An *in silico* Approach to the β-defensin Structure-Activity Problem

Hayden Lewis Eastwood

Submitted in partial fulfilment of the degree of Doctor of Philosophy

April 2008

Declaration

This thesis is submitted in part fulfillment of the requirement of the degree of Doctor of philosophy at the University of Edinburgh. Unless otherwise stated, the work is original and has not been previously submitted, in whole or in part, for any degree, or any other University.

Hayden Eastwood

April 2008

Acknowledgements

I am indebted first and foremost to my supervisor Perdita Barran. Without her none of this would have been possible. Perdi, Thank you for the tremendous freedom and flexibility I was afforded to pursue the science that I was interested in. And thank you for working so tirelessly to secure funding for me. (In that regard I am also indebted to Peter Tasker, who, for reasons still unknown, decided that I was worth all that money).

I owe a really big thank you to David Wales for compiling GMIN on our cluster (I know from experience that compiling is frequently a nightmare). David, a hearty thanks to you for having me to stay at your lab in Cambridge; I learned a great deal during my visits there, though mostly during the morning coffee session! And thanks for Telluride, the best, most enjoyable conference I've ever been to!

I am also indebted to the other members of Team Defensin: Julia Dorin, Dom Campopiano, John Govan, Duhsan Urin, Karen Taylor and everyone else, thank you. It's because of people like you that people like me have jobs. Your knowledge of biology and your dedication to experimentation were invaluable to me.

I enjoyed my PhD first and foremost because of who I worked with; the science was just a bonus. So thanks to all you guys in room 5. Hannah, thank you for distracting me with gossip when I really wasn't in the mood. Jason, thank you for making me realize that I actually don't know much about anything, including my field of study (not to mention film noire, Byzantine architecture, Satre and fricken impressionists). Pete, thanks for being lively and cheerful. Wut,

thanks for keeping me on the straight and narrow academically. Fi, thanks for the coffee breaks. All of you guys were great.

Thank you to my funding bodies: EPSRC and Edinburgh University Chemistry Department. Again, without the financial support none of this would have been possible.

Finally, a big thanks to Annette Burgesse, Amanda Ewing and everyone else that made the chemistry department work. Thank you for being patient with me and my frustrating reluctance to fulfill basic bureaucratic requirements.

This work is dedicated to Dan, whom I still miss every day.

"Mum, I think it's very unlikely that you might not win"

Contents

Abstract

β-defensins are a family of cationic, cysteine-rich antimicrobial peptide (AMP) components of the innate immune response to infection. They are expressed both inducibly and constitutively within vertebrates, insects and plants and antimicrobial action is observed against (both gram positive and gram negative) bacteria and a subset of enveloped viruses. The antimicrobial phenomenon is thought to result from membrane permeablisation that depends on key, electrostatic binding events between defensin and pathogen cell surface. This thesis tackles, *in silico*, two components of this structure-activity problem: That of rationally predicting β-defensin structure, and that of elucidating the first (presumed) binding events between β-defensin and pathogen cell surface. Preliminary results suggest that successful *in silico* folding requires a mobile disulphide bond strategy to circumvent kinetic trapping of intermediate states, and that the mechanism of pathogenic binding involves a complex interplay of hydrogen bonding, as well as productive electrostatic interactions.

Chapter One

Introduction

1.1. Defensins and immunity

Living creatures have had to evolve sophisticated immunity to counter the ever present threat of microbes and viruses. A central aspect of the innate* immune system in vertebrates (and other life forms) is a deadly repertoire of antimicrobial peptides that counter pathogenic invasion and infection[1-8]. Defensins[9,10] are one member of this repertoire that exhibit antimicrobial activity in phagocytes, inflammatory body fluids and epithelial secretions[9,10]. The peptide family is small (35-45 amino acids in length), cysteine rich, highly cationic and sequence heterogeneous[11] (classification is therefore based on characteristic disulphide cross linkages between cysteine residues rather than sequence).

The mechanism of antimicrobial action of this class of protein is not well understood though current evidence supports a sequential killing mechanism in which initial binding to cell surface components is followed by lethal membrane destabilisation[12-14].

1.2. Why study defensins?

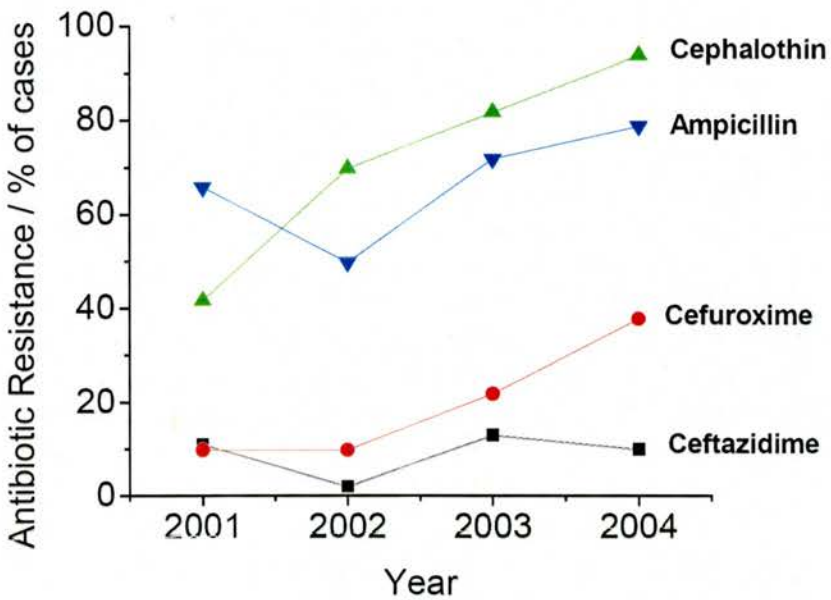
There is an urgent need for new and better antibiotics to keep pace with the increasing problem of multidrug resistance in hospitals worldwide (*Figure 1.2.1*). Defensins are of interest because insights into their antimicrobial mechanism may pave the way to new and better antibiotics.

* Living organisms possess two strategies for dealing with the threat of pathogenic infection: innate immunity and adaptive immunity. The former strategy involves killing the pathogen before entry to the host and the latter involves adapting to an infection that has overcome the innate barrier.

Furthermore, defensins are already present within human serum and treatments based on their mechanism of action may have fewer adverse side effects than current therapies.

From a protein folding standpoint defensins are interesting because their small size and conformational restriction makes them (in principle) computationally tractable (compared to most biological proteins, which are much larger). Furthermore, the peptide family possesses α-helical and β-sheet secondary structure[16-19], a property that permits the mechanism of secondary structural formation to be studied. Lastly, there are almost no computational studies of disulphide-bridged systems. Defensins are a good model system with which to investigate cross-linked, conformationally restricted folding.

Figure 1.2.1. *Selected data of antibiotic resistance in American hospitals between the years of 2001 and 2004. There is a marked increase in resistance for Cephalothin, Ampicillin and Cefuroxime. (Adapted from Szeto et al.[20])*



1.3. Summary of research aims

The ultimate, long-term goal of this research is to understand the molecular basis of defensin antimicrobial activity so that a new and effective antibiotic can be developed. Understanding this process depends, in turn, on understanding the initial interactions between defensin and pathogen.

A secondary (but related) aim is to develop *in silico* folding methodologies that predict the global energy minimum for the defensin family of proteins; doing so will increase the academic understanding of protein folding and reduce the requirement of expensive NMR and crystallographic procedures (that are currently used to elucidate defensin structural characteristics).

Chapter Two and Chapter Three introduce the defensin problem and outline the methodology for tackling the problem. Chapter Four and Chapter Five deal with rational structure prediction of β-defensins whilst Chapter Six and Chapter Seven focus on elucidating the mechanism of defensin membrane permeation. These research aims are summarised in *Table 1.3*.

Table 1.3. *Overview of Global Research Aims*

Aim	Notes
Rational structure prediction of Defensins.	1. Small size of defensins and restricted conformational space makes defensins ideal candidates for rational structure prediction. 2. Structures are currently solved with NMR or crystallography. An efficient protein folding algorithm will relieve dependence on expensive and time consuming experimental procedures.
Elucidate mechanism of antimicrobial action by studying interactions of HBD1, HBD2 and HBD3 with suspected binding partners.	1. A primary aim is to establish molecular basis of defensin killing mechanism so that a new and effective antibiotic can be developed. 2. The problem of membrane permeablisation is simplified into a number of initial, presumed events. The first (and primary focus of this work) is the initial binding of defensin with Lipid A (a suspected cell surface defensin binding partner).

References

(1) Yan, Q. Y.; Bennick, A. *Biochemical Journal* **1995**, *311*, 341-347.

(2) Fox, R. O.; Richards, F. M. *Nature* **1982**, *300*, 325-330.

(3) Bals, R.; Weiner, D. J.; Moscioni, A. D.; Meegalla, R. L.; Wilson, J. M. *Infection and Immunity* **1999**, *67*, 6084-6089.

(4) Xiao, Y. J.; Dai, H.; Bommineni, Y. R.; Soulages, J. L.; Gong, Y. X.; Prakash, O.; Zhang, G. L. *Febs Journal* **2006**, *273*, 2581-2593.

(5) Matsuzaki, K.; Harada, M.; Funakoshi, S.; Fujii, N.; Miyajima, K. *Biochimica Et Biophysica Acta* **1991**, *1063*, 162-170.

(6) Dimarcq, J. L.; Bulet, P.; Hetru, C.; Hoffmann, J. *Biopolymers* **1998**, *47*, 465-477.

(7) Bellm, L.; Lehrer, R. I.; Ganz, T. *Expert Opinion on Investigational Drugs* **2000**, *9*, 1731-1742.

(8) Mygind, P. H.; Fischer, R. L.; Schnorr, K. M.; Hansen, M. T.; Sonksen, C. P.; Ludvigsen, S.; Raventos, D.; Buskov, S.; Christensen, B.; De Maria, L.; Taboureau, O.; Yaver, D.; Elvig-Jorgensen, S. G.; Sorensen, M. V.; Christensen, B. E.; Kjaerulff, S.; Frimodt-Moller, N.; Lehrer, R. I.; Zasloff, M.; Kristensen, H. H. *Nature* **2005**, *437*, 975-980.

(9) Ganz, T.; Selsted, M. E.; Szklarek, D.; Harwig, S. S. L.; Daher, K.; Bainton, D. F.; Lehrer, R. I. *Journal of Clinical Investigation* **1985**, *76*, 1427-1435.

(10) Selsted, M. E.; Harwig, S. S. L.; Ganz, T.; Schilling, J. W.; Lehrer, R. I. *Journal of Clinical Investigation* **1985**, *76*, 1436-1439.

(11) Maxwell, A. I.; Morrison, G. M.; Dorin, J. R. *Molecular Immunology* **2003**, *40*, 413-421

(12) Shai, Y. *Biophysical Journal* **2001**, *80*, 2A-3A.

(13) Shai, Y.; Oren, Z. *Peptides* **2001**, *22*, 1629-1641.

(14) Ludtke, S. J.; He, K.; Heller, W. T.; Harroun, T. A.; Yang, L.; Huang, H. W. *Biochemistry* **1996**, *35*, 13723-13728.

(15) A. Pigache, P. C. F.-Y. D. **2004**.

(16) Hill, C. P.; Yee, J.; Selsted, M. E.; Eisenberg, D. *Science* **1991**, *251*, 1481-1485.

(17) Szyk, A.; Wu, Z. B.; Tucker, K.; Yang, D.; Lu, W. Y.; Lubkowski, J. *Protein Science* **2006**, *15*, 2749-2760.

(18) Schibli, D. J.; Hunter, H. N.; Starner, T. D.; Tack, B. F.; McCray, P. B.; Wiencek, J. M.; Vogel, H. J. *Journal of Biological Chemistry* **2002**, *277*, 8279-8289.

(19) Bauer, F.; Schweimer, K.; Kluver, E.; Conejo-Garcia, J. R.; Forssmann, W. G.; Rosch, P.; Adermann, K.; Sticht, H. *Protein Science* **2001**, *10*, 2470-2479.

(20) Szeto, C.; Chow, V.; Chow, K.; Lai, R.; Leung, B.; Kwan, B.; Li, P. *Kidney Int.* **2006**, *7*, 1107-1109.

Chapter Two

Background

2.1. Introduction

The research presented in this thesis bridges the fields of quantum mechanics, molecular modelling, biochemistry and immunology. A brief introduction to each area is given in this Chapter and Chapter Three so that the defensin problem can be considered in its full context. Defensins, aspects of immunology and bacteria are discussed here (below) and the available methodology is discussed in Chapter Three.

2.2. Defensins

Defensins are a family of evolutionarily related cysteine rich antimicrobial peptides that exhibit characteristic β-sheet secondary structure and a specific disulphide bridge connectivity[1]. The polypeptides are expressed in vertebrates, insects and plants and are an important component of the innate immune response to invading pathogens.

Defensins were first discovered in the early 1960s when studies of guinea pig and rabbit leukocyte lysates revealed a set of broad spectrum antimicrobial peptides that were defined by their unusually high gel electrophoretic mobility[2,3]. Structurally similar peptides were later discovered in human leukocytes in the mid 1980s and the term "defensin" coined as a result of their perceived role in the host defence mechanism[1].

2.2.1. Classification and structure

Mammalian defensins are classified into three subfamilies, α, β and θ on the basis of differing disulphide connectivity (*Table 2.2.1.1* and *Figure 2.2.1.2*)[1,4,5]. The θ family is further differentiated from the α and β forms by the fact that the *N* terminus is covalently attached to the *C* terminus, rendering each member of the family a circular peptide[6,7]. Interestingly, this species is created post-translationally via splicing of two partial α defensin peptides, each of which contributes 3 cysteines[8], but is not present functionally in human.

The reason for the sequence diversity variation across the defensin family is not perfectly understood. Both α and β-defensins have been shown to be subject to positive selection pressure and so, presumably, have evolved in response to species-specific evolutionary pressures.

Save for the cysteine network already discussed, β-defensins exhibit low sequence homology[9]. The little homology that does exist is confined to three glycine residues and a single, positively charged lysine or arginine residue that trails the final cysteine (*Figure 2.1.1.1*). Analysis of *phi-psi* torsion angles indicates that the glycines permit the existence of steep, characteristic β-defensin folds.

Figure 2.2.1.1. *The amino acid sequence alignments of a variety of human and murine β-defensins[9]. Black shading denotes heavily conserved regions and grey areas denote less conserved regions. Uncoloured regions show no significant homology. Note that the glycine residue immediately preceding cysteine 4 is highly conserved as is the double glycine pair between cysteines 1 and 2 (though to a lesser extent). The red ribbon denotes α-helical regions and the green arrows denote β-sheet regions.*

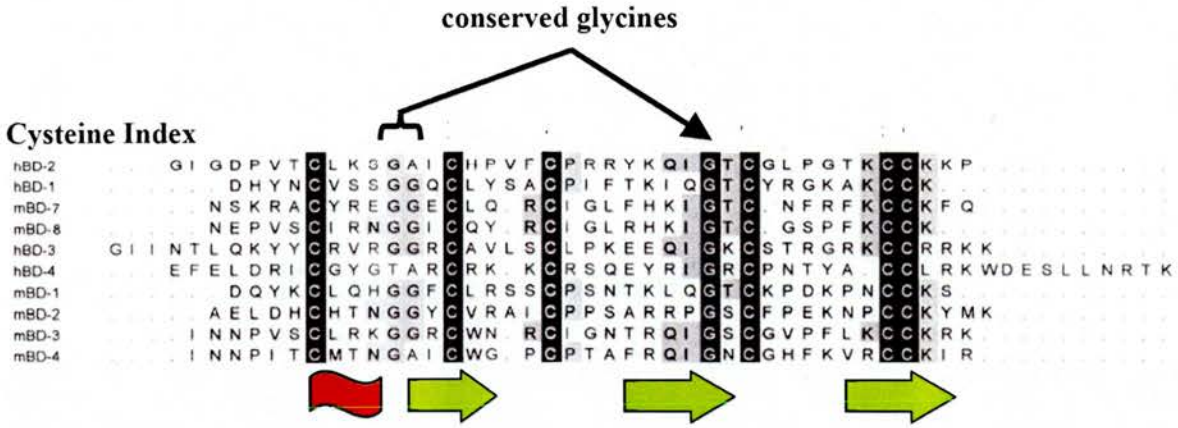
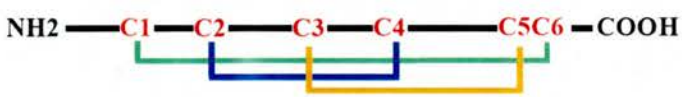
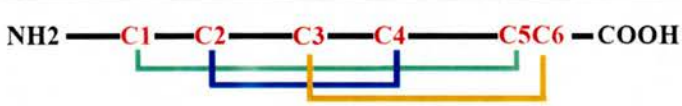
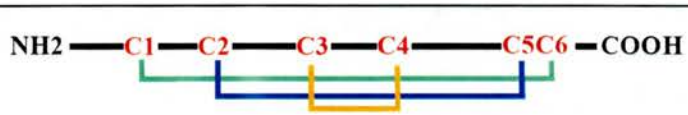


Table 2.2.1.1 *The schematic figures illustrate the key differences in cysteine connectivity across the subfamily. α defensins have a 1-6, 2-4, 3-5 connectivity, β defensins have a 1-5, 2-4, 3-6 connectivity, and θ defensins (which are circular), have a 1-6, 2-5, 3-4 connectivity.*

Subfamily	Connectivity
α	
β	
θ	

2.2.2. The role of disulphide bridges

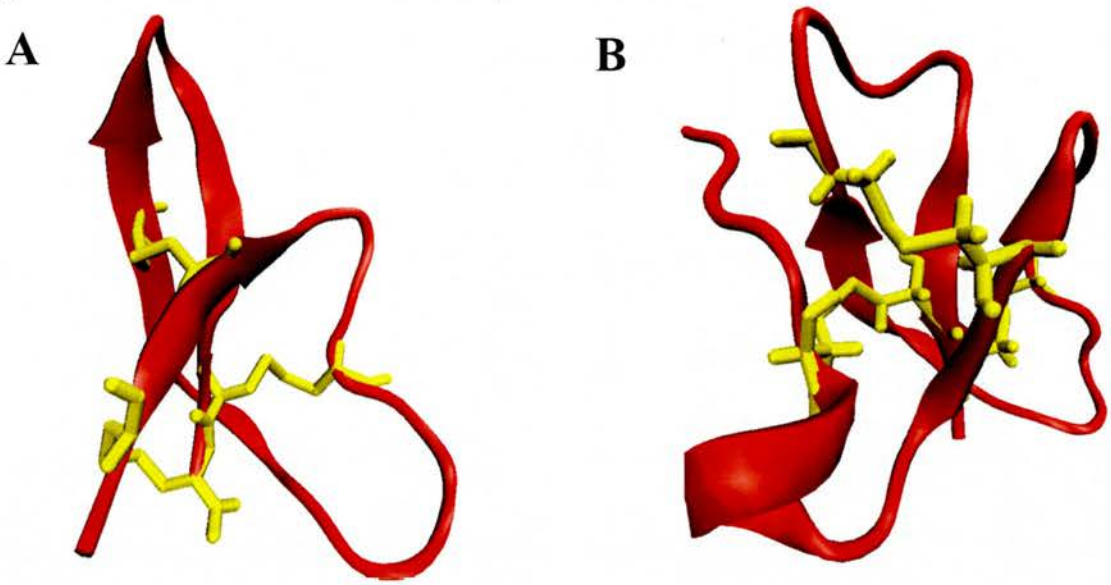
Defensins exhibit low sequence homology with one another (typically less than 35%) but have disulphide bridge networks that are highly conserved[97,98]. The central dogma of biology is that structure is related to function and the initial assumption with defensins was that their disulphide bridge connectivity (and associated structure) was a direct requirement for their antimicrobial activity. The emerging picture however, challenges this biological assumption. The first evidence of this contradiction was born out by Wu *et al* who, in 2003, illustrated that peptide fragments sharing sequence homology with portions of HBD3 were antimicrobially active[11]. Interestingly, the killing ability of these peptides correlated strongly with cationic charge. Furthermore, HBD3 molecules that had been entirely reduced* exhibited similar levels of killing to that of the canonically connected form[11]. This finding implies that charge, and not structure, is important for antimicrobial action. However, in contrast to this, human neutrophils α-defensins are known to be inactive in their reduced form[12]. These findings raise two important questions: Do α and β-defensins have different mechanisms of action? And, why have disulphide bridges been conserved in β-defensins if their presence isn't required for antimicrobial activity?

When considering the answers to these problems it is important to remember that defensins exist and participate in a complex and intermeshed biological environment. The disulphide bonds might be required (for example) to protect the backbone from proteolysis during biosynthesis, or to perform other secondary functions[11,13] (*Section 2.2.6*).

Figure 2.2.2.1 shows the disulphide bridge networks in HBD1 and HNP3.

* The term "reduced" refers to a cysteine that is covalently bonded to hydrogen, rather than being involved in a disulphide bridge.

Figure 2.2.2.1. *Typical structures for an α and β-defensin shown in a cartoon representation. Arrows represent β-sheets and ribbons represent α-helices. Disulphide cross linking is shown in yellow with a "liquorice" representation. A, The Crystal structure of the α-defensin HNP3*[10]*. B, A representative native NMR structure of the β-defensin HBD1*[9]*.*



2.2.3. Where are defensins expressed?

Within vertebrates, defensins are expressed both intracellularly and extracellularly. The specific locations and relative intensities of expression vary across mammalian species, a phenomenon which is thought to result from pathogen specific evolutionary pressures. Rabbits for example, are the only animals known to have high levels of α-defensins in their alveolar macrophages[14], whilst the leukocytes of mice (in contrast to all other examined species) exhibit no defensin expression whatsoever[15].

a) Intracellular secretion

In an intracellular setting defensins form part of a cocktail of antimicrobial ingredients that phagocytes utilise to kill engulfed organisms. The defensins are stored in intracellular compartments (or "granules") where they become highly

concentrated (often in excess of 10mg ml^{-1}). When a foreign particle is engulfed the intracellular granules fuse with the vacuole housing the invading microbe and the internal contents of the granule released in order to obliterate the pathogen[1,16-18].

b) Extracellular secretion

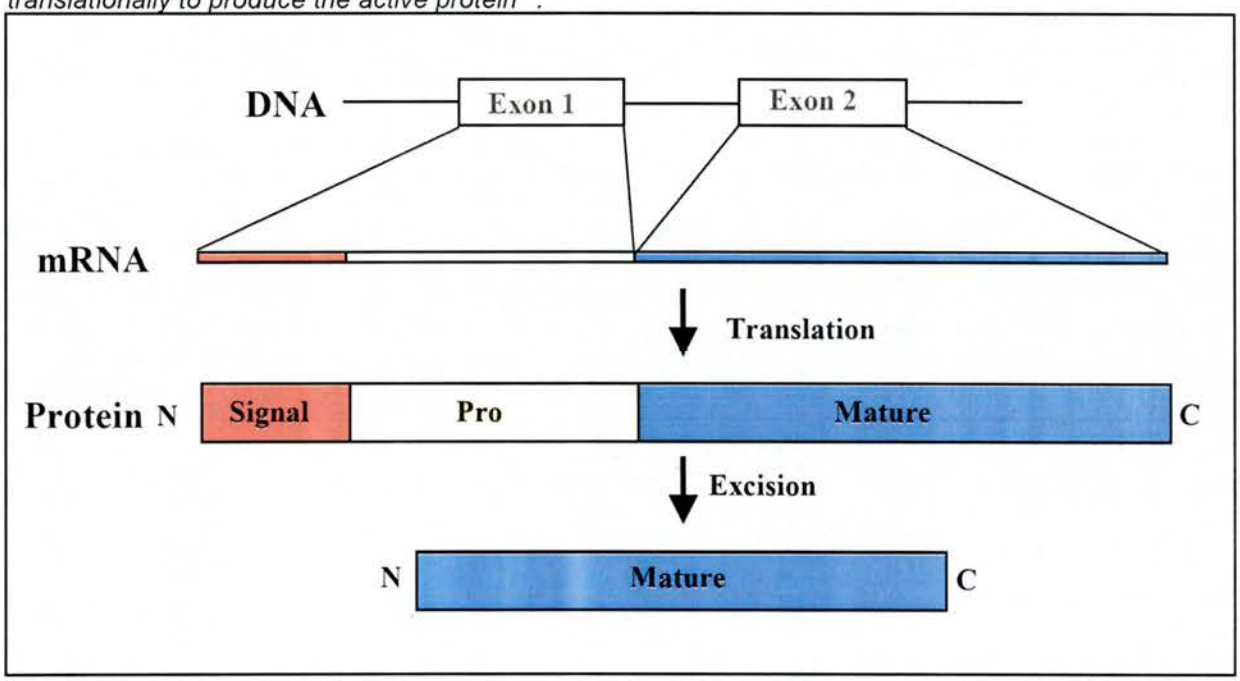
Extracellular secretion is known to occur on skin, in the respiratory system, in the reproductive system and at base levels within most (if not all) mucosal secretions. Keratinocytes within human skin secrete at least 3 types of β-defensins (HBD1,2 and 3)[19 20 21]. The respiratory system is protected by the same three defensins, with sub epithelial glands of the respiratory system exhibiting similar patterns of expression to skin[22-25]. In the digestive system Human Paneth cells (specialized epithelial cells that line the crypts of the small intestine) secrete α-defensins in response to bacterial challenge[26]. Indeed, work by Wilson *et al.* illustrate that knockout mice for α-defensin activity are less able to control ingested bacterial challenges from *E.coli* and *S.enterica*[27].

2.2.4. Genetic control and synthesis

Genes for both α and β-defensins are housed in the p23 region of chromosome 8[28-32]. Bioinformatic exploration of the human genome has resulted in other defensin gene clusters being found in other chromosome locations[33]. Notably Motzkus *et al.*, in 2006, synthesised a novel β-defensin (DEFB123) from bioinformatic information[34].

Defensins are synthesised as inactive precursor molecules that consist of a signal sequence, a "propeptide" region and a "normal" defensin region[35] (*Figure 2.2.4.1*). Post-translational modification of the inactive precursor yields the biologically functional species.

Figure 2.2.4.1. *A schematic representation of a defensin being transcribed and translated. The mRNA is produced from various exons and the signal and pro regions removed post-translationally to produce the active protein*[35].



Defensins exhibit both constitutive and inducible expression in a manner that often depends on their specific localisation. HBD1, for example, is expressed constitutively in airway epithelia[36] but expressed inducibly in monocytes[37]. Similarly, expression of HBD2-4 is induced by proinflammatory cytokines such as IL-1β , LPS and TNF-α[25 38,39]. Recent work has shown that the mRNA of HBD2 in oral epithelial cells is upregulated in response to microbial proteases[40], a finding that indicates the existence of other modes of regulation.

2.2.5. Mechanism of antimicrobial action

Understanding the molecular basis for defensin antimicrobial activity is a central aim of this research and Chapter Six and Chapter Seven treat the problem in greater depth. The challenge is formidable. Different defensins kill different organisms under different conditions, implying that there is more than one killing pathway. Accordingly (and frustratingly) insights gained from the mechanism of one defensin do not necessarily map to the killing mechanism of another. Most antimicrobial peptides exhibit high levels of cationic charge and there is consensus that this property is connected with their ability to attack and permeablise pathogenic membranes[7,41-44], though some antimicrobial peptides such as alamethicin are known to kill despite being charge neutral[45]. There are two models for β-defensin antimicrobial activity (*Figure 2.2.5.1*):

a) The carpet model[46]

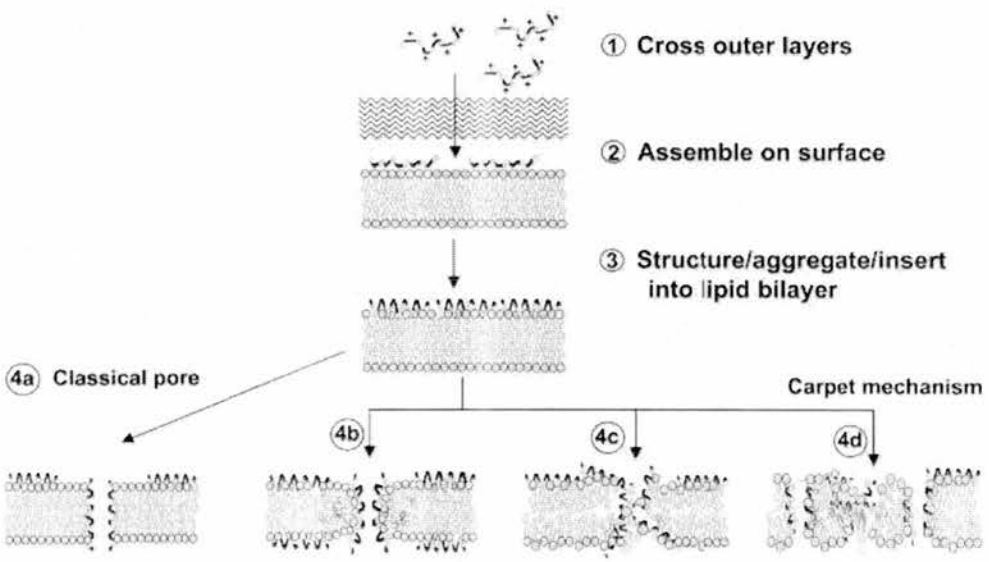
In this model defensins are presumed to accumulate on the target surface. When a threshold concentration is reached the "carpet" destabilises the macromolecular structure of the underlying membrane. Peptides do not necessarily insert into the hydrophobic membrane core and ion channels do not necessarily form.

b) The pore model[47]

In this model defensins are presumed to oligomerise on the cell surface and generate pores that lethally disrupt the ionic balance of the pathogen. A dynamic process is thought to follow initial binding in which structural reorganisations of the peptide complex drives lipid insertion and subsequent pore formation. The hypothesis is born out of experimental evidence of pore formation in other peptide systems[47-49].

Both these models assume that the defensins permeate the outer membrane freely. This is a plausible conjecture because the pores in the peptidoglycan component of the outer membrane are known to be in the region of 2-3 nm[50], large enough to permit such passage.

Figure 2.2.5.1. *Models of pore formation by defensins adapted from Sahl et al.*[52]*. Step 1: The outer membrane of the bacteria is crossed. Step 2: The defensins are electrostatically attracted to the anionic bacterial membrane surface. Step 3: An amphipathic structure capable of membrane insertion is formed. Step 4: The membrane is permeablised via a carpet or pore mechanism.*



Whilst both these mechanisms (*Figure 2.2.5.1*) are plausible, neither provides detail or quantitive measure (or calculation) of the physical and thermodynamic process at a molecular level. Are the carpet and pore models energetically realistic given that small, highly cationic monomers must pack in close proximity? There is tentative evidence that dimerisation is possible (at least in some defensins). Indeed, recent electrostatic potential calculations on a variety of defensin homologues suggest that HBD2 is able to exist dimerically[51]. Similarly, the 5-cysteine β-defensin related protein, Defr1, has been experimentally shown to dimerise covalently via a single, intermolecular disulphide bridge[118]. The 6 cysteine analogue of this

protein is known to dimerise non-covalently under mass-spectrometric and collision Ion Dissociation (CID) conditions[118]. Finally, NMR diffusion spectroscopy of HBD3 suggests symmetric, non-covalent dimers can form via the B2 strand of the β-sheet in conditions of 0.1 M and *pH* 8[119]. This same study does, however, indicate that HBD1 and HBD2 remain in monomeric form under these conditions. It is not yet clear whether the local chemical milieu surrounding the bacterial surface permits dimerisation (or higher order oligomerisation).

2.2.6. Secondary immune functions

Defensins exist within a complex soup of interacting immune particles and whilst they are highly antimicrobial, they are also extensively involved in regulating and directing immune (and other) processes. For example, they stimulate proliferation of epithelial cells[53], interfere with signal transduction pathways[54], upregulate cytokine and adhesion molecule expression[55] and directly recruit other immune particles (a process known immunologically as "chemoattraction" or "chemotaxis")[11]. This latter role is known to require intact disulphide bridges[11]. There is even some evidence that defensins selectively target and destroy tumour cells[56].

2.3. Other antimicrobial peptides

Living organisms have evolved an enormous repertoire of defences to deal with the milieu of environmental pathogens. Defensins are (unsurprisingly) not the only component of this repertoire. Some specific examples (*Figure 2.3.1.1*) of other antimicrobial peptides are presented here:

2.3.1. Cathelicidins[57]

These are a related family of antimicrobial peptides expressed in Leukocytes and epithelial cells of the testes, skin and gastrointestinal tract within a variety of mammals (including humans).

Like defensins they are highly charged, active against both gram-negative and gram-positive bacteria and upregulated in response to inflammatory cytokines. Unlike defensins however, they have no cysteine residues. In solution they adopt an unstructured coil, but quickly fold into an alpha helical form on entering hydrophobic environments.

2.3.2. Histatins[58]

These, like defensins and cathelicidins, are a family of small, cationic peptides (24-38 amino acids in length). They are histidine rich, possess no cysteine residues and are constitutively expressed in human saliva. Aqueous environments favour a non-structured linear form whilst non-aqueous environments favour an alpha-helical form[59]. Furthermore histatins are known to perform many secondary functions[60-64]. For example, they downregulate inflammatory cytokines and fibroblasts[60], as well as inhibit various disease-causing bacterial enzymes[61].

2.3.3. Magainins[47]

Magainin antibiotics are a family of immunogenic, cationic peptides (of 21 to 26 amino acids) found in the skin and intestines of the African Clawed frog (*Xenopus laevis*). Their sigmoidal relationship between peptide concentration and killing indicate a highly cooperative membrane permeablisation process[65-67]. Furthermore, some members of the family are known to kill synergistically with other members[68]. In solution they adopt a non-structured coil whilst in non-aqueous solvents they exist in extended alpha-helical configurations.

