	
G10			
D11	5.6	5.3	
L12	5.3	6.0	8.5
V13	4.8		
K14	5.4	5.9	
W15	5.1	7.2	
I16	5.1		
I17	4.6		
D18	4.9	5.2	
T19	6.0	4.4	
V20	5.7	4.0	
N21	5.3	4.6	
K22	5.9	6.1	11.5
F23	7.4	7.8	
T24	7.4	7.9	
K25	6.9	7.0	
K26	7.6	8.1	

Table A.5: ^{1}H NMR assignments* of δ-toxin in DMSO at 303 K

	H^N	H^α		H^β	others	
N-formyl					HCO	8.009
M1	8.317	4.421	$H^{\beta2/\beta3}$	1.904/1.769	$H^{\gamma2/\gamma3}$	2.431
A2	8.162	4.260	H^β	1.227		
Q3	8.050	4.198	$H^{\beta2/\beta3}$	1.846/1.705	$H^{\gamma2/\gamma3}$	2.071
D4	7.957	4.550	$H^{\beta2/\beta3}$	2.449/ 2.382		
I5	8.288	4.169	H^β	1.875		
T8	8.130	3.691	H^β	3.674	$H^{\gamma2}$	1.122
I9	7.804	4.450	H^β	2.472		
L12	7.585	3.888	$H^{\beta2/\beta3}$	1.659/1.517	H^γ	1.289
V13	7.977	4.172	H^β	1.617	$H^{\gamma1}$ (R)	1.356
					$H^{\gamma2}$ (S)	0.808
K14	7.772	4.031	$H^{\beta2/\beta3}$	1.804	$H^{\gamma2/\gamma3}$	1.116
					$H^{\delta2/\delta3}$	1.500
					$H^{\varepsilon2/\varepsilon3}$	2.259
W15	8.003	4.421	$H^{\beta2/\beta3}$	3.196/2.982	$H^{\delta1}$	7.166
					$H^{\varepsilon1}$	10.925
					$H^{\varepsilon3}$	7.494
					$H^{\eta2}$	7.020
					$H^{\zeta2}$	7.292
					$H^{\zeta3}$	6.940
I16	7.649	4.116	H^β	1.763		
I17	7.872	4.072	H^β	1.734	$H^{\gamma2}$	1.069
					$H^{\gamma12/\gamma13}$	1.465
T19	7.755	4.169	H^β	4.098		
V20	7.892	3.911	H^β	2.092	$H^{\gamma1}$ (R)	0.858
					$H^{\gamma2}$ (S)	0.797
N21	7.954	4.101	$H^{\beta2/\beta3}$	1.586/1.488		
K22	8.256	4.101	$H^{\beta2/\beta3}$	1.863	$H^{\gamma2/\gamma3}$	1.172
					$H^{\delta2/\delta3}$	1.451
					$H^{\varepsilon2/\varepsilon3}$	2.705
F23	8.276	4.482	$H^{\beta2/\beta3}$	3.095/2.854	$H^{\delta1/\delta2}$	7.236
					$H^{\varepsilon1/\varepsilon2}$	7.157
T24	7.555	4.263	H^β	4.084		
K25	8.036	4.005	$H^{\beta2/\beta3}$	2.033	$H^{\gamma2/\gamma3}$	0.826
K26	8.434	4.447	$H^{\beta2/\beta3}$	2.525/2.461		

* Chemical shift values are in ppm.

Table A.6: Convergence of properties characterizing performed simulations

Label*	Potential energy $10^4\ V$ /KJmol^{-1}	Kinetic energy $10^4\ E_{kin}$ /KJmol^{-1}	Temperature T /K	Volume V_{box} /nm^3	Density d /Kgm^{-3}
W_1M	-32.8762 ± 3	5.8708 ± 2	298.13 ± 0.01	243.57 ± 0.01	979.25 ± 0.02
W_2Ma	-40.4384 ± 4	7.2213 ± 2	298.14 ± 0.01	300.46 ± 0.01	982.40 ± 0.02
W_2Mb	-40.4365 ± 4	7.2216 ± 2	298.15 ± 0.01	300.52 ± 0.01	982.19 ± 0.02
W_2Mc	-40.4360 ± 3	7.2217 ± 2	298.15 ± 0.01	300.55 ± 0.01	982.11 ± 0.02
W_4M	-49.6662 ± 4	8.8694 ± 2	298.14 ± 0.01	370.91 ± 0.01	987.09 ± 0.02
M_1M	-12.8173 ± 4	2.6685 ± 1	297.87 ± 0.01	241.51 ± 0.01	792.40 ± 0.02
M_2Ma	-15.6265 ± 4	3.2422 ± 1	297.88 ± 0.01	291.44 ± 0.01	798.48 ± 0.02
M_2Mb	-15.6275 ± 25	3.2422 ± 1	297.88 ± 0.01	291.54 ± 0.05	798.20 ± 0.13
M_2Mc	-15.6267 ± 5	3.2424 ± 1	297.90 ± 0.01	291.53 ± 0.01	798.23 ± 0.02
M_4M	-19.8997 ± 7	4.1072 ± 1	297.89 ± 0.01	365.63 ± 0.01	807.25 ± 0.02
D_1M	-8.7111 ± 3	2.0309 ± 1	297.91 ± 0.01	211.38 ± 0.01	1105.70 ± 0.03
D_2Ma	-10.7149 ± 7	2.4771 ± 1	297.91 ± 0.01	255.36 ± 0.01	1108.06 ± 0.02
D_2Mb	-10.7049 ± 12	24772 ± 1	297.92 ± 0.01	255.36 ± 0.01	1108.05 ± 0.03
D_2Mc	-10.7103 ± 41	24773 ± 1	297.93 ± 0.01	255.39 ± 0.01	1107.94 ± 0.03
D_4M	-13.5932 ± 6	31056 ± 1	297.93 ± 0.01	315.43 ± 0.01	1111.71 ± 0.02
B_2Md	-51.9175 ± 108	8.1960 ± 2	302.95 ± 0.01	418.82 ± 0.06	988.31 ± 0.14
B_3M	-42.9489 ± 64	6.5234 ± 1	302.94 ± 0.01	349.73 ± 0.03	1000.93 ± 0.08

Analysis was performed in the time window 20 to 50 ns on neat solvent simulations whereas on bilayer systems we discarded the first half of the trajectory. * Identification of molecular dynamics simulation.

Table A.7: Minimum distance between periodic images of δ-toxin chains

Label*	chain†	d_{min}^{PBC} /nm‡	$\overline{d_{min}^{PBC}}$ /nm¶
W_1M	A	3.40	4.40 ± 0.06
W_2Ma	A	3.94	4.78 ± 0.11
	B	4.35	5.13 ± 0.05
W_2Mb	A	3.59	4.66 ± 0.05
	B	2.93	3.99 ± 0.03
W_2Mc	A	2.79	4.26 ± 0.18
	B	3.42	4.54 ± 0.17
W_4M	A	4.28	4.96 ± 0.02
	B	3.46	4.44 ± 0.06
	C	4.13	5.08 ± 0.04
	D	3.13	4.26 ± 0.05
M_1M	A	2.55	3.63 ± 0.15
M_2Ma	A	2.10	4.06 ± 0.14
	B	1.68	3.77 ± 0.15
M_2Mb	A	2.10	3.54 ± 0.05
	B	2.27	3.53 ± 0.06
M_2Mc	A	2.57	3.97 ± 0.14
	B	2.66	3.96 ± 0.21
M_4M	A	2.69	4.53 ± 0.09
	B	3.37	4.68 ± 0.13
	C	2.89	4.02 ± 0.07
	D	2.22	4.07 ± 0.09
D_1M	A	0.81	2.43 ± 0.14
D_2Ma	A	1.74	3.39 ± 0.27
	B	0.81	2.87 ± 0.86
D_2Mb	A	1.77	3.36 ± 0.11
	B	1.05	2.93 ± 0.22
D_2Mc	A	1.84	3.01 ± 0.09
	B	2.30	3.50 ± 0.11
D_4M	A	1.89	3.43 ± 0.15
	B	1.53	3.72 ± 0.25
	C	2.37	3.90 ± 0.10
	D	2.40	4.05 ± 0.14
B_2Md	A	1.18	1.99 ± 0.23
	B	1.39	2.78 ± 0.04
B_3M	A	1.63	2.55 ± 0.02
	B	1.51	3.28 ± 0.05
	C	1.29	2.77 ± 0.03

* Identification of molecular dynamics simulation.
† Peptide chain identification. ‡ Shortest periodic distance.
¶ Average image periodic distance; analysis was performed in the time window 20 to 50 ns on neat solvent simulations whereas on bilayer systems we discarded the first half of trajectory.

Table A.8: Analysis of distance restraints violations for the 10 best structures of each cluster resulting from the docking runs

			Unambiguous Restraints		Ambiguous Restraints	
Docking Run	Cluster*	E_{tot} †	$E_{unambig}$ ‡	Viol¶	E_{ambig} §	Viol‖
NearNI+epr	1	0.03 ± 0.02	0 ± 0	0 ± 0	0.03 ± 0.02	0.8 ± 0.9
NearNI-epr	1	0.03 ± 0.02			0.03 ± 0.02	1.1 ± 0.5
NonNI+epr	1	5.07 ± 0.14	0.005 ± 0.005	0.2 ± 0.4	5.06 ± 0.14	2.3 ± 0.6
	2	7.57 ± 1.30	0.020 ± 0.007	0.8 ± 0.7	7.55 ± 1.30	4.7 ± 0.6
NonNI-epr	1	4.65 ± 0.75	0.004 ± 0.005	0.1 ± 0.3	4.64 ± 0.75	4.4 ± 0.5
	2	5.57 ± 1.97	0.025 ± 0.017	0.8 ± 0.4	5.55 ± 2.55	4.7 ± 0.6
	3	6.32 ± 1.60	0.015 ± 0.011	0.5 ± 0.5	6.31 ± 1.60	4.6 ± 0.5

* Cluster identification is in accordance with table 6.2, page 166; † Average total distance restraints energy, $E_{tot} = E_{unambig} + E_{ambig}$. All energy terms in this table are in Kcal mol^{-1};
‡ Average energy contribution from unambiguous distance restrains;
¶ Number of unambiguous distance restraints violations averaged over the ensemble;
§ Average energy contribution from ambiguous distance restrains; ‖ Number of ambiguous distance restraints violations averaged over the ensemble. Violations > 0.3 Å are reported.

Table A.9: Analysis of distance restraints violations for the 100 best docked structures of the cluster with lowest mean intermolecular energy

			Unambiguous Restraints		Ambiguous Restraints	
Docking Run	Cluster*	E_{tot} †	$E_{unambig}$ ‡	Viol¶	E_{ambig} §	Viol‖
NearNI+epr	1	0.02 ± 0.02	0 ± 0	0 ± 0	0.02 ± 0.02	0 ± 0
NearNI-epr	1	0.03 ± 0.02			0.03 ± 0.02	0 ± 0
NonNI+epr	1	4.84 ± 0.35	0.004 ± 0.005	0 ± 0	4.83 ± 0.35	2.0 ± 0.0
NonNI-epr	1	5.36 ± 1.45	0.008 ± 0.011	0 ± 0	5.35 ± 1.45	3.7 ± 0.8

* Cluster identification is in accordance with table 6.2, page 166;
† Average total distance restraints energy, $E_{tot} = E_{unambig} + E_{ambig}$. All energy terms in this table are in Kcal mol^{-1}; ‡ Average energy contribution from unambiguous distance restrains;
¶ Number of unambiguous distance restraints violations averaged over the ensemble;
§ Average energy contribution from ambiguous distance restrains;
‖ Number of ambiguous distance restraints violations averaged over the ensemble. Violations > 0.3 Å are reported.

Table A.10: Structural statistics of the 100 best docking model structures for each HADDOCK docking run

	NearNI		NonNI	
	NearNI+epr	NearNI-epr	NonNI+epr	NonNI-epr
Backbone RMSD (Å) with respect to mean				
Flexible interface backbone*	0.44 ± 0.12	0.45 ± 0.13	0.48 ± 0.17	0.59 ± 0.18
All backbone	0.49 ± 0.14	0.50 ± 0.15	0.58 ± 0.28	0.69 ± 0.30
Number of ambiguous interaction restraints (AIRs)				
From native monomer	11	11		
From non-native monomer			9	9
Total AIRs	22	22	18	18
Number of unambiguous interaction restraints				
Between native monomers	4			
Between non-native monomers			29	26
CNS intermolecular energies after water refinement†				
E_{vdw} (Kcal mol-1)	−130 ± 7	−129 ± 7	−83 ± 6	−66 ± 5
E_{elec} (Kcal mol-1)	−230 ± 35	−229 ± 29	−425 ± 33	−400 ± 25
Buried surface area (Å^2)‡	2480 ± 59	2478 ± 54	2175 ± 75	1829 ± 54
RMSD from idealized covalent geometry				
Bonds (Å)	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000
Angles (°)	0.392 ± 0.006	0.392 ± 0.004	0.386 ± 0.007	0.386 ± 0.008
Impropers (°)	0.380 ± 0.010	0.379 ± 0.010	0.394 ± 0.011	0.390 ± 0.012
Ramachandran analysis				
Residues in the favoured region (%)	85.3 ± 1.6	85.4 ± 1.7	83.6 ± 2.4	87.8 ± 2.4
Residues in additionally allowed regions (%)	14.5 ± 1.7	14.4 ± 1.7	16.4 ± 2.4	12.2 ± 2.4
Residues in generously allowed regions (%)	0.2 ± 0.2	0.2 ± 0.2	0.0 ± 0.0	0.0 ± 0.0
Residues in disallowed regions (%)	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Structural statistics of the 100 lowest energy structures (cluster 1) are based on structures obtained after flexible docking with HADDOCK followed by refinement in explicit water using ambiguous interaction restraints derived from H/D exchange data (NearNI-epr, NonNI-epr runs), and unambigous interaction restraints derived from EPR data (NearNI+epr, NonNI+epr runs). Hydrogen bonds restraints were also applied on the non-native monomers (see Computational Methods);
* The flexible interface is defined from the active and passive residues used in the definition of the ambiguous interactions restraints ±2 sequential residues; † The non-bonded energies were calculated with the OPLS parameters using a 8.5 Å cut-off;
‡ The buried surface area was calculated with CNS using a 1.4 Å radius water probe and 0.055 Å ofaccuracy.

Table A.11: Intermolecular contacts statistics calculated over the ensemble of the 100 best docked structures of the cluster with lowest mean intermolecular energy

Interacting Residues				Hydrogen-Bonds		
Donor		Acceptor				
Residue	Chain	Residue	Chain	M-M	M-S	S-S
A. Common contacts between runs NearNI+epr and NearNI-epr*						
Glu 92	A	Val 94	B	97/94		
Val 94	A	Glu 92	B	89/89		
Arg 103	A	Tyr 114	B			94/97
Thr 118	A	Thr 118	B	100/100		
Ala 120	A	Tyr 116	B	99/97		
Val 122	A	Tyr 114	B	63/69		
Glu 92	B	Val 94	A	91/95		
Val 94	B	Glu 92	A	90/88		
Arg 103	B	Tyr 114	A			100/100
Thr 118	B	Thr 118	A	100/100		
Ala 120	B	Tyr 116	A	99/99		
Val 122	B	Tyr 114	A	63/65		
B. Specific contacts for NearNI+epr run						
Glu 89	A	Thr 96	B		51	
Thr 116	A	Ala 120	B	56		
C. Common contacts between runs NonNI+epr and NonNI-epr†						
Arg 104	B	Asp 18	A			51/50
D. Specific contacts for NonNI+epr run						
Leu 12	A	Val 16	B	51		
Val 14	A	Val 14	B	100		
Val 32	A	Val 30	B	100		
Arg 104	A	Asp 18	B		85	82
Leu 12	B	Val 16	A	50		
Val 14	B	Val 14	A	100		
Val 32	B	Val 30	A	100		
Arg 104	B	Asp 18	A		88	86
E. Specific contacts for NonNI-epr run						
Leu 12	A	Tyr 78	B		85	
Val 16	A	Val 14	B	88		
Val 30	A	Val 30	B	92		
Val 32	A	Val 28	B	66		
Leu 12	B	Tyr 78	A		84	
Val 16	B	Val 14	A	94		
Val 28	B	Val 32	A	50		
Val 30	B	Val 30	A	93		
Val 32	B	Val 28	A	61		

Intermolecular contacts were analyzed with DIMPLOT (Wallace et al., 1995) and are reported if present in at least 50 of the 100 best structures. The occurrence of main-chain/main-chain (M-M), main-chain/side-chain (M-S), and side-chain/side-chain (S-S) hydrogen bonds is reported, over the ensemble of the 100 lowest energy structures (cluster 1). Chain A and B correspond to the two monomers involved in the docking procedure;

* The occurrence is reported as NearNI+epr/NearNI-epr;

† The occurrence is reported as NonNI+epr/NonNI-epr.

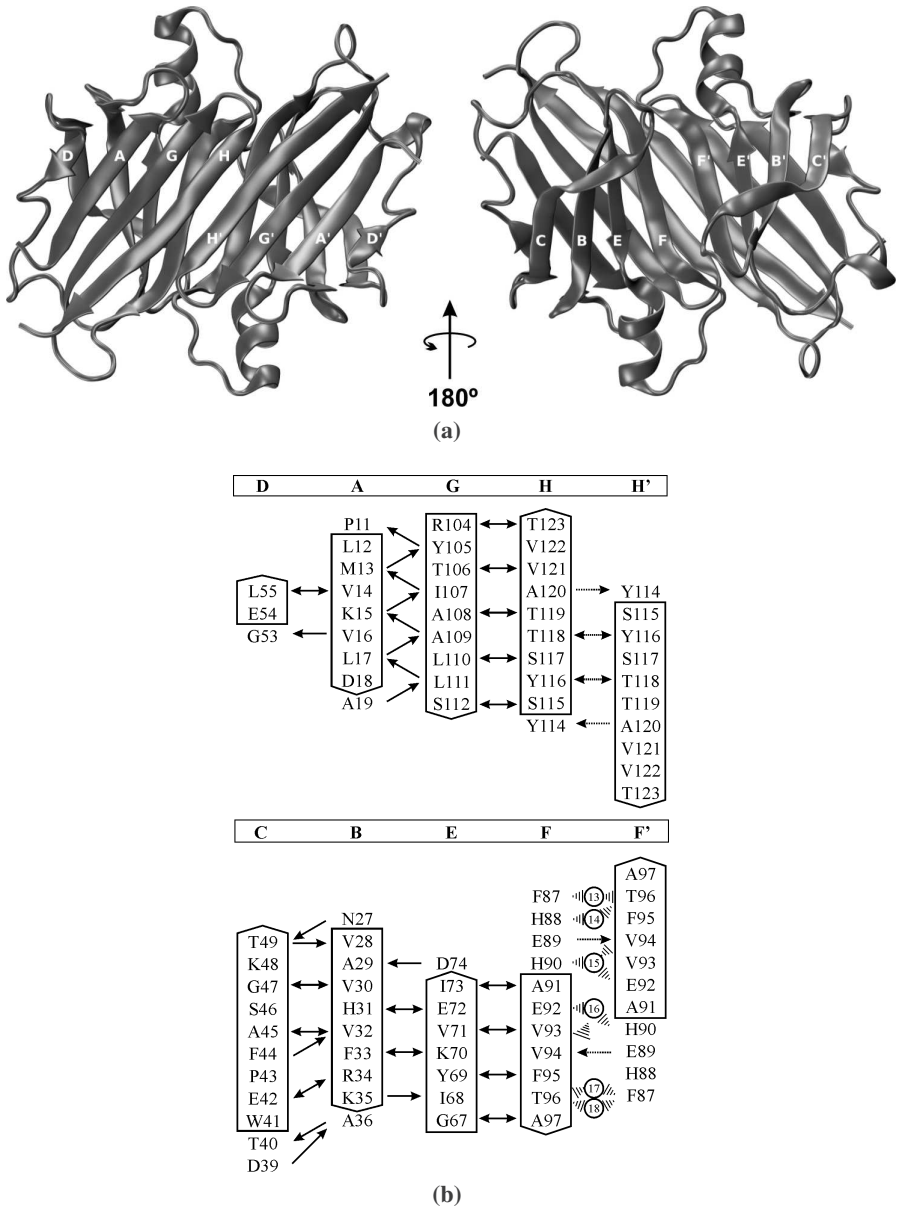
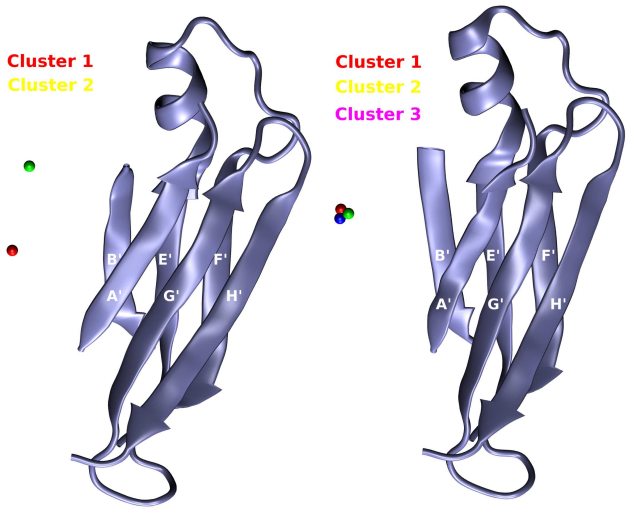


Figure A.1: **(a)** Three-dimensional structure of WT TTR (PDB entry, 1F41) in the dimeric form. The eight β-strands are named from *A* to *H*. **(b)** Schematic representation of the main-chain hydrogen bonds within the WT TTR monomer and at the monomer-monomer interface. Arrows point from the hydrogen bond donor to the acceptor. Double-headed arrows indicate two hydrogen bonds between the pair of residues. The open circles symbolize water molecules involved in indirect hydrogen bonds between strands FF'. The identification of the water molecule in the crystallographic structure is also represented.