2.3.4. Alamethicin[45,69]

Alamethicin is a helical 20 amino acid peptide product of the fungus *Trichoderma viride*. This peptide gives rise to voltage-dependent conductance across lipid bilayers through a channel

formation process that involves cooperative insertion of peptide monomers. Unlike its mammalian analogues, it attacks both prokaryotic and eukaryotic cells. Interestingly, the molecule is charge neutral.

2.3.5. Melittin[69]

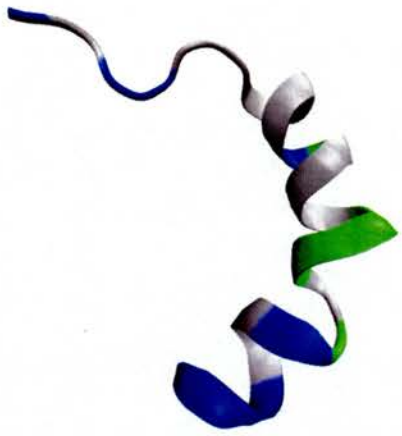
There are a whole host of other toxins that indiscriminately permeablise both prokaryotic and eukaryotic membranes. Melittin, for example, is the primary polypeptide component of European Bee venom. NMR and crystal studies indicate that the peptide adopts two alpha helical regions that hinge between amino acids 10 and 12[70,71]. This "hinging" ability in conjunction with the hydrophobic packing of non-polar groups gives rise to a distorted or "bent" helix. There is no evidence that this peptide oligomerises, and indeed, the most recent study confirms that melittin remains monomeric within liquid crystalline bilayers[72].

The above peptides have two unifying factors; they are all less than 40 amino acids in length and all possess some degree of helicity in non-polar environments. The other properties bear exceptions. Most are highly cationic, though one (alamethicin) is neutral. Some almost certainly oligomerise, others certainly don't. Some are highly selective for prokaryotic cells, whilst others attack all membranes indiscriminately. Thus, which properties are responsible for antimicrobial action? Do they all have similar modes of action? Is charge important for discriminating self over non-self, or for killing, or for both? There are many questions that need to be answered, and further treatment of this matter outside of the defensin problem is beyond the scope of this introduction.

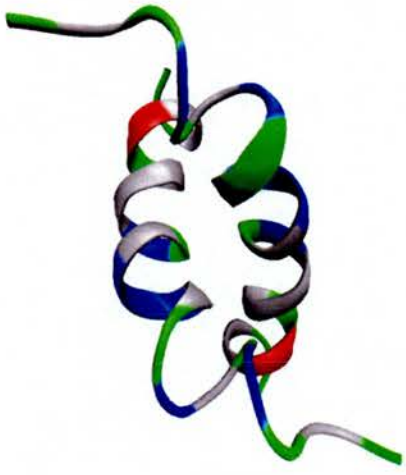
Figure 2.3.1.1. *Structures of selected antimicrobial peptides shown in cartoon representation. Grey shading refers to non-polar residues, green to polar residues, blue to cationic residues and red to anionic residues.* ***A*** *NMR structure of melittin in methanol[71]. **B** Crystal structure of cathelicidin[73], **C** Dimerised NMR magainins[74]. **D** Crystal structure of a lamethicin[69].*

A

B



C



D

2.4. Molecular modelling

In this thesis molecular modelling is divided into two separate problems; that of rational structure prediction of defensins, and that of molecular dynamical interactions with the target membrane (or its components). Molecular modelling techniques are widely used to tackle protein problems[75-81, 82-89] and are a central research theme (Chapter Three, Chapter Four, Chapter Five, Chapter Six and Chapter Seven).

2.5. Bacterial membrane structure

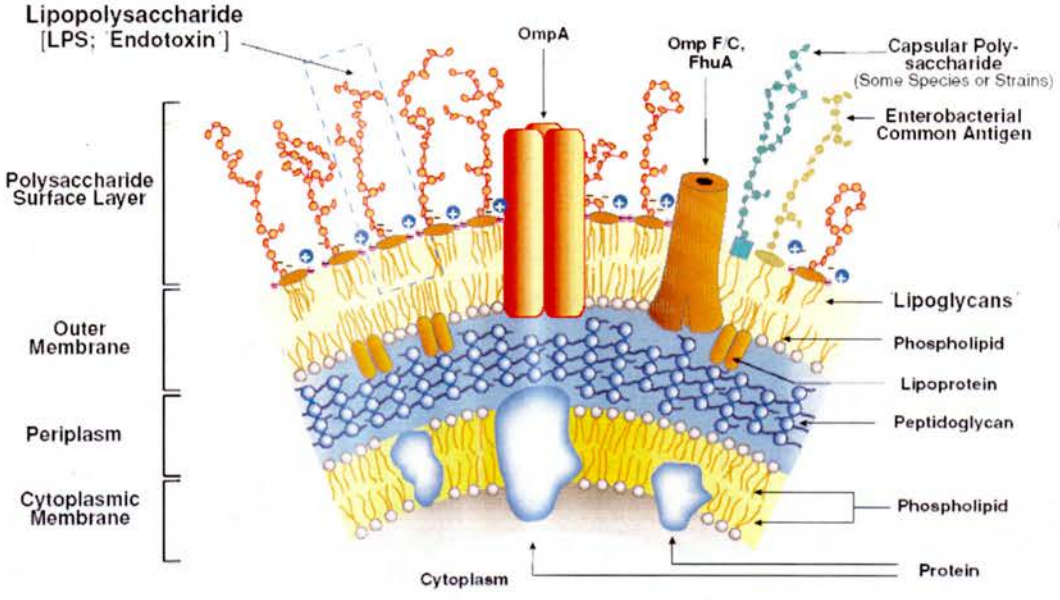
The first point of contact between defensins and bacteria occurs at the multi-layered cell envelope that surrounds the bacteria. Defensins must traverse this layer to target the cytoplasmic membrane. Accordingly, a mechanism of action must begin with accounting for the binding event that occurs at the interface of the outer membrane and the external environment. An understanding of surface components of bacteria is needed and as such a brief outline of bacterial membrane structure is provided.

All bacteria can broadly be divided into two categories, "gram-positive" and "gram-negative", on the basis of key differences in cell wall structure. Gram-positive bacteria are characterised by a double layer membrane whilst gram-negative bacteria are characterised by a triple-layered membrane. Both cell wall types are here discussed in more detail.

2.5.1. Gram-negative bacteria[120]

Gram-negative bacteria are characterised by the presence of a triple layered membrane (*Figure 2.5.1.1*)

Figure 2.5.1.1. *Schematic representation of gram negative cell wall structure (taken from review article by Alexander and* Rietschel[90]*).*



Aspects of each layer of the membrane are discussed below:

a) The inner cytoplasmic membrane

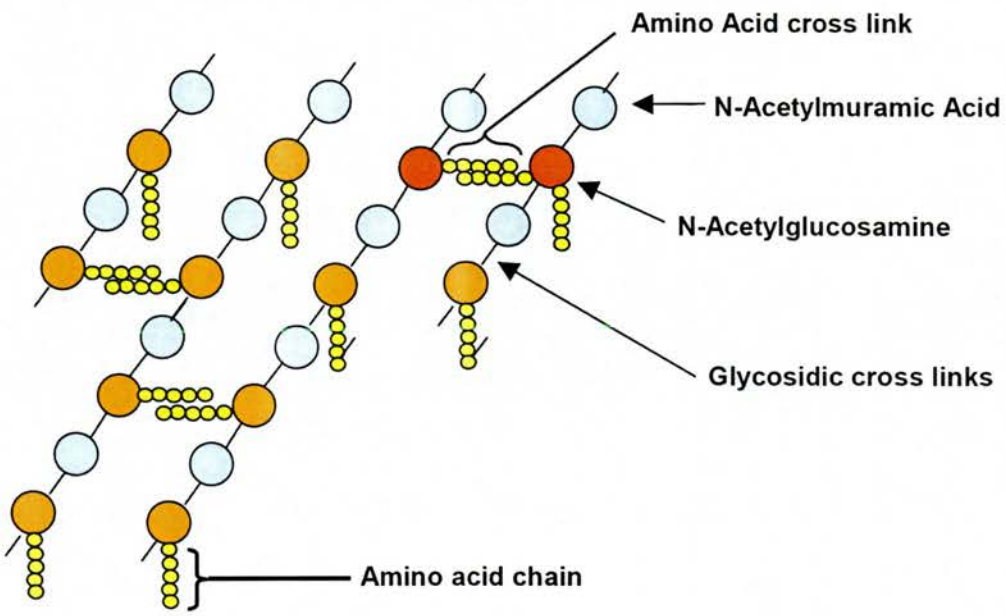
The inner layer is a cytoplasmic phospholipid bilayer awash with an array of proteins required for recognition and transport of various metabolites and synthetic products into and out of the cell. Its role is highly specialised to dealing with selective reabsorption.

b) The peptidoglycan layer

Bacteria have been forced to develop means of survival in an external environment over which they have almost no control. One strategy for dealing with this has been to develop a mechanically durable and versatile cell wall. Many of these mechanical

properties arise from the presence of peptidoglycan, a macromolecular scaffold constructed from a polymer of N-acetylmuramic (NAM) acid and N-acetylglucosamine (NAG) (*Figure 2.5.1.2*). The peptidoglycan layer surrounds the entire cell and confers great strength and versatility to the bacterium.

Figure 2.5.1.2. *Schematic structure of a portion of gram-positive peptidoglycan.*



In gram negative bacteria the peptidoglycan layer is approximately 2-10nm thick[91] and forms almost 10% of the dry weight of an individual bacterium[92]. Whilst the structure is mechanically strong, recent evidence suggest that it is in fact also highly flexible and capable of expanding and contracting in response to changes in osmotic pressure[93]. Peptidoglycan layers have also been found to have large pores approximately 5nm in diameter (but varying with osmotic conditions) which permit the passage of all biologically relevant molecules.

c) The outer membrane

Finally, surrounding the inner membrane and peptidoglycan matrix is a second phospholipid membrane festooned with an array of surface markers and ion channels. Since it forms the first site of contact with the external world, the outer membrane possesses many of the properties responsible for the infective and pathogenic capability of the organism. One such marker, with particular clinical relevance is Lipopoysacharide (LPS) (commonly referred to as Endotoxin). It is one of many markers responsible for aspects of virulence and pathological immune responses[94-98]. LPS is particularly relevant to the study of defensins and as such it will be given further consideration.

d) The periplasmic space[99]

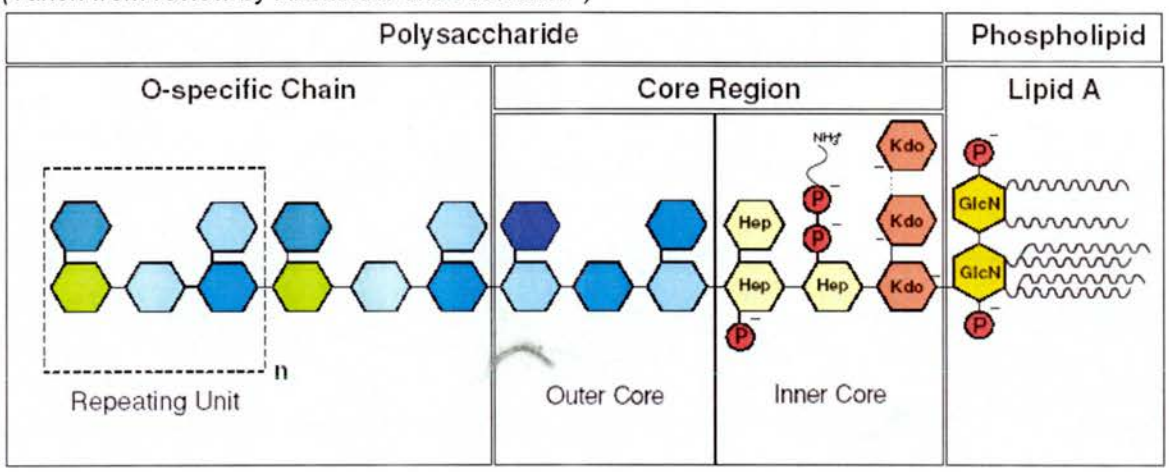
Eukaryotic cells possess an abundance of specialised intracellular organelles for compartmentalising potentially damaging reactions. Bacteria do not possess such organelles and as such potentially damaging reactions must take place outside the cell. Digestive enzymes are translocated from the interior of the bacterium to a small gap, known as the periplasmic space, which exists between the inner membrane and the peptidoglycan layer. The digested products then diffuse (or are actively absorbed) across the inner cytoplasmic membrane.

e) The structure and function of LPS

LPS consists of a phospholipid region and a polysaccharide region. The phospholipid region contains a functionally vital component known as Lipid A[101,113,114]. The polysaccharide region is itself often considered as two distinct entities; the "core region" and the "O-specific chain", each of which is synthesised

independently[90,111,112]. The O-region of the molecule is known to be highly variable, whilst Lipid A and specific parts of the core region remain largely conserved[100]. *Figure 2.5.1.3* summarises the structure of the molecule.

Figure 2.5.1.3. *Structure of LPS showing O-specific chain, Core region and Lipid A region. (Taken from review by Alexander and Rietschel[90])*



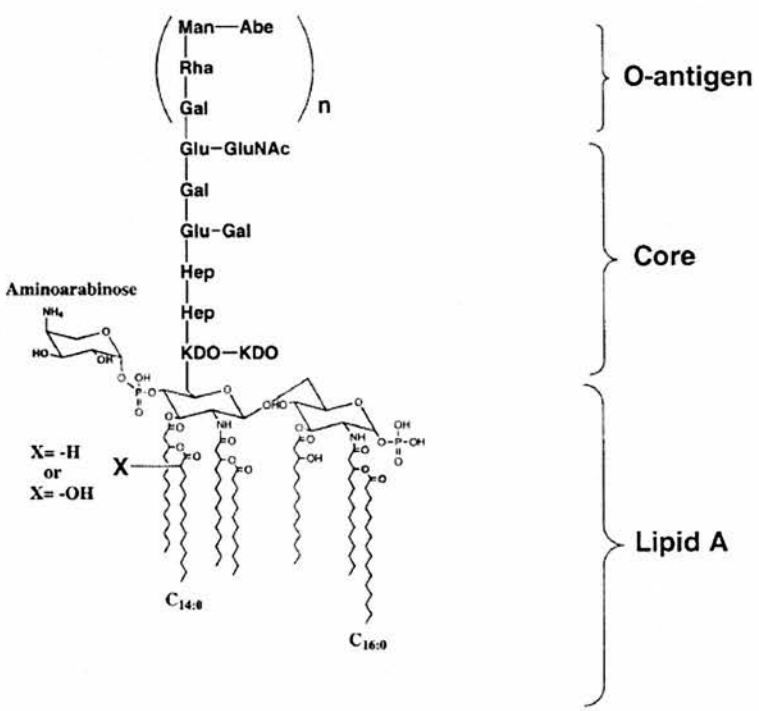
i. Lipid A

Lipid A is the primary immunogenic stimulant of the molecule and a number of modifications to it are known to impart resistance to defensin attack. One such modification in *Salmonella typhimurium* arises by addition of 4-deoxy-4-aminoarabinose (*Ara4N*) to diglucosamine phosphate residues[101] (*Figure 2.5.1.4*). In addition to strict genetic contributions, serum factors are also known to produce an assortment of LPS phenotypes. For example, some *S. typhimurium* isolates alter their LPS structures in response to changes in serum levels of Mg^{2+} ions[102,103]. Though many structural changes involve the incorporation of cationic species to the saccharide region, some alterations involve changes to the membrane bound component[104]. Indeed, certain strains of

S. Typhimurium are characterized by the presence of palmitate in the lipid portion. Surprisingly, this alteration also confers resistance to defensins and other cationic peptides.

The species *P. aeruginosa* is known to possess similar responses to environmental conditions. Interestingly, lipid A in serotypes associated with Cystic Fibrosis were shown to exhibit both altered acyl groups and high levels of substitution with *Ara4N*, much like those alterations exhibited by S. *Typhimurium*[105].

Figure 2.5.1.4. *Structure of Lipid A region showing O-antigen, Core and Lipid A. The region of Aminoarabinose incorporation is also illustrated. (Illustration taken from Ernst et. al*[106]*.)*



ii. The Core Region

The inner core region of LPS tends to be more conserved than the outer core region and, thus, it is likely that the inner core possesses traits essential to LPS function. In general, the two regions can be distinguished from one another on the basis of their monosaccharide composition[100]. The outer core consists predominantly of hexoses such as D-glucose, D-galactose, D-glucosamine, N-acetylglucosamine or N-acetylgalactosamine. In contrast, the inner core is characterized by the presence of 3-deoxy-D-*manno*-octulosonic acid (dOclA) (usually also termed '2-keto-3-deoxyoctulosonic acid' (Kdo)) and L- or D-glycero-D-*manno*-heptose (L,D-Hep). These groups are occasionally modified by additional anionic substituents such as phosphate (P), diphosphate (PP) or diphosphoethanolamine (PPEtn) groups[100]. Current proposals posit that the sugar acid Kdo forms a characteristic and essential component of the inner core region of bacterial LPS and accordingly provides a diagnostic basis for characterization[100, 107]. In the vast majority of gram negative bacteria, a Kdo residue links the polysaccharide moiety to the lipid A domain via a specific ketosidic bond that is unusually sensitive to chemical cleavage under acidic conditions[108 109]. Mutants associated with changes in the core region are also known to confer resistance to defensins[110].

iii. The O-specific Region

The O-specific chain forms the most structurally heterogeneous portion of LPS. Indeed, differences in the O-specific region frequently provide the very basis for classifying serotypes. The synthesis of the O-specific chain in enterobacteria is

determined by the *wb** (or *rfb*) gene cluster[111]. Bacterial isolates possessing a dysfunctional (or absent) *wb** locus synthesize LPS that possess absent or malformed O-specific chains. These isolates are capable of developing and multiplying *in vitro* and accordingly demonstrate that the O-chain is not essential for bacterial viability. However, experiments under *in vivo* conditions have shown that a variety of pathogenic enterobacteria require O-specific chains to prevent uptake by phagocytes and attack by serum complement[90]. It is posited that this large and relatively cumbersome region of the molecule provides a shield to attack by elements of the host immune system[90].

In addition, a variety of strains from several pathogenic species such as *H. pylori*, *N. gonorrhoeae*, *N. meningitidis* and *H. influenzae* have been found to possess terminal ligosaccharide structures that closely resemble human glycosphingolipids. This mimicry probably arises due to the presence of Nacetylneuraminic acid (NeuNAc) or L-fucose (L-Fuc), which are also present on mammalian cells[112].

iv. Concluding Remarks for LPS

The study of LPS has been hampered by the fact that it exhibits enormous heterogeneity across species and serotype. Furthermore, culturing populations of bacteria with homogenous LPS is exceedingly difficult because the consistency of the LPS is highly dependent on the growth medium. Indeed, subtle changes in external conditions can tilt the microbial environment towards favouring a mixed bag of LPS structures. Accordingly, developing a practical means of

testing molecular modelling results for a specific serotype against an experiment that utilises a homogenous serotype is time consuming, fiddly and expensive.

A central aim of this research is to understand the interactions between defensins and lipid A, and to understand the behaviour of cationic Lipid A mutations. Both these problems are dealt with extensively in Chapter Six and Chapter Seven.

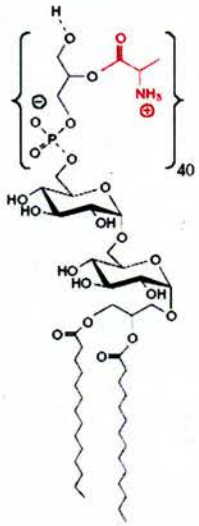
2.5.2. Gram-positive bacteria[120]

The gram-positive membrane consists only of two layers; the inner phospholipid bilayer and the outer and relatively much larger peptidoglycan matrix. Protruding from the surface are hundreds of different surface markers, of which the most abundant and specific to the species are teichoic acids. Like LPS, teichoic acids impart polyanionic properties on the bacterial surface. Whilst both molecules possess some similar functional properties, they are also markedly different in a variety of respects. For example, gram negative bacteria are known to be far more resistant to intestinal environments in comparison to gram positive bacteria, an observation which is assigned to the increased shielding properties of LPS components in comparison to TA[90].

a) Teichoic Acids

TA constitute between 30% and 60% of the cell wall. Like LPS, TA residues possess a fatty acid membrane, a core sugar unit and a polymer chain of variable length (*Figure 2.5.2.1)*. The repeat region consists of a series of highly anionic phosphodiester linked molecules of either 1,3- glycerol-phosphate (-Gro-P-) or 1,5 D-ribitol-phosphate (-Rbo-P-)[115 116].

Figure 2.5.2.1. *Teichoic acid structure of S.aureus (Adapted from review article by Peschel[117]).*



Lipoteichoic acid
(*Staphylococcus aureus*)

b) Peptidoglycan

The nature and consistency of peptidoglycan is very similar for both membrane types. However, gram-positive peptidoglycan is often characterised by a greater degree of cross-linking.

References

(1) Ganz, T.; Selsted, M. E.; Szklarek, D.; Harwig, S. S. L.; Daher, K.; Bainton, D. F.; Lehrer, R. I. *Journal of Clinical Investigation* **1985**, *76*, 1427-1435.

(2) Zeya, H. I.; Spitznagel, J. K. *Science* **1963**, *142*, 1085-&.

(3) Lehrer, R. I.; Selsted, M. E.; Szklarek, D.; Fleischmann, J. *Infection and Immunity* **1983**, *42*, 10-14.

(4) Ganz, T.; Selsted, M. E.; Szklarek, D.; Lehrer, R. I. *Federation Proceedings* **1985**, *44*, 907-907.

(5) Ganz, T.; Cardaci, C.; Selsted, M. E.; Lehrer, R. I. *Clinical Research* **1986**, *34*, A458-A458.

(6) Bals, R.; Lang, C.; Weiner, D. J.; Vogelmeier, C.; Welsch, U.; Wilson, J. M. *Clinical and Diagnostic Laboratory Immunology* **2001**, *8*, 370-375.

(7) Boniotto, M.; Antcheva, N.; Zelezetsky, I.; Tossi, A.; Palumbo, V.; Falzacappa, M. V. V.; Sgubin, S.; Braida, L.; Amoroso, A.; Crovella, S. *Biochemical Journal* **2003**, *374*, 707-714.

(8) Tang, Y. Q.; Yuan, J.; Osapay, G.; Osapay, K.; Tran, D.; Miller, C. J.; Ouellette, A. J.; Selsted, M. E. *Science* **1999**, *286*, 498-502.

(9) Bauer, F.; Schweimer, K.; Kluver, E.; Conejo-Garcia, J. R.; Forssmann, W. G.; Rosch, P.; Adermann, K.; Sticht, H. *Protein Science* **2001**, *10*, 2470-2479.

(10) Hill, C. P.; Yee, J.; Selsted, M. E.; Eisenberg, D. *Science* **1991**, *251*, 1481-1485.

(11) Wu, Z. B.; Hoover, D. M.; Yang, D.; Boulegue, C.; Santamaria, F.; Oppenheim, J. J.; Lubkowski, J.; Lu, W. Y. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 8880-8885.

(12) Daher, K. A.; Selsted, M. E.; Lehrer, R. I. *Journal of Virology* **1986**, *60*, 1068-1074.

(13) Nabil, K.; Rihn, B.; Jaurand, M. C.; Vignaud, J. M.; Ripoche, J.; Martinet, Y.; Martinet, N. *Biochemical Journal* **1997**, *326*, 377-383.

(14) Selsted, M. E.; Ouellette, A. J. *Nature Immunology* **2005**, *6*, 551-557.

(15) Eisenhauer, P. B.; Lehrer, R. I. *Infection and Immunity* **1992**, *60*, 3446-3447.

(16) Lehrer, R. I.; Ganz, T. *Current Opinion in Immunology* **2002**, *14*, 96-102.

(17) Ganz, T. *Infection and Immunity* **1987**, *55*, 568-571.

(18) Chalifour, A.; Jeannin, P.; Gauchat, J. F.; Blaecke, A.; Malissard, M.; N'Guyen, T.; Thieblemont, N.; Delneste, Y. *Blood* **2004**, *104*, 1778-1783.

(19) Becker, M. N.; Diamond, G.; Verghese, M. W.; Randell, S. H. *Journal of Biological Chemistry* **2000**, *275*, 29731-29736.

(20) Ali, R. S.; Falconer, A.; Ikram, M.; Bissett, C. E.; Cerio, R.; Quinn, A. G. *Journal of Investigative Dermatology* **2001**, *117*, 106-111.

(21) Harder, J.; Bartels, J.; Christophers, E.; Schroder, J. M. *Journal of Biological Chemistry* **2001**, *276*, 5707-5713.

(22) Harder, J.; Bartels, J.; Christophers, E.; Schroder, J. M. *Nature* **1997**, *387*, 861-861.

(23) Travis, S. M.; Singh, P. K.; Welsh, M. J. *Current Opinion in Immunology* **2001**, *13*, 89-95.

(24) Singh, P. K.; Tack, B. F.; McCray, P. B.; Welsh, M. J. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **2000**, *279*, L799-L805.

(25) Singh, P. K.; Jia, H. P.; Wiles, K.; Hesselberth, J.; Liu, L. D.; Conway, B. A. D.; Greenberg, E. P.; Valore, E. V.; Welsh, M. J.; Ganz, T.; Tack, B. F.; McCray, P. B. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 14961-14966.

(26) Ayabe, T., Satchell, D. P.; Wilson, C. L.; Parks, W. C.; Selsted, M. E.; Ouellette, A. J. *Nature Immunology* **2000**, *1*, 113-118.

(27) Wilson, C. L.; Ouellette, A. J.; Satchell, D. P.; Ayabe, T.; Lopez-Boado, Y. S.; Stratman, J. L.; Hultgren, S. J.; Matrisian, L. M.; Parks, W. C. *Science* **1999**, *286*, 113-117.

(28) Liu, L. D.; Zhao, C. Q.; Heng, H. H. Q.; Ganz, T. *Genomics* **1997**, *43*, 316-320.

(29) Sparkes, R. S.; Kronenberg, M.; Heinzmann, C.; Daher, K. A.; Klisak, I.; Ganz, T.; Mohandas, T. *Genomics* **1989**, *5*, 240-244.

(30) Hannigan, G. E.; Bayani, J.; Weksberg, R.; Beatty, B.; Pandita, A.; Dedhar, S.; Squire, J. *Genomics* **1997**, *42*, 177-179.

(31) Liu, L.; Wang, L. N.; Jia, H. P.; Zhao, C. Q.; Heng, H. H. Q.; Schutte, B. C.; McCray, P. B.; Ganz, T. *Gene* **1998**, *222*, 237-244.

(32) Mars, W. M.; Patmasiriwat, P.; Maity, T.; Huff, V.; Weil, M. M.; Saunders, G. F. *Journal of Biological Chemistry* **1995**, *270*, 30371-30376.

(33) Schutte, B. C.; Mitros, J. P.; Bartlett, J. A.; Walters, J. D.; Jia, H. P.; Welsh, M. J.; Casavant, T. L.; McCray, P. B. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 14611-14611.

(34) Motzkus, D.; Schulz-Maronde, S.; Heitland, A.; Schulz, A.; Forssmann, W. G.; Jubner, M.; Maronde, E. *Faseb Journal* **2006**, *20*, 1701-+.

(35) Lehrer, R. I.; Lichtenstein, A. K.; Ganz, T. *Annual Review of Immunology* **1993**, *11*, 105-128.

(36) Harder, J.; Meyer-Hoffert, U.; Wehkamp, K.; Schwichtenberg, L.; Schroder, J. M. *Journal of Investigative Dermatology* **2004**, *123*, 522-529.

(37) Duits, L. A.; Ravensbergen, B.; Rademaker, M.; Hiemstra, P. S.; Nibbering, P. H. *Immunology* **2002**, *106*, 517-525.

(38) Hancock, R. E. W.; Diamond, G. *Trends in Microbiology* **2000**, *8*, 402-410.

(39) Harder, J.; Meyer-Hoffert, U.; Teran, L. M.; Schwichtenberg, L.; Bartels, J.; Maune, S.; Schroder, J. M. *American Journal of Respiratory Cell and Molecular Biology* **2000**, *22*, 714-721.

(40) Chung, W. O.; Hansen, S. R.; Rao, D.; Dale, B. A. *Journal of Immunology* **2004**, *173*, 5165-5170.

(41) Tran, D.; Tran, P. A.; Tang, Y. Q.; Yuan, J.; Cole, T.; Selsted, M. E. *Journal of Biological Chemistry* **2002**, *277*, 3079-3084.

(42) Lehrer, R. I.; Szklarek, D.; Ganz, T.; Selsted, M. E. *Infection and Immunity* **1985**, *49*, 207-211.

(43) Lehrer, R. I.; Ganz, T.; Szklarek, D.; Selsted, M. E. *Journal of Clinical Investigation* **1988**, *81*, 1829-1835.

(44) Lehrer, R. I.; Barton, A.; Daher, K. A.; Harwig, S. S. L.; Ganz, T.; Selsted, M. E. *Journal of Clinical Investigation* **1989**, *84*, 553-561.

(45) Bechinger, B. *Journal of Membrane Biology* **1997**, *156*, 197-211.

(46) Shai, Y.; Oren, Z. *Peptides* **2001**, *22*, 1629-1641.

(47) Ludtke, S. J.; He, K.; Heller, W. T.; Harroun, T. A.; Yang, L.; Huang, H. W. *Biochemistry* **1996**, *35*, 13723-13728.

(48) Gazit, E.; Shai, Y. *Journal of Biological Chemistry* **1995**, *270*, 2571-2578.

(49) Gazit, E.; Boman, A.; Boman, H. G.; Shai, Y. *Biochemistry* **1995**, *34*, 11479-11488.

(50) Demchick, P.; Koch, A. L. *Journal of Bacteriology* **1996**, *178*, 768-773.

(51) Suresh, A.; Verma, C. *Bmc Bioinformatics* **2006**, *7*, -.

(52) Sahl, H. G.; Pag, U.; Bonness, S.; Wagner, S.; Antcheva, N.; Tossi, A. *Journal of Leukocyte Biology* **2005**, *77*, 466-475.

(53) Murphy, C. J.; Foster, B. A.; Mannis, M. J.; Selsted, M. E.; Reid, T. W. *Journal of Cellular Physiology* **1993**, *155*, 408-413.

(54) Charp, P. A.; Rice, W. G.; Raynor, R. L.; Reimund, E.; Kinkade, J. M.; Ganz, T.; Selsted, M. E.; Lehrer, R. I.; Kuo, J. F. *Biochemical Pharmacology* **1988**, *37*, 951-956.

(55) Chaly, Y. V.; Paleolog, E. M.; Kolesnikova, T. S.; Tikhonov, I. I.; Petratchenko, E. V.; Voitenok, N. N. *European Cytokine Network* **2000**, *11*, 257-266.

(56) Lichtenstein, A.; Ganz, T.; Selsted, M. E.; Lehrer, R. I. *Blood* **1986**, *68*, 1407-1410.

(57) Turner, J.; Cho, Y.; Dinh, N. N.; Waring, A. J.; Lehrer, R. I. *Antimicrobial Agents and Chemotherapy* **1998**, *42*, 2206-2214.

(58) Mackay, B. J.; Pollock, J. J.; Iacono, V. J.; Baum, B. J. *Infection and Immunity* **1984**, *44*, 688-694.

(59) Raj, P. A.; Marcus, E.; Sukumaran, D. K. *Biopolymers* **1998**, *45*, 51-67.

(60) Murakami, Y.; Xu, T.; Helmerhorst, E. J.; Ori, G.; Troxler, R. F.; Lally, E. T.; Oppenheim, F. G. *Oral Microbiology and Immunology* **2002**, *17*, 143-149.

(61) Gusman, H.; Travis, J.; Helmerhorst, E. J.; Potempa, J.; Troxler, R. F.; Oppenheim, F. G. *Infection and Immunity* **2001**, *69*, 1402-1408.

(62) Jensen, J. L.; Lamkin, M. S.; Oppenheim, F. G. *Journal of Dental Research* **1992**, *71*, 1569-1576.

(63) Iontcheva, I.; Oppenheim, F. G.; Troxler, R. F. *Journal of Dental Research* **1997**, *76*, 734-743.

(64) Yan, Q. Y.; Bennick, A. *Biochemical Journal* **1995**, *311*, 341-347.

(65) Cruciani, R. A.; Stanley, E. F.; Zasloff, M.; Lewis, D. L.; Barker, J. L. *Biophysical Journal* **1988**, *53*, A9-A9.

(66) Matsuzaki, K.; Harada, M.; Funakoshi, S.; Fujii, N.; Miyajima, K. *Biochimica Et Biophysica Acta* **1991**, *1063*, 162-170.

(67) Matsuzaki, K.; Murase, O.; Miyajima, K. *Biochemistry* **1995**, *34*, 12553-12559.

(68) Westerhoff, H. V.; Zasloff, M.; Rosner, J. L.; Hendler, R. W.; Dewaal, A.; Gomes, A. V.; Jongsma, A. P. M.; Riethorst, A.; Juretic, D. *European Journal of Biochemistry* **1995**, *228*, 257-264.

(69) Fox, R. O.; Richards, F. M. *Nature* **1982**, *300*, 325-330.

(70) Terwilliger, T. C.; Weissman, L.; Eisenberg, D. *Biophysical Journal* **1982**, *37*, 353-361.