(a) **(b)**

Figure A.2: Centers of mass of the non-native monomer from each cluster superimposed onto the cartoon representation of the other non-native monomer for **(a)** NonNI+epr and **(b)** NonNI-epr HADDOCK runs. The centroids were calculated for the structure with the lowest intermolecular energy of each cluster. The numbering of the clusters corresponds to that in table 6.2, chapter 6.

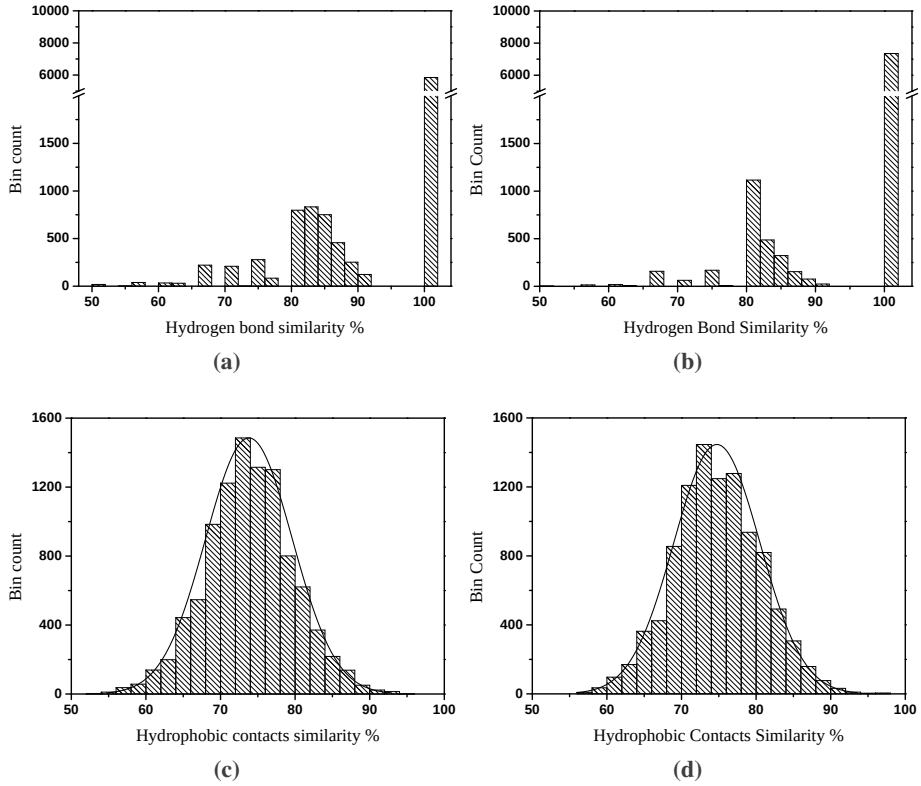


Figure A.3: Statistical histograms showing the distribution of the similarity of intermolecular contacts for the NearNI. The hydrogen-bonds (a,b) and the hydrophobic contacts (c,d) were calculated and compared, between the NearNI in the dimeric docked structures and in the resulting protofilaments. More than 60% of the protofilaments maintain all the main-chain/main-chain hydrogen bonds as in the near-native dimers used to build the two protofilaments sets, namely **(a)** protofil-epr and **(b)** protofil+epr. The hydrophobic contacts, at the residue level, have a gaussian distribution with a mean similarity value of 75% in both **(c)** protofil-epr and **(d)** protofil+epr sets.

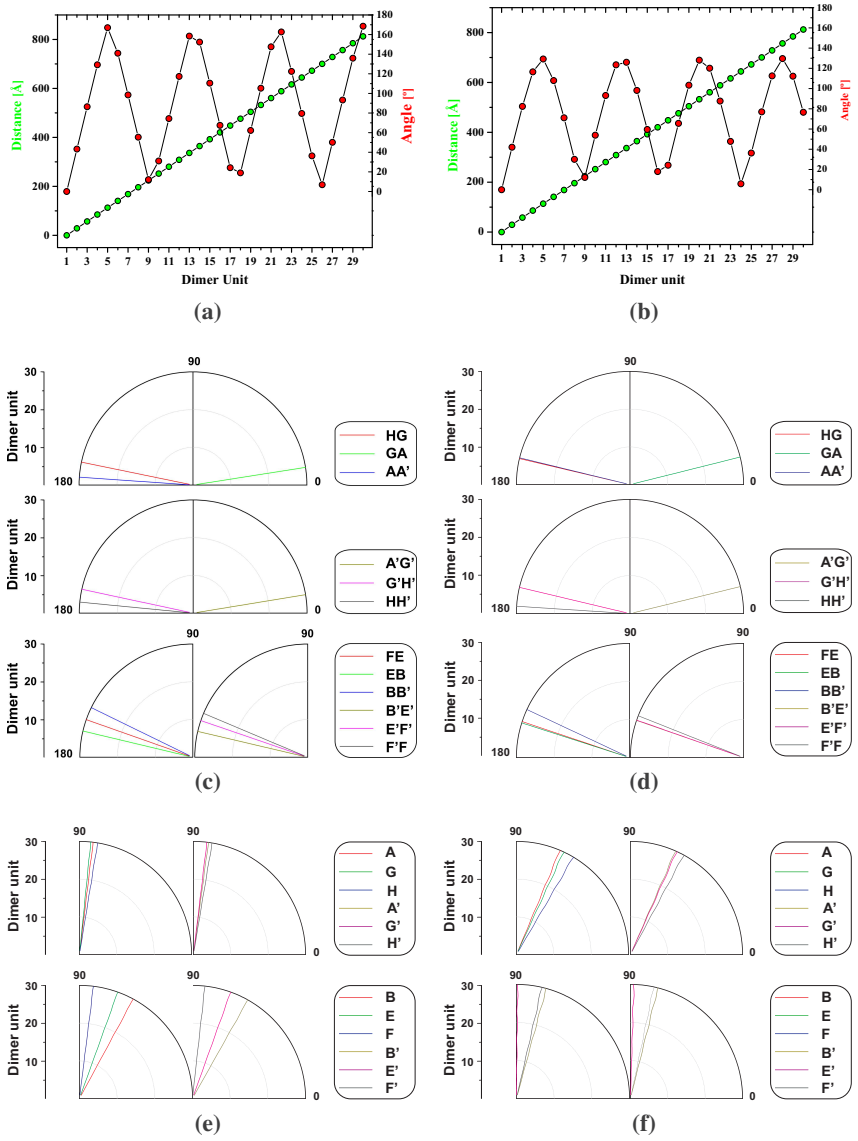


Figure A.4: Geometric characterization of the representative protofilaments for protofil+epr (a,c,e) and protofil-epr sets (b,d,f). **(a),(b)** The linearity and the helical twist of the chosen protofilament is shown. Both protofilaments show a helical periodicity of about 8 dimers. **(c),(d)** The angles between adjacent β-strands were determined as well as **(e),(f)** between each strand and the protofilament major axis. The polar plots were made for protofilaments with 10 dimeric units.

Appendix B

van der Waals interactions

Electrostatic interactions cannot account for all of the non-bonded interactions in a system. As an example, all of the multipole moments of a rare gas atom are zero, thus there can be no dipole-dipole or dipole-induced dipole interactions. Nevertheless, there clearly must be interactions between the atoms, since rare gas atoms show deviations from ideal gas behavior. These deviations were described by van der Waals, thus the forces that give rise to such deviations are often referred to as van der Waals forces (Leach, 1998).

B.1 London dispersive interactions

All atoms attract each other if they are not too close. This attractive contribution is due to dispersive forces. This type of interactions were explained using quantum mechanics by London, and are sometimes referred to as the London force. The dispersive force is due to instantaneous dipoles which arise from fluctuations in the electrons clouds. An instantaneous dipole in a molecule can in turn induce a dipole in neighboring atoms, giving rise to an attractive inductive effect. This fluctuating dipole-induced dipole interaction energy becomes more negative as the distance r between the atoms decreases; the force is always attractive.

A simple model to explain the dispersive interaction was proposed by Drude (Leach, 1998). The Drude model predicts that the interaction energy depends on the inverse sixth power of the distance between the two atoms, and that the magnitude of the interaction is proportional to the polarizability of each of the groups involved (Tinoco Jr. et al., 1995).

B.2 van der Waals repulsive contribution

All atoms or molecules repel at short distances. The energy of two atoms increase very rapidly if they are squeezed into the same space. The repulsion has a quantum mechanical origin and can be understood in terms of the Pauli principle, which formally prohibits any two electrons in a system from having the same set of quantum numbers. The interaction is due to electrons with the same spin. The short-range repulsive force comes from the electrostatic repulsion between pairs of electrons by forbidding them to occupy the same internuclear region. At very short internuclear r separations, the interaction energy varies as $1/r$ due to this nuclear repulsion, but at large separations the energy decays exponentially (Leach, 1998).

B.3 Modeling van der Waals interactions

Though van der Waals interactions can be determined from quantum mechanics, their large number in a bio-system has motivated scientists to use simple empirical expressions to increase the calculation speed. Briefly, these mathematical expressions sum up the attractive and repulsive energy components to learn the net force on the interacting parts. The best known of the van der Waals potential functions is the Lennard-Jones 12-6 function, which takes the following form for the interaction between two atoms i and j

$$V_{LJ}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma}{r}\right)_{ij}^{12} - \left(\frac{\sigma}{r}\right)_{ij}^{6}\right] \tag{B.1}$$

The Lennard-Jones 12-6 potential contains two adjustable parameters: the collision diameter σ, which is the separation for which the energy is zero, and the well depth ε. This potential is characterized by an attractive tail at large separations that varies as r^{-6}, the same power-law relationship described by the Drude model, and a steeply rising repulsive wall at distances less than $r \sim \sigma$ of the form r^{-12}. There is also a negative well, responsible for the cohesion in condensed phases. All these components can be visualized in figure B.1. The r^{-6} variation is of the same power-law relationship than the theoretical treatment of the dispersive forces. Though there are no strong theoretical arguments in favor of the repulsive term r^{-12}, it has been widely used to describe the repulsive van der Waals contribution.

The Lennard-Jones potential may also be written in the following form

$$V_{LJ} = \frac{C12_{ij}}{r_{ij}^{12}} - \frac{C6_{ij}}{r_{ij}^{6}} \tag{B.2}$$

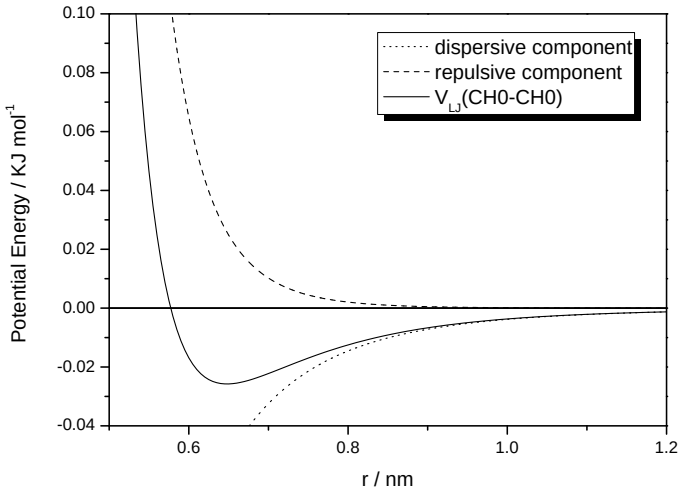


Figure B.1: Lennard-Jones potential V_{LJ} for two C-ester atoms (CH0) in a lipid system. The repulsive and attractive components of the potential function are also depicted. The parameters used are $\sigma = 0.5773$ nm and $\varepsilon = 0.02577$ KJ mol^{-1} (Chandrasekhar et al., 2003).

The van der Waals constants $C6_{ij}$ and $C12_{ij}$ depend on both atoms i and j. In a polyatomic system it should be expected to have several interactions between different types of atoms. Some of these cross interactions can be defined explicitly in the force field, or computed according to combination rules (van der Spoel et al., 2005). The GROMOS van der Waals parameters for an atom pair i and j, are derived from single atom van der Waals parameters (van Gunsteren et al., 1996) using the following relations

$$C6_{ij} = \sqrt{C6_{ii}C6_{jj}} \tag{B.3}$$

$$C12_{ij} = \sqrt{C12_{ii}C12_{jj}} \tag{B.4}$$

where $C6_{xx}$ is equal to $4\varepsilon_x\sigma_x^6$ and $C12_{xx}$ is equal to $4\varepsilon_x\sigma_x^{12}$, with $x = i$ or j. The mathematical relationships B.3 and B.4 were used to build four variations of

the GROMOS96 45A3 force field set[1] used in chapter 4. The only difference between them is the definition of the CH0 atom type used to model the ester-carbon of lipid molecules.

[1]For the GROMOS force fields the non-bonded atom type information and the van der Waals interaction parameters can be found in the interaction function parameter file (ifp). The following changes have been made to the original version of ifp45a3.dat : Atom type 6 (Terminal nitrogen, NH2) interacts with atom types 5 (Peptide nitrogen, NH), 6, 7 (Terminal nitrogen, NH3), 9 (Arg NH, NH2), and 10 (Arg NE, NH) through polar interaction, i.e. the repulsive $\sqrt{C12}$ parameters between these atom types are taken from column 5 in Table 2.5.6.2.2 (GROMOS 45A3 normal van der Waals parameters Table, in force field documentation) instead of column 4 (see van Gunsteren et al., 1996, pg. II-33). Chris Oostenbrink is acknowledged for sharing a debugged version of ifp45a3.dat. The GROMOS to GROMACS parameters conversion was accomplished with a home built script.

Agradecimentos

As minhas últimas palavras serão talvez as mais importantes desta tese, os agradecimentos. Todos sabemos o quanto difícil é realizar uma tarefa sózinhos, especialmente quando essa tarefa envolve diversas áreas do saber.

Anos faz, no meio de umas tantas estantes de livros pesadas, na velhinha biblioteca da Zoologia, comecei um percurso temido por muitos. Entrar no mundo da Bioquímica Estrutural. E foi aí que me perdi também. Um professor deu-me o seu voto de confiança então, e assim descobri a Ressonância Magnética Nuclear, e já agora para que se conste, o que são as proteínas. Longe de imaginar o que viria no depois, mas a tarefa lá chegou ao seu términus. Pelo meio ficam muitos obstáculos, e quase consigo sentir o sabor de os ter contornado. Agradeceço ao Dr. Rui Brito a sua alegria, perspicácia e sabedoria, em me colocar sempre de alerta ao caminho a seguir.

Passei horas a fio com o grande Ruizão, ora perdidos no "shiming" do espectrómetro, ora em pedaços desconexos de código. Devo a este senhor uma educação equilibrada na arte da administração *nix, e uma mente mais apurada no exercício da abstracção.

A Daniela está comigo desde o início, 1992. Uma pessoa brilhante no laboratório experimental, e genial no manejo da linguística. Muito obrigado DaniSan.

Agradeço à Babi Jú a alegria do dia a dia, um sopro bem fresco no sufoco diário de quem passa o tempo a falar com hardware. Ao bárbaro Conan tenho a agradecer o caos completo. Ajudou-me a destruir e a perverter, mesmo que por breves momentos, os dias lógicos e racionais. E do caos se fez luz Bruno.

Muitas faces pulularam as salas do grupo, não me lembro de todas. À Cândida pela matemática perdida na dislexia. Ao Carlos pela iniciativa temível. Ao Pedro pela ponderação metálica. À Elsa pela honestidade incisiva. À Zaida pelo misticismo. Catarina, pelo aroma feliz. À Frangolho pelo sorriso atraente. À Ana pela singeleza original.

Às minhas amigas de Biologia. Por me apresentarem os olhos meigos de quem sabe dar prioridade ao paraíso que nos abriga, do nada.

Às minhas amigas de Química. Obrigado por terem tornado as ruas da

cidade menos desertas, e os comboios para a Universidade uma fatalidade suportável.

Aos meus amigos Engenheiros. Pela descongestão de um dia normal.

Ao Paulo, por teres compartilhado o tecto, o som, a leitura e a filosofia de estar. Pelo canto-chão das coisas.

Ao meu Lèon Salgado. Pelo desembaraço e pela sagacidade de alguém que mora em paz no topo do mundo.

Ao Senhor João. Pelas braçadas conjuntas, faz tempo. Pelas suas mãos, por dar um corpo final a esta tese.

À minha cidade de Coimbra. Por guardares eternamente calada os nossos segredos. Pelas lágrimas de saudade que continuas a infligir. Sempre.

Aos meus avós eternos. Sacrificaram o tudo pelos filhos, providenciando as ferramentas de trabalho das gerações vindouras. Não estou aqui por acaso. Aprendi a subir escadas com o meu avô, e a cair também. Equilíbrio e técnica por passo dado. Aprendi a fazer a troca do milho pela farinha com a minha avó. Atenção e responsabilidade por saco vendido. Divago nas suas vozes e bebo da sua sapiência, mesmo longe.

Aos meus pais, pelo berço. Sei que não percebem muito bem o que faço, mas dão sempre importância ao meu dia a dia. Aos meus irmãos favoritos, pela companhia das palavras permanentes de carinho.

As minhas últimas palavras são dedicadas à minha princesa, Aida. Pelas minhas dores e achaques nina e pelo tempo que nos privei. Pela alegria de seres Mãe. Pai.

Bibliography

Adcock, S. and McCammon, J. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106(5):1589–1615.

Allen, M. P. and Tildesley, D. J. (1996). *Computer Simulation of Liquids*. Oxford University Press, Oxford, UK.

Almeida, P. F., Vaz, W. L., and Thompson, T. E. (1992). Lateral diffusion in the liquid phases of dimyristoylphosphatidylcholine/cholesterol lipid bilayers: a free volume analysis. *Biochemistry*, 31(29):6739–6747.

Altieri, A. S., Hinton, D. P., and Byrd, R. A. (1995). Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *Journal of the American Chemical Society*, 117(28):7566–7567.

Andreu, D. and Rivas, L. (1998). Animal antimicrobial peptides: an overview. *Biopolymers*, 47(6):415–433.

Andrusier, N., Mashiach, E., Nussinov, R., and Wolfson, H. J. (2008). Principles of flexible protein-protein docking. *Proteins*, 73(2):271–289.

Anézo, C., de Vries, A., Höltje, H.-D., Tieleman, D., and Marrink, S.-J. (2003). Methodological issues in lipid bilayer simulations. *Journal of Physical Chemistry B*, 107(35):9424–9433.

Anfinsen, C. B. (1972). The formation and stabilization of protein structure. *Biochem J*, 128(4):737–749.

Antalek, B. (2002). Using pulsed gradient spin echo NMR for chemical mixture analysis: How to obtain optimum results. *Concepts in Magnetic Resonance*, 14(4):225–258.

Antalek, B. (2007). Using PGSE NMR for chemical mixture analysis: quantitative aspects. *Concepts In Magnetic Resonance Part A*, 30A(5):219–235.

Antalek, B., Hewitt, J. M., Windig, W., Yacobucci, P. D., Mourey, T., and Le, K. (2002). The use of PGSE NMR and DECRA for determining polymer composition. *Magnetic Resonance in Chemistry*, 40(13):S60–S71.

Aragon, S. and Hahn, D. K. (2006). Precise boundary element computation of protein transport properties: diffusion tensors, specific volume, and hydration. *Biophys J*, 91(5):1591–1603.

Argiolas, A. and Pisano, J. J. (1983). Facilitation of phospholipase A2 activity by mastoparans, a new class of mast cell degranulating peptides from wasp venom. *J Biol Chem*, 258(22):13697–13702.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 98(18):10037–10041.

Barsukov, I. L. and Lian, L.-Y. (1993). *NMR of Macromolecules. A pratical approach.*, volume 134 of *The Practical Approach Series*, chapter Structure determination from NMR data. I. Analysis of NMR data, pages 315–357. Oxford University Press, Oxford.

Bax, A. and Davis, D. G. (1985). MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance*, 65(2):355–360.

Bechinger, B. (1999). The structure, dynamics and orientation of antimicrobial peptides in membranes by multidimensional solid-state NMR spectroscopy. *Biochim Biophys Acta*, 1462(1-2):157–183.

Bechinger, B. and Lohner, K. (2006). Detergent-like actions of linear amphipathic cationic antimicrobial peptides. *Biochim Biophys Acta*, 1758(9):1529–1539.