(71) Bazzo, R.; Tappin, M. J.; Pastore, A.; Harvey, T. S.; Carver, J. A.; Campbell, I. D. *European Journal of Biochemistry* **1988**, *173*, 139-146.

(72) John, E.; Jahnig, F. *Biophysical Journal* **1991**, *60*, 319-328.

(73) Xiao, Y. J.; Dai, H.; Bommineni, Y. R.; Soulages, J. L.; Gong, Y. X.; Prakash, O.; Zhang, G. L. *Febs Journal* **2006**, *273*, 2581-2593.

(74) Hara, T.; Kodama, H.; Kondo, M.; Wakamatsu, K.; Takeda, A.; Tachi, T.; Matsuzaki, K. *Biopolymers* **2001**, *58*, 437-446.

(75) Anfinsen, C. B. *Science* **1973**, *181*, 223-230.

(76) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *Journal of the American Chemical Society* **2002**, *124*, 11258-11259.

(77) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 14812-14816.

(78) Carr, J. M.; Wales, D. J. *Journal of Chemical Physics* **2005**, *123*, -.

(79) Kamiya, N.; Mitomo, D.; Shea, J. E.; Higo, J. *Journal of Physical Chemistry B* **2007**, *111*, 5351-5356.

(80) Headgordon, T.; Stillinger, F. H. *Biopolymers* **1993**, *33*, 293-303.

(81) Dressel, F.; Kobe, S. *Chemical Physics Letters* **2006**, *424*, 369-373.

(82) Khandelia, H.; Kaznessis, Y. N. *Peptides* **2005**, *26*, 2037-2049.

(83) Khandelia, H.; Kaznessis, Y. N. *Journal of Physical Chemistry B* **2005**, *109*, 12990-12996.

(84) Berneche, S.; Nina, M.; Roux, B. *Biophysical Journal* **1998**, *75*, 1603-1618.

(85) Pasenkiewicz-Gierula, M.; Murzyn, K.; Rog, T.; Czaplewski, C. *Acta Biochimica Polonica* **2000**, *47*, 601-611.

(86) Lin, J. H.; Baumgaertner, A. *Biophysical Journal* **2000**, *78*, 1714-1724.

(87) Sachs, J. N.; Tieleman, D. P.; Engelman, D. M. *Biophysical Journal* **2007**, 424a-424a.

(88) Faramarz Mehrnejad, H. N.-M., Bijan Ranjbar *PROTEINS: Structure, Function, and Bioinformatics* **2007**, *67*, 931-940.

(89) Iovino, M.; Falconi, M.; Marcellini, A.; Desideri, A. *Journal of Peptide Research* **2001**, *58*, 45-55.

(90) Alexander, C.; Rietschel, E. T. *Journal of Endotoxin Research* **2001**, *7*, 167-202.

(91) Beveridge, T. J. *Journal of Bacteriology* **1999**, *181*, 4725-4733.

(92) Vollmer, W.; Holtje, J. V. *Journal of Bacteriology* **2004**, *186*, 5978-5987.

(93) Barnickel, G.; Labischinski, H.; Bradaczek, H.; Giesbrecht, P. *European Journal of Biochemistry* **1979**, *95*, 157-165.

(94) Luderitz, O.; Galanos, C.; Risse, H. J.; Ruschman.E; Schlecht, S.; Schmidt, G.; Schulteh.H; Wheat, R.; Westphal, O.; Schlossh.J *Annals of the New York Academy of Sciences* **1966**, *133*, 349-&.

(95) Kim, Y. B.; Watson, D. W. *Journal of Bacteriology* **1967**, *94*, 1320-&.

(96) Kasai, N.; Nowotny, A. *Journal of Bacteriology* **1967**, *94*, 1824-&.

(97) Galanos, C. *Zeitschrift Fur Immunitats-Forschung Experimentelle Und Klinische Immunologie* **1975**, *149*, 214-229.

(98) Wiemann, B.; Starnes, C. O. *Pharmacology & Therapeutics* **1994**, *64*, 529-564.

(99) Beacham, I. R. *International Journal of Biochemistry* **1979**, *10*, 877-883.

(100) Holst, O.; Ulmer, A. J.; Brade, H.; Flad, H. D.; Rietschel, E. T. *Fems Immunology and Medical Microbiology* **1996**, *16*, 83-104.

(101) Nummila, K.; Kilpelainen, I.; Zahringer, U.; Vaara, M.; Helander, I. M. *Molecular Microbiology* **1995**, *16*, 271-278.

(102) Guo, L.; Lim, K. B.; Gunn, J. S.; Bainbridge, B.; Darveau, R. P.; Hackett, M.; Miller, S. I. *Science* **1997**, *276*, 250-253.

(103) Gibbons, H. S.; Kalb, S. R.; Cotter, R. J.; Raetz, C. R. H. *Molecular Microbiology* **2005**, *55*, 425-440.

(104) Guo, L.; Lim, K. B.; Poduje, C. M.; Daniel, M.; Gunn, J. S.; Hackett, M.; Miller, S. I. *Cell* **1998**, *95*, 189-198.

(105) Ernst, R. K.; Yi, E. C.; Guo, L.; Lim, K. B.; Burns, J. L.; Hackett, M.; Miller, S. I. *Science* **1999**, *286*, 1561-1565.

(106) Ernst, R. K.; Guina, T.; Miller, S. I. *Journal of Infectious Diseases* **1999**, *179*, S326-S330.

(107) Rietschel, E. T.; Brade, H.; Holst, O.; Brade, L.; MullerLoennies, S.; Mamat, U.; Zahringer, U.; Beckmann, F.; Seydel, U.; Brandenburg, K.; Ulmer, A. J.; Mattern, T.; Heine, H.; Schletter, J.; Loppnow, H.; Schonbeck, U.; Flad, H. D.; Hauschildt, S.; Schade, U. F.; DiPadova, F.; Kusumoto, S.; Schumann, R. R. *Pathology of Septic Shock* **1996**, *216*, 39-81.

(108) Droge, W.; Lehmann, V.; Luderitz, O.; Westphal, O. *European Journal of Biochemistry* **1970**, *14*, 175-&.

(109) Brade, H.; Galanos, C. *Infection and Immunity* **1983**, *42*, 250-256.

(110) McCoy, A. J.; Liu, H. J.; Falla, T. J.; Gunn, J. S. *Antimicrobial Agents and Chemotherapy* **2001**, *45*, 2030-2037.

(111) Pridmore, A. C.; Wyllie, D. H.; Abdillahi, F.; Steeghs, L.; van der Ley, P.; Dower, S. K.; Read, R. C. *Journal of Infectious Diseases* **2001**, *183*, 89-96.

(112) Moran, A. P.; Prendergast, M. M.; Appelmelk, B. J. *Fems Immunology and Medical Microbiology* **1996**, *16*, 105-115.

(113) Ferguson, A. D.; Hofmann, E.; Coulton, J. W.; Diederichs, K.; Welte, W. *Science* **1998**, *282*, 2215-2220.

(114) Ferguson, A. D.; Welte, W.; Hofmann, E.; Lindner, B.; Holst, O.; Coulton, J. W.; Diederichs, K. *Structure* **2000**, *8*, 585-592.

(115) Neuhaus, F. C.; Linzer, R.; Reusch, V. M. *Annals of the New York Academy of Sciences* **1974**, *235*, 502-518.

(116) Baddiley, J. *Essays Biochem* **1972**, *8*, 35–77.

(117) Peschel, A. *Trends in Microbiology* **2002**, *10*, 179-186.

(118) Campopiano, D. J.; Clarke, D. J.; Polfer, N. C.; Barran, P. E.; Langley, R. J.; Govan, J. R. W.; Maxwell, A.; Dorin, J. *The Journal of Biological Chemistry* **2004**, *279(47)*, 48671-48679.

(119) Schibli, D. J.; Hunter, H. N.; Aseyev, V.; Starner, T. D.; Wiencek, J. M.; McCray, P. B.; Tacks, B. F.; Vogel, H. J. *The Journal of Biological Chemistry* **2002**, *277(10)*, 8279-8289.

(120) White, D. *The physiology and biochemistry of Prokaryotes* **2006**. Oxford University Press.

Chapter Three

General Methodology

3.1. Introduction

Chapter Two outlines the β-defensin problem and the background needed to understand it. This Chapter discusses the techniques used to tackle the problem. Molecular mechanical approximations are outlined, as well as a repertoire of rational structure prediction algorithms and free-energy methods. Unless otherwise stated, the methods are implemented with the *AMBER 9*[1] force field.

In an ideal world of infinite computational resources researchers would have the luxury of solving protein folding (and interaction) problems with high-level quantum mechanical treatments. The reality is that only the smallest molecules (less than one hundred atoms) can conceivably be tackled in this fashion, and often on time scales that are biologically irrelevant. The molecular modeller is thus forced to make drastic simplifications. Frequently these are:

1. Ignore any electronic fluctuations: instead, calculate an approximate charge for each atomic centre based on a single, plausible conformation and assume that this fixed charge approximates the true, time-averaged electronic distribution (*Section 3.2*).
2. Treat atoms as semi-hard spheres of differing radii (*Section 3.3*).
3. Approximate the interactions of these atoms with one another as the sum of energetic contributions from van der Waals, electrostatic interactions, dihedral twisting and bond angle terms (*Section 3.3*)

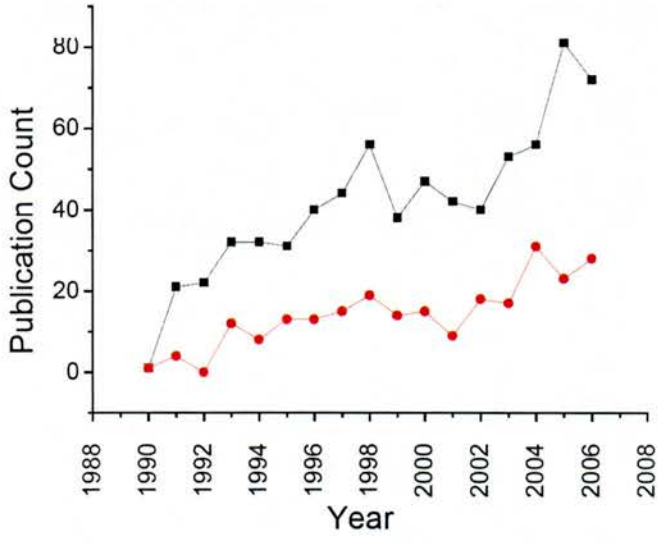
In order to achieve *step 3* the modeller requires a set of point charges for each of the atomic centres (to calculate the electrostatic contributions), and a set of constants of proportionality and atomic radii (to describe the remaining interactions).

3.1.1. Choice of forcefield

The AMBER force field and software implementation is chosen for this research for two principle reasons:

1. AMBER is widely used (*Figure 3.1.1.1*) and has in many cases provided excellent agreement with experiment.
2. The generalised AMBER force field (GAFF) in conjunction with the "Antechamber" procedure allows novel residues to be developed in an easy, transparent (and often reproducible) manner[2-5].

Figure 3.1.1.1. *Publications within the "Web of science" database listed under the searches of "AMBER force-field" (black) and "CHARMM force-field" (red) between 1990 and 2006.*



3.2. Obtaining point charges: The RESP procedure

The electrostatic properties of a molecule are intricately connected to its distribution of electrons and nuclei. The challenge is to simplify the dynamic electrostatic potential of this complex system into a single, point-charge representation of a static state. Unfortunately however, a static charge state cannot be calculated unambiguously from a quantum mechanical wavefunction and the modeller is thus forced to:

1. Calculate the electrostatic potential for a given system in a given configuration using quantum mechanical techniques.
2. Sample points from this potential and balance the corresponding charge arising from these points with a new, single point-charge.

There are many ways of performing these two steps and there is no agreement as to which strategy is most representative of a real system. The strategy used in this research is that of the *RESP* procedure implemented in *AMBER 9*[6].

3.2.1. Deriving charges from the molecular electrostatic potential

The electrostatic potential at a point in space is defined as:

"*The work done in bringing a unit positive charge from infinite displacement to that point.*"

Under this definition nuclei generate a positive (repulsive) force, whilst electrons generate a negative (attractive) force. A value for the electrostatic potential, φ, in a given configuration can be obtained from the wavefunction for the system[7]:

$$\phi(\mathbf{r}) = \phi_{nucl}(\mathbf{r}) + \phi_{elec}(\mathbf{r}) = \sum_{A=1}^{M} \frac{Z_A}{\mathbf{r} - R_A} - \int \frac{d\mathbf{r}' \rho(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} \qquad (1)$$

Where **r** represents a given point in Cartesian space for a set of M nuclei of charge Z_A (and displacement R_A) interacting with a set of electrons with distribution density ρ.

Equation 1 is a continuous function for which there is no analytic solution. For the purpose of modelling it is necessary to represent this function as a collection of discrete points, the superposition of which approximate the underlying behavior. One such widely used method for doing this is the Restrained Electrostatic Potential fit procedure[6,8].

3.2.2. Restrained Electrostatic Potential (RESP) charge derivation

The RESP procedure is one of a myriad of tools used to bridge the gap between quantum and Newtonian mechanics. The method offers a means of mapping the quantum mechanically derived electrostatic potential onto a set of point charges for each of the atomic centres.

a) The charge fitting procedure[6,8-10]

The charge fitting process involves spherically encasing a set of j atoms, each with a set of i points. The electrostatic potential is then computed for each of the atomic centres such that a distant dependent charge distribution is generated. A single point charge q_j is subsequently fitted to the centre of each of the j atoms such that the initial electrostatic potential arising from the i points is exactly balanced. This can be conveyed mathematically as a least squares problem:

$$R^2 = \sum_i \left[V_i^{initial} - V_i^{calculated} \right]^2 \qquad (2)$$

Where R is the objective function, $V^{initial}$ is the initial set of discretised points and $V^{calculated}$ is the electrostatic potential arising after the point charge q_j has been fitted:

$$V_i^{calculated} = \sum_j \frac{q_j}{r_{ij}} \qquad (3)$$

At the minimum point of the function,

$$\frac{\partial \left(R^2 \right)}{\partial q_j} = 0 \qquad (4)$$

where,

$$\frac{\partial \left(R^2 \right)}{\partial q_j} = -2 \sum_i \left[\frac{V_i^{initial} - V_i^{calculated}}{r_{ij}} \right] \qquad (5)$$

Equation 5 above can be recast in an appropriate matrix representation[11] and solved for q.

b) Adding restraints

A traditional unrestrained least squares fit of the charge problem (as described above) frequently overestimates the bond strength in species involving hydrogen bonding[6]. This overestimation can effectively be reduced by including a hyperbolic penalty function within the fitting procedure:

$$R^2_{adjusted} = R^2_{original} + R^2_{restrnt} \qquad (6)$$

Where the hyperbolic function R takes on the form:

$$R^2_{rest} = k_{rest} \sum_j \left[\left(q_i^2 + b^2 \right)^{\frac{1}{2}} - c \right] \qquad (7)$$

The "tightness" of the hyperbola and its minimum are determined by b and the extent to which the restraint affects the overall fit is given by k_{rest}. The restrained electrostatic fit is then solved in matrix form in the same manner as the normal ESP method.

c) Calculating the constants of proportionality for the forcefield equation

The *AMBER* suite of programs contains a "generalized force field" in which the parameters for many atomic connectivities have been obtained empirically or calculated[2,12]. Thus, for many novel systems (for which there are no "library" residues) one important challenge is to assign the charge state.

3.3. The forcefield equation[1]

The sum of atomic interactions can be expressed mathematically as:

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedral} \frac{V_n}{2} [1 + \cos(n\varphi - \gamma)] + \sum_{\substack{non-bonded \\ interactions \\ i>j}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{DR_{ij}} \right] \qquad (8)$$

The term E_{total} represents the energy of the system. The term $\sum_{bonds} K_r(r-r_{eq})^2$ represents the stretching of covalent bonds from the equilibrium bond length r_{eq} with a force constant K_r. The term $\sum_{angles} K_\vartheta(\theta-\theta_{eq})^2$ refers similarly to displacement from the equilibrium angle θ. The term $\sum_{dihedrals} \frac{V_n}{2}[1+\cos(n\phi-\gamma)]$ encompasses the twisting of angles from the equilibrium position φ. The term $\sum \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}$ models the van der Waals interactions between atoms i and j, and $\sum_{i<j} \frac{q_i q_j}{DR_{ij}}$ models electrostatic interactions between point charges, q_i and q_j.

3.3.1. Bond energy

This energy term is approximated as a quadratic function. The trough of the function corresponds to the equilibrium bond length and displacement from the equilibrium results in the energy of the system increasing exponentially. In reality, increasing the distance between bonded atoms past a certain point results in diminished attraction, and as such the quadratic function is only valid when considering small displacements from the equilibrium.

3.3.2. Bond angle

The bond angle function measures the extent to which energy varies as the bond angle between two directly connected atoms changes. This is a quadratic function that describes fluctuations accurately about small displacements from equilibrium.

3.3.3. Dihedral bond angle

The dihedral bond angle is a measure of twisting about the equilibrium bond. This function varies with the cosine of the dihedral angle. The energy function of the system accordingly oscillates between a maximum and minimum with every 360 degree change in angle.

3.3.4. Van der Waals and electrostatic interactions

The remaining interactions are approximated with the final term of *equation 8*. Whilst it captures good general behaviour of atomic interactions, it tends to exaggerate attractive forces in the region of the minimum. The choice of dielectric constant employed to calculate the interaction energy depends on the environment of the simulation. Values range from approximately 80 in water, to 1-4 in the interior of a protein.

3.3.5. The Generalized Born model[14-19]

Explicit water particles can often make simulations prohibitively time consuming. In many cases it is far more computationally efficient to replace discrete water molecules with a continuum that possesses the bulk properties of water. Continuum approaches involve approximating the solvation free energy by assuming that it can be decomposed into "electrostatic" and "non-electrostatic" terms:

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{np} \qquad (9)$$

Where ΔG_{solv} is the solvation free energy, and where ΔG_{elec} and ΔG_{np} refer to the electrostatic and non-electrostatic components respectively.

a) The non-electrostatic term, ΔG_{np}

This term is assumed to result from van der Waals interactions between solvent and solute combined with the unfavourable cost of disturbing the bulk water environment. Within AMBER, ΔG_{np} is taken to be proportional to the total solvent accessible surface area (SA) of the molecule, with a constant of proportionality that is derived experimentally[13].

b) The electrostatic term, ΔG_{elec}

The electrostatic component of the solvation free energy is obtained from the difference in electrostatic potential between the solvated and de-solvated forms:

$$\Delta G_{elec} = \frac{1}{2}\sum_i q_i(\phi_{sol}(\mathbf{r}_i) - (\phi_{vac}(\mathbf{r}_i)) \qquad (10)$$

Where ϕ_{sol} and ϕ_{vac} refer to the electrostatic potential of solvated and gas-phase atomic partial charges, q_i, for a position vector **r**. Partial charges in the interior of the molecule are assigned a dielectric constant of 1, and those that are solvent exposed assigned a value equal to that of water.

The electrostatic potential can be solved to arbitrary accuracy by numerically solving the Poisson-Boltzmann (PB) equation[14]. Whilst accurate, this numerical method is computationally very expensive and instead, a fast, analytical generalized Born[15-19] (GB) approximation to the PB equation is often used:

$$\Delta G_{elec} \approx \Delta G_{GB} \qquad (11)$$

Where

$$\Delta G_{GB} = -\sum_i \frac{q_i^2}{2R_i}\left(1-\frac{1}{\varepsilon_W}\right) - \frac{1}{2}\sum_{ij,i\neq j}\frac{q_i q_j}{f^{GB}\left(r_{ij}, R_i, R_j\right)}\left(1-\frac{1}{\varepsilon_W}\right) \qquad (12)$$

The term r_{ij} refers to the distance between atoms i and j, and the term R refers to the *Born Radius*, a measure of the degree to which an ion is buried in the molecular interior. Thus, deeply buried atoms have large effective radii, whilst those closer to the exterior have small radii. In the limit of complete solvation, the Born Radius reduces to the van der Waals radius.

The function f_{GB} varies in accordance with the Born radius and inter-atomic distance as follows[1]:

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp\left(\frac{-r_{ij}^2}{4R_i R_j}\right)\right]^{\frac{1}{2}} \qquad (13)$$

In summary the GB approximation is significant for three reasons:

1. It increases the rate of phase space sampling because the viscosity of water is not considered.
2. The rate of simulation is accelerated because explicit water interactions are ignored.
3. Good agreement has often been observed between GB and explicitly solvated and experimental models[19-23].

One downside of the GB approximation is that it does not implicitly consider hydrogen bonding. Thus, in systems where hydrogen bonding dominates there is a danger that the GB model fails to represent adequately the non-electrostatic solvation term. In this thesis the GB model is reserved for rational structure prediction, where rapid phase space sampling is given higher priority than accurately modelling each visited state.

3.4. Molecular Dynamics (MD) simulations

Equation 8 sets out an approximation for the interaction energy of the molecular system. It is frequently useful however, to know how these interactions direct the time evolution of the system.

3.4.1. Newton's equation of motion[24]

The differential of the potential energy function with respect to the parameters (coordinates, bond angle, dihedral angle) determines the force acting on the system in any given configuration. The acceleration that results from the applied force is described by Newton's second law of motion:

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \qquad (14)$$

In which a particle i, of mass m_i moves under a force F in the x direction.

Equation 8 describes a coupled system in which changing the position of one atom effects the interactions experienced by each of the other atoms. Coupled systems of this nature cannot be

solved with standard analytical techniques and the modeller is forced (instead) to resort to *finite difference* techniques.

In the finite difference approach, the integration process is split into a number of small stages, each of which is separated in time by a fixed quantity, δt. Knowledge of the positions and velocities of a particle at a time t allows a subsequent set of positions and velocities to be established at a time ($t + \delta t$). Forcefield software assumes that the acceleration, velocity and positions of the particles in the system can be approximated using a Taylor series expansion[24]:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) + \frac{1}{6}\delta t^3\mathbf{b}(t) + \ldots \quad (15)$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \delta t\mathbf{a}(t) + \frac{1}{2}\delta t^2\mathbf{b}(t) + \frac{1}{6}\delta t^3\mathbf{c}(t) + \ldots \quad (16)$$

$$\mathbf{a}(t+\delta t) = \mathbf{a}(t) + \delta t\mathbf{b}(t) + \frac{1}{2}\delta t^2\mathbf{c}(t) + \ldots \quad (17)$$

$$\mathbf{b}(t+\delta t) = \mathbf{b}(t) + \delta t\mathbf{c}(t) + \ldots \quad (18)$$

In the equations above, v refers to velocity, t to time, a to the first derivative (acceleration) and b to the second derivative (the rate of change of acceleration). Thus the velocity and position of a particle at a time frame ($t+\delta t$) can be calculated from the position of that same particle at the previous time (t). The integral is solved within AMBER using the velocity Verlet integration algorithm[1].

The potential energy surface that arises from the particle in the new position ($t+\delta t$) permits a new set of forces to be calculated for the system. Repeated cycles of force evaluation followed by integration allow a number of time dependent trajectories to be simulated for the system.

3.4.2. Temperature regulation

There are two approaches applied in this research to the problem of temperature regulation. The first is the Langevin[25-27] dynamics approach in which adjustments to temperature are based on randomized collisions with imaginary solvent particles. The second is the Berendsen algorithm[28], in which the system is coupled to an imaginary heat bath.

a) Langevin dynamics[25-27]

In the Langevin model the force acting on a given particle is assumed to arise from three sources; interactions with other particles (via the normal force-field equation), friction (arising from solvent viscosity) and random collisions with surrounding solvent particles. The Langevin equation therefore expresses the relationship between these competing forces as follows:

$$m_i \frac{d^2 x_i(t)}{dt^2} = \mathbf{F}_i\{x_i(t)\} - \mathbf{F}_i^{fric}\{x_i(t)\} + \mathbf{R}_i(t) \qquad (19)$$

Where i denotes a partcle of mass m and position x at time t. The term $\mathbf{F}$ is the force arising from the normal force-field equation, the term $\mathbf{F}^{fric}$ is the Frictional component:

$$\mathbf{F}_i^{fric}\{x_i(t)\} = -\gamma_i \frac{dx_i(t)}{dt} m_i \qquad (20)$$

In which γ is the collision frequency[26] and the term **R**, the random collision component. The Langevin equation within AMBER is integrated with a simple leap-frog strategy[27].

b) The Berendsen thermostat[28]

The Berendsen thermostat regulates temperature by periodically rescaling the velocities of the particles within the system so that the average temperature over many time steps is pushed towards a "target" temperature. Thus, during each iteration of the Berendsen algorithm, the "new" velocities, v^{new}, are equal to the factor-modified "old" velocities, v^{old}, as follows:

$$v_i^{new} = \lambda v_i^{old} \qquad (21)$$

Where λ is a scaling factor based on the difference between the "actual" ($T_{(t)}$) and "desired" (T_0) temperature of the system, as well as the extent of coupling to an imaginary heat bath (μ). Thus,

$$\lambda = \left[1 + \frac{\Delta t}{\mu_T}\left(\frac{T_o}{T(t)} - 1\right)\right]^{\frac{1}{2}} \qquad (22)$$

In which Δt refers to the integration time step. The smaller μ is, the tighter is the coupling to the heat bath.

The temperature of the system is calculated directly from the kinetic energy of the system as follows:

$$T(t) = \frac{2E_K(t)}{Dk_B} \tag{23}$$

Where D is the number of degrees of freedom and k_B is the Boltzmann constant.

c) Which temperature regulator is best?

The Berendsen thermostat, due to its simplistic velocity rescaling procedure, relies on atomic collisions to equilibrate velocity within the simulation box. This has the disadvantage that "hot-spots" can occur in isolated areas, particularly with explicit solvent simulations. The random collisional component of the Langevin method prevents the occurrence of these "hot-spots" and is thus the preferred temperature method for explicit solvent simulations.

3.5. Energy minimization

It is frequently necessary to find a local minimum in configuration space in algorithms such as "Basin Hopping" (*Section 3.6.2*), or when generating initial structures from which to begin MD simulations. The challenge with minimization is to find a set of coordinates that satisfy the condition:

$$\frac{\partial E}{\partial x_i} = \frac{f(x+\delta x) - f(x)}{\delta x} = 0 \tag{24}$$

3.5.1. Steepest descent

The steepest descent method is an iterative process in which the positions of the atoms in the system are progressively altered until a minimum is found. The updated atomic positions are obtained during each iteration through a vector translation $\boldsymbol{S}_k$ as follows:

$$\mathbf{s}_k = \frac{-\mathbf{g}_k}{|\mathbf{g}_k|} \tag{25}$$

Where $\boldsymbol{g}_k$ represents the potential energy gradient.

Each new $\boldsymbol{S}_k$ vector specifies a direction perpendicular to the previous vector, and whilst this is beneficial for initial minimization, the algorithm is inefficient for many cycles.

Once a direction in which to move the particles in the system has been established, it is then necessary to consider how far in this direction the system should go. This is typically done by a line search or an arbitrary step approach.

3.5.2. Conjugate gradient descent

The conjugate gradient descent algorithm is computed as follows:

$$\mathbf{v}_{k+1} = -\mathbf{g}_k + \lambda_k \mathbf{v}_k \tag{26}$$

Where $\mathbf{v}_{k+1}$ refers to the new direction in which to adjust the coordinates, $\mathbf{X}_k$ is the starting configuration, $\mathbf{g}_k$ the potential energy gradient, and $\mathbf{v}_k$ the previous conjugate gradient direction. The function thus gives rise to a direction vector that is weighted to the previous vector by a constant of proportionality.

3.5.3. Which method is best?

The steepest descent method is often employed during the first few cycles of minimization because it can descend high energy 'hills' more rapidly than a conjugate gradient method. Following this "initial" minimization, the conjugate gradient method is often used because it avoids inefficient "zig-zagging".

3.6. Conformational sampling

Chapter Four and Chapter Five deal with rational structure prediction. This section outlines some of the existing structure prediction methods used in this research.

3.6.1. Simulated Annealing (SA)[29]

The principle of simulated annealing is this:

1. Begin with a molecule in a given configuration. Simulate the dynamics of the molecule at a high temperature (often ~ 800-2000K) for a given period of time.
2. Progressively cool the molecule to 0K.
3. Obtain an output structure and use this as a starting structure for step 1.
4. Repeat steps 1-3 for a given number of cycles.

At high temperatures, the molecule is presumed to sample a large potential energy surface. As the molecule cools it begins to assume a shape that reflects an increasingly restricted potential energy landscape. At the final temperature (0 K) the molecule is presumed to have collapsed into a structure that corresponds to a minimum energy pocket. It is hoped that this energy pocket

corresponds to (or lies close to) a global energy minimum for the system that maps to a biologically meaningful three-dimensional structure.

3.6.2. Basin Hopping (BH)[30-34]

The principle of BH is similar to that of a traditional Monte Carlo (MC)[54] algorithm in the sense that changes to configuration space occur randomly, abruptly and in a time independent manner. However, a key difference with the BH model is that the potential energy space is reconfigured to a transformed energy landscape in which the energy value, E, at each point in configuration space, X, is set to the value for the minimized structure starting from that point[30]. This transformed energy can be described succinctly as[30]:

$$\widetilde{E}(X) = \min\{E(X)\} \tag{27}$$

Where "min" refers to the minimized energy starting from the configuration X.

Thus the "transformation" converts an initial bumpy and undulating landscape into a set of "square" energy minima (*Figure 3.7.2.1*).

Transitions in configuration space are then attempted using a metropolis-type criterion within the framework of the transformed surface with the advantage that high energy states are less penalised by Boltzmann weighting than ordinary MC attempts (*Figure 3.7.2.2*).

Figure 3.7.2.1 *A schematic representation of the transformed Energy surface (dotted line) shown in comparison to the non-transformed surface (solid line).Adapted from Wales*[30]

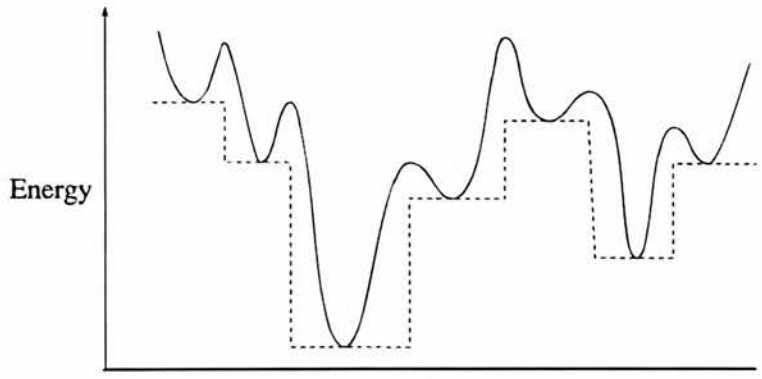
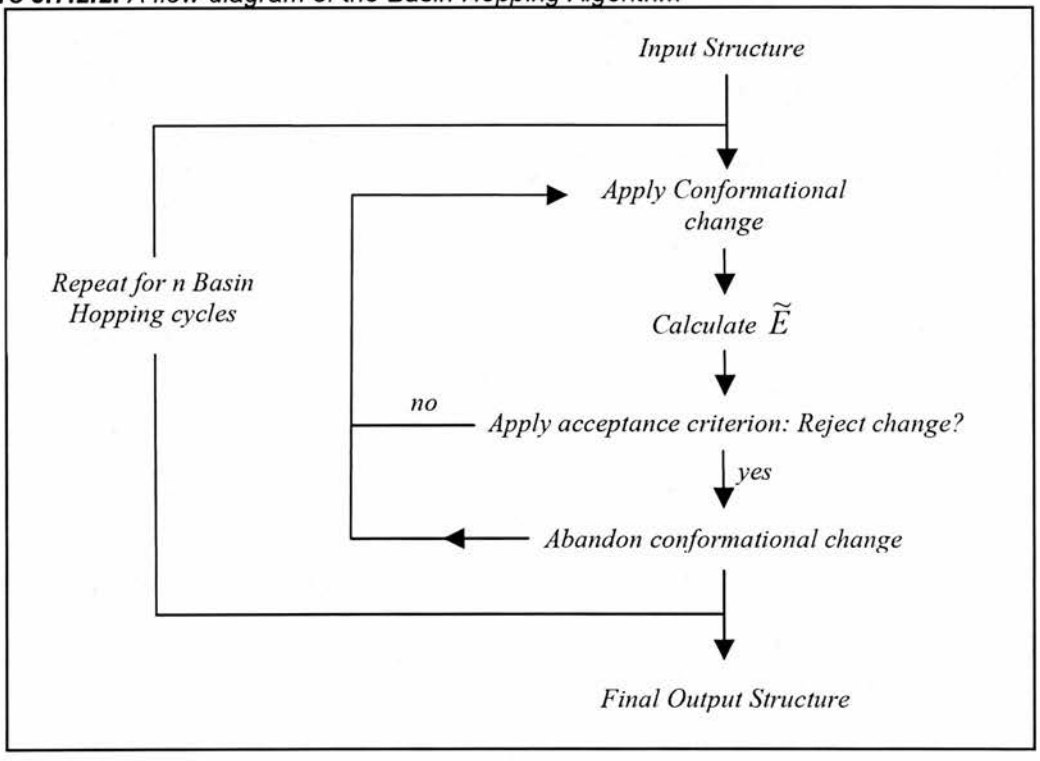


Figure 3.7.2.2. *A flow diagram of the Basin Hopping Algorithm*



3.6.3. Replica Exchange (RE)

One pitfall of conventional MD and MC methods is that configurations often become trapped in local regions of PE space. Such drawbacks are inherent to canonical ensemble simulations because each state exists only in relation to an energy dependent Boltzmann distribution. In a Replica Exchange (RE) Simulation, a number of simulations are performed concurrently at different target temperatures[35-38]. After a specified number of MD steps, each simulation is afforded the opportunity of swapping its atomic configuration with that of a neighbouring simulation with a probability determined by a metropolis-type function. Configurations that would otherwise be trapped within a local energy pocket are thus given the opportunity to perform a random walk to a new and unexplored region of PE space.

a) RE theory[35]

An RE system consists of M non-interacting copies (or replicas) at M-different temperatures (T_m) within a canonical ensemble. Each state, **X**, (describing all the replicas at any given moment) can be thought of in terms of the coordinates (q) and momenta (p) of each of the associated replicas. Mathematically **X** can be stated as:

$$x_n = (q_i, p_i) \qquad (i = 1 \rightarrow m) \qquad (28)$$

Where the subscript i refers to the ith replica (of M possible replica states) and the value n refers to the nth state, x.

b) Describing the RE system[35-38]

Each replica within every state **X** exists in isolation. Accordingly, the weight required to specify succinctly each state is merely the product of individual Boltzmann factors associated with each of the M replicas.