Bechinger, B. and Seelig, J. (1991). Conformational changes of the phosphatidylcholine headgroup due to membrane dehydration. A ^{2}H-NMR study. *Chem Phys Lipids*, 58(1-2):1–5.

Benyamini, H., Gunasekaran, K., Wolfson, H., and Nussinov, R. (2005). Fibril modelling by sequence and structure conservation analysis combined with protein docking techniques: β(2)-microglobulin amyloidosis. *Biochim Biophys Acta*, 1753(1):121–130.

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

Berendsen, H. J. and Hayward, S. (2000). Collective protein dynamics in relation to function. *Curr Opin Struct Biol*, 10(2):165–169.

Berendsen, H. J. C., Postma, J., van Gunsteren, W., and Hermans, J. (1981). *Intermolecular forces*, chapter Interaction models for water in relation to protein hydration, pages 331–342. Reidel, D. Publishing Company.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690.

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

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res*, 28(1):235–242.

Bernheimer, A. W. and Rudy, B. (1986). Interactions between membranes and cytolytic peptides. *Biochim Biophys Acta*, 864(1):123–141.

Birkbeck, T. H. and Freer, J. H. (1988). Purification and assay of staphylococcal δ-lysin. *Methods Enzymol*, 165:16–22.

Bladon, C. M., Bladon, P., and Parkinson, J. A. (1992). δ-toxin and analogues as peptide models for protein ion channels. *Biochem Soc Trans*, 20(4):862–864.

Blake, C. and Serpell, L. (1996). Synchrotron X-ray studies suggest that the core of the transthyretin amyloid fibril is a continuous β-sheet helix. *Structure*, 4(8):989–998.

Brand, T., Cabrita, E., and Berger, S. (2005). Intermolecular interaction as investigated by NOE and diffusion studies. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 46:159–196.

Branden, C. and Tooze, J. (1999). *Introduction to Protein Structure*. Garland Publishing, Inc., second edition.

Brito, R., Damas, A., and Saraiva, M. (2003). Amyloid formation by transthyretin: from protein stability to protein aggregation. *Curr Med Chem - Immun Endoc & Metab Agents*, 3(4):349–360.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, 54(Pt 5):905–921.

Bryson, B. (2004). *A short history of nearly everything*. Transworld Publishers.

Bundi, A. and Wüthrich, K. (1979). ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X_L-Ala-OH. *Biopolymers*, 18(2):285–297.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Callaghan, P. (1991). *Principles of Nuclear Magnetic Resonance Microscopy*. Clarendon Press, Oxford.

Cantor, C. R. and Schimmel, P. R. (1980). *Biophysical Chemistry. Part II: Techniques for the study of biological structure and function*. W. H. Freeman and Company. 11^{th} printing.

Cardoso, I., Goldsbury, C. S., Müller, S. A., Olivieri, V., Wirtz, S., Damas, A. M., Aebi, U., and Saraiva,

M. J. (2002). Transthyretin fibrillogenesis entails the assembly of monomers: a molecular model for in vitro assembled transthyretin amyloid-like fibrils. *J Mol Biol*, 317(5):683–695.

Carr, H. and Purcell, E. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94:630–638.

Cavanagh, J., Fairbrother, W. J., Palmer III, A., and Skelton, N. (1996). *Protein NMR Spectroscopy : Principles and Practice*. Academic Press, Inc., San Diego, California, USA.

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

Chandrasekhar, I. and van Gunsteren, W. F. (2001). Sensitivity of molecular dynamics simulations of lipids to the size of the ester carbon. *Current Science*, 81(10):1325–1327.

Chiti, F. and Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, 75:333–366.

Chiu, S. W., Clark, M., Balaji, V., Subramaniam, S., Scott, H. L., and Jakobsson, E. (1995). Incorporation of surface tension into molecular dynamics simulation of an interface: a fluid phase lipid bilayer membrane. *Biophys J*, 69(4):1230–1245.

Cho, C., Urquidi, J., Singh, S., and Robinson, G. (1999). Thermal offset viscosities of liquid H_2O, D_2O, and T_2O. *Journal of Physical Chemistry B*, 103(11):1991–1994.

Clore, G. M. and Kuszewski, J. (2002). χ_1 rotamer populations and angles of mobile surface side chains are accurately predicted by a torsion angle database potential of mean force. *J Am Chem Soc*, 124(12):2866–2867.

Collings, A. F. and Mills, R. (1970). Temperature-dependence of self-diffusion for benzene and carbon tetrachloride. *J Chem Soc Faraday Trans*, 66:2761–2766.

Conte, L. L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol*, 285(5):2177–2198.

Cordeiro, Y., Kraineva, J., Suarez, M. C., Tempesta, A. G., Kelly, J. W., Silva, J. L., Winter, R., and Foguel, D. (2006). Fourier transform infrared spectroscopy provides a fingerprint for the tetramer and for the aggregates of transthyretin. *Biophys J*, 91(3):957–967.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197.

Correia, B. E., Loureiro-Ferreira, N., Rodrigues, J. R., and Brito, R. M. M. (2006). A structural model of an amyloid protofilament of transthyretin. *Protein Sci*, 15(1):28–32.

Crank, J. (1975). *The mathematics of diffusion*. Oxford University Press, Oxford, England, second edition.

Cruz, P. J. (2007). Plasticidade estrutural do peptídeo δ-toxina em diversos ambientes químicos. Tese de licenciatura. Departamento de Química, Universidade de Coimbra, Portugal.

Damberg, P., Jarvet, J., and Gräslund, A. J. (2001). Accurate measurement of translational diffusion coefficients: a practical method to account for nonlinear gradients. *J Magn Reson*, 148(2):343–348.

Danielsson, J., Jarvet, J., Damberg, P., and Graslund, A. (2002). Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer Aβ(1-40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. *Magnetic Resonance In Chemistry*, 40:S89–S97.

Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald - an N.Log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98(12):10089–10092.

Dathe, M. and Wieprecht, T. (1999). Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. *Biochim Biophys Acta*, 1462:71–87.

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., and Mark, A. E. (1999). Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38(1-2):236–240.

Davis, J. E., Rahaman, O., and Patel, S. (2009). Molecular dynamics simulations of a DMPC bilayer using nonadditive interaction models. *Biophys J*, 96(2):385–402.

de Vries, S. J., van Dijk, A. D. J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., and Bonvin, A. M. J. J. (2007). HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins: Structure, Function, and Bioinformatics*, 69(4):726–733.

DeLano, W. L. (2002). The PyMOL molecular graphics system.

Derrick, T., McCord, E., and Larive, C. (2002). Analysis of protein/ligand interactions with NMR diffusion measurements: the importance of eliminating the protein background. *J Magn Reson*, 155(2):217–225.

Dhople, V. M. and Nagaraj, R. (1993). δ-toxin, unlike melittin, has only hemolytic activity and no antimicrobial activity: rationalization of this specific biological activity. *Bioscience Reports*, 13(4):245–250.

Dill, K. A., Ozkan, S. B., Weikl, T. R., Chodera, J. D., and Voelz, V. A. (2007). The protein folding problem: when will it be solved? *Curr Opin Struct Biol*, 17(3):342–346.

Dobson, C. M. (2004). Experimental investigation of protein folding and misfolding. *Methods*, 34(1):4–14.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). Pdb2pqr: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*, 32(Web Server issue):W665–W667.

Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). Haddock: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*, 125(7):1731–1737.

Dominguez, C., Bonvin, A. M. J. J., Winkler, G. S., van Schaik, F. M. A., Timmers, H. T. M., and Boelens, R. (2004). Structural model of the UbcH5BCNOT4 complex revealed by combining NMR, mutagenesis, and docking approaches. *Structure*, 12(4):633–644.

Douliez, J.-P., Ferrarini, A., and Dufourc, E.-J. (1998). On the relationship between C-C and C-D order parameters and its use for studying the conformation of lipid acyl chains in biomembranes. *The Journal of Chemical Physics*, 109(6):2513–2518.

Drazin, P. G. (2002). *Introduction to hydrodynamic stability*, volume 32 of *Cambridge texts in applied mathematics*. Cambridge University Press, Cambridge, United Kingdom.

Dubin, S. B., Clark, N. A., and Benedek, G. B. (1971). Measurement of the rotational diffusion coefficient of lysozyme by depolarized light scattering: sonfiguration of lysozyme in solution. *The Journal of Chemical Physics*, 54(12):5158–5164.

Eanes, E. D. and Glenner, G. G. (1968). X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem*, 16(11):673–677.

Easteal, A. and Woolf, L. (1985). (P, Vm, T, X) measurements for ((1-X)H2O+Xch3Oh) in the range 278-K to 323-K and 0.1 to 280 Mpa .1. Experimental results, isothermal compressibilities, thermal expansivities, and partial molar volumes. *Journal of Chemical Thermodynamics*, 17(1):49–62.

Efimova, Y. M., Haemers, S., Wierczinski, B., Norde, W., and van Well, A. A. (2007). Stability of globular proteins in H2O and D2O. *Biopolymers*, 85(3):264–273.

Ehrenstein, G. and Lecar, H. (1977). Electrically gated ionic channels in lipid bilayers. *Q Rev Biophys*, 10(1):1–34.

Einstein, A. (1905). The motion of elements suspended in static liquids as claimed in the molecular kinetic theory of heat. *Annalen Der Physik*, 17(8):549–560.

Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences*, 26(10):597–604.

Eneqvist, T., Andersson, K., Olofsson, A., Lundgren, E., and Sauer-Eriksson, A. E. (2000). The beta-slip: a novel concept in transthyretin amyloidosis. *Mol Cell*, 6(5):1207–1218.

Ernst, R. R. (1997). *Nobel Lectures, Chemistry 1991-1995*, chapter Nuclear Magnetic Resonance Fourier Transform Spectroscopy. World Scientific Publishing Co., Singapore.

Esturau, N., Sánchez-Ferrando, F., Gavin, J. A., Roumestand, C., Delsuc, M.-A., and Parella, T. (2001). The use of sample rotation for minimizing convection effects in self-diffusion NMR measurements. *Journal of Magnetic Resonance*, 153(1):48–55.

Fahy, E., Sud, M., Cotter, D., and Subramaniam, S. (2007). Lipid MAPS online tools for lipid research. *Nucleic Acids Res*, 35(Web Server issue):W606–W612.

Feynman, R. P., Leighton, R. B., and Sands, M. (1963). *The Feynman Lectures on Physics*, volume I.

Addison-Wesley, Reading, Massachusetts.

Fitton, J. (1981). Physicochemical studies on delta haemolysin, a staphylococcal cytolytic polypeptide. *FEBS Lett*, 130(2):257–260.

Fitton, J. E., Dell, A., and Shaw, W. V. (1980). The amino acid sequence of the delta haemolysin of Staphylococcus aureus. *FEBS Lett*, 115(2):209–212.

Foster, I. (2004). *The Grid: a new infrastructure for the* 21^{st} *century science*, chapter 2, pages 51–63. Wiley Series in Communications Networking & Distributed Systems. John Wiley & Sons Ltd.

Freddolino, P. L., Liu, F., Gruebele, M., and Schulten, K. (2008). Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys J*, 94(10):L75–L77.

Freddolino, P. L., Park, S., Roux, B., and Schulten, K. (2009). Force field bias in protein folding simulations. *Biophys J*, 96(9):3772–3780.

Freeman, R. (1988). *A handbook of Nuclear Magnetic Resonance*. Longman Scientific & Technical, Essex, England.

Freer, J. H. and Birkbeck, T. H. (1982). Possible conformation of delta-lysin, a membrane-damaging peptide of Staphylococcus aureus. *J Theor Biol*, 94(3):535–540.

Frishman, D. and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins: structure, function and genetics*, 23(4):566–579.

Gales, L., Cardoso, I., Fayard, B., Quintanilha, A., Saraiva, M. J., and Damas, A. M. (2003). X-ray absorption spectroscopy reveals a substantial increase of sulfur oxidation in transthyretin upon fibrillization. *J Biol Chem*, 278(13):11654–11660.

Geerke, D. P., Oostenbrink, C., van der Vegt, N. F. A., and van Gunsteren, W. F. (2004). An effective force field for molecular dynamics simulations of dimethyl sulfoxide and dimethyl sulfoxide/water mixtures. *The Journal of Physical Chemistry B*, 108(4):1436–1445.

Glasoe, P. K. and Long, F. A. (1960). Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.*, 64(1):188–190.

Glass, W. (1992). The resonance report. NMR tube specifications and quality. Technical report, Wilmad Glass, Buena, NJ, USA.

Goldsteins, G., Andersson, K., Olofsson, A., Dacklin, I., Edvinsson, A., Baranov, V., Sandgren, O., Thylén, C., Hammarstrom, S., and Lundgren, E. (1997). Characterization of two highly amyloidogenic mutants of transthyretin. *Biochemistry*, 36(18):5346–5352.

Goux, W. J., Verkruyse, L. A., and Saltert, S. J. (1990). The impact of Rayleigh-Benard convection on NMR pulsed-field-gradient diffusion measurements. *Journal of Magnetic Resonance*, 88(3):609–614.

Gregory, S. M., Cavenaugh, A., Journigan, V., Pokorny, A., and Almeida, P. F. F. (2008). A quantitative model for the all-or-none permeabilization of phospholipid vesicles by the antimicrobial peptide cecropin A. *Biophys J*, 94(5):1667–1680.

Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R. R. (1988). Clean TOCSY for proton spin system identification in macromolecules. *Journal of the American Chemical Society*, 110(23):7870–7872.

Guntert, P. and Wüthrich, K. (1992). Flatt - a new procedure for high-quality base-line correction of multidimensional nmr-spectra. *Journal Of Magnetic Resonance*, 96(2):403–407.

Gustavsson, A., Engström, U., and Westermark, P. (1991). Normal transthyretin and synthetic transthyretin fragments form amyloid-like fibrils in vitro. *Biochem Biophys Res Commun*, 175(3):1159–1164.

Habermann, E. (1972). Bee and wasp venoms. *Science*, 177(46):314–322.

Hahn, E. (1950). Spin echoes. *Physical Review*, 80:580–594.

Hess, B. (2002). Determining the shear viscosity of model liquids from molecular dynamics simulations. *The Journal of Chemical Physics*, 116(1):209–217.

Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472.

Högberg, C.-J. and Lyubartsev, A. P. (2006). A molecular dynamics investigation of the influence of hydration and temperature on structural and dynamical properties of a dimyristoylphosphatidylcholine bilayer. *J Phys Chem B*, 110(29):14326–14336.

Holz, M. and Weingärtner, H. (1991). Calibration in accurate spin-echo self-diffusion measurements using ^{1}H and less-common nuclei. *Journal of Magnetic Resonance*, 92:115–125.

Hörnberg, A., Eneqvist, T., Olofsson, A., Lundgren, E., and Sauer-Eriksson, A. E. (2000). A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. *J Mol Biol*, 302(3):649–669.

Hubbard, S. J. and Argos, P. (1994). Cavities and packing at protein interfaces. *Protein Sci*, 3(12):2194–2206.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14(1):33–38.

Inouye, H., Domingues, F. S., Damas, A. M., Saraiva, M. J., Lundgren, E., Sandgren, O., and Kirschner, D. A. (1998). Analysis of X-ray diffraction patterns from amyloid of biopsied vitreous humor and kidney of transthyretin (ttr) Met30 familial amyloidotic polyneuropathy patients: axially arrayed TTR monomers constitute the protofilament. *Amyloid*, 5(3):163–174.

Jackson, M. and Mantsch, H. H. (1991). Beware of proteins in DMSO. *Biochim Biophys Acta*, 1078(2):231–235.

Jansen, R., Dzwolak, W., and Winter, R. (2005). Amyloidogenic self-assembly of insulin aggregates probed by high resolution atomic force microscopy. *Biophys J*, 88(2):1344–1353.

Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *The Journal of Chemical Physics*, 71(11):4546–4553.

Jerschow, A. and Müller, N. (1998). Convection compensation in gradient enhanced nuclear magnetic resonance spectroscopy. *Journal of Magnetic Resonance*, 132(1):13–18.

Johnson Jr., C. (1999). Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34:203–256.

Jones, J. A., Wilkins, D. K., Smith, L. J., and Dobson, C. M. (1997). Characterisation of protein unfolding by NMR diffusion measurements. *Journal of Biomolecular NMR*, 10:199–203.

Jorgensen, W. L. and Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666.

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

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

Karplus, M. (1959). Contact electron-spin coupling of nuclear magnetic moments. *The Journal of Chemical Physics*, 30(1):11–15.

Karplus, M. and Kuriyan, J. (2005). Molecular dynamics and protein function. *Proc Natl Acad Sci U S A*, 102(19):6679–6685.

Keeler, J. (2006). *Understanding NMR Spectroscopy*. John Wiley & Sons Ltd, West Sussex, England.

Kelley, N. W., Huang, X., Tam, S., Spiess, C., Frydman, J., and Pande, V. S. (2009). The predicted structure of the headpiece of the Huntingtin protein and its implications on huntingtin aggregation. *J Mol Biol*.

Kerssebaum, R. and Salnikov, G. (2006). *DOSY and diffusion by NMR*. Bruker BioSpin GmbH, Rheinstetten, Germany. Version 2.0.0.

Kreger, A. S., Kim, K.-S., Zaboretzky, F., and Bernheimer, A. W. (1971). Purification and properties of staphylococcal delta hemolysin. *Infect Immun*, 3(3):449–465.

Krishnan, V. V. (1997). Determination of oligomeric state of proteins in solution from pulsed-field-gradient self-diffusion coefficient measurements. A comparison of experimental, theoretical, and hard-sphere approximated values. *Journal of Magnetic Resonance*, 124(2):468–473.

Kučerka, N., Liu, Y., Chu, N., Petrache, H. I., Tristram-Nagle, S., and Nagle, J. F. (2005). Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys J*, 88(4):2626–2637.

Lai, Z., Colón, W., and Kelly, J. W. (1996). The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry*, 35(20):6470–6482.

Laidman, J., Forse, G. J., and Yeates, T. O. (2006). Conformational change and assembly through edge beta strands in transthyretin and other amyloid proteins. *Acc Chem Res*, 39(9):576–583.

Lamb, D. M., Grandinetti, P. J., and Jonas, J. (1987). Fixed field gradient NMR diffusion measurements using bessel function fits to the spin-echo signal. *Journal of Magnetic Resonance*, 72(3):532–539.

Lashuel, H. A., Lai, Z., and Kelly, J. W. (1998). Characterization of the transthyretin acid denaturation pathways by analytical ultracentrifugation: implications for wild-type, V30M, and L55P amyloid fibril formation. *Biochemistry*, 37(51):17851–17864.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). Procheck: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2):283–291.

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

Leach, A. R. (1998). *Molecular Modelling : Principles and Applications*. Addison Wesley Longman Limited, Essex, England.

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

Lee, K. H., Fitton, J. E., and Wüthrich, K. (1987). Nuclear magnetic resonance investigation of the conformation of δ-haemolysin bound to dodecylphosphocholine micelles. *Biochim Biophys Acta*, 911(2):144–153.

Lemaître, G. (1931). The beginning of the world from the point of view of quantum theory. *Nature*, 127:706.

Levinthal, C. (1968). Are there pathways for protein folding?. *Journal de Chimie Physique et de Physico-Chimie Biologique*, 65:44–45.

Lian, L.-Y. and Roberts, G. C. K. (1993). *NMR of Macromolecules. A pratical approach.*, volume 134 of *The Practical Approach Series*, chapter Effects of chemical exchange on NMR spectra, pages 153–182. Oxford University Press, Oxford.

Lide, D. R., editor (1997). *CRC Handbook of Chemistry and Physics*. CRC Press, Boca Raton, 78th edition.

Lindahl, E. and Edholm, O. (2001). Molecular dynamics simulation of NMR relaxation rates and slow dynamics in lipid bilayers. *The Journal of Chemical Physics*, 115(10):4938–4950.

Lindahl, E., Hess, B., and van der Spoel, D. (2001). Gromacs 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7(8):306–317.

Lindahl, E. R. (2008). Molecular dynamics simulations. *Methods Mol Biol*, 443:3–23.

Lobo, V. M. M. Mutual and self-diffusion coefficients: a common confusion. To be published.

Loening, N. M. and Keeler, J. (1999). Measurement of convection and temperature profiles in liquid samples. *Journal of Magnetic Resonance*, 139(2):334–341.

Longsworth, L. G. (1960). The mutual diffusion of light and heavy water. *Journal of Physical Chemistry*, 64(12):1914–1917.

Loureiro-Ferreira, N. (1997). Estudo do efeito isotópico na estabilidade proteica. Tese de licenciatura. Departamento de Bioquímica, Universidade de Coimbra, Portugal.

Loureiro-Ferreira, N., Wassenaar, T., de Vries, S., van Dijk, M., van der Schot, G., van der Zwan, J., Boelens, R., Giachetti, A., Carotenuto, D., Rosato, A., Bertini, I., Herrmann, T., Bagaria, A., Zharavin, V., Jonker, H. R. A., Güntert, P., Schwalbe, H., Vranken, W., Dal Pra, S., Mazzucato, M., Frizziero, E., Traldi, S., Verlato, M., and Bonvin, A. M. J. J. (2010). e-NMR gLite grid enabled infrastructure. In *Proceedings of the 4th Iberian Grid Infrastructure Conference (IBERGRID'2010)*. Braga, Portugal.