The weight factor for each state **X** is:

$$W_{X_n} = \exp\left\{-\sum_{i=1}^{M}\frac{1}{k_B T_i}\mathrm{H}(q_i, p_i)\right\} \qquad (29)$$

Where k_B refers to the Boltzmann factor and H refers to the Hamiltonian for the kinetic and potential energy terms of the system for a configuration q and set of momenta p.

In the instance of an exchange event, a pair of replicas within a state **X** swap their coordinate configurations. The momenta associated with the swap must be rescaled to suit the new temperature environment. The procedure is demonstrated in the following simple example.

Consider two replicas with the following attributes:

Replica 1: $R_1 = (q_1, p_1)$ (T = 300 K)

Replica 2: $R_2 = (q_2, p_2)$ (T = 250 K)

In the event of an exchange, replica R_1 adopts the configuration of replica R_2 and replica R_2 adopts the configuration of replica R_1. The momenta information associated with each configuration is then be rescaled as follows:

$$\text{Replica 1: } P_1^{new} = \sqrt{\frac{300}{250}} \times p_2 \qquad (30)$$

$$\text{Replica 2: } P_2^{new} = \sqrt{\frac{300}{250}} \times p_1 \qquad (31)$$

Where P_1^{new} refers to replica 1 that has adopted the scaled momenta associated with replica 2 (P_2) and where P_2^{new} refers to replica 2 that has adopted the scaled momenta associated with replica 1 (P_1).

After the momenta reassignments each replica exists in terms of contributions from both the coordinates and momenta as follows:

$$\text{Replica 1: } R_1 = (q_2, \sqrt{\frac{300}{250}} \times p_2) \qquad (32)$$

$$\text{Replica 2: } R_2 = (q_1, \sqrt{\frac{250}{300}} \times p_1) \qquad (33)$$

c) Determining the exchange probability

The transition probability $a_{(\mathbf{x}_i \to \mathbf{x}_j)}$ for an exchange between 2 replicas in a state **X** is determined in relation to the **X** weighting factor. The probability of exchange from a

state *T=300* to a state *T=200* must be equal to the exchange probability of the state *T=200* to *T=300* and as such the following mathematical condition must be fulfilled[35]:

$$W_{\mathbf{x}_i} a_{(\mathbf{x}_i \to \mathbf{x}_j)} = W_{\mathbf{x}_j} a_{(\mathbf{x}_j \to \mathbf{x}_i)} \tag{34}$$

In which $a_{(\mathbf{x}_i \to \mathbf{x}_j)}$ Refers to a transition probability from a state **X***i* to a state **X***j*.

Substituting for W_x, the rescaled velocities and the Hamiltonian energy for the system the following transition condition is obtained[35]:

$$\frac{a_{(\mathbf{x}_i \to \mathbf{x}_j)}}{a_{(\mathbf{x}_j \to \mathbf{x}_i)}} = \exp\left\{-\left(\beta_j - \beta_i\right)\left(E\left(q_i\right) - E\left(q_j\right)\right)\right\} \tag{35}$$

$$= \exp\left\{-\Delta\right\}$$

where $\Delta = \left(\beta_j - \beta_i\right)\left(E\left(q_i\right) - E\left(q_j\right)\right)$

and Where $\beta_i = \dfrac{1}{K_B T_i}$

This transition condition (a) can be satisfied with the Metropolis criterion:

$$a_{(\mathbf{x}_i \to \mathbf{x}_j)} = 1 \qquad (\text{for } \Delta \le 0)$$

$$a_{(\mathbf{x}_i \to \mathbf{x}_j)} = \exp\left\{-\Delta\right\} \qquad (\text{for } \Delta > 0)$$

The transition criterion describes an exchange probability that decays exponentially with increasing temperature disparity. Exchanges are therefore most likely to occur between states that are immediately adjacent in temperature space. The temperature of each simulation must be chosen carefully; too great a difference will result in limited exchange events, whilst too small a difference will result in exchange between configurations trapped in the same region of space.

d) Concluding remarks

RE simulations are particularly suitable for parallelised computer runs because each replica can be simulated by an individual processor[35]. Thus given a parallel system, the rate at which a given PE surface can be sampled is increased by a factor proportional to the number of processors in use. Many standard MD and MC techniques are difficult to parallelise because simulation setup requires clumsy information flow between computing nodes (for which information assimilation and exchange is frequently a limiting performance factor.) RE is therefore a methodology that provides for both better and faster sampling than the traditional MD and MC methodologies.

3.7. Free energies of binding

Chapter Seven deals with molecular docking of β-defensins with candidate binding partners. This section accordingly outlines the methodology for computing the binding-free energy of docked structures. The method of choice is that of the approximate MM-PBSA approach developed by Kollman *et al.* in 2000[39], and implemented in AMBER 9.

3.7.1. The MM-PBSA method[39]

The MM-PBSA method, whilst prone (in principle) to error has in practice produced good agreement with experiment, particularly with regards to binding free energies[40-49]. For example, Kollman *et al.* report that the MM-PBSA method accurately predicts the absolute binding free energy of an HIV reverse transcriptase inhibitor, TIBO, with calculated values of -13.2 kcal/mol, versus -11.6 kcal/mol for the experimental procedure[42].

With the MM-PBSA approach, a typical MD simulation is first run with the full set of MD ingredients; explicit solvation, counter ions and periodic boundary conditions. A set of representative snapshots of the molecular trajectory are then extracted and each molecule of the snapshot stripped of surrounding water molecules and counter-ions. The free energy of each trajectory is then approximated with a Poisson-Boltzmann surface area[50] term that replaces the explicit water as follows:

$$\overline{G} = \overline{E}_{MM} + \overline{G}_{PBSA} - TS_{MM} \qquad (36)$$

Where $\overline{G}$ is the calculated average free energy and where $\overline{E}_{MM}$ is the average molecular mechanical energy:

$$\overline{E}_{MM} = \overline{E}_{bond} + \overline{E}_{angle} + \overline{E}_{tors} + \overline{E}_{vdw} + \overline{E}_{elec} \qquad (37)$$

(The subscript terms on the right hand side refer to bond, angle, torsional, van der Waals and electrostatic terms respectively[51])

$\overline{G}_{PBSA}$ refers to the free energy of solvation, which can itself be split into two terms:

$$\overline{G}_{PBSA} = G_{PB} + G_{NP} \qquad (38)$$

Where G_{PB} is the electrostatic contribution to the solvation free energy[52], obtained through a numerical solution of the Poisson–Boltzmann (PB) equation, and where G_{NP} is the non-polar contribution, proportional to the solvent-exposed surface area[52,53]. The term S_{MM} refers to the solute entropy which, in AMBER 9, is obtained with normal mode analysis[51].

3.7.2. Calculating binding free energies

The free energy of association of two molecules, A and B, can now be expressed as follows:

$$A + B \rightarrow AB \qquad (39)$$

Where A and B refer to the free species, and AB to the bound species.

A thermodynamic cycle (*Figure 3.7.2.1*) of the process relates this simple expression to the molecular mechanical properties of the system as follows[42]:

$$\begin{array}{ccccc} A_{aqu} & + & B_{aqu} & \xrightarrow{\Delta G_{binding}} & AB_{aqu} \\ \downarrow -\Delta G^{A}_{solv} & & \downarrow -\Delta G^{B}_{solv} & & \downarrow \Delta G^{AB}_{solv} \\ A_{gas} & + & B_{gas} & \xrightarrow{\Delta G_{gas}} & AB_{gas} \end{array}$$

Where,

$$\begin{aligned}\Delta G_{binding} &= \Delta G_{gas} - \Delta G^{A}_{solv} - \Delta G^{B}_{solv} + \Delta G^{AB}_{solv} \\ &= \Delta H_{gas} - T\Delta S - \Delta G^{A}_{PBSA} - \Delta G^{B}_{PBSA} + \Delta G^{AB}_{PBSA} \\ &= \Delta H_{gas} - T\Delta S + \Delta\Delta G_{PB} + \Delta\Delta G_{SA} \end{aligned} \quad (40)$$

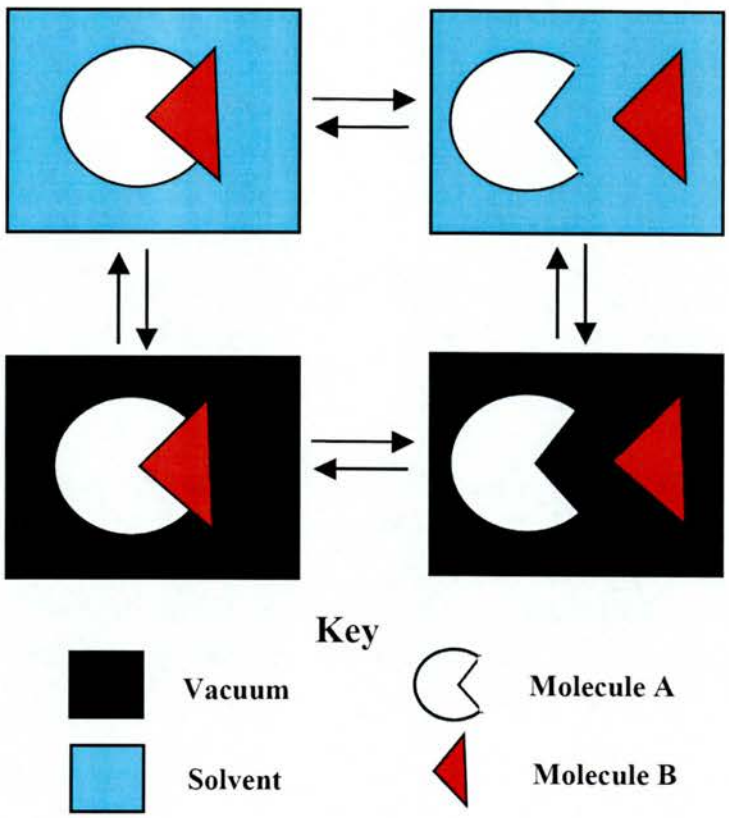
$$\Delta H_{gas} \approx \Delta E_{gas} = \Delta E_{intra} + \Delta E_{electrostatic} + \Delta E_{vdW} \quad (41)$$

$$\Delta\Delta G_{PB} = \Delta G^{AB}_{PB} - (\Delta G^{A}_{PB} + \Delta G^{B}_{PB}) \quad (42)$$

$$\Delta\Delta G_{SA} = \Delta G^{AB}_{SA} - (\Delta G^{A}_{SA} + \Delta G^{B}_{SA}) \quad (43)$$

Thus, a variety of docking attempts are made (each in a different orientation) with two molecules, A and B, using a variety of MD based sampling strategies (discussed in Chapter Seven). The MM-PBSA approach is then used to rank the "quality" of these docked structures via a comparison of relative free energies of binding. In addition to this, absolute binding energies for a selection of defensins are also calculated so that the binding efficiency of HBD1, 2 and 3 can be ranked.

Figure 3.7.2.1. *Schematic diagram of thermodynamic cycle to calculate free energy of binding.*



References

(1) Pearlman, D. A., Case, D. A., Caldwell, J. W., Cheatham, T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., Kollman, P. A. **2002**.

(2) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157-1174.

(3) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *Journal of Computational Chemistry* **2000**, *21*, 132-146.

(4) Jakalian, A.; Jack, D. B.; Bayly, C. I. *Journal of Computational Chemistry* **2002**, *23*, 1623-1641.

(5) A. Pigache, P. C. F.-Y. D. **2004**.

(6) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *Journal of Physical Chemistry* **1993**, *97*, 10269-10280.

(7) Gedeck, P.; Schindler, T.; Alex, A.; Clark, T. *Journal of Molecular Modelling* **2000**, *6*, 452-466.

(8) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *Journal of the American Chemical Society* **1993**, *115*, 9620-9631.

(9) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *Journal of Computational Chemistry* **1995**, *16*, 1357-1377.

(10) Woods, R. J.; Chappelle, R. *Journal of Molecular Structure-Theochem* **2000**, *527*, 149-156.

(11) Besler, B. H.; Merz, K. M.; Kollman, P. A. *Journal of Computational Chemistry* **1990**, *11*, 431-439.

(12) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2005**, *26*, 114-114.

(13) Weiser, J.; Shenkin, P. S.; Still, W. C. *Journal of Computational Chemistry* **1999**, *20*, 217-230.

(14) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Journal of Physical Chemistry* **1996**, *100*, 16385-16398.

(15) Cramer, C. J.; Truhlar, D. G. *Chemical Reviews* **1999**, *99*, 2161-2200.

(16) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *Journal of the American Chemical Society* **1990**, *112*, 6127-6129.

(17) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *Journal of Physical Chemistry B* **1997**, *101*, 1190-1197.

(18) Jayaram, B.; Liu, Y.; Beveridge, D. L. *Journal of Chemical Physics* **1998**, *109*, 1465-1471.

(19) Bashford, D.; Case, D. A. *Annual Review of Physical Chemistry* **2000**, *51*, 129-152.

(20) Tsui, V.; Case, D. A. *Journal of the American Chemical Society* **2000**, *122*, 2489-2498.

(21) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins-Structure Function and Bioinformatics* **2004**, *55*, 383-394.

(22) Onufriev, A.; Bashford, D.; Case, D. A. *Journal of Physical Chemistry B* **2000**, *104*, 3712-3720.

(23) Calimet, N.; Schaefer, M.; Simonson, T. *Proteins-Structure Function and Genetics* **2001**, *45*, 144-158.

(24) Leach, A. R.. *Molecular Modelling: Principles and Applications (2nd Ed.)* . **1996**. Prentice Hall.

(25) Pastor, R. W.; Brooks, B. R.; Szabo, A. *Molecular Physics* **1988**, *65*, 1409-1419.

(26) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523-535.

(27) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. *Journal of Chemical Physics* **2001**, *114*, 2090-2098.

(28) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684-3690.

(29) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671-680.

(30) Wales, D. J.; Doye, J. P. K. *Journal of Physical Chemistry A* **1997**, *101*, 5111-5116.

(31) Doye, J. P. K.; Wales, D. J.; Miller, M. A. *Journal of Chemical Physics* **1998**, *109*, 8143-8153.

(32) Hodges, M. P.; Wales, D. J. *Chemical Physics Letters* **2000**, *324*, 279-288.

(33) Hernandez-Rojas, J.; Breton, J.; Llorente, J. M. G.; Wales, D. J. *Chemical Physics Letters* **2005**, *410*, 404-409.

(34) Carr, J. M.; Wales, D. J. *Journal of Chemical Physics* **2005**, *123*, -.

(35) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96-123.

(36) Rhee, Y. M.; Pande, V. S. *Biophysical Journal* **2003**, *84*, 775-786.

(37) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Journal of Chemical Physics* **2003**, *118*, 6664-6675.

(38) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Journal of Chemical Physics* **2003**, *118*, 6676-6688.

(39) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Accounts of Chemical Research* **2000**, *33*, 889-897.

(40) Weinzinger, P.; Hannongbua, S.; Wolschann, P. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2005**, *20*, 129-134.

(41) Rizzo, R. C.; Gochin, M.; Kuntz, I. D. *Abstracts of Papers of the American Chemical Society* **2003**, *226*, U445-U445.

(42) Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A. *Journal of the American Chemical Society* **2001**, *123*, 5221-5230.

(43) Massova, I.; Kollman, P. A. *Perspectives in Drug Discovery and Design* **2000**, *18*, 113-135.

(44) Yan, S. X.; Geacintov, N. E.; Broyde, S. *Abstracts of Papers of the American Chemical Society* **2001**, *222*, U397-U398.

(45) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. *Biopolymers* **2003**, *68*, 16-34.

(46) Pricl, S.; Ferrone, M.; Fermeglia, M.; Lodi, A.; Mazzurco, A. *Abstracts of Papers of the American Chemical Society* **2003**, *225*, U260-U260.

(47) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. *Journal of Medicinal Chemistry* **2005**, *48*, 4040-4048.

(48) Wojciechowski, M.; Fogolari, F.; Baginski, M. *Biophysical Journal* **2005**, *88*, 61a-61a.

(49) Page, C. S.; Bates, P. A. *Journal of Computational Chemistry* **2006**, *27*, 1990-2007.

(50) Luo, R.; David, L.; Gilson, M. K. *Journal of Computational Chemistry* **2002**, *23*, 1244-1253.

(51) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *Journal of the American Chemical Society* **1998**, *120*, 9401-9409.

(52) Sitkoff, D.; Sharp, K. A.; Honig, B. *Journal of Physical Chemistry* **1994**, *98*, 1978-1988.

(53) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305-320.

(54) Carlacci, L.; Englander, S.W. Abstracts of Papers of the American Chemical Society **1996**, *211*, 66.

Chapter Four

Rational Structure Prediction with Static Disulphide Bridges

4.1. Introduction

This Chapter and Chapter five deal with techniques for rationally predicting the structure of defensins. Methodologies that invoke a static disulphide description are considered in this Section, whilst those involving mobile-disulphide descriptions are discussed in Chapter Five (the need for a mobile disulphide procedure is discussed in due course).

4.1.1. The protein folding problem

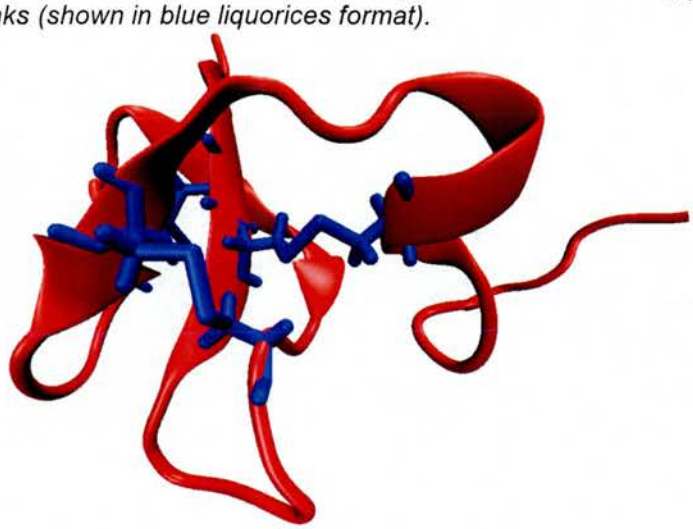
Levinthal in 1968 showed that a random global minimum search within a small biological system possessing just two degrees of freedom per residue would require a folding time scale greater than the known age of the universe[1]. His simple observation demonstrated that biological folding does not involve a random search. Later work provided evidence for the existence of a folding funnel[2-5] in which partially folded[6,7] (or nucleated) intermediates arise at various folding stages. A growing repertoire of algorithms have been (and are being) developed for solving this problem[3,8-22]. This Chapter discusses the implementation of some popular folding techniques to the defensin class of problem: simulated annealing (SA)[21-45], replica exchange (RE)[20,41,46-91] and Basin Hopping (BH)[3,9-19], as well as some novel folding methodologies.

4.1.2. Interpreting pictorial representations of defensins

Defensin illustrations in this thesis are provided in a cartoon format in which disulphide bridges are (for clarity purposes) not shown. It is, however, important to realize that disulphide bridges

are present and that their existence kinetically traps the peptide backbone into specific loops and secondary structural folds (*Figure 4.1.2.1*). Discussions of defensin structure in both this Chapter and Chapter Five often refer to the torsional angles of these loop residues, and hence, indirectly refer to the disulphide bridges that make these tight turns energetically possible.

***Figure 4.1.2.1**. A cartoon representation of HBD3 showing characteristic α-helix, β-sheets and disulphide bridge cross-links (shown in blue liquorices format).*



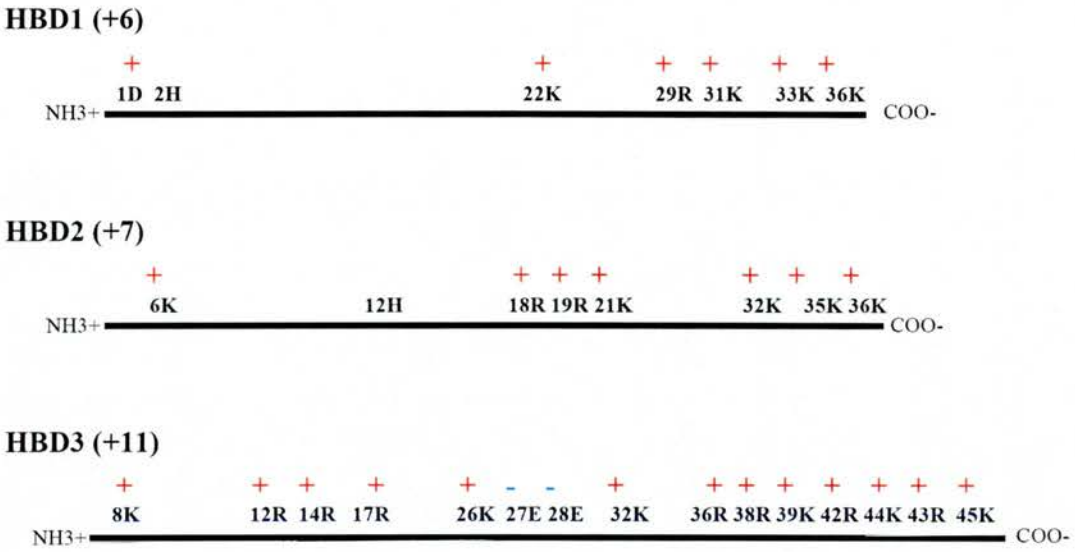
4.2. System set-up

There are three central facets to the rational structure prediction problem. The first is the problem of assigning the amino acid charges within the polypeptide; which amino acids possess what charge during which conformations? The second problem (which is of particular relevance to cross-bridged species like defensins) lies in the choice of starting structure; what is the best "initial" configuration from which to begin the folding process? The third and final problem rests with how to measure the "fitness" of a rationally folded structure – A strict force-field energy comparison (with its inherent approximations) can sometimes be an imperfect means of measuring structural fitness.

4.2.1. Charge assignment

The issue of charge assignment is contentious in the field of protein folding. Predicting the local chemical environment of an amino acid, and hence predicting its likely charge state remains an open problem. Nevertheless, methods do exist for approximating protonation states within a variety of systems. Case *et al* for example, detail a "constant *pH*" method in which the polypeptide is subjected to Generalized Born solvated MD followed by periodic Metropolis Monte Carlo transitions in protonation state[92-95]. The sampling methodology permits the transition free energy (for a given change in protonation state) to be calculated during a variety of conformations, and hence the *pKa* of the selected amino acid estimated[92,93]. *Figure 4.2.1.1* and *Figure 4.2.1.2* summarise the predicted protonation states and corresponding protonation assignments for HBD1, HBD2 and HBD3.

Figure 4.2.1.1. *The charge assignments for the protonation sites within HBD1, HBD2 and HBD3. ARG is not supported by the constant pH algorithm, and it is accordingly assigned a positive charge owing to large resonance stabilisation of the proton-absent form. The protein backbone is shown as a thick, black line. Charge assignments ("+" or "-") are denoted on the 1 letter amino acid code.*



4.2.2. Generating initial geometries

The rational structure prediction algorithms outlined below (unless otherwise stated) employ the following procedure for generating initial structures:

1. A protein sequence is read into the *xleap* component of AMBER using the "sequence" command. This process introduces a default set of β-sheet angles to the entire sequence with the result that the side chains alternate above and below the plane of the backbone.
2. Disulphide bridges within the sequence defined in (1) are bonded in their canonical form. This results in bonds that are widely displaced from their equilibrium value, sometimes by as much as 30 Angstroms.

3. The structure in step (2) is minimized with *sander* using the standard minimization settings (*Section 4.9*). This results in a compact structure in which the disulphide groups are connected at their equilibrium distance.

There are some obvious drawbacks with generating preliminary structures in this fashion, the most obvious being that conformational freedom is reduced. The problem of disulphide bridges and how to deal with them is a recurring theme (and central problem) of this research and is considered in greater depth in the proceeding Chapter.

4.2.3. Fitness criterion

Unless otherwise stated all fitness measurements encompass an energetic and torsional comparison with the target structure*. The torsional component is obtained with a harmonic-type energy function for sequence stretches that are expected to exhibit secondary structure in the natively folded state. The total "fitness", $E_{fitness}$, is then determined as follows:

$$E_{fitness} = E_{folded} + E_{tor} \qquad (1)$$

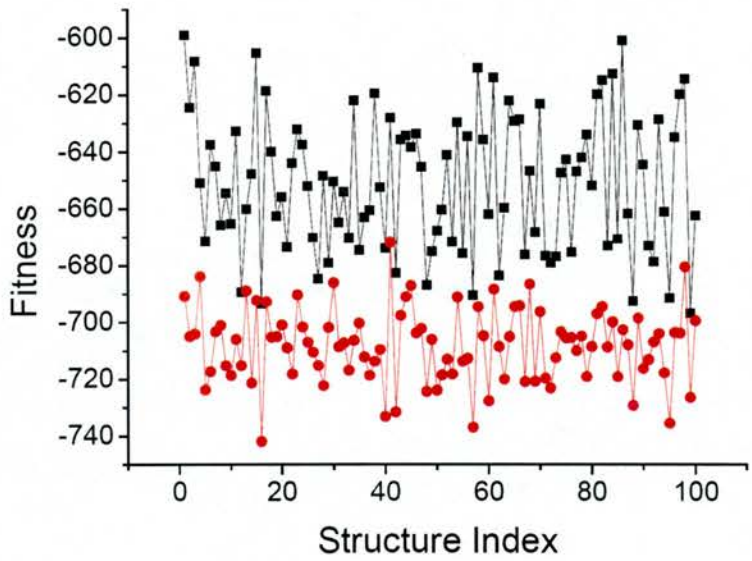
Where E_{folded} is the unbiased single point energy of the rationally folded molecule, and where E_{tor} refers to the torsional bias term:

$$E_{tor} = K\sum_{i}^{n}(\varphi_i - \varphi_{i_{eq}}^{eq})^2 \qquad (2)$$

* The torsional weighting is used because an unbiased energy comparison in some cases fails to distinguish native from non-native structures. Accordingly, the penalty term better separates "good" structures from "bad" structures.

Where K refers to a constant, φ to the torsional angle for amino acid i and $\varphi_{i_{eq}}^{eq}$ to the corresponding native state torsional value (*Figure 4.2.3.1*).

Figure 4.2.3.1. *The unbiased energy (black) versus the torsionally penalized energy (red).*



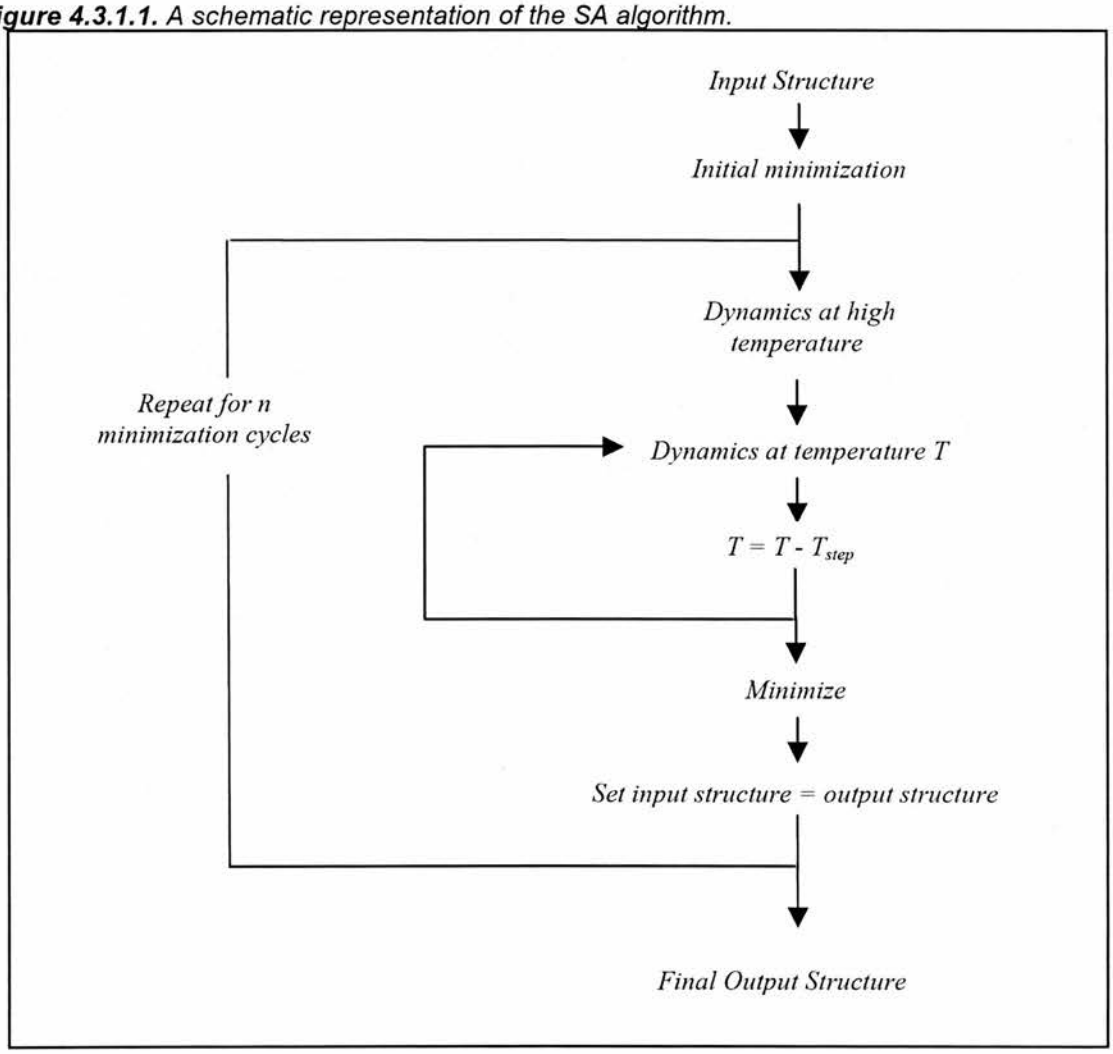
4.3. Simulated Annealing (SA)

SA is historically an early folding strategy[36] and one that has enjoyed success with small systems such as met-enkephalin[40] and fragments of BPTI[41]. However, a dearth of literature publications for systems larger than 20 residues indicates that this is the upper limit of this strategy. Does this limit hold when the peptide in question is conformationally restricted (as is the case with defensins)? Do such restrictions to the potential energy surface make the sampling problem tractable?

4.3.1. Method

MD simulated annealing is applied to initial model constructs as described in Chapter Three using standard settings (*Section 4.9*). The SA protocol is implemented with a combination of Fortran 90 and *cshell* (*Figure 4.3.1.1*).

Figure 4.3.1.1. *A schematic representation of the SA algorithm.*



4.3.2. Results and discussion

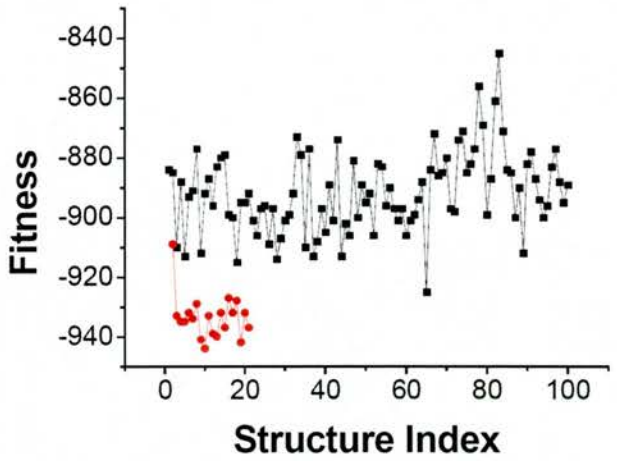
The standard fitness criterion reveals that SA (applied as a standalone procedure) does not find the global fold for this class of protein. Indeed, the fitness measure of the calculated structures

reveals significant divergence from the experimental native NMR species (*Figure 4.3.2.1.*) Furthermore, many of the required *phi-psi* angles are never sampled during the simulation (*Figure 4.3.2.2.*), an observation that lends credence to the notion that a restricted conformational space (contrary to initial hopes) inhibits rather than aids sampling. Indeed, key residues such as VAL 14 within HBD2 never sample the required torsional space (*Figure 4.3.2.3 and Figure 4.3.2.4*). A likely explanation for this is that the loop regions of the molecule remain kinetically trapped by the disulphide cross-linkages in the starting configuration, such that escape to a global (or less frustrated) minimum is kinetically impossible, even at very high temperatures.