Ludtke, S. J., He, K., Heller, W. T., Harroun, T. A., Yang, L., and Huang, H. W. (1996). Membrane pores induced by magainin. *Biochemistry*, 35(43):13723–13728.

MacKerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, W., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, 102(18):3586–

3616.

Mackerell, A. D. (2004). Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem*, 25(13):1584–1604.

Makin, O. S. and Serpell, L. C. (2005). Structures for amyloid fibrils. *FEBS J*, 272(23):5950–5961.

Mao, X.-A. and Kohlmann, O. (2001). Diffusion-broadened velocity spectra of convection in variable-temperature BP-LED experiments. *Journal of Magnetic Resonance*, 150(1):35–38.

Marassi, F. M., Opella, S. J., Juvvadi, P., and Merrifield, R. B. (1999). Orientation of cecropin a helices in phospholipid bilayers determined by solid-state NMR spectroscopy. *Biophys J*, 77(6):3152–3155.

Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wüthrich, K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J Mol Biol*, 280(5):933–952.

Marvin (2007). Marvin - ChemAxon. [Computer program]. Release 4.1.13.

Mathai, J. C., Tristram-Nagle, S., Nagle, J. F., and Zeidel, M. L. (2008). Structural determinants of water permeability through the lipid membrane. *J Gen Physiol*, 131(1):69–76.

McDonald, I. K. and Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J Mol Biol*, 238(5):777–793.

Méndez, R., Leplae, R., Lensink, M. F., and Wodak, S. J. (2005). Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins: Structure, Function, and Bioinformatics*, 60(2):150–169.

Mills, R. (1973). Self-diffusion in normal and heavy water in the range 1-45 degrees. *The J Physical Chemistry*, 77(5):685–688.

Miyamoto, S. and Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962.

Mor, A., Nguyen, V. H., Delfour, A., Migliore-Samour, D., and Nicolas, P. (1991). Isolation, amino acid sequence, and synthesis of dermaseptin, a novel antimicrobial peptide of amphibian skin. *Biochemistry*, 30(36):8824–8830.

Morris, G. (2002). Guest editor's foreword: NMR and diffusion. *Magn Reson Chem*, 40:S2.

Murday, J. S. (1973). Measurement of magnetic field gradient by its effect on the NMR free induction decay. *Journal of Magnetic Resonance*, 10(2):111–120.

Nagle, J. F. and Tristram-Nagle, S. (2000a). Lipid bilayer structure. *Curr Opin Struct Biol*, 10(4):474–480.

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

Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. ., Riekel, C., Grothe, R., and Eisenberg, D. (2005). Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, 435(7043):773–778.

Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Letters*, 229(2):317–324.

Nilsson, M. and Morris, G. A. (2005). Improving pulse sequences for 3D DOSY: Convection compensation. *Journal of Magnetic Resonance*, 177(2):203–211.

Nizet, V., Ohtake, T., Lauth, X., Trowbridge, J., Rudisill, J., Dorschner, R. A., Pestonjamasp, V., Piraino, J., Huttner, K., and Gallo, R. L. (2001). Innate antimicrobial peptide protects the skin from invasive bacterial infection. *Nature*, 414(6862):454–457.

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

Olofsson, A., Ippel, H. J., Baranov, V., Hörstedt, P., Wijmenga, S., and Lundgren, E. (2001). Capture of a dimeric intermediate during transthyretin amyloid formation. *J Biol Chem*, 276(43):39592–39599.

Olofsson, A., Ippel, J. H., Wijmenga, S. S., Lundgren, E., and Ohman, A. (2004). Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy. *J Biol Chem*, 279(7):5699–5707.

Orädd, G., Lindblom, G., and Westerman, P. W. (2002). Lateral diffusion of cholesterol and dimyristoylphosphatidylcholine in a lipid bilayer measured by pulsed field gradient NMR spectroscopy. *Biophys*

J, 83(5):2702–2704.

Ousterhout, J. K. (1994). *Tcl and the Tk toolkit*. Addison-Wesley.

Pace, C. N. and Schmid, F. X. (1997). *Protein structure. A pratical approach.*, volume 174 of *The Practical Approach Series*, chapter How to determine the molar absorvance coefficient of a protein, pages 253–259. IRL Press, Oxford University Press, Oxford, 2nd edition.

Pardi, A., Billeter, M., and Wüthrich, K. (1984). Calibration of the angular dependence of the amide proton-C^{α} proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein. use of $^{3}J_{HN\alpha}$ for identification of helical secondary structure. *J Mol Biol*, 180(3):741–751.

Park, K. D. and Lee, Y. J. (2006). Slice-selected LED and BPPLED: application of slice selection to DOSY. *Magn Reson Chem*, 44(9):887–891.

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

Pepys, M. B. (2006). Amyloidosis. *Annu Rev Med*, 57:223–241.

Peschier, L. J. C., Bouwstra, J. A., de Bleyser, J., Junginger, H. E., and Leyte, J. C. (1996). Cross-relaxation effects in pulsed-field-gradient stimulated-echo measurements on water in a macromolecular matrix. *Journal of Magnetic Resonance, Series B*, 110(2):150–157.

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

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem*, 26(16):1781–1802.

Pokorny, A. and Almeida, P. F. F. (2004). Kinetics of dye efflux and lipid flip-flop induced by δ-lysin in phosphatidylcholine vesicles and the mechanism of graded release by amphipathic, α-helical peptides. *Biochemistry*, 43(27):8846–8857.

Pokorny, A. and Almeida, P. F. F. (2005). Permeabilization of raft-containing lipid vesicles by δ-lysin: a mechanism for cell sensitivity to cytotoxic peptides. *Biochemistry*, 44(27):9538–9544.

Pokorny, A., Birkbeck, T., and Almeida, P. (2002). Mechanism and kinetics of delta-lysin interaction with phospholipid vesicles. *Biochemistry*, 41(36):11044–11056.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical recipes in FORTRAN 77: the art of scientific computing*, volume 1. Cambridge University Press, New York, USA; Melbourne, Australia, 2nd edition.

Price, W. S. (1996). Gradient NMR. *Annual Reports on NMR Spectroscopy*, 32:51–142.

Price, W. S. (1997). Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part 1. Basic theory. *Concepts in Magnetic Resonance*, 9(5):299–336.

Price, W. S. (1998a). Probing molecular dynamics in biochemical and chemical systems using pulsed field gradient NMR diffusion measurements. In AttaUr-Rahman, editor, *Recent advances in analytical techniques*, pages 31–72. Amsterdam.

Price, W. S. (1998b). Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part II. Experimental aspects. *Concepts in Magnetic Resonance*, 10(4):197–237.

Price, W. S. (1999). Water signal suppression in NMR spectroscopy. *Annual Reports on NMR Spectroscopy*, 38:289–354.

Price, W. S. (2000). NMR gradient methods in the study of proteins. *Annu Rep Prog Chem, Sect C*, 96:3–53.

Price, W. S. (2006). Protein association studied by NMR diffusometry. *Current Opinion in Colloid & Interface Science*, 11(1):19–23.

Price, W. S., Stilbs, P., Jönsson, B., and Söderman, O. (2001). Macroscopic background gradient and radiation damping effects on high-field PGSE NMR diffusion measurements. *J Magn Reson*, 150(1):49–56.

Price, W. S., Tsuchiya, F., and Arata, Y. (1999). Lysozyme aggregation and solution properties studied using PGSE NMR diffusion measurements. *Journal of the American Chemical Society*, 121(49):11503–11512.

Quintas, A., Saraiva, M. J., and Brito, R. M. (1999). The tetrameric protein transthyretin dissociates to a non-native monomer in solution. A novel model for amyloidogenesis. *J Biol Chem*, 274(46):32943–32949.

Quintas, A., Vaz, D. C., Cardoso, I., Saraiva, M. J., and Brito, R. M. (2001). Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *J Biol Chem*, 276(29):27207–27213.

Raiford, D. S., Fisk, C. L., and Becker, E. D. (1979). Calibration of methanol and ethylene glycol nuclear magnetic resonance thermometers. *Analytical Chemistry*, 51(12):2050–2051.

Rance, M., Sørensen, O., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983). Improved spectral resolution in cosy ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun*, 117(2):479–485.

Redfield, C. (1993). *NMR of macromolecules. A pratical approach.*, volume 134 of *The Practical Approach Series*, chapter Resonance assignment strategies for small proteins, pages 71–99. Oxford University Press, Oxford.

Reif, M. M., Kräutler, V., Kastenholz, M. A., Daura, X., and Hünenberger, P. H. (2009). Molecular dynamics simulations of a reversibly folding beta-heptapeptide in methanol: influence of the treatment of long-range electrostatic interactions. *J Phys Chem B*, 113(10):3112–3128.

Saberwal, G. and Nagaraj, R. (1994). Cell-lytic and antibacterial peptides that act by perturbing the barrier function of membranes: facets of their conformational features, structure-function correlations and membrane-perturbing abilities. *Biochim Biophys Acta*, 1197(2):109–131.

Schlick, T. (2002). *Molecular Modeling and Simulation: An Interdisciplinary Guide*, volume 21 of *Interdisciplinary Applied Mathematics*. Springer Verlag.

Schuler, L. D., Daura, X., and van Gunsteren, W. F. (2001). An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry*, 22(11):1205–1218.

Schwieters, C. D. and Clore, G. M. (2001a). Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *Journal of Magnetic Resonance*, 152(2):288–302.

Schwieters, C. D. and Clore, G. M. (2001b). The VMD-XPLOR visualization package for NMR structure refinement. *Journal of Magnetic Resonance*, 149(2):239–244.

Schwieters, C. D., Kuszewski, J. J., and Marius Clore, G. (2006). Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 48(1):47–62.

Serag, A. A., Altenbach, C., Gingery, M., Hubbell, W. L., and Yeates, T. O. (2001). Identification of a subunit interface in transthyretin amyloid fibrils: evidence for self-assembly from oligomeric building blocks. *Biochemistry*, 40(31):9089–9096.

Serag, A. A., Altenbach, C., Gingery, M., Hubbell, W. L., and Yeates, T. O. (2002). Arrangement of subunits and ordering of beta-strands in an amyloid sheet. *Nat Struct Biol*, 9(10):734–739.

Serpell, L. C., Sunde, M., Benson, M. D., Tennent, G. A., Pepys, M. B., and Fraser, P. E. (2000). The protofilament substructure of amyloid fibrils. *J Mol Biol*, 300(5):1033–1039.

Serpell, L. C., Sunde, M., Fraser, P. E., Luther, P. K., Morris, E. P., Sangren, O., Lundgren, E., and Blake, C. C. (1995). Examination of the structure of the transthyretin amyloid fibril by image reconstruction from electron micrographs. *J Mol Biol*, 254(2):113–118.

Shai, Y. (1999). Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochim Biophys Acta*, 1462(1-2):55–70.

Shai, Y. (2002). Mode of action of membrane active antimicrobial peptides. *Biopolymers*, 66(4):236–248.

Söderman, O. and Stilbs, P. (1994). NMR studies of complex surfactant systems. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 26(Part 5):445–482.

Sørland, G. H., Seland, J. G., Krane, J., and Anthonsen, H. W. (2000). Improved convection compensating pulsed field gradient spin-echo and stimulated-echo methods. *Journal of Magnetic Resonance*, 142(2):323–325.

States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982). A two-dimensional nuclear overhauser experiment

with pure absorption phase in four quadrants. *Journal of Magnetic Resonance*, 48(2):286–292.

Steiner, H., Hultmark, D., Engström, A., Bennich, H., and Boman, H. G. (1981). Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature*, 292(5820):246–248.

Stejskal, E. O. and Tanner, J. E. (1965). Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *The Journal of Chemical Physics*, 42(1):288–292.

Stilbs, P. (1987). Fourier transform pulsed-gradient spin-echo studies of molecular diffusion. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 19(1):1–45.

Stryer, L. (1995). *Biochemistry*. W. H. Freeman and Company, fourth edition.

Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol*, 273(3):729–739.

Sutcliffe, M. J. (1993). *NMR of macromolecules. A pratical approach.*, volume 134 of *The Practical Approach Series*, chapter Structure determination from NMR data. II. Computational approaches., pages 359–390. Oxford University Press, Oxford.

Sutherland, W. (1905). A dynamical theory of diffusion for non-electrolytes and the molecular mass of albumin. *Philosophical Magazine*, 9(49–54):781–785.

Swain, M., Silva, C. G., Loureiro-Ferreira, N., Ostropytskyy, V., Brito, J., Riche, O., Stahl, F., Dubitzky, W., and Brito, R. M. (2010). P-found: Grid-enabling distributed repositories of protein folding and unfolding simulations for data mining. *Future Generation Computer Systems*, 26:424–433.

Talbot, J., Thiaudière, E., Vincent, M., Gallay, J., Siffert, O., and Dufourcq, J. (2001). Dynamics and orientation of amphipathic peptides in solution and bound to membranes: a steady-state and time-resolved fluorescence study of staphylococcal delta-toxin and its synthetic analogues. *Eur Biophys J*, 30(2):147–161.

Tame, J. R. H. (2005). Scoring functions–the first 100 years. *J Comput Aided Mol Des*, 19(6):445–451.

Tanford, C. (1968). Protein denaturation. *Adv Protein Chem*, 23:121–282.

Tappin, M., Pastore, A., Norton, R., Freer, J., and Campbell, I. (1988). High-resolution ^{1}H NMR study of the solution structure of delta-hemolysin. *Biochemistry*, 27(5):1643–1647.

Tartaglia, G. G. and Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. *Chemical Society Reviews*, 37(7):1395–1401.

Thiaudière, E., Siffert, O., Talbot, J., Bolard, J., Alouf, J., and Dufourcq, J. (1991). The amphiphilic α-helix concept. Consequences on the structure of staphylococcal δ-toxin in solution and bound to lipids. *Eur J Biochem*, 195(1):203–213.

Thomas, D. J., Mitschang, L., Simon, B., and Oschkinat, H. (1999). Signal selection in high-resolution NMR by pulsed field gradients: II. The design of gradient pulse sequences. *Journal of Magnetic Resonance*, 137(1):10–24.

Thompson, S. J., Hattotuwagama, C. K., Holliday, J. D., and Flower, D. R. (2006). On the hydrophobicity of peptides: Comparing empirical predictions of peptide log P values. *Bioinformation*, 1(7):237–241.

Thylén, C., Wahlqvist, J., Haettner, E., Sandgren, O., Holmgren, G., and Lundgren, E. (1993). Modifications of transthyretin in amyloid fibrils: analysis of amyloid from homozygous and heterozygous individuals with the met30 mutation. *EMBO J*, 12(2):743–748.

Tinoco Jr., I., Sauer, K., and Wang, J. C. (1995). *Physical Chemistry : Principles and Applications In Biological Sciences.* Prentice Hall, New Jersey, USA, third edition.

Tycko, R. (2004). Progress towards a molecular-level structural understanding of amyloid fibrils. *Curr Opin Struct Biol*, 14(1):96–103.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., and Markley, J. L. (2008). Biomagresbank. *Nucl. Acids Res.*, 36(suppl. 1):D402–408.

van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., and Berendsen, H. J. C. (2005). *Gromacs User Manual version 3.3*.

van Dijk, A. D. J., Boelens, R., and Bonvin, A. M. J. J. (2005). Data-driven docking for the study of biomolecular complexes. *FEBS J*, 272(2):293–312.

van Dusschoten, D., De Jager, P. A., and Van As, H. (1995). Extracting diffusion constants from echo-time-dependent PFG NMR data using relaxation-time information. *Journal of Magnetic Resonance, Series A*, 116(1):22–28.

van Gunsteren, W., Billeter, S., Eising, A., Hünenberger, P., Krüger, P., Mark, A., Scott, W., and Tironi, I. (1996). *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* Hochschulverlag ETH-Zürich/Biomos, Zürich/Groningen.

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

Vishnyakov, A., Lyubartsev, A., and Laaksonen, A. (2001). Molecular dynamics simulations of dimethyl sulfoxide and dimethyl sulfoxide-water mixture. *J. Phys. Chem. A*, 105(10):1702–1710.

Wade, D., Boman, A., Wåhlin, B., Drain, C. M., Andreu, D., Boman, H. G., and Merrifield, R. B. (1990). All-D amino acid-containing channel-forming antibiotic peptides. *Proc Natl Acad Sci U S A*, 87(12):4761–4765.

Waldeck, A. R., Kuchel, P. W., Lennon, A. J., and Chapman, B. E. (1997). NMR diffusion measurements to characterise membrane transport and solute binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 30(1-2):39–68.

Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*, 8(2):127–134.

Walser, R., Mark, A. E., van Gunsteren, W. F., Lauterbach, M., and Wipff, G. (2000). The effect of force-field parameters on properties of liquids: parametrization of a simple three-site model for methanol. *The Journal of Chemical Physics*, 112(23):10450–10459.

Wang, D. J. and Leigh, J. S. (1994). Wireless precision piezoelectric thermometer using an RF excitation-detection technique with an NMR probe. *Journal of Magnetic Resonance, Series B*, 105(1):25–30.

Wang, X. and Morden, K. M. (1997). *Antibacterial peptide protocols*, volume 78 of *Methods in molecular biology*, chapter NMR characterization of amphipathic helical peptides, pages 85–112. Humana Press Inc., Totowa, New Jersey.

Wassenaar, T., van Dijk, M., Loureiro-Ferreira, N., van der Schot, G., de Vries, S., van der Zwan, J., Boelens, R., Giachetti, A., Carotenuto, D., Rosato, A., Bertini, I., Jonker, H., Bagaria, A., Zharavin, V., Güntert, P., Schwalbe, H., Vranken, W., Verlato, M., Dal Pra, S., Mazzucato, M., and Bonvin, A. (2010). e-NMR: Structural Biology on the Grid. *(submitted)*.

Wassenaar, T. A. (2006). *Molecular Dynamics of Sense and Sensibility in Processing and Analysis of Data*. PhD thesis, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, The Netherlands.

Weber, W., Hünenberger, P., and McCammon, J. (2000). Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: Influence of artificial periodicity on peptide conformation. *J Phys Chem B*, 104(15):3668–3675.

Welch, J. (1998a). *^{UNITY} INOVA, UNITY$_{plus}$ and UNITY acessory instalation. UNITY-series NMR spectrometer systems*. Varian Associates, Inc., USA. Publication No. 01-999033-00, Rev. C0898.

Welch, J. (1998b). *User guide: liquids NMR. Varian NMR spectrometer systems with VNMR 6.1B software*. Varian Associates, Inc., USA. Publication No. 01-999084-00, Rev. A1298.

Welch, J. (1998c). *VNMR Command and Parameter Reference. VNMR 6.1B software*. Varian Associates, Inc., USA. Publication No. 01-999087-00, Rev. A1298.

White, S. H. and Wimley, W. C. (1998). Hydrophobic interactions of peptides with membrane interfaces. *Biochim Biophys Acta*, 1376(3):339–352.

White, S. H. and Wimley, W. C. (1999). Membrane protein folding and stability: physical principles. *Annual Review of Biophysics and Biomolecular Structure*, 28(1):319–365.

Wiehe, K., Peterson, M. W., Pierce, B., Mintseris, J., and Weng, Z. (2008). Protein-protein docking: overview and performance analysis. *Methods Mol Biol*, 413:283–314.

Wilson, E. O. (1998). *Consilience: The Unity of Knowledge*. Alfred A. Knopf.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995). ^{1}H, ^{13}C and ^{15}N random

coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR*, 5(1):67–81.

Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol*, 222(2):311–333.

Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, 31(6):1647–1651.

Wu, D. H., Chen, A. D., and Johnson, C. S. (1995). An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *Journal of Magnetic Resonance, Series A*, 115(2):260–264.

Wüthrich, K. (1986). *NMR of proteins and nuclei acids*, volume 51. John Wiley & Sons, Inc.

Wüthrich, K. (2003). *Les Prix Nobel. The Nobel Prizes 2002*, chapter NMR Studies of Structure and Function of Biological Macromolecules. Nobel Foundation, Stockholm.

Wüthrich, K., Wider, G., Wagner, G., and Braun, W. (1982). Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J Mol Biol*, 155(3):311–319.

Xiang, Z., Soto, C. S., and Honig, B. (2002). Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. *Proc Natl Acad Sci U S A*, 99(11):7432–7437.

Yadav, N. N., Torres, A. M., and Price, W. S. (2008). An improved approach to calibrating high magnetic field gradients for pulsed field gradient experiments. *Journal of Magnetic Resonance*, In Press, Accepted Manuscript.

Zasloff, M. (1987). Magainins, a class of antimicrobial peptides from xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc Natl Acad Sci U S A*, 84(15):5449–5453.

Zhou, X. (1995). Calculation of translational friction and intrinsic viscosity. II. Application to globular proteins. *Biophys J*, 69(6):2298–2303.

Zuiderweg, E. R. P., Boelens, R., and Kaptein, R. (1985). Stereospecific assignments of ^{1}H-NMR methyl lines and conformation of valyl residues in the *lac* repressor headpiece. *Biopolymers*, 24(4):601–611.