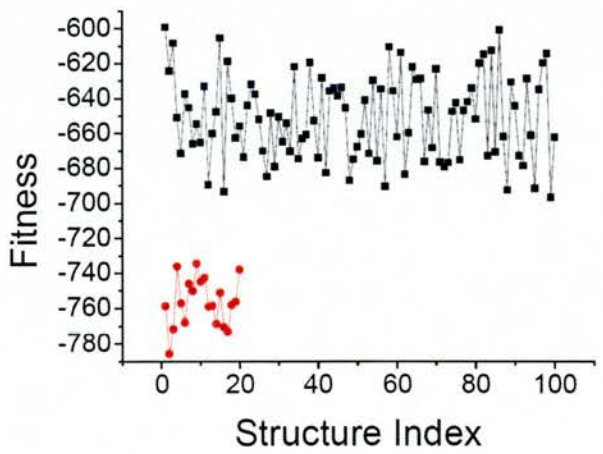
Work presented elsewhere in this Chapter (and additionally in Chapter Five) attempts to deal with the problem of kinetic trapping.

Figure 4.3.2.1. *Simulated annealing energies (black) versus 20 minimized native NMR structures (Red) for HBD1-3.*

HBD1



HBD2



HBD3

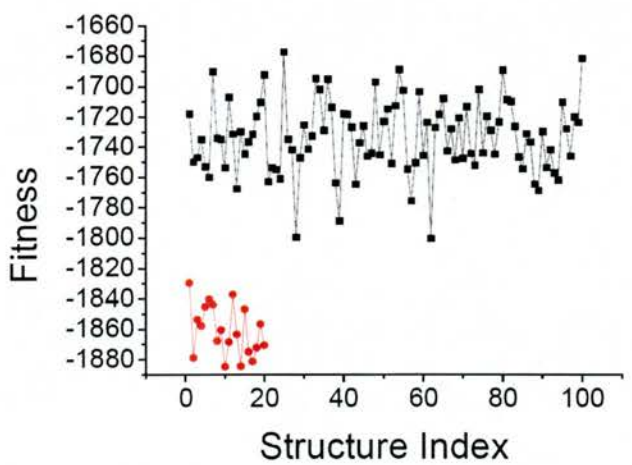
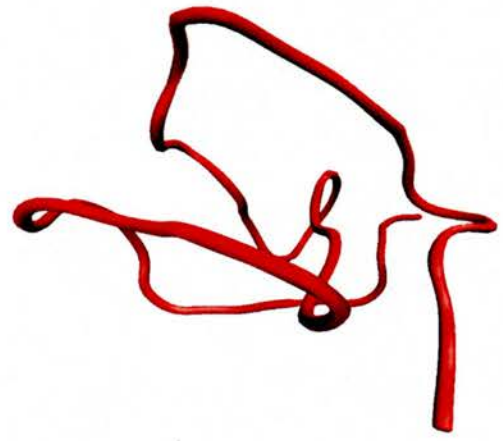
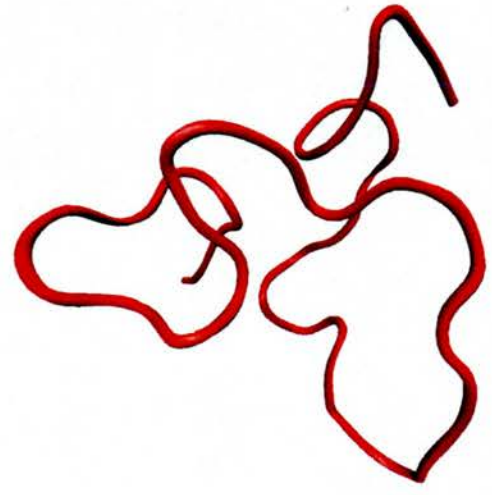


Figure 4.3.2.2. *Simulated annealing structures shown in "cartoon" mode for the fittest predictions of HBD1-3. There is no discernible secondary structure.*

HBD1



HBD2



HBD3

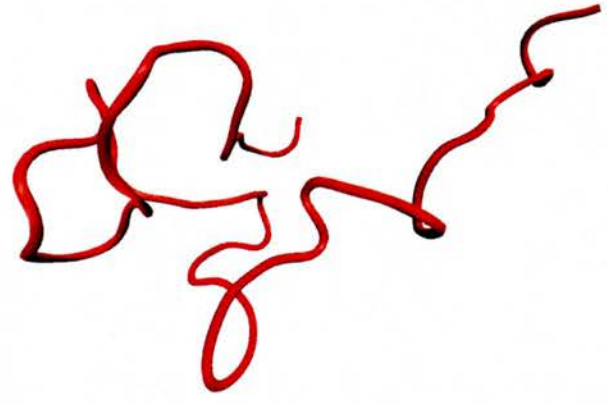
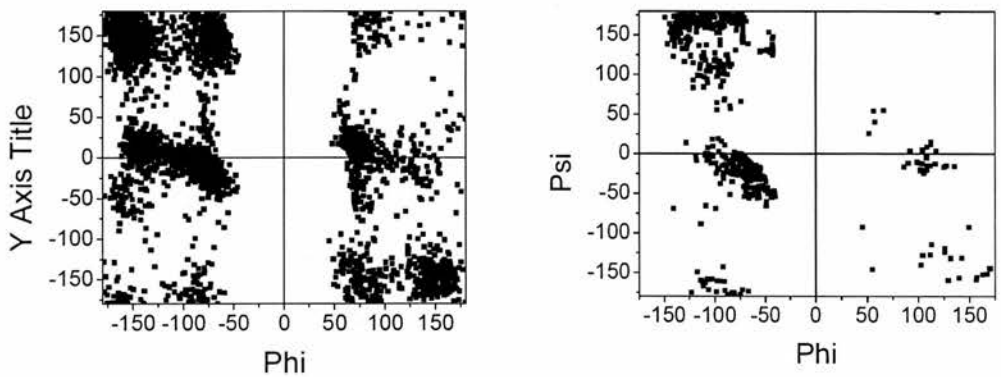
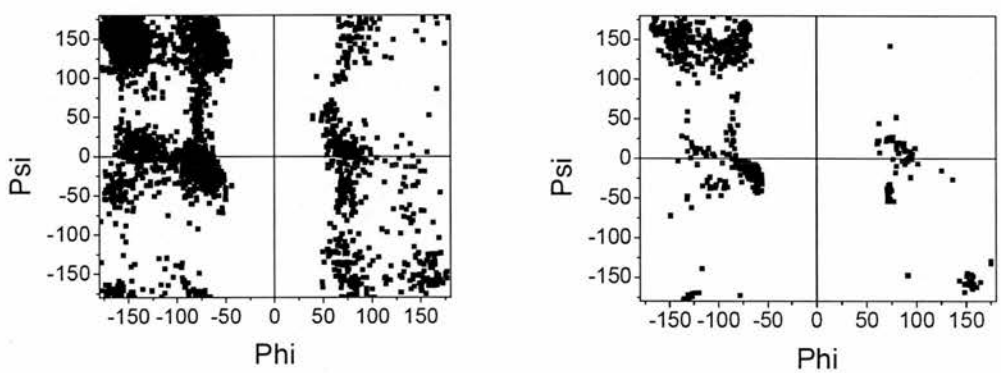


Figure 4.3.2.3. *Scatter plots of torsional angles for each of the final SA structures (left) compared to 20 native NMR structures (right) for each of HBD1-3.*

HBD1



HBD2



HBD3

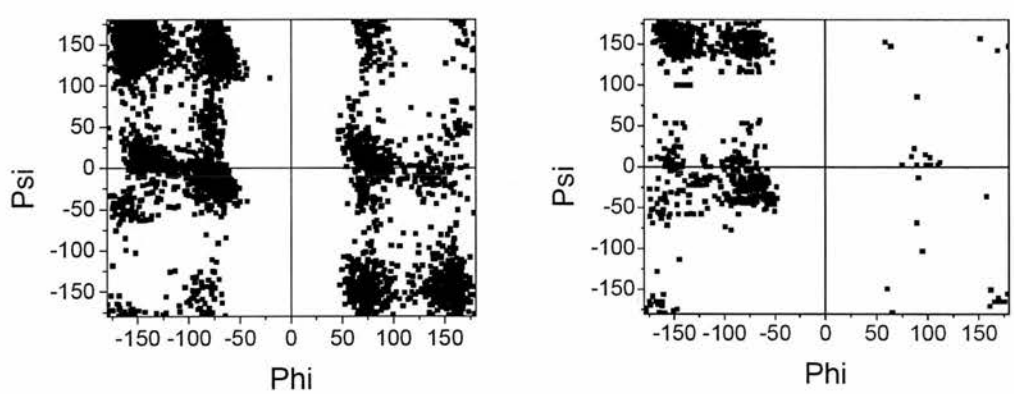
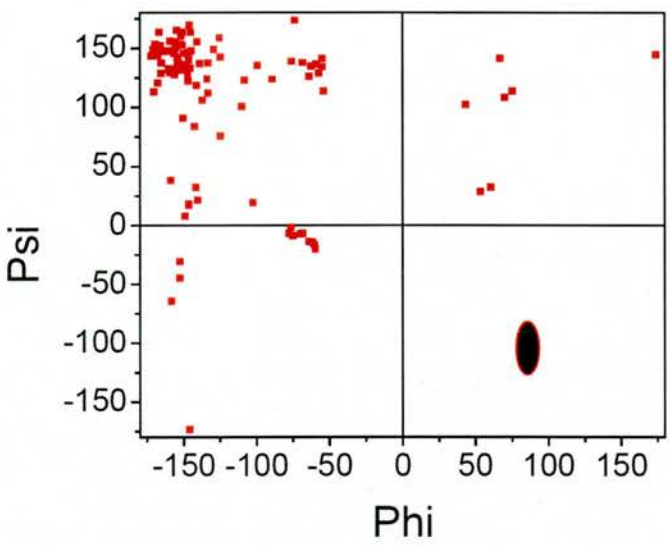
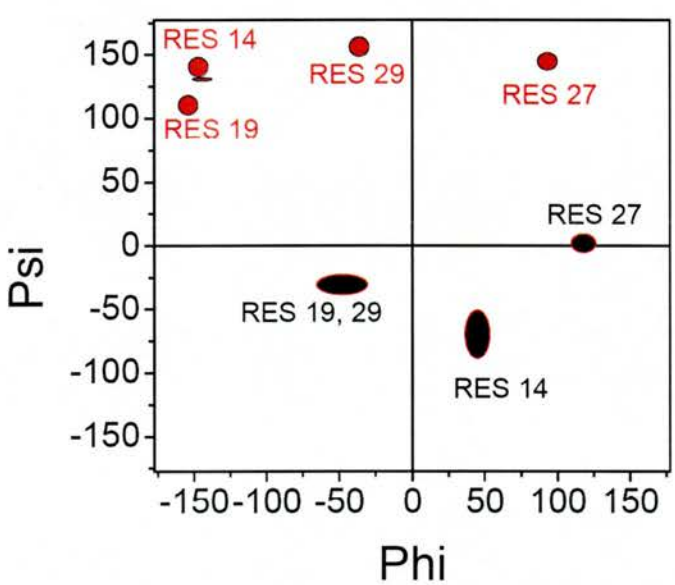


Figure 4.3.2.4. *A: Phi and Psi angles for residue 14 (Valine) of HBD2 for both simulated annealing and native NMR structures. B: SA torsional structures shown schematically for a variety of other residues (NMR is shown in red and SA in black). The absence of sampling in the required torsional space in both cases suggests the existence of kinetic trapping. This problem is common to all of the defensins examined.*

A



B



4.4. MD with altered Hamiltonian states

A central problem with rational structure prediction is the oft inadequate sampling of phase space. Indeed the complexity of the defensin problem appears to be related (in part) to the inability of algorithms to address the kinetic trapping problem. In this section a methodology is outlined for increasing phase space sampling by progressively and artificially "flattening" the potential energy landscape.

4.4.1. Method

With this strategy the system Hamiltonian is modified so that the PE surface is effectively flattened. This is done by zeroing the disulphide bridge and dihedral force constants and by truncating the non-bonded cut off at 4 Angstroms. The disulphide bonding and dihedral twisting terms are therefore present in the Hamiltonian description of the molecule only in name, for they provide no energetic penalty to deviations from the equilibrium position. From this vantage point a burst phase of MD (as it will hereon be referred to) commences in which the molecule is permitted to sample the flattened landscape in a comparatively uninhibited manner. Following this initial burst, a set of linear and non-linear transformations restore the Hamiltonian to its original state (*Figure 4.4.1.1* and *Figure 4.4.1.2*).

The non-linear disulphide bridge transformation is applied iteratively as follows:

$$F_{i,n} = \frac{F_i^o}{\exp[Kn + C]} \qquad (3)$$

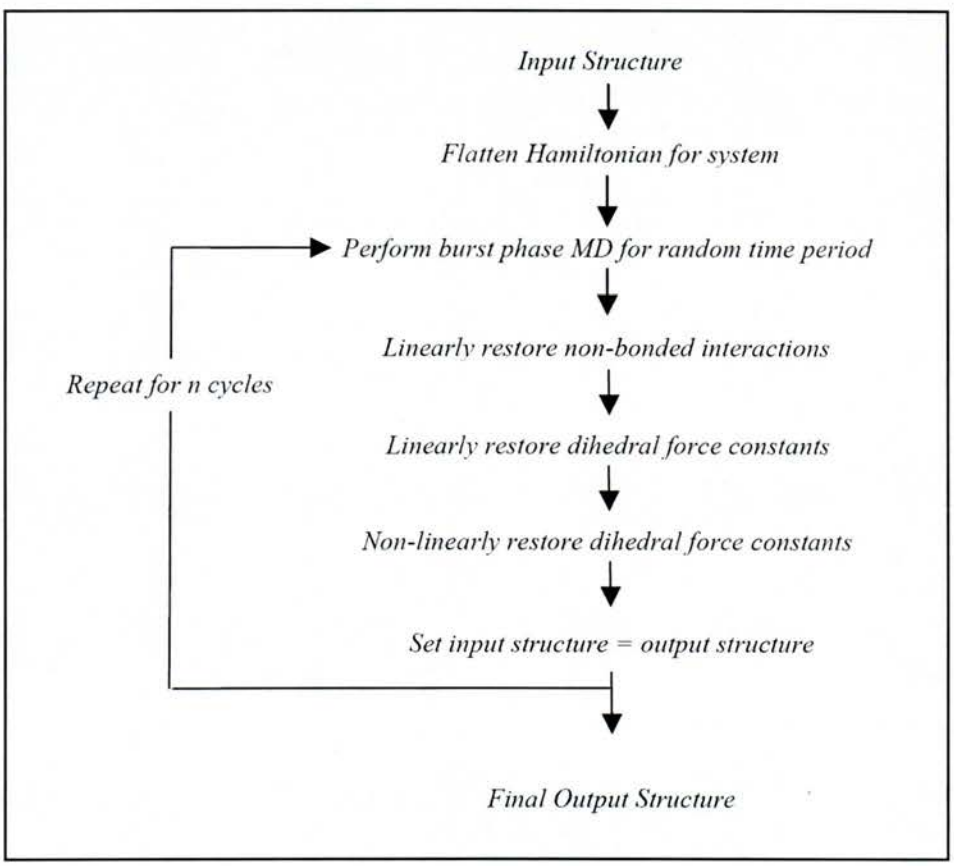
Where $F_{i,n}$ refers to the force constant to be used in the Hamiltonian description for atom pair i during iteration n. K and C are constants that are chosen to produce an intuitively desirable rate of restoration and F_i^o refers to the final intended force constant.

Figure 4.4.1.1 *An illustration of the modifications that flatten energy landscape during the "burst" phase. The Sections highlighted in red are the components of the force field that are affected; non-bonded interactions are truncated at 4 angstroms and dihedral force constants and disulphide bond force constants are set to zero. These modifications are subsequently restored to their original values with a set of linear and non-linear transformations.*

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 +$$

$$\sum_{dihedral} \frac{V_n}{2}[1 + \cos(n\varphi - \gamma)] + \sum_{\substack{non-bonded \\ interactions \\ i>j}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}} \right]$$

Figure 4.4.1.2. *A schematic representation of modified Hamiltonian sampling simulation.*



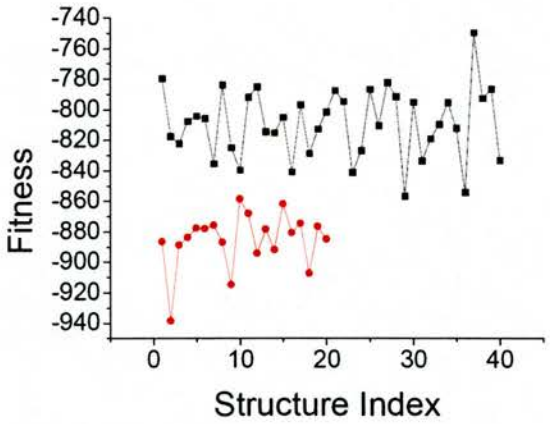
4.4.2. Results and discussion

The modified Hamiltonian method, should, in principle avoid the problem of kinetic trapping associated with the SA procedure. This is not born out in practice. In fact, the energetic discrepancy (*Figure 4.4.2.1*) and extent of torsional sampling (*Figure 4.4.2.3* and *Figure 4.4.2.4*) appear to resemble that obtained via SA. However, despite this apparent similarity, some α-helical structure is observed in HBD3, and a good "general" topology can be visually discerned (*Figure 4.4.2.2*).

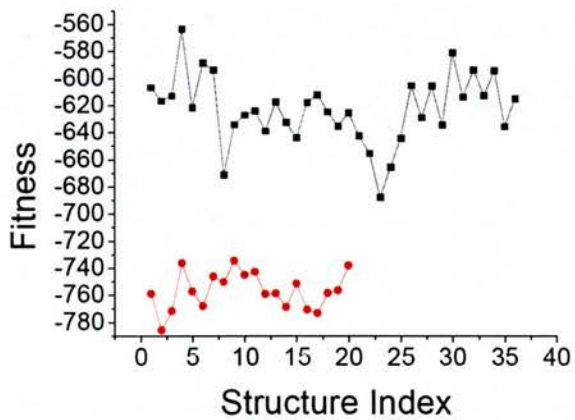
It is not clear why this method has failed to produce more effective sampling. One possibility is that the non-linear restoration phase results in structural collapse that is too rapid for desirable secondary structural elements to form. Indeed, intermolecular bonds are approximated with a harmonic potential for which small changes in the magnitude of force constant produce an exponential change in energy. Future improvements to the procedure will focus on defining inter-disulphide constraints with a linear energy penalty so that structural collapse can be generated more gradually.

Figure 4.4.2.1. *the Fitness measure for the modified Hamiltonian technique (black) versus 20 NMR structures (red) for each of HBD1-3.*

HBD1



HBD2



HBD3

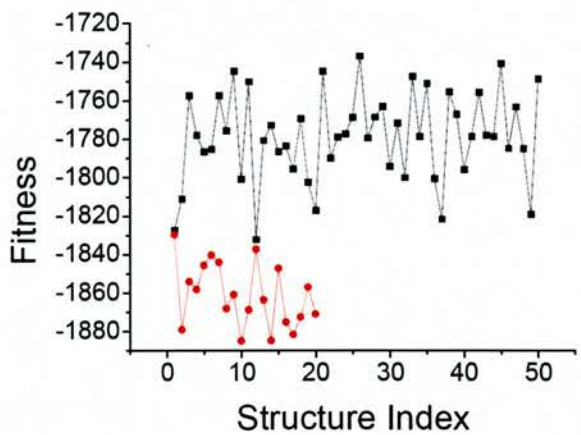
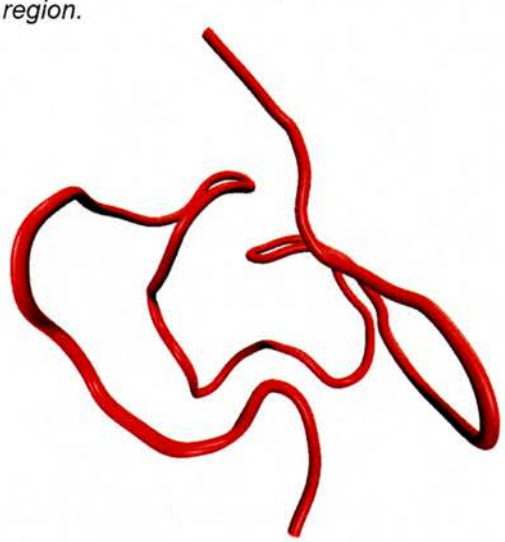
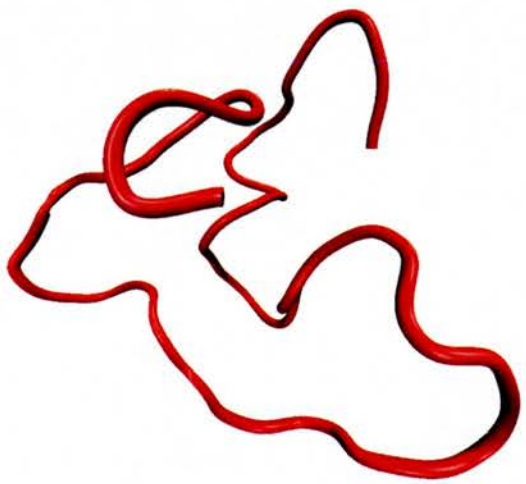


Figure 4.4.2.2. *Fittest structures for each of HBD1-3. Cartoon representations of each defensin show no discernible secondary structural motif, except for HBD3, which shows desirable α-helical content in the N-terminal region.*

HBD1



HBD2



HBD3

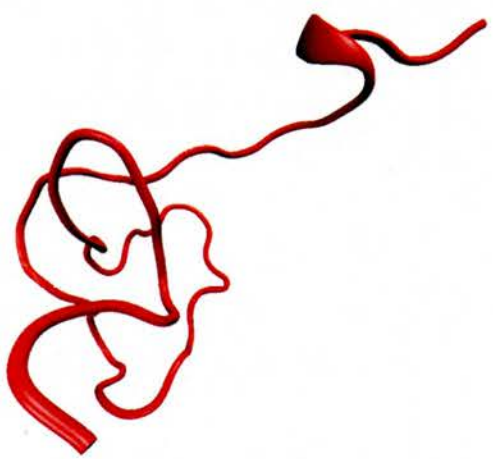
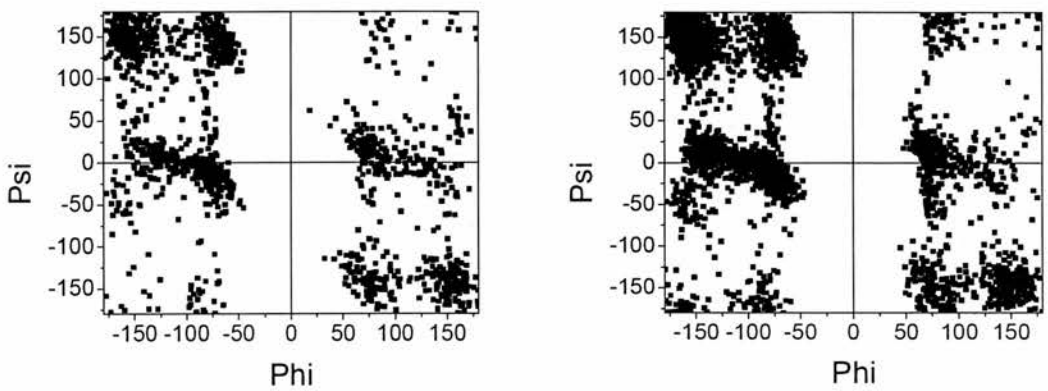
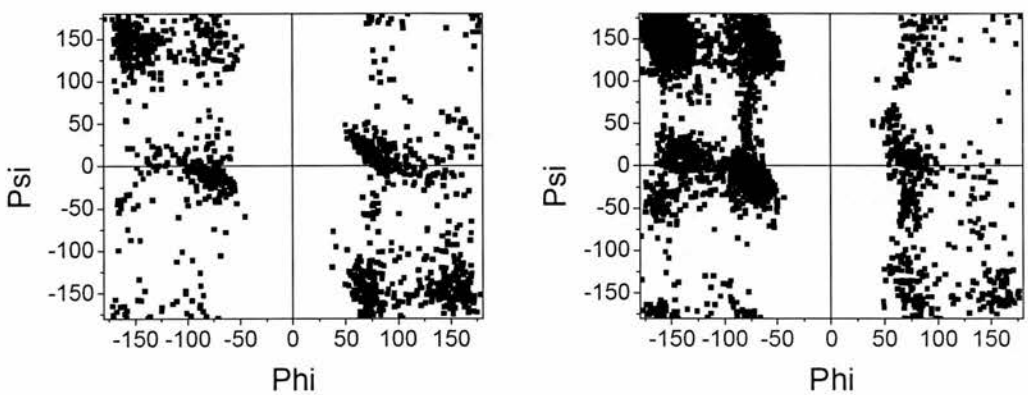


Figure 4.4.2.3. *A scatter plot of torsional angles for each of HBD1-3 for both the modified Hamiltonian method (left) and SA derived structures (right).*

HBD1



HBD2



HBD3

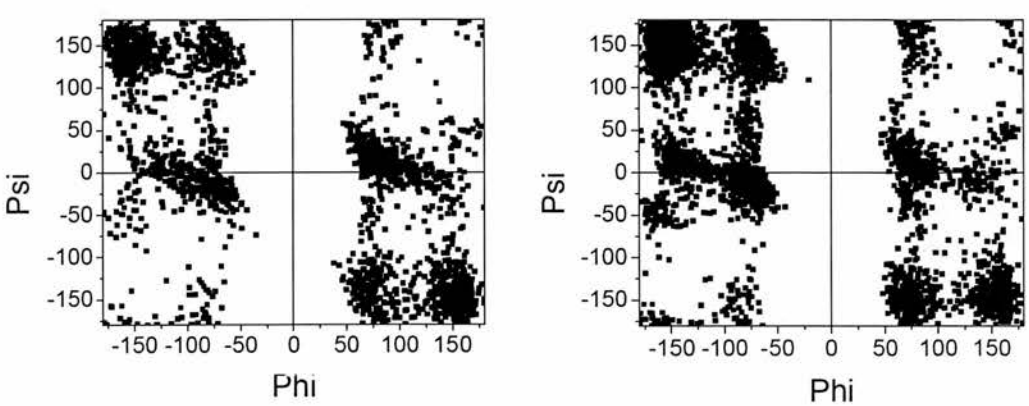
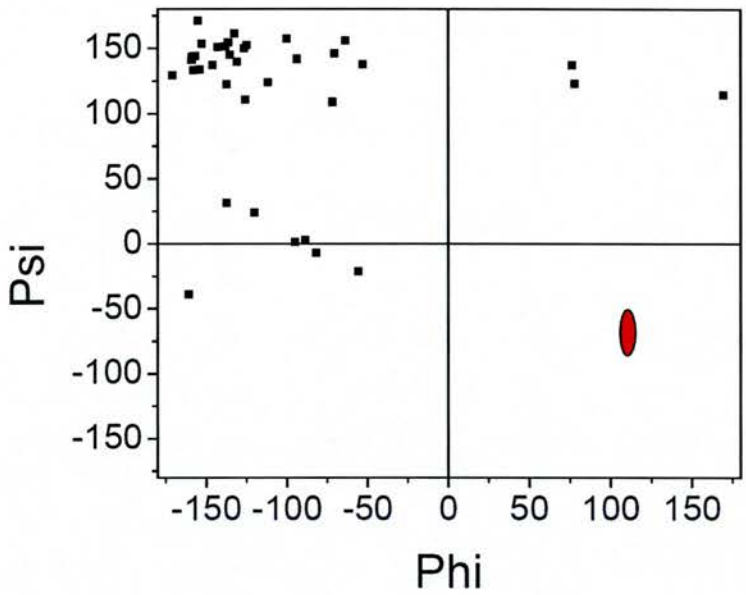
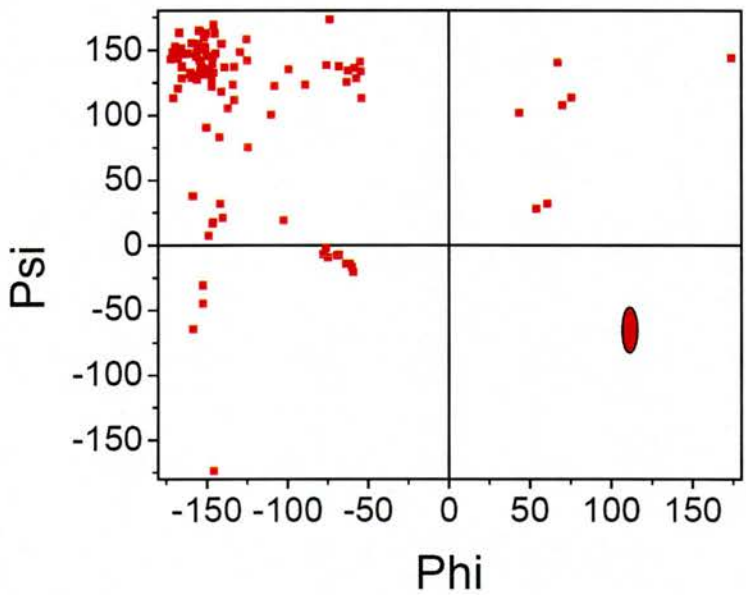


Figure 4.4.2.4. *A: Torsional scatter plot for HBD2 residue VAL 14 (shown in black) against a schematic representation of residue VAL 14 in 20 NMR structures (shown as a red blob). B: Torsional sampling for SA calculated structures for the same residue. The absence of required torsional angles suggests inefficient sampling or the existence of kinetic trapping for both methods.*

A



B



4.5. Basin Hopping (BH)

The BH strategy has enjoyed significant success with cluster problems, which, despite being governed by a multiple funnel topography, are nevertheless tractable with this algorithm[10-14,16-18]. Biological folding problems have also been tackled with this algorithm within a coarse grain force-field framework[9] with some success. The BH procedure is attractive in the context of the defensin problem because the "transformed" PES permits (in principle) efficient reorganisation of disulphide-bridge-loop-regions.

4.5.1. Method

Burst phase dynamics are performed with the *sander* component of AMBER and BH runs are performed with GMIN version 2.2.31. Where necessary, *cshell* scripting in conjunction with the AMBER module *ptraj* are used to interconvert AMBER and GMIN formats.

A central problem with the basin hopping task rests with the choice of parameters; with more than 30 continuous variables, what is the most suitable choice? This optimisation problem is outside the scope of this research and, as such, only the *temp* and *step* parameters are optimised. The remaining parameters are chosen on the basis of their empirical success with other systems[9].

Optimisation is achieved with the small, 20 amino acid test-system test of trp-cage as follows:

a) Optimising *temp* and *step*

The *temp* and *step* parameters are optimised by systematically changing the selected variable whilst holding the remaining variables fixed. Each stepwise change in magnitude is followed by a BH run of 3000 steps and the "fitness" of the parameter choice assessed by tabulating the energetic output (*Figure 4.5.1.1* and *4.5.1.2*).

Figure 4.5.1.1. *Relative energies of the lowest minimum obtained using different step sizes. A step size of 10 corresponds with the lowest energy minimum.*

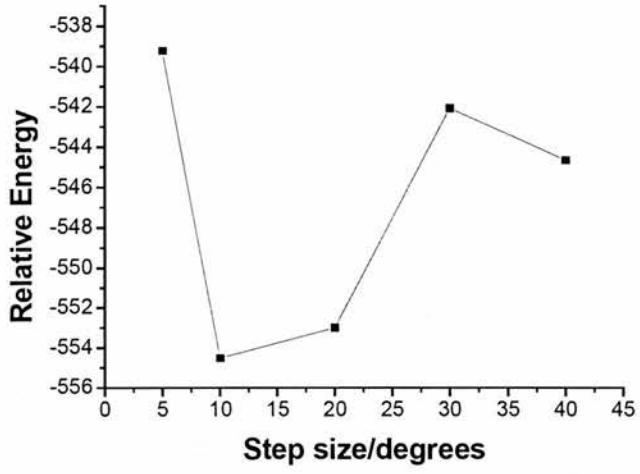
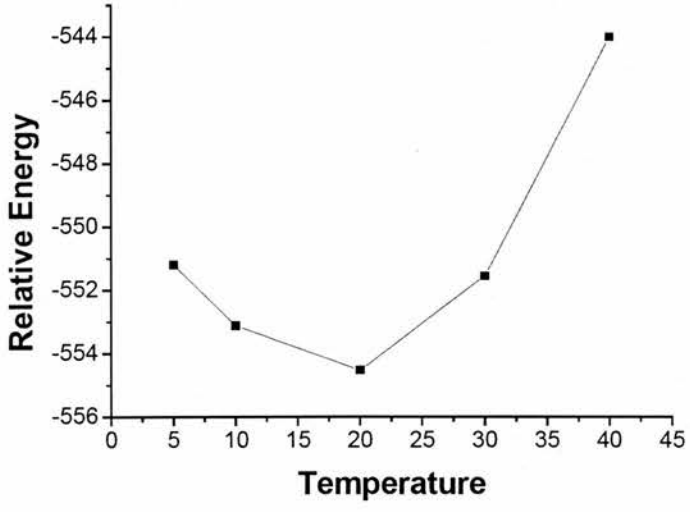


Figure 4.5.1.2. *Relative energies of the lowest energy minima obtained from different temperatures. A temperature of 20 degrees corresponds with locating the best local minimum.*

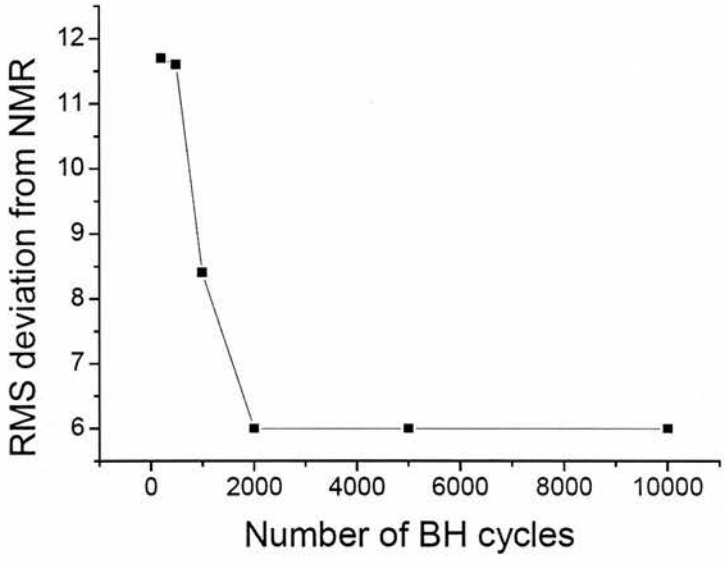


b) Choosing the number of BH iterations

The BH algorithm, during the first one thousand (or more) iterations, generates rapid structural collapse. Following this initial event subsequent cycles appear only to generate local sampling. Employing 3000 – 4000 BH cycles produces structures that are

not significantly lower in energy (or different in structure) to those obtained with 20 000 cycles.

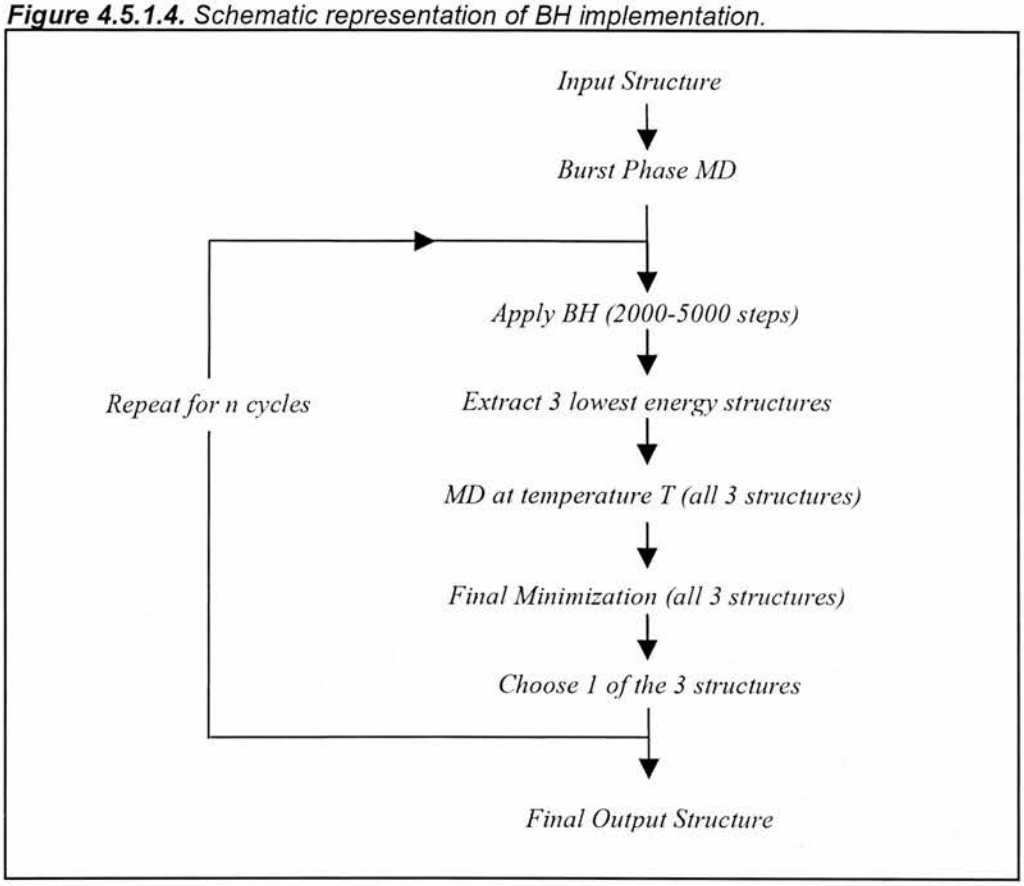
Figure 4.5.1.3. *RMS deviation of BH calculated Trp-cage structures compared to that of the NMR derived structure[96].Rapid initial collapse is followed by non-productive local sampling.*



The strategy in this research focuses on combining aspects of the BH directed structural collapse with traditional MD in order to optimise both global and local sampling (*Figure 4.5.1.4*) as follows:

1. A "burst phase" of MD (as outlined in *Section 4.3.*) is used to sample phase space in the non-bonded state.
2. BH is invoked (for 2000 to 5000 steps) with the canonical disulphide bridge connectivity to direct global collapse.
3. A brief MD equilibration is performed to permit local kinetic sampling.
4. A thorough minimization is performed to locate the closest minimum.

Figure 4.5.1.4. *Schematic representation of BH implementation.*



4.5.2. Results and discussion*

BH as a standalone procedure has enjoyed significant success with a variety of systems but no such success is observed here with the defensin family of proteins. Indeed, rapid structural collapse occurs, but the rearrangements required for the formation of native structure do not (results not shown). Why is this? The most intuitive explanation is that folding requires kinetically trapped loop regions to be re-arranged into "more native" structures during the folding process. In many cases, this event (or these events) would require between 6 and 12 amino acids moving in unison and (possibly) passing through one another. The probability of

* GMIN version 2.2.1 used in this research encounters significant problems when performing torsional variations to N-terminal proline. As such, there are no publishable results for HBD2 (which contains an N-terminal proline).

such enormous cooperative movement occurring as a result of random changes to randomly selected torsional angles is exceedingly small. Thus, BH as a standalone procedure does not appear to be effective in solving disulphide-cross-linked systems of this ilk. Furthermore, it remains a distinct possibility that the BH parameter set is not appropriately optimised. GMIN versions 2.2.1 and higher allow for cooperative changes to selected residues, and more productive sampling might be obtained with appropriate changes in the nature (and magnitude) of this cooperative change.

The hybrid BH-MD method aims to circumvent the problem of local trapping by seeding BH from a variety of distinct and extended geometries. The benefits and pitfalls of this strategy are here discussed with reference to HBD3. Whilst the hybrid method fails to generate the intended secondary structural motifs (*Figure 4.5.2.1* and *Figure 4.5.2.2*), it nevertheless avoids the trapping problem observed with both the SA and modified Hamiltonian approach. For example, the residue CYS 23 within HBD3 samples the correct torsional space with BH-MD but remains kinetically trapped in the case of SA (*Figure 4.5.2.3*). Furthermore, scatter plots of dihedral sampling reveal that the hybrid method visits more favourable regions of phase space than SA (*Figure 4.5.2.3*).

The BH-MD method avoids the problem of kinetic trapping, but alas, it does not locate the global fold. Why the method fails is not clear. One reason may lie with the assumption that all disulphide bridges form at the same moment in time. In reality the nucleated structure on which the global fold ultimately depends may only form within a partially cross-linked complex[100]. The effect of dynamically forming and breaking disulphide bonds is considered in greater depth in Chapter Five.

Figure 4.5.2.1. *The fitness of the BH-MD generated HBD3 structure (black) versus that of 20 native NMR structures (red).*

HBD3

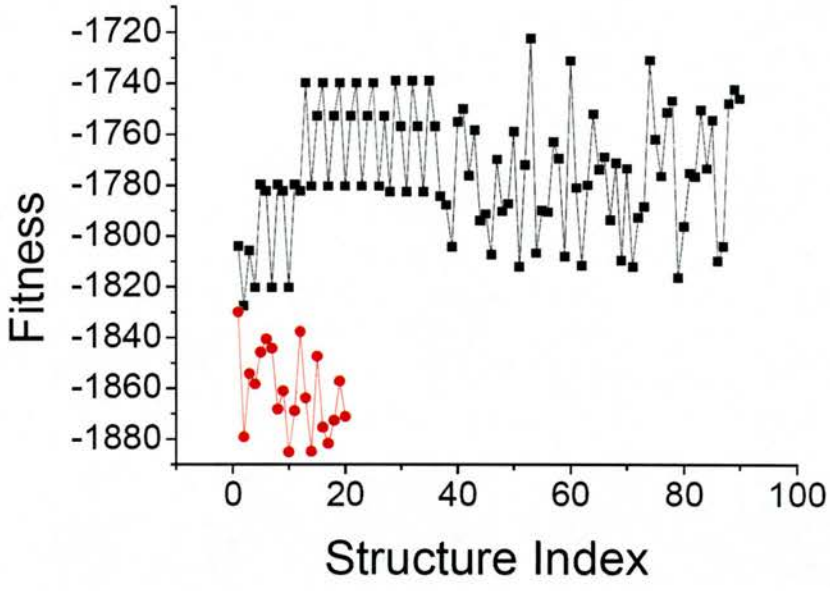


Figure 4.5.2.2. *A cartoon representation of the fittest BH-MD generated structure of HBD3. No native secondary structural motifs are detected.*

HBD3

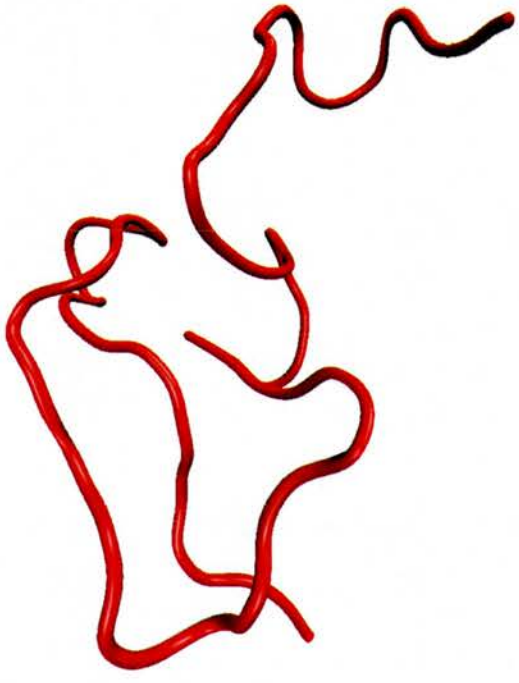
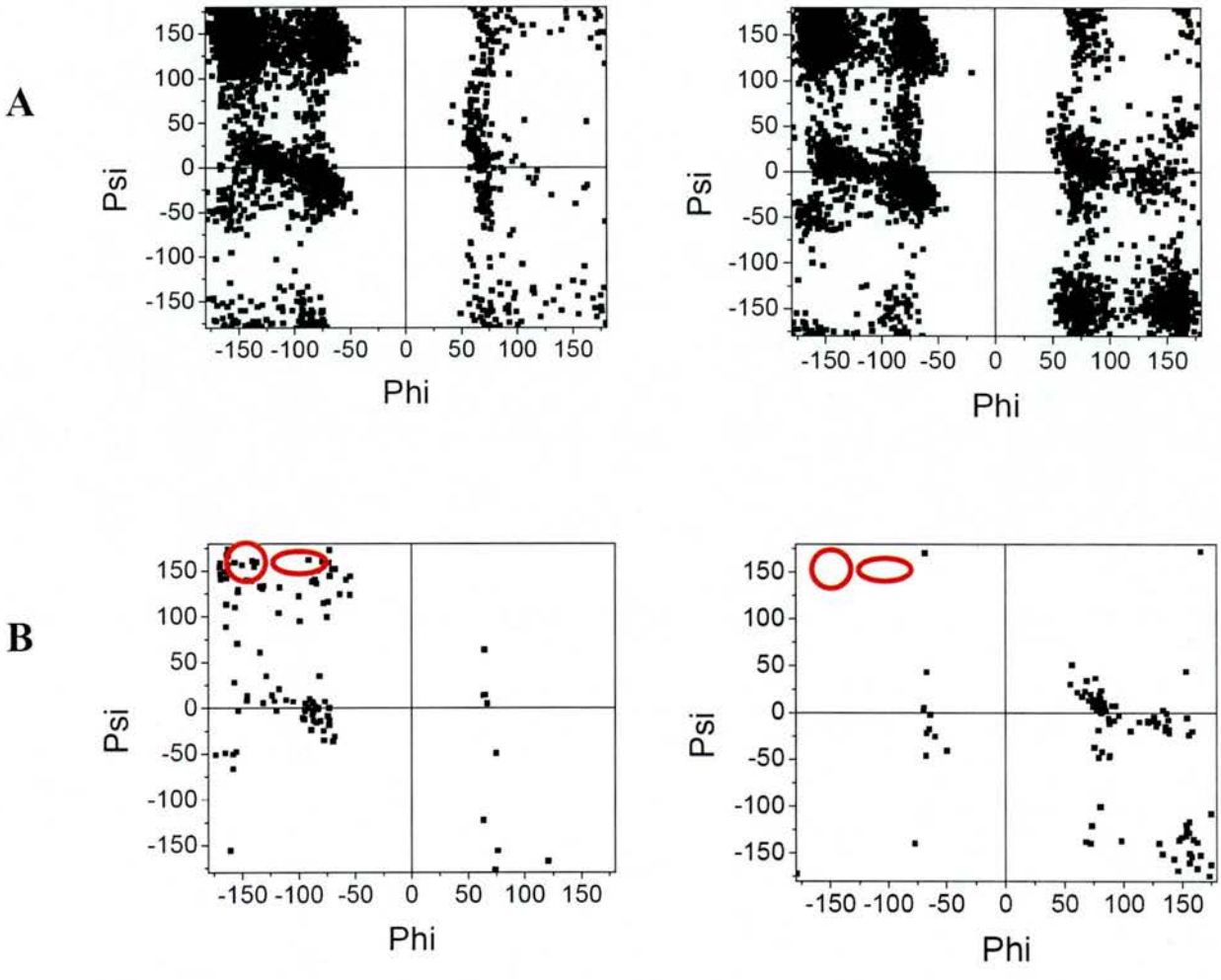


Figure 4.5.2.3. *A. Torsional scatter plot for HBD3 showing the full range of visited dihedral angles for the BH-MD method (left) in comparison to that of SA (right). The BH-MD method preferentially samples energetically favourable regions of phase space compared to SA. B. A torsional scatter plot for the residue CYS 23 for both the BH-MD (left) and SA (right) method. The torsional regions sampled by the native NMR structures are shown schematically in red and illustrate that the BH-MD method avoids kinetic trapping in undesirable regions of configuration space.*



4.6. Replica Exchange (RE)[20,47,48,52-57,59,61,62,64-66,70-75,79-83,85,89,91]

Conventional simulation methods within the canonical ensemble are often subject to the pitfall of local trapping. This problem can often be overcome with generalized ensemble techniques such as RE in which random walks through PE space are permitted[20]. Indeed, a variety of distinct (and otherwise unvisited) conformations can often be observed with RE[77,79,81,84,89,90]. The method has accordingly enjoyed significant success with modelling aggregation problems[87] and protein folding problems[20,48,49,53,54,58,59,61,64,75,80,83,89,90]. For example, Sanbonmatsu *et. al* demonstrate that RE sampling for the system of met-enkaphalin is 5 times more extensive than with normal constant temperature MD[84], a phenomenon that permits helical to non-helical transitions to be identified.

4.6.1. Method

REMD (as outlined in Chapter Three) is performed using 64 replicas with standard MD settings (*Section 4.9*) and Langevin dynamics. The temperature of each replica is assigned on the basis of the following exponential function:

$$T_i = T_o \exp(Ki) \tag{4}$$

Where T_i refers to the temperature of replica i, T_o to the initial temperature and K to a constant (found by trial and error) that is scaled by a factor of i. The temperature spacing is optimised by examining the exchange rate for stepwise variations in K. RE studies of other systems ordinarily choose a temperature spacing that gives rise to an acceptance ratio that lies between 0.15 and 0.75[20,41,46-91]. Accordingly, the parameters in equation 4 are selected so that the acceptance ratio falls in between these limits (*Figure 4.6.1.1*). The initial temperature is set to 278 K, and the

constant, *K*, chosen to be 0.03 via a trial and error process (*Figure 4.6.1.2*). All the initial structures are seeded from a single, *xleap* generated geometry and exchange attempts are made every 2ps (Other studies have made use of lower exchange periods[49,83], but this system, with a more extreme temperature range is thought to require a slightly longer equilibration period between exchange events).

Figure 4.6.1.1. *The exchange rate for HBD2 for a 2ns REMD run. The temperature spacing is chosen to maintain an acceptance ratio between 0.2 and 0.3.*

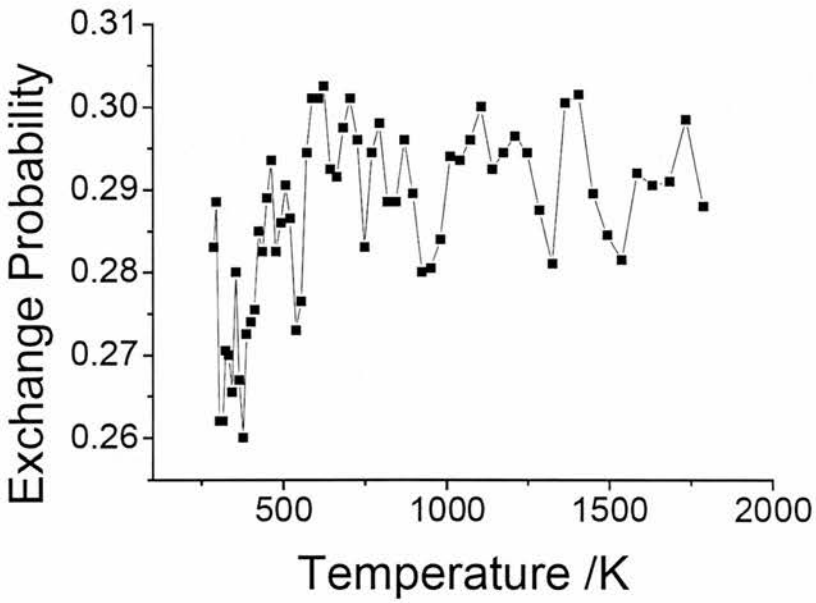
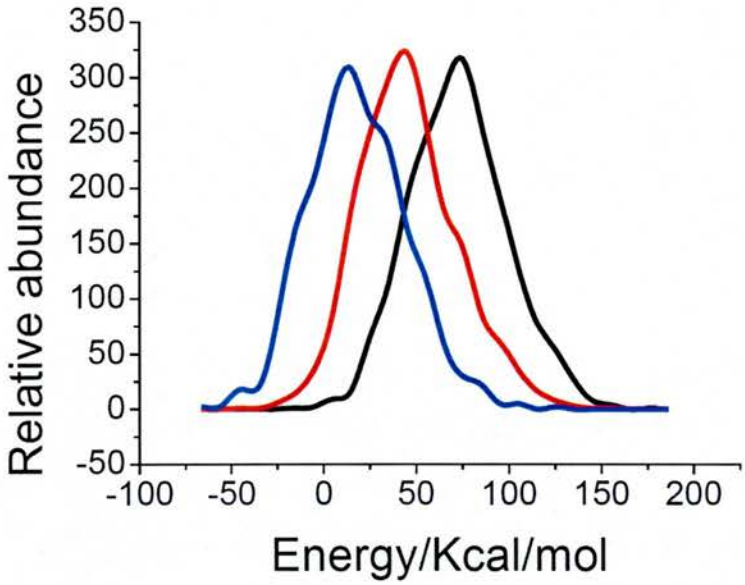
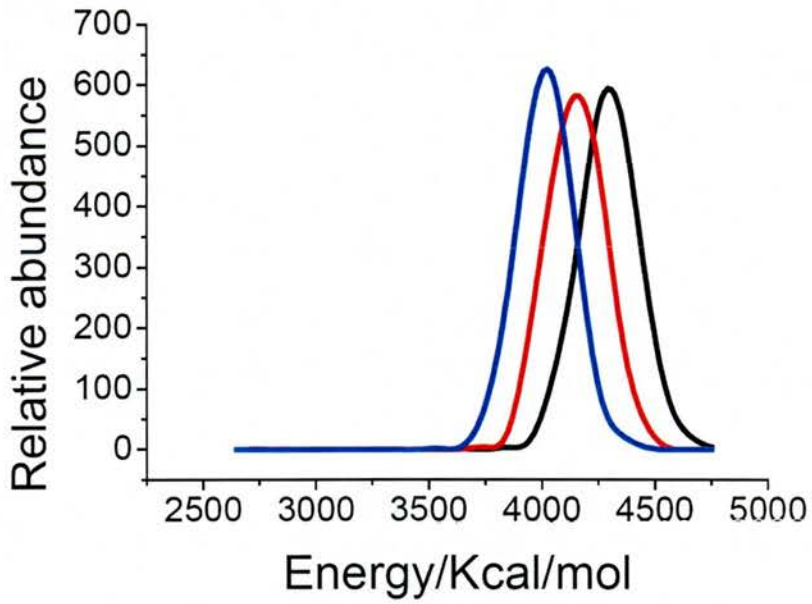


Figure 4.6.1.2. *Energy distributions for replicas within HBD2 at different temperatures (K). A: Temperatures 278 (blue), 287 (red) and 295 (black). B: Temperatures 1683 (blue), 1734 (red), 1787 (black). The extensive overlap shows that exchange is possible between adjacent temperatures.*

A



B



4.6.2. Results and discussion

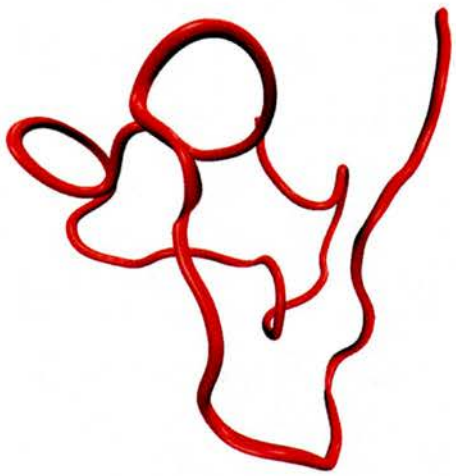
REMD, in principle, allows kinetic barriers to be overcome because replicas are permitted random walks through potential energy space with every exchange event. However, this increased sampling does not (in this case) translate to better structures (*Figure 4.6.2.1*). Indeed, only HBD2 exhibits any discernible secondary structure, and even so, only in the form of a partial N-terminal α-helix.

The observed torsional sampling at first glance looks promising (*Figure 4.6.2.2*). However, this apparently extensive sampling is likely to be an artefact of sampling geometries from conformationally diverse, high energy states. Closer inspection of the dihedral angles of individually kinetically trapped residues reveals that the correct conformational space for many key loop residues is never visited. This kinetic trapping quagmire is (again) well illustrated by the residue VAL 14 of HBD2 (*Figure 4.6.2.3*).

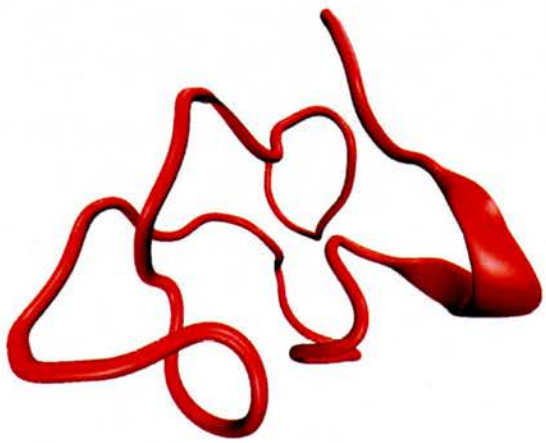
This REMD study is very much a preliminary investigation and further optimisation and experimentation is required. Future work will centre on seeding simulations from a variety of different starting geometries so that loop regions that are trapped in different regions of potential energy space can interchange. A more advanced RE procedure for dealing with dynamic bond formation and breakage is discussed in Chapter Five.

Figure 4.6.2.1. *Cartoon representations of replica exchange generated structures for each of HBD1-3. The structures are representative and sampled from the temperature of 287 K. HBD2 is the only molecule to exhibit secondary structure.*

HBD1



HBD2



HBD3

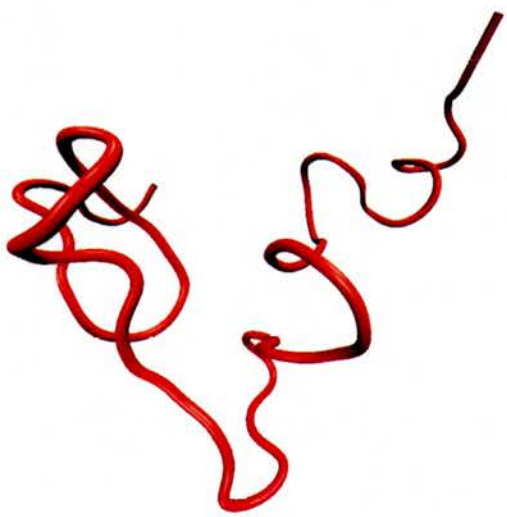
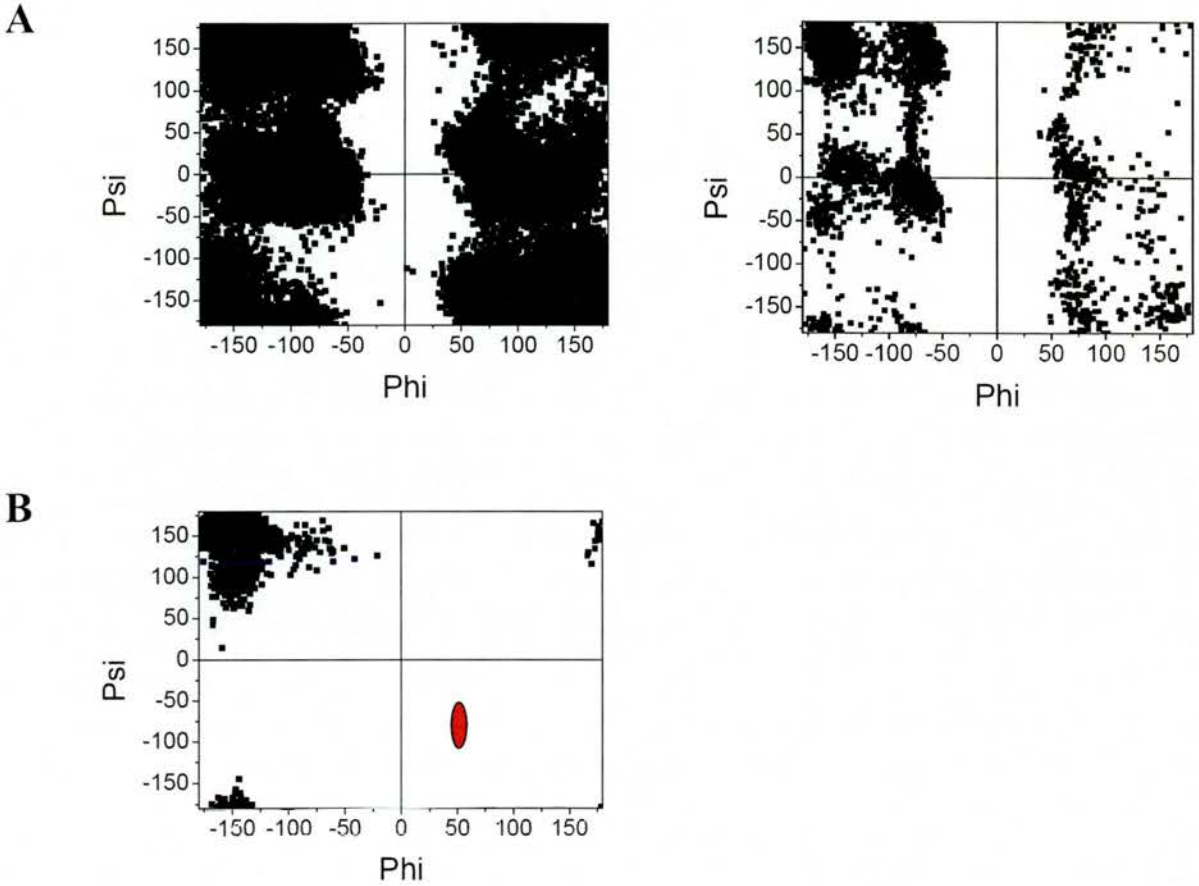


Figure 4.6.2.2. *Torsional scatter plots for HBD2. A: All residues for REMD HBD2 (left) vs SA (right). B: The residue VAL 14 shown with a schematic representation of the native NMR sampled dihedral space. The absence of sampling in the required region shows that REMD fails to relieve kinetic trapping.*



4.7. Backbone mapping

Genetic and bioinformatic studies indicate that defensins are likely to have evolved from a single ancestor[97-99]. Divergent evolution from this molecular godfather is presumed to have produced the overwhelming diversity of defensins that are now ubiquitous to the animal and plant kingdoms[99]. Defensins, whilst dissimilar in sequence, tend to exhibit similar folds to one another, especially with regards to the various loop regions and bends that intersperse the various secondary structural elements. How much of the relative ancestral fold is maintained across the defensin family?

This Section investigates the outcome of superimposing the structure of related and unrelated defensins onto one another. Can HBD1 be folded by mapping it onto the backbone of HBD2? If not, is the resulting structure at least a better geometry from which to begin the folding process?

4.7.1. Method

The mapping procedure is implemented as follows:

1. The *pdb* file of each of the template structures is cleaved so that only the backbone atoms remain, and often a single atom within the side chain.
2. The modified *pdb* file from step (1) is loaded into *xleap* using the "loadpdbusingseq" command (see *Section 4.9* for command details). This permits a least squares fit of the sequence of interest onto the template *pdb* coordinates.
3. The structure from step (2) is minimized with standard settings in a wholly non-bonded form.
4. The disulphide absent form of the structure in step (3) is canonically bonded within *xleap* and re-minimized with sander.

The resulting structure is heated from 0K to 300K and 1ns of implicitly solvated production MD is performed with standard MD settings (*Section 4.9*).

With this procedure HBD1 is superimposed onto HBD2, MBD7 and the murine orthologue, MBD8[1] (*Figure 4.7.1.1*). HBD3 in turn is superimposed onto its murine orthologue, DEFB14 (*Figure 4.7.1.2*).

[1] Defensins are often named with conflicting nomenclature. Murine β-defensins, for example, are sometimes given the abbreviation MBD, and at other times given a DEFB prefix.

Figure 4.7.1.1. *The backbone mapping sequence alignment for HBD1 superimposed onto HBD2, MBD7 and MBD8. Sequence similarity is marked in yellow, red, green and blue.*

HBD1	HBD2 Align 1	HBD2 Aign2	MBD7	MBD8
ASP	PRO		ASN	GLU
HIS	VAL	PRO	SER	PRO
TYR	THR	VAL	LYS	VAL
ASN	CYX	THR	ARG	SER
CYX	LEU	CYS	ALA	CYX
VAL	LYS	LEU	CYS	ILE
SER	SER	LYS	TYR	ARG
SER	GLY	SER	ARG	ASN
GLY	ALA	GLY	GLU	GLY
GLY	ILE	ALA	GLY	GLY
GLN	CYX	ILE	GLY	ILE
CYS	HIE	CYS	GLU	CYS
LEU	PRO	HIE	CYS	GLN
TYR	VAL	PRO	LEU	TYR
SER	PHE	VAL	GLN	ARG
ALA	CYX	PHE	ARG	CYS
CYS	PRO	CYS	CYS	ILE
PRO	ARG	PRO	ILE	GLY
ILE	ARG	ARG	GLY	LEU
PHE	TYR	ARG	LEU	ARG
THR	LYS	TYR	PHE	HIS
LYS	GLN	LYS	HIS	LYS
ILE	ILE	GLN	LYS	ILE
GLN	GLY	ILE	ILE	GLY
GLY	THR	GLY	GLY	THR
THR	CYX	THR	THR	CYS
CYS	GLY	CYS	CYS	GLY
TYR	LEU	GLY	ASN	SER
ARG	PRO	LEU	PHE	PRO
GLY	GLY	PRO	ARG	PHE
LYS	THR	GLY	PHE	LYS
ALA	LYS	THR	LYS	CYX
LYS	CYX	LYS	CYX	CYX
CYS	CYX	CYS	CYX	LYS
CYS	LYS	CYS	LYS	
LYS	LYS	LYS	PHE	
	PRO	LYS	GLN	

Figure 4.7.1.2. *The backbone mapping sequence alignment for DEFB14 superimposed onto HBD3. The yellow colouring shows that there is high sequence homology.*

DEFB14	HBD3
PHE	GLY
LEU	ILE
PRO	ILE
LYS	ASN
THR	THR
LEU	LEU
ARG	GLN
LYS	LYS
PHE	TYR
PHE	TYR
CYS	CYS
ARG	ARG
ILE	VAL
ARG	ARG
GLY	GLY
GLY	GLY
ARG	ARG
CYS	CYS
ALA	ALA
VAL	VAL
LEU	LEU
ASN	SER
CYS	CYS
LEU	LEU
GLY	PRO
LYS	LYS
GLU	GLU
GLU	GLU
GLN	GLN
ILE	ILE
GLY	GLY
ARG	LYS
CYS	CYS
SER	SER
ASN	THR
SER	ARG
GLY	GLY
ARG	ARG
LYS	LYS
CYS	CYS
CYS	CYS
ARG	ARG
LYS	ARG
LYS	LYS
LYS	LYS

4.7.2. Results and discussion

The method of backbone superposition yields some surprising results. The structure of DEFB14 for example (*Table 4.7.2.1. and Figure 4.7.2.2*) produces the expected native folds (although this cannot be confirmed until an experimental DEFB14 structure is published). In some respects this result is unsurprising for the system of DEFB14 because the two molecules share a sequence

similarity of approximately 75%. Superpositions of HBD1 onto HBD2, MBD7 and MBD8 produce less favourable results, although in some cases partially formed β-sheets and α-helices are observed (*Table 4.7.2.1.* and *Figure 4.7.2.2-5*). For each of these HBD1 superpositions the secondary structures are not located in the correct regions of the molecule, but, despite this, there is some evidence for partially correct folds. The residue LYS 10 (for example) exists in a tightly confined and unusual region of torsional space within the native state. Torsional values for the superimposed structures of HBD1 (onto both alignments of HBD2) are very similar to the native NMR case (*Figure 4.7.2.2*). This observation provides support for the notion that backbone superposition (at least in the case of a reasonably related protein) provides a better starting configuration than a randomly chosen geometry. This may be of particular benefit because the mapping procedure (without extensive MD equilibration) is brief and capable of generating starting structures within a matter of minutes.

Table 4.7.2.1. *All atom RMS deviations (in Angstroms) for superpositions of HBD1.*

Alignment	**RMS/Angstroms**
HBD1 on HBD2 (1)	8.31
HBD1 on HBD2 (2)	9.09
HBD1 on MBD7	10.99
HBD1 on MBD8	10.18

Figure 4.7.2.2. *A Ramachandran plot of torsion angles for HBD1 (A) and HBD3 (B). The red outlines in both cases denote a schematic range of values present within the solved native NMR structures. A) Superpositions involving each of HBD2 (both alignments), MBD7 and MBD8 are shown alongside the torsional angle of GLY 10 within a randomly generated xleap geometry (shown in blue). B) Torsional angles for residue CYS 23 within HBD3 and the superimposed structure of DEFB14 on HBD3.*

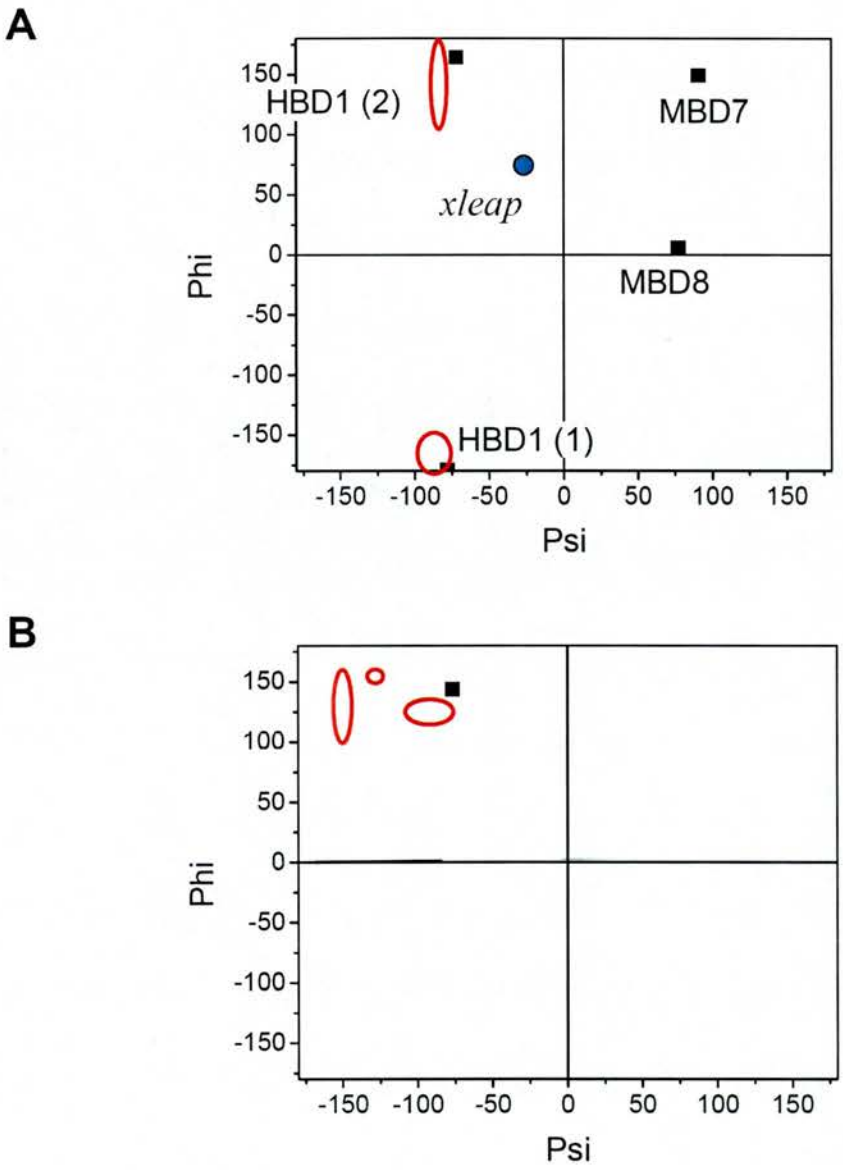


Figure 4.7.2.2. *The backbone mapping structure of DEFB14 superimposed onto native HBD3. The cartoon depicts the characteristic α and β-folds. There is unfortunately no experimental structure to validate this structure.*

Figure 4.7.2.3. *The backbone mapping structure of HBD1 superimposed onto native MBD7 (A) and native MBD8 (B). For each alignment there is some partial secondary structure formation: β-sheet structure in alignment A, and α-helical structure in alignment B.*

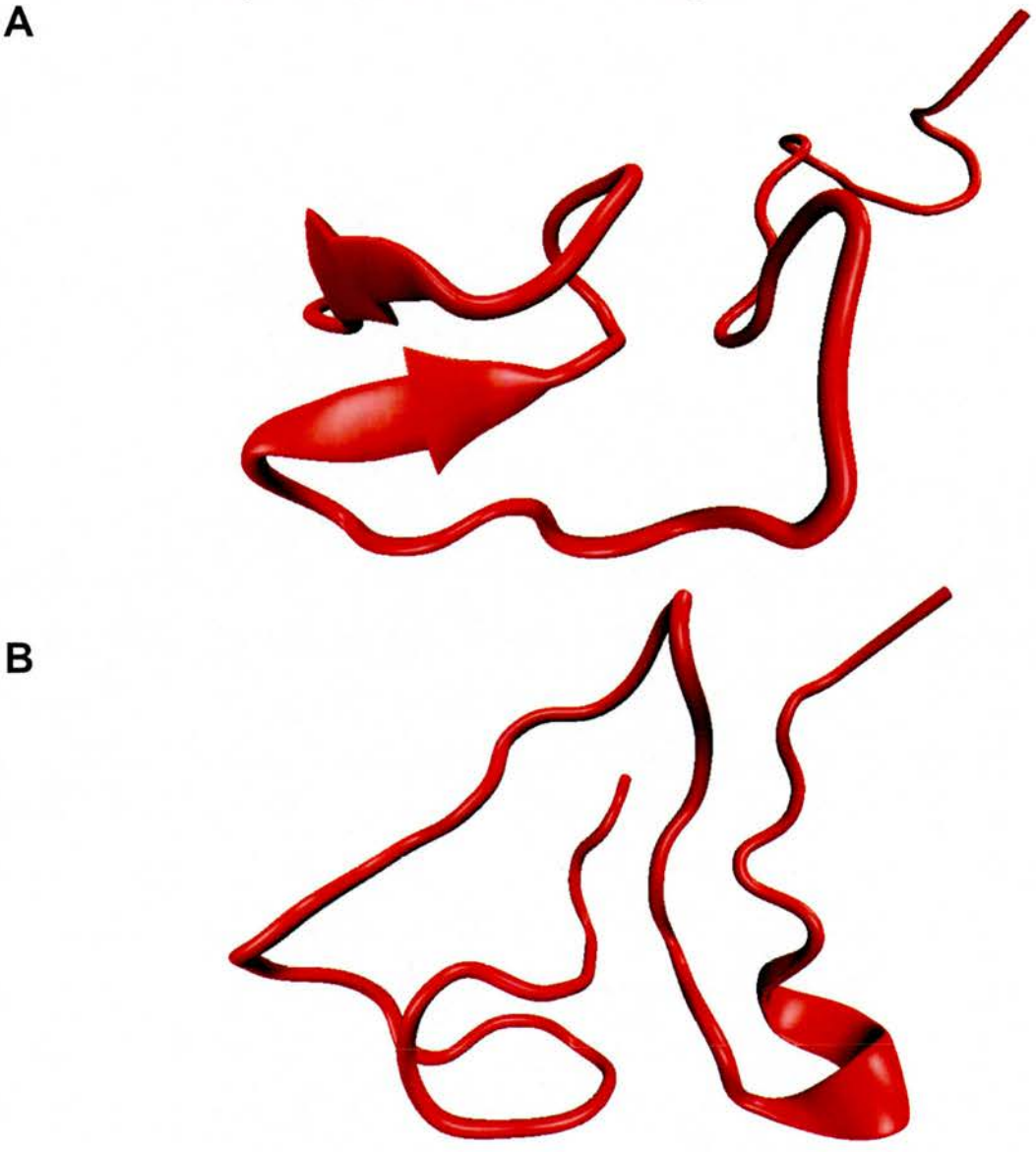


Figure 4.7.2.4. *The backbone mapping structure of HBD1 superimposed onto native HBD2 with 2 alignments (A and B). Alignment B appears to have better secondary structure than A.*

4.7.3. Conclusion

In cases in which the defensin to be folded shares homology with the template molecule, the mapping procedure is similar in many respects to that of homology modelling. Indeed, homology modelling requires that regions of like sequence share like structure. Rationally folded DEFB14 unsurprisingly possesses similar characteristics to the published homology model. Superposition onto structures that share low sequence homology does not give rise to the global fold. However, it remains a possibility that these partially formed configurations capture at least some of the "correct" folds and, hence, provide a better in *silico* starting geometry than an arbitrary *xleap* generated configuration.

4.8. Chapter conclusion

The prevailing evidence is that disulphide cross-linkages do not conformationally restrict defensins in a useful way. On the contrary, they appear to trap the molecule in local, frustrated minima from which escape to the global minimum is (for all intensive purposes) impossible. Key to solving the defensin structure prediction problem is a means for knowing when and under what conditions to break and form disulphide cross-bridges. Chapter Five discusses novel procedures for achieving this, and draws conclusions about the nature and tractability of this problem.

4.9. Standard simulation settings

All dynamics and minimizations are run (unless otherwise specified) with the settings in *Table 4.9.3*. Computer code for implementing PE sampling is written in *cshell* or FORTRAN (See CD for code).

Table 4.9.1. *Simulated Annealing Settings*

Property	Value	Explanation
High temperature	800K	A high temperature allows the potential energy space to be explored efficiently.
Low temperature	0K	Final temperature that traps molecule in minimum energy pocket. This may or may not be the global minimum.
Hot MD time	30 ps	The length of time that the system is maintained at the *high temperature*.
Cool step	Non-linear function	An exponential cooling step is thought to provide better sampling than a linear one.

Table 4.9.2. *Standard xleap commands.*

Aim	**Example command**	**Comment**
Generate amino acid sequence with name "seq".	seq = (NALA ARG ARG CLYS)	The sequence consists of N-terminal ALA, ARG,ARG and C-terminal LYS)
Bond Disulphide atoms within molecule "seq".	bond seq.13.SG seq.34.SG	In this case the "SG" atom of residue 13 and 34 are bonded to one another.
Save coordinate and topology files.	saveamberparm seq seq.parm seq.crd	"seq" identifies the molecule to be saved, "seq.parm" refers to the topology file (or Hamiltonian) and "seq.crd" refers to the coordinate file.
Load pdb file using existing sequence with name "seq".	seq = loadpdbusingseq seq.pdb (NALA ARG ARG CLYS)	This command is often useful when superimposing a modified sequence onto a pre-existing set of coodinates.
Save pdb file to disk	savepdb seq seq.pdb	"seq.pdb" refers to the name of the pdb file to be saved.

Table 4.9.3. *Standard MD settings*

Parameter	**Setting**	**Explanation**
nstlim	1000	Number of femtoseconds for MD run
igb	0 or 2	Set the implicit solvent environment (0 = gas-phase simulation) 2 = standard Born solvent simulation.
dt	0.001 or 0.002	Time step for simulation in picoseconds.
ntt	1 (3 when gb=2)	ntt=1: Employ constant temperature using weak coupling algorithm[16]. ntt=3: Employ Langevin dynamics.
temp0	n/a	Specify the target temperature for an MD run. This value is varied depending on the simulation. For simulated annealing this is varied from 800K to 0K. For standard

		dynamics this is set to 300k.
ntf	2	Ignore contributions from Hydrogen when forming interaction matrix.
ntc	2	Constrain bonds involving Hydrogen.

Table 4.9.4. *Standard Basin Hopping settings*

Parameter	**Setting**	**Explanation**
sloppyconv	0.01	RMS convergence criterion for basin hopping quenches.
tightconv	0.001	Tolerance for the RMS force in the final set of quenches.
acceptratio	0.3	The acceptance ration for the MC exploration of the transformed surface.
ediff	0.01	Quench minima only considered to be different if they are of within 0.01(or less) units of one another.
Updates	100	The number of previous steps saved in the LBFGS routine.
maxit	250 500	Both integers specify the maximum number of iterations permitted with the conjugate gradient quenches. However, the 1st integer (250) applies to sloppy quench runs where as the 2nd integer (500) applies to the final quenches.
temperature	20	The temperature at which the MC runs are conducted.
steps	3000 1.0	The first integer determines the length of the MC runs. The 2nd integer refers to the chosen annealing procedure.
step	10 0.4	The 1st integer specifies the maximum step size (degrees) and the 2nd integer specifies the level of binding energy tolerance between different atoms.
chpmax	0.4	Charmm keyword. Specifies the maximum allowed probability for twisting an angle.
chpmin	0.2	Charmm keyword: The minimum allowed probability for twisting an angle.
chnmax	10	Charmm keyword: The maximum allowed number of angles to be twisted.
chnmin	0	Charmm keyword: The minimum allowed probability for twisting an angle.

Table 4.9.5. *Standard minimization settings*

Parameter	Setting	Explanation
maxcycl	50 000	Maximum number of allowed minimization cycles.
drms	0.005	The convergence criteria determined by the root mean square force (kcal/mol)
ntpr	1000	Sample energy information every 1000 steps.
ntmin	1	Perform steepest descend minimization for n cycles and then switch to conjugate gradient descent.
ncycl	100	Perform steepest descent minimization for 100 cycles.
igb	0 or 1	Default minimization in gas-phase or Born solvent environment.

Table 4.9.6.*Table of constant pH results for HBD1.*

HBD1	Fraction protonated	Predicted Pka	Shift
AS4 1	0.001	3.708	-3.292
HIP 2	0.232	6.481	-0.519
TYR 3	0.999	9.906	2.906
TYR 14	0.999	10.229	3.229
LYS 22	1.000	10.310	3.310
TYR 28	0.998	9.614	2.614
LYS 31	0.998	9.684	2.684
LYS 33	0.998	9.670	2.670
LYS 36	0.999	9.960	2.960

Table 4.9.7. *Table of constant pH results for HBD2.*

HBD2	Fraction protonated	Predicted Pka	Shift
LYS 6	0.995	9.289	2.289
HIP 12	0.232	6.481	-0.519
TYR 20	0.998	9.774	2.774
LYS 21	0.995	9.269	2.269
LYS 32	0.997	9.588	2.588
LYS 35	0.998	9.691	2.691
LYS 36	1.000	10.356	3.356

Table 4.9.7. *Table of constant pH results for HBD3.*

HBD3	Fraction protonated	Predicted Pka	Shift
LYS 8	0.997	9.465	2.465
TYR 9	0.993	9.167	2.167
TYR 10	0.978	8.638	1.638
LYS 26	0.999	9.935	2.935
GL4 27	0.000	3.114	-3.886
GL4 28	0.000	n/a	n/a
LYS 32	0.972	8.540	1.540
LYS 39	0.990	8.987	1.987
LYS 44	0.997	9.580	2.580
LYS 45	0.999	10.004	3.004

References

(1) Levinthal.C *Journal De Chimie Physique Et De Physico-Chimie Biologique* **1968**, *65*, 44-45.

(2) Kamiya, N.; Mitomo, D.; Shea, J. E.; Higo, J. *Journal of Physical Chemistry B* **2007**, *111*, 5351-5356.

(3) Wales, D. J. *Physical Biology* **2005**, *2*, S86-S93.

(4) Yon, J. M. *Cellular and Molecular Life Sciences* **1997**, *53*, 557-567.

(5) Zdanowski, K.; Dadlez, M. *Journal of Molecular Biology* **1999**, *287*, 433-445.

(6) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *Journal of Molecular Biology* **1995**, *254*, 260-288.

(7) Neira, J. L.; Davis, B.; Ladurner, A. G.; Buckle, A. M.; Gay, G. D.; Fersht, A. R. *Folding & Design* **1996**, *1*, 189-208.

(8) Hao, M. H.; Pincus, M. R.; Rackovsky, S.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 9614-9631.

(9) Carr, J. M.; Wales, D. J. *Journal of Chemical Physics* **2005**, *123*, 234901.

(10) Doye, J. P. K.; Wales, D. J.; Miller, M. A. *Journal of Chemical Physics* **1998**, *109*, 8143-8153.

(11) Hernandez-Rojas, J.; Breton, J.; Llorente, J. M. G.; Wales, D. J. *Chemical Physics Letters* **2005**, *410*, 404-409.

(12) Hernandez-Rojas, J.; Breton, J.; Llorente, J. M. G.; Wales, D. J. *Journal of Physical Chemistry B* **2006**, *110*, 13357-13362.

(13) Hernandez-Rojas, J.; Breton, J.; Llorente, J. M. G.; Wales, D. J. *Journal of Chemical Physics* **2004**, *121*, 12315-12322.

(14) Hodges, M. P.; Wales, D. J. *Chemical Physics Letters* **2000**, *324*, 279-288.

(15) Naumkin, F. Y.; Wales, D. J. *Computer Physics Communications* **2002**, *145*, 141-155.

(16) Naumkin, F. Y.; Wales, D. J. *Molecular Physics* **1998**, *93*, 633-648.

(17) Rapacioli, M.; Calvo, F.; Spiegelman, F.; Joblin, C.; Wales, D. J. *Journal of Physical Chemistry A* **2005**, *109*, 2487-2497.

(18) Wales, D. J.; Doye, J. P. K. *Journal of Physical Chemistry A* **1997**, *101*, 5111-5116.

(19) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368-1372.

(20) Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141-151.

(21) Burgess, E. M. *Abstracts of Papers of the American Chemical Society* **1988**, *195*, 41-Csec.

(22) Carlacci, L.; Englander, S. W. *Abstracts of Papers of the American Chemical Society* **1996**, *211*, 66-Comp.

(23) Agostini, F. P.; Soares-Pinto, D. D. O.; Moret, M. A.; Osthoff, C.; Pascutti, P. G. *Journal of Computational Chemistry* **2006**, *27*, 1142-1155.

(24) Bezuglov, A.; Vargas, J. E. *Foundations of Intelligent Systems, Proceedings* **2006**, *4203*, 642-651.

(25) Bohm, M.; Stadlthanner, K.; Gruber, P.; Theis, F. J.; Lang, E. W.; Tome, A. M.; Teixeira, A. R.; Gronwald, W.; Kalbitzer, H. R. *Ieee Transactions on Biomedical Engineering* **2006**, *53*, 810-820.

(26) Boisbouvier, J.; Blackledge, M.; Sollier, A.; Marion, D. *Journal of Biomolecular Nmr* **2000**, *16*, 197-208.

(27) Chou, K. C.; Carlacci, L. *Protein Engineering* **1991**, *4*, 661-667.

(28) Collura, V.; Higo, J.; Garnier, J. *Protein Science* **1993**, *2*, 1502-1510.

(29) Dinola, A.; Roccatano, D.; Berendsen, H. J. C. *Proteins-Structure Function and Genetics* **1994**, *19*, 174-182.

(30) Ferguson, D. M.; Garrett, D. G. *Abstracts of Papers of the American Chemical Society* **1996**, *211*, 32-Comp.

(31) Garrett, D. G.; Kastella, K.; Ferguson, D. M. *Journal of the American Chemical Society* **1992**, *114*, 6555-6556.

(32) Goodsell, D. S.; Morris, G. M.; Olson, A. J. *Journal of Molecular Recognition* **1996**, *9*, 1-5.

(33) Goodsell, D. S.; Olson, A. J. *Proteins-Structure Function and Genetics* **1990**, *8*, 195-202.

(34) Grigoriev, I. V.; Rakhmaninova, A. B.; Mironov, A. A. *Journal of Biomolecular Structure & Dynamics* **1998**, *16*, 115-122.

(35) Hansmann, U. H. E.; Okamoto, Y. *Physica A* **1994**, *212*, 415-437.

(36) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671-680.

(37) Liu, Z. J.; Mao, F. L.; Li, W. Z.; Han, Y. Z.; Lai, L. H. *Journal of Molecular Modelling* **2000**, *6*, 1-8.

(38) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. *Journal of Computer-Aided Molecular Design* **1996**, *10*, 293-304.

(39) Ogura, T.; Sato, C. *Journal of Structural Biology* **2006**, *156*, 371-386.

(40) Okamoto, Y. *Progress of Theoretical Physics Supplement* **2000**, 301-310.

(41) Okamoto, Y. *International Journal of Modern Physics C* **1999**, *10*, 1571-1582.

(42) Sankararamakrishnan, R.; Konvicka, K.; Mehler, E. L.; Weinstein, H. *International Journal of Quantum Chemistry* **2000**, *77*, 174-186.

(43) Slusarz, M. J.; Sikorska, E.; Slusarz, R.; Ciarkowski, J. *Journal of Medicinal Chemistry* **2006**, *49*, 2463-2469.

(44) Snow, M. E. *Journal of Computational Chemistry* **1992**, *13*, 579-584.

(45) Wang, Y.; Li, W. *Chemical Research in Chinese Universities* **2005**, *21*, 73-77.

(46) Ash, W. L.; Stockner, T.; Tieleman, D. P. *Biophysical Journal* **2005**, *88*, 152a-152a.

(47) Baumketner, A.; Shea, J. E. *Journal of Molecular Biology* **2007**, *366*, 275-285.

(48) Baumketner, A.; Shea, J. E. *Theoretical Chemistry Accounts* **2006**, *116*, 262-273.

(49) Beck, D. A. C.; White, G. W. N.; Daggett, V. *Journal of Structural Biology* **2007**, *157*, 514-523.

(50) Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. *Journal of Chemical Physics* **2004**, *121*, 10748-10756.

(51) Faraldo-Gomez, J. D.; Kutluay, E.; Jogini, V.; Zhao, Y. X.; Heginbotham, L.; Roux, B. *Journal of Molecular Biology* **2007**, *365*, 649-662.

(52) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins-Structure Function and Bioinformatics* **2004**, *56*, 310-321.

(53) Furlan, S.; La Penna, G.; Perico, A.; Cesaro, A. *Macromolecules* **2004**, *37*, 6197-6209.

(54) Jang, S.; Kim, E.; Pak, Y. *Proteins-Structure Function and Bioinformatics* **2006**, *62*, 663-671.
(55) Jas, G. S.; Kuczera, K. *Biophysical Journal* **2004**, *87*, 3786-3798.
(56) Kawashima, Y.; Sasaki, Y. C.; Sugita, Y.; Yoda, T.; Okamoto, Y. *Molecular Simulation* **2007**, *33*, 97-102.
(57) Kawashima, Y.; Sugita, Y.; Yoda, T.; Okamoto, Y. *Chemical Physics Letters* **2005**, *414*, 449-455.
(58) Kokubo, H.; Okamoto, Y. *Chemical Physics Letters* **2004**, *383*, 397-402.
(59) Kubitzki, M. B.; de Groot, B. L. *Biophysical Journal* **2007**, *92*, 4262-4270.
(60) La Penna, G.; Mitsutake, A.; Masuya, M.; Okamoto, Y. *Chemical Physics Letters* **2003**, *380*, 609-619.
(61) Li, P. C.; Huang, L.; Makarov, D. E. *Journal of Physical Chemistry B* **2006**, *110*, 14469-14474.
(62) Li, W. F.; Zhang, J.; Wang, W. *Proteins-Structure Function and Bioinformatics* **2007**, *67*, 338-349.
(63) Liu, Z. W.; Pophristic, V. *Abstracts of Papers of the American Chemical Society* **2006**, *231*, -.
(64) Lou, H. F.; Cukier, R. I. *Journal of Physical Chemistry B* **2006**, *110*, 24121-24137.
(65) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96-123.
(66) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Journal of Chemical Physics* **2003**, *118*, 6664-6675.
(67) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Journal of Chemical Physics* **2003**, *118*, 6676-6688.
(68) Mobley, D. L.; Chodera, J.; Dill, K. *Biophysical Journal* **2005**, *88*, 332a-332a.
(69) Moskovsky, A. A.; Vanovschi, V. V.; Konyukhov, S. S.; Nemukhin, A. V. *International Journal of Quantum Chemistry* **2006**, *106*, 2208-2213.
(70) Yang, L; Grubb, M. P.; Gao, Y, Q; *Journal of Chemical Physics* **2007**, *126*, 125102.
(71) Murata, K.; Sugita, Y.; Okamoto, Y. *Journal of Theoretical & Computational Chemistry* **2005**, *4*, 411-432.
(72) Murata, K.; Sugita, Y.; Okamoto, Y. *Journal of Theoretical & Computational Chemistry* **2005**, *4*, 433-448.
(73) Nagasima, T.; Sugita, Y.; Mitsutake, A.; Okamoto, Y. *Computer Physics Communications* **2002**, *146*, 69-76.
(74) Nguyen, P. H.; Mu, Y. G.; Stock, G. *Proteins-Structure Function and Bioinformatics* **2005**, *60*, 485-494.
(75) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. *Proteins-Structure Function and Bioinformatics* **2005**, *61*, 795-808.
(76) Nishino, M.; Sugita, Y.; Yoda, T.; Okamoto, Y. *Febs Letters* **2005**, *579*, 5425-5429.
(77) Nymeyer, H.; Garcia, A. E. *Biophysical Journal* **2004**, *86*, 624a-624a.
(78) Paschek, D.; Garcia, A. E. *Biophysical Journal* **2005**, *88*, 41a-41a.
(79) Paschek, D.; Garcia, A. E. *Physical Review Letters* **2004**, *93*, -.
(80) Pitera, J. W.; Swope, W. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 7587-7592.
(81) Rao, F.; Caflisch, A. *Journal of Chemical Physics* **2003**, *119*, 4035-4042.
(82) Ravindranathan, K. P.; Gallicchio, E.; Friesner, R. A.; McDermott, A. E.; Levy, R. M. *Journal of the American Chemical Society* **2006**, *128*, 5786-5791.
(83) Rhee, Y. M.; Pande, V. S. *Biophysical Journal* **2003**, *84*, 775-786.
(84) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins-Structure Function and Genetics* **2002**, *46*, 225-234.
(85) Suenaga, A. *Journal of Molecular Structure-Theochem* **2003**, *634*, 235-241.
(86) Sun, Y.; Latour, R. A. *Abstracts of Papers of the American Chemical Society* **2006**, *231*, -.
(87) Tsai, H. H. G.; Nussinov, R. *Biophysical Journal* **2005**, *88*, 401a-401a.

(88) Yamamoto, R.; Kob, W. *Physical Review E* **2000**, *61*, 5473-5476.

(89) Yoda, T.; Sugita, Y.; Okamoto, Y. *Proteins-Structure Function and Bioinformatics* **2007**, *66*, 846-859.

(90) Yoshida, K.; Yamaguchi, T.; Okamoto, Y. *Chemical Physics Letters* **2005**, *412*, 280-284.

(91) Zhang, W.; Wu, C.; Duan, Y. *Journal of Chemical Physics* **2005**, *123*, 154105.

(92) Mongan, J.; Case, D. A.; McCammon, J. *Protein Science* **2004**, *13*, 219-219.

(93) Mongan, J.; Case, D. A.; McCammon, J. A. *Journal of Computational Chemistry* **2004**, *25*, 2038-2048.

(94) Mongan, J. T.; Case, D. A. *Abstracts of Papers of the American Chemical Society* **2005**, *229*, U768-U768.

(95) Mongan, J.; Case, D. A.; McCammon, J. A. *Biophysical Journal* **2005**, *88*, 515a-515a.

(96) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nature Structural Biology* **2002**, *9*, 425-430.

(97) Morrison, G. M.; Semple, C. A. M.; Kilanowski, F. M.; Hill, R. E.; Dorin, J. R. *Molecular Biology and Evolution* **2003**, *20*, 460-470.

(98) Semple, C. A. M.; Maxwell, A.; Gautier, P.; Kilanowski, F. M.; Eastwood, H.; Barran, P. E.; Dorin, J. R. *Bmc Evolutionary Biology* **2005**, *5*, -.

(99) Semple, C. A. M.; Taylor, K.; Eastwood, H.; Barran, P. E.; Dorin, J. R. *Biochemical Society Transactions* **2006**, *34*, 257-262.

(100) Darby, N. J.; Vanmierlo, C. P.; Creighton, T. E. *Febs Letters* **1991**, *279*, 61-64.

Chapter Five

Rational Structure Prediction:

Probabilistic Disulphide Bridge Algorithms

5.1. Introduction

Chapter Four provides some evidence for the existence of kinetic trapping within the *in silico* folding environment. This Chapter discusses procedures for overcoming this trapping problem by rearranging the offending disulphide bridges dynamically during MD or BH.

When considering the defensin disulphide bridge problem it is instructive to examine systems for which folding has been largely characterised. Bovine Pancreatic Trypsin Inhibitor (BPTI), for example, is a 58 residue, 6 cysteine protein[1] known to fold via partially disulphide-bridged intermediates[2-14]. The reduced form of the protein is known to adopt a predominantly denatured configuration[13] in which single pair disulphide bridging is observed for each of the 15 possible pairings[14]. During this phase, the first, crucial, partially folded kinetic intermediate, CYS 30–CYS 51 (hereon referred to as 30-51) forms[15]. The second disulphide bridge, 5–55, can then form by direct, sequential canonical bonding, or via an alternative (kinetically favoured) set of incorrectly paired states that eventually exchange to produce the correct pairing[15,16]. Once these first two canonical pairings have formed (5-55, 30-51), the last remaining cysteine residues (14 and 38)[13] are brought together at the surface of the molecule, and their bonding to one another is then rapid.

In *vitro* defensin folding requires redox conditions[17,18], a fact that suggests a BPTI-like, dynamic-bond folding pathway with many intermediates, and not a two state, nucleation-condensation mechanism[19-23] common to many other small, globular proteins.

Some key questions raised for the defensin folding problem are thus:

1. At what point during the folding process do canonical connectivity's arise?
2. Do non-canonical connectivity's arise during the folding process? If so, do they form essential folding intermediate states?
3. At what stage during the folding process can disulphide bridges be regarded as "glued together"?
4. Do initial folding events partition the potential energy surface in a manner that directs future productive folding?

These questions cannot be adequately answered with current forcefield methodologies because the system Hamiltonian in such procedures describes a fixed set of inter-atomic bonds that do not (and cannot) exchange during the MD process. Indeed, current simulations of systems such as BPTI invoke single Hamiltonian descriptions of the entire folding process[24-26].

The challenge, as set out in this chapter, is to develop a procedure that forms and breaks disulphide bridges with an appropriate criterion.

5.2. Probabilistic bonding method

Each procedure that invokes the mobile disulphide bridge method employs a probabilistic bonding algorithm that is here discussed.

5.2.1. Dynamic bond formation and cleavage

The AMBER suite of programs employs a single Hamiltonian operator for establishing the energy of a system for a given set of coordinates. The challenge for a mobile disulphide bridge algorithm is thus to:

1. Modify this single Hamiltonian operator so that it can at any time correspond to a differently bonded (or non-bonded) state.
2. Know when to break and form bonds (and hence know when to modify the Hamiltonian)

Points 1 and 2 (above) are tackled by generating *a priori* the entire repertoire of Hamiltonian states for each of the 8 conceivable canonical bonding connectivities (*Table 5.2.1.1.*). These are then toggled between one another with a probability P_{ab} that varies as a non-linear function of the Euclidean displacement, D_{ab}, between bonded (or potentially bonded) atom pairs a and b:

$$P_{ab} = \frac{1}{1+\exp(A_{ab}D_{ab} - B_{ab})} \qquad (1)$$

Where

$$D_{ab} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2} \qquad (2)$$

In which x, y and z refer to the coordinates of atoms a and b, and where the term B_{ab} is a bias term that shifts the function along the horizontal axis; it can be interpreted as the distance at which the probability of bond formation is 0.5. The term A_{ab} determines the rate of change of probability for a given change in distance and can be regarded in a crude sense as the functions "steepness". The function is sigmoidal in nature and therefore increases linearly

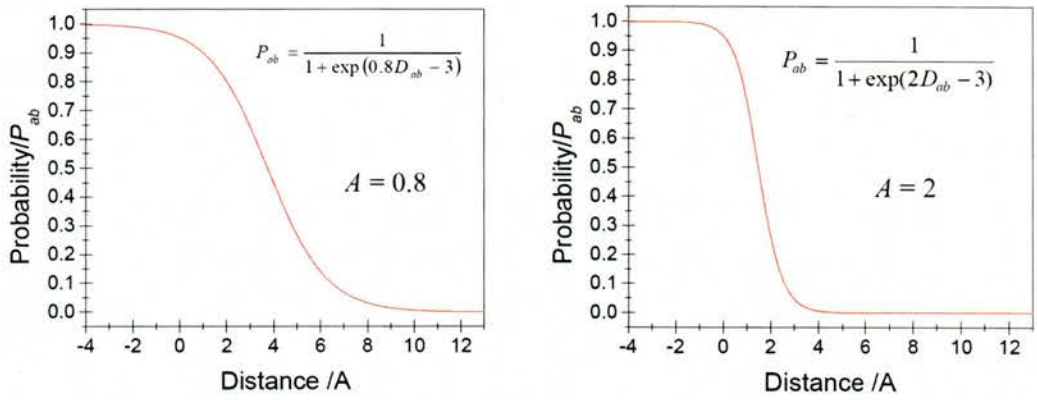
over small displacements from the centre point and non-linearly at the "extreme" ends (*Figure 5.2.1.2*).

Equation 1 is applied to snapshots of the coordinates at given intervals (usually 1 – 2 picoseconds) so that the Euclidean distance between bonding pairs can be tracked, and bonds formed or broken accordingly. Each change in bonding state is mapped to a corresponding Hamiltonian for the existing set of coordinates and velocities.

Table 5.2.1.1. *The conceivable disulphide connectivity's for a typical defensin. The bonding configuration "1-5", "2-4" and "3-6" refer to the cysteine indexes (beginning from the N-terminus). The shaded boxes indicate that a bond exists between the specified pair. The Hamiltonians (*$\mathbf{H}_i$*) corresponding to each connectivity are generated prior to simulation start.*

	S-S connectivity: Make a bond?		
$\mathbf{H}_i$	**1-5**	**2-4**	**3-6**
1	no	no	no
2	yes	no	no
3	no	yes	no
4	no	no	yes
5	yes	yes	no
6	no	yes	yes
7	yes	no	yes
8	yes	yes	yes

***Figure 5.2.1.2**. Example plots of probabilistic sigmoidal function (P_{ab}) vs Distance. Left: A= 0.8, B = -3. Right: A=2, B=-3. Comparison of the two graphs illustrates that the value of A effects probability gradient.*



The sigmoidal function is chosen because:

1. It is conveniently constrained to values between 0 and 1 and therefore lends itself well to probabilistic treatment.
2. The function dies off exponentially with increasing displacement in a manner similar to dipole-dipole interactions.
3. The function contains parameters A_{ab} and B_{ab} that can be optimized for the system at hand.

5.2.2. Dealing with bonds that "never form"

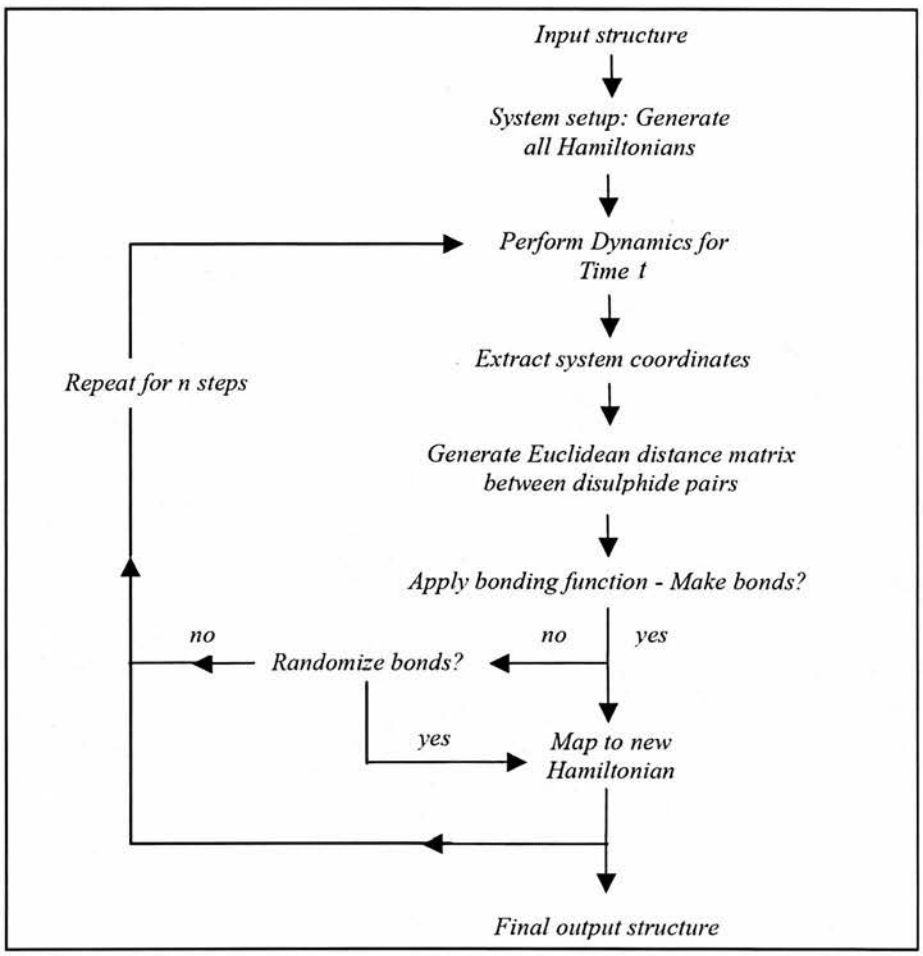
MD runs frequently give rise to bonding pairs that (if left undisturbed) remain far apart in Euclidean space for the duration of the simulation. A randomizing term is present in the algorithm to combat this phenomenon so that bonding can be "forced" upon systems that would otherwise remain in a non-bonded state. A brief conjugate gradient descent minimization is performed in order to alleviate non-equilibrium bond-lengths.

5.2.3. Implementation

The probabilistic bonding paradigm is implemented with an individually compiled Fortran 90 program in which system calls* to the AMBER suite of programs are invoked at appropriate junctures of program flow. In this manner, the Hamiltonians for the system are generated *a priori* via the *tleap* component of AMBER, whilst coordinate and velocity information are generated with *sander*. The probabilistic function makes use of the coordinate and velocity output files following pre-defined trajectory time-lengths and any changes in bonding state are mapped to the correct Hamiltonian. The overall procedure is summarized in *Figure 5.2.3.1*.

* The "system call" commands exist within Fortran 90 to permit system (and other) software to be run and terminated within the program flow.

Figure 5.2.3.1. *A schematic representation of the mobile disulphide bridge algorithm. Normal MD is followed by applying the probabilistic bonding procedure to a snapshot of the coordinates. A decision is then made as to which Hamiltonian to use.*



5.3. Simulated Annealing

5.3.1. Method

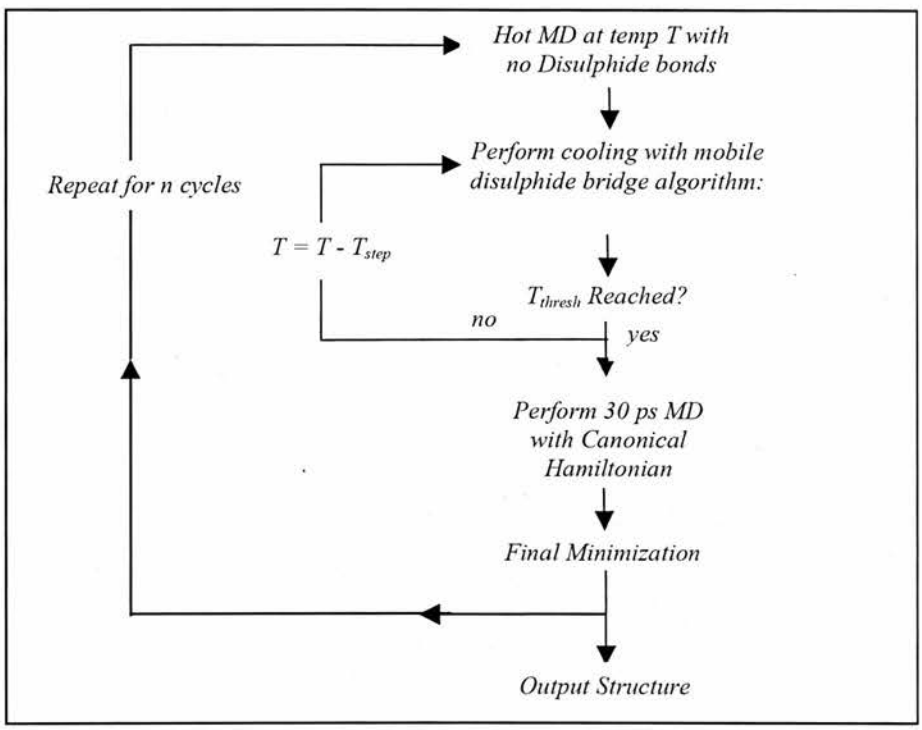
The annealing algorithm for the mobile disulphide case is identical to that presented in Chapter Four, except that:

1. A randomized heating temperature, cool step and heating period are used*.
2. During the heating step a wholly non-bonded Hamiltonian (as opposed to the canonical one) is employed to permit unhindered sampling of phase space.
3. The mobile disulphide bridge algorithm is applied during the cooling phase.
4. Following the final cooling phase, the canonical Hamiltonian (if it hasn't already arisen) is invoked for a 30 ps interval prior to final minimization.

The modified SA algorithm is illustrated in *Figure 5.3.1.1.* The standard settings for the procedure are summarised in *Section 5.7.*

* The extent of this "randomisation" is set in the input file to lie within a user-defined range.

Figure 5.3.1.1. *A schematic summary of the mobile disulphide bond SA procedure. The scheme is similar to that of standard SA, except for the heating phase in which the wholly non-bonded Hamiltonian is invoked, and the last phase, which requires a further 30 ps of equilibration prior to final minimization.*



5.3.2. Results and discussion

The mobile disulphide bridge SA procedure gives rise to some desirable secondary structural motifs in the system of HBD2 (*Figure 5.3.2.1* and *Figure 5.3.2.2*), but not in HBD1 or HBD3 (results not shown). The reason for this failure (and partial success) is not clear. Indeed, analysis of the torsional sampling for key residues of HBD2 reveals that kinetic trapping is (to a large extent) circumvented for the residue of VAL 14 (*Figure 5.3.2.3*). It remains a possibility that the problem lies with the SA procedure, which has not enjoyed success for systems larger than 20 residues[12,27-33]. Another possibility is that the procedure fails to capture vital, non-canonical intermediates. Future work will centre on permitting non-canonical bonding and (rather than randomising formation and breakage) directing disulphide pairing in a concerted fashion.

Figure 5.3.2.1. *A mobile bond-SA predicted structure of HBD2. There is some partial correct β-sheet formation.*

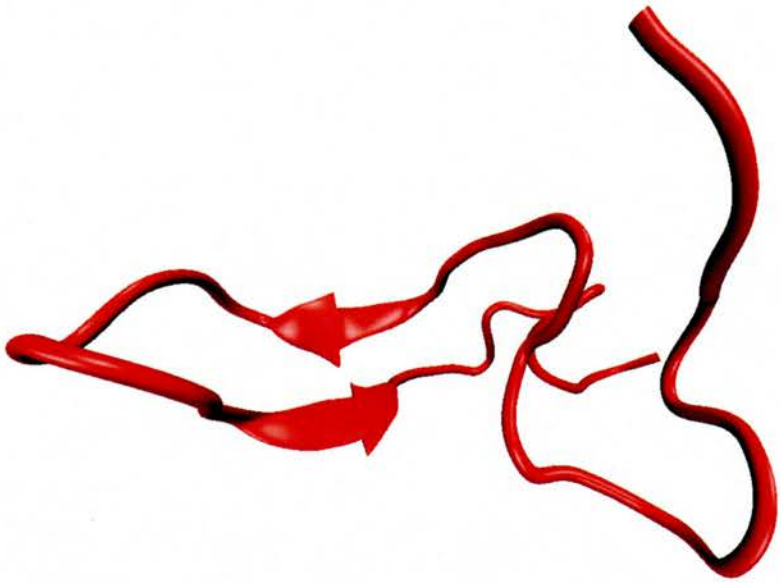


Figure 5.3.2.2. *Structural fitness measure of mobile disulphide SA generated HBD2 (black) versus 20 native NMR structures (red).*

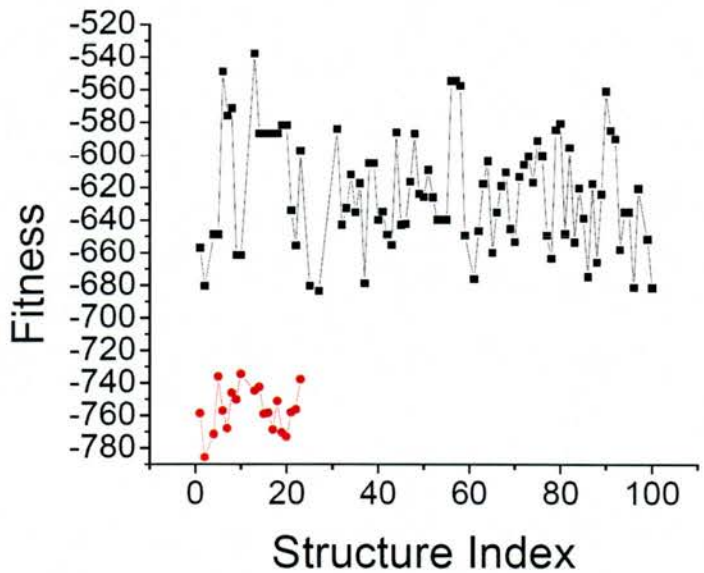
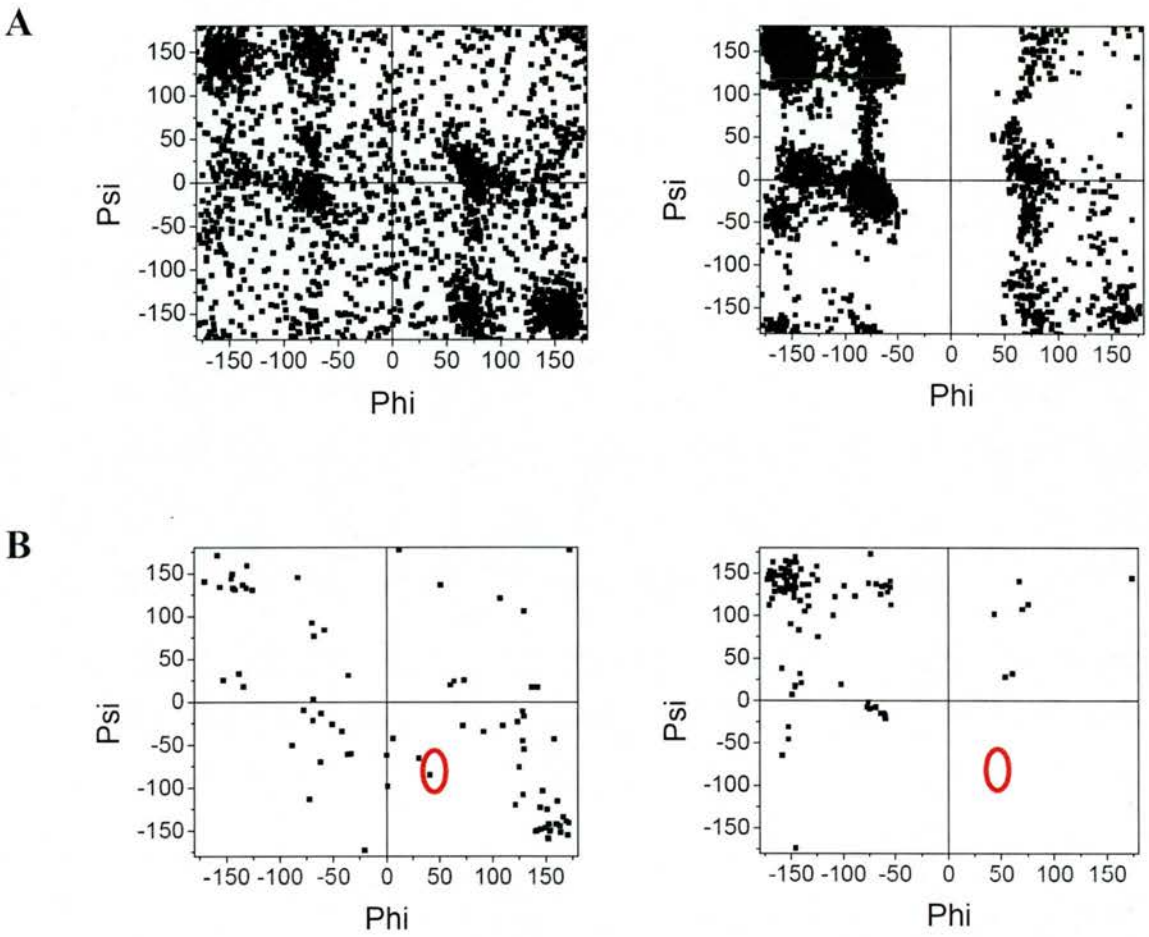


Figure 5.3.2.3. *A: Torsional scatter plots of all residues of HBD2 for mobile disulphide bridge SA (left) and standard SA (right). B: Torsional scatter plots for the residue VAL14 of HBD2 for mobile disulphide SA (left) versus normal SA (right). The torsional region sampled by the native NMR structures is shown in red on both plots. The sampling distribution suggests that the mobile disulphide procedure avoids the problem of kinetic trapping.*



5.4. "*In vivo*" BH folding

Experimental and theoretical studies of proteins such as Barnase and Chymotrypsin inhibitor 2 (CI2) indicate that small molecule folding occurs via nucleated intermediates[22,34-44] in a two state[36,45] or three state process[11].

In this Section the possibility of nucleation events directing (correct) disulphide-bridge and loop formation is investigated with a dynamic disulphide-bridge-BH algorithm. With this

procedure an initial nucleating amino acid sequence is sequentially grown and successively optimised with BH. It is hoped that preliminary folding events provide a molecular scaffold for subsequent folding events in a manner that directs better loop formation. This strategy therefore remains a possibility for both folding proteins, and also for generating families of differently configured (and kinetically trapped) loop regions (the structures of which can then be used by other algorithms).

5.4.1. Method

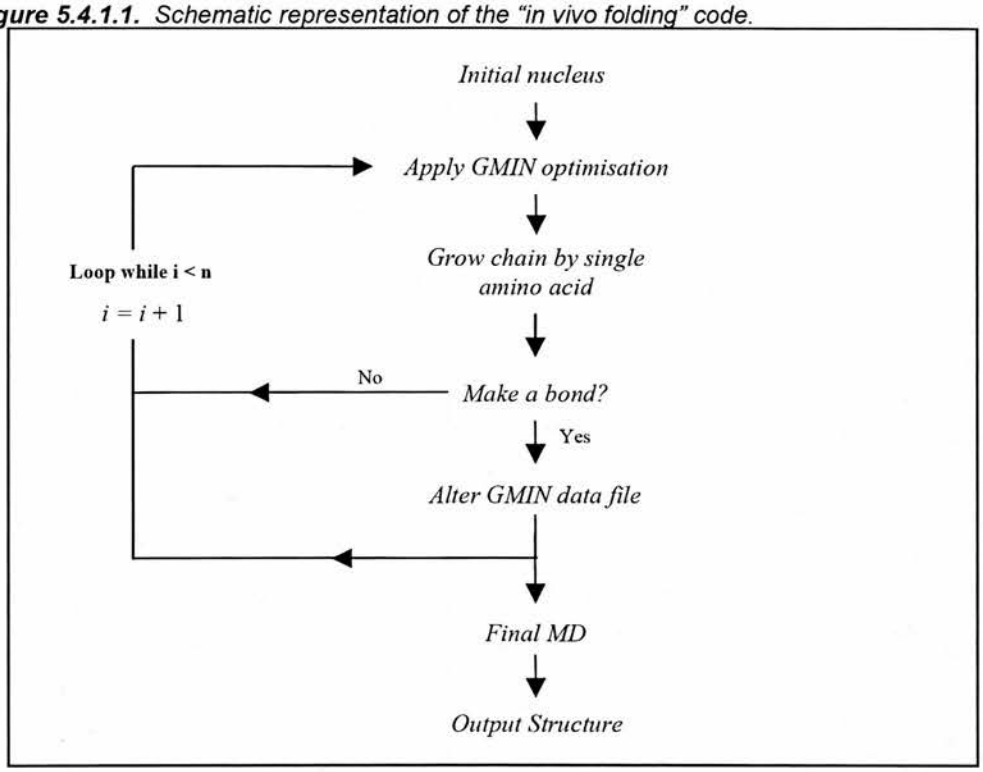
A schematic representation of the process is given in *Figure 5.4.1.1*. Bonds are formed and broken with the standard procedure following every elongation phase. Geometric optimization is carried out at each stage with GMIN 2.2.1 using 200-500 BH steps. Bonding within this framework is achieved by altering the GMIN input file with an appropriate *cshell* script (that invokes the same probabilistic bonding algorithm as with the MD case).

5.4.1. Results and discussion

Preliminary results with this method do not support the notion that productive nucleation is possible from an N-terminal residue seed (*Figure 5.4.2.1*). Indeed, both HBD1 and HBD3 fail to exhibit correct torsional sampling for residues in key loop regions** (*Figure 5.4.2.2*). However, despite this undesirable result, it remains a possibility that there could be beneficial nucleation from different sequence seeds. For example, it is conceivable that seeding an "*in vivo*" simulation from a hydrophobic portion of the sequence could give rise to a productive hydrophobic core, as observed in the two-state hydrophobic collapse model[46-48]). The GMIN parameter set may not be well optimized for this class of problem, and thus, future work will also focus on developing more applicable parameter sets.

** There are no results for HBD2 for the reason stated in chapter 4; namely that GMIN has some trouble dealing with N-terminal proline torsional variations.

Figure 5.4.1.1. *Schematic representation of the "in vivo folding" code.*

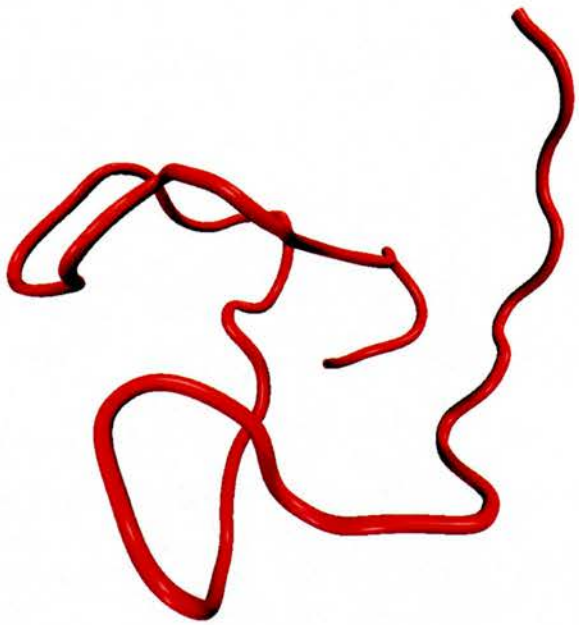


5.4.2. Conclusion

A "seeding" approach may offer an additional means of generating geometrically favourable starting configurations with correct (or partially correct) relative folds. This method is attractive because simulation time is in the order of hours.

Figure 5.4.2.1. *A cartoon representation of HBD1 and HBD3 following the "in vivo" folding procedure. No discernible secondary structure is detected.*

HBD1



HBD3

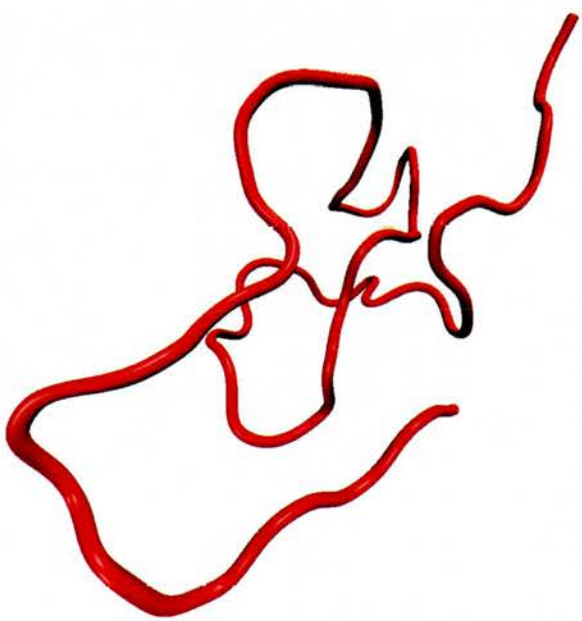
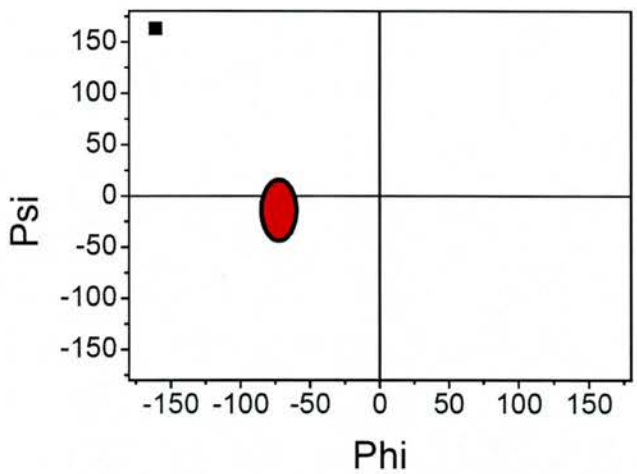
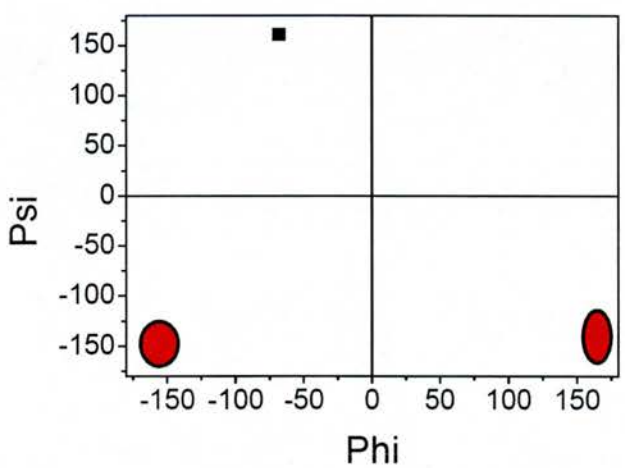


Figure 5.4.2.2. *Torsional dihedral angles for the residue LYS 31 (of HBD1) and GLY 31 (of HBD3). The native NMR sampled regions are shown schematically in red and illustrate that the required loop region folds are not located.*

HBD1



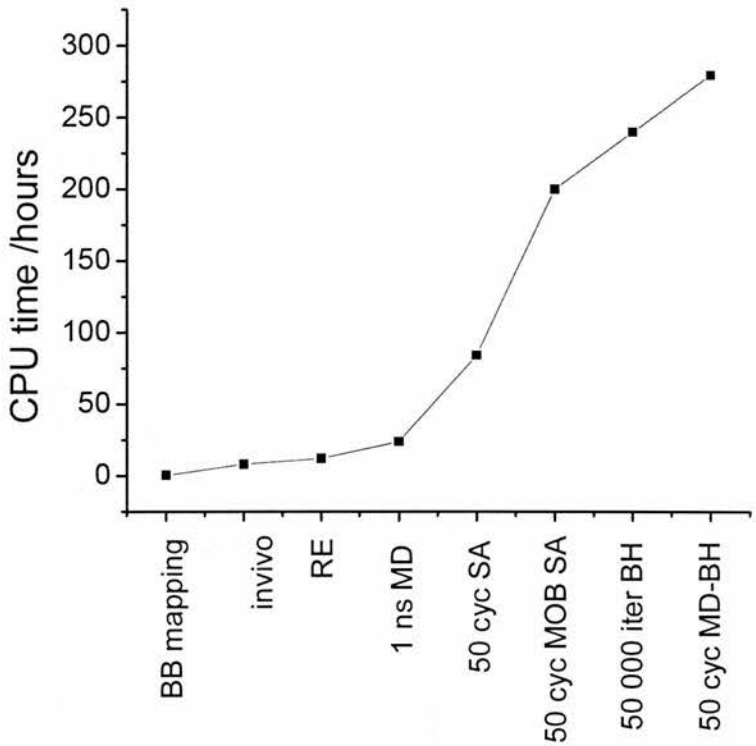
HBD3



5.5. Computational resource comparison of folding algorithms

Computational resource availability is frequently a limiting factor in solving any protein folding problem. Indeed, were infinite computer power in existence, the field of protein folding would not exist because the global minimum would be located by simple, exhaustive sampling. A brief comparison of the resource consumption of each algorithm is here presented so that the algorithm in question can be weighed alongside its computational efficiency (*Figure 5.5.1.1*).

Figure 5.5.1.1. *A relative comparison of computational resource use for different folding algorithms. Keywords: "BB" – Backbone, "cyc" – cycles, "iter" – iterations, "MOB" – Mobile disulphide bridge procedure. All non-RE resource use refers to total CPU time within a Dual core, Xeon, 2.8GHZ, 1 MB RAM system. RE is conducted on IBM Bluegene and the time quote in this instance refers to the total time taken for a 512 processor submission.*



5.6. Conclusion

The in *silico* evidence suggests strongly that the failure of static-bond folding algorithms to solve the defensin problem lies with their inability to rearrange incorrect disulphide cross-linked loop-regions into native (or "more" native) loops and folds. The dynamic-bond folding strategy, whilst being unable to locate the global structure, nevertheless alleviates the problem of kinetic trapping ubiquitous to static-bond procedures. Indeed, the ability to rearrange bonds dynamically will likely be an essential requirement of future cross-linked protein folding strategies. The work presented here is a first attempt at modelling dynamic bond formation and deals only with canonical pairings. Future work will centre on including non-canonical connectivities into the repertoire of Hamiltonian states.

5.6.1. Are cross-links a help or a hindrance?

It was originally hoped that the innate cross-linking within the β-defensin family of proteins would reduce the available conformational space, and hence make the folding problem easier to conquer. The results presented here at first glance indicate the opposite; that a restricted conformational space inhibits beneficial sampling. However, the true picture is likely more complex than this. A well designed algorithm should be able to take advantage of both restricted and unrestricted conformational space by selectively invoking dynamic and static descriptions of the molecular system. In this way, dynamic bond descriptions would sample globally by allowing different loop configurations to be explored, and static-bond descriptions would sample locally *within* a given, beneficially restricted region. Cross-linking, invoked at the right moment, would therefore likely provide a useful means of selectively ignoring undesirable regions of potential energy space. The probabilistic SA algorithm presented in this chapter provides the first combined dynamic-static approach to tackling the problem and future work will focus on improving the basis for switching between global and local sampling.

5.6.2. Future work

There may not be a single best strategy for folding defensins. Indeed, the answer may lie with combining many strategies. For example, it may well be that backbone mapping provides a good initial structure from which to begin, and that subsequent algorithms can then solve this partially native structure. Future methodologies (as stated in *Section 5.6.1*) will focus on combining aspects of both dynamic and static bond descriptions. For example, one idea (in developmental progress) is to invoke a dynamic-static disulphide bridge depiction of the molecular system within a RE framework. In this scheme every replica but one is permitted to interchange between canonical and non-canonical connectivities using the dynamic bond scheme. The remaining replica is then fixed with the correct, static, canonical pairing. In this way, replicas at high temperatures (and with dynamic disulphide bridge

arrangements) are permitted random walks to a low energy, canonically connected replica in which restricted local sampling can occur. A modified RE procedure of this nature requires a hybrid Hamiltonian description of the system in which exchange events are evaluated with a single, wholly non-bonded system Hamiltonian, whilst production MD is generated with Hamiltonians that corresponded to the respective canonical and non-canonical pairings*.

5.7. Standard simulation settings

All MD and minimizations are run (unless otherwise specified) with the settings in *Table 5.7.2* and *Table 5.7.3*. Computer code for implementing PE sampling is written in *cshell* or FORTRAN. The CD contains all the written code.

***Table 5.7.1**. Standard xleap commands. See Appendix for further xleap explanations.*

Aim	Example command	Comment
Generate amino acid sequence with name "seq".	seq = (NALA ARG ARG CLYS)	The sequence consists of N-terminal ALA, ARG,ARG and C-terminal LYS)
Bond Disulphide atoms within molecule "seq".	bond seq.13.SG seq.34.SG	In this case the "SG" atom of residue 13 and 34 are bonded to one another.
Save coordinate and topology files.	saveamberparm seq seq.parm seq.crd	"seq" identifies the molecule to be saved, "seq.parm" refers to the topology file (or Hamiltonian) and "seq.crd" refers to the coordinate file.
Load pdb file using existing sequence with name "seq".	seq = loadpdbusingseq seq.pdb (NALA ARG ARG CLYS)	This command is often useful when superimposing a modified sequence onto a pre-existing set of coodinates.
Save pdb file to disk	savepdb seq seq.pdb	"seq.pdb" refers to the name of the pdb file to be saved.

* AMBER 9 possesses a hybrid Hamiltonian facility that has historically been invoked to perform replica exchange in an explicit solvent environment but with an implicit solvent exchange criterion[49].

Table 5.7.2 *Standard MD settings*

Parameter	Setting	Explanation
nstlim	1000	Number of femtoseconds for MD run
igb	0 or 2	Set the implicit solvent environment (0 = gas-phase simulation) 2 = standard Born solvent simulation.
dt	0.001 or 0.002	Time step for simulation in picoseconds.
ntt	1 (3 when gb=2)	ntt=1: Employ constant temperature using weak coupling algorithm[16]. ntt=3: Employ Langevin dynamics.
temp0	Dynamic setting	Specify the target temperature for an MD run. This value is varied depending on the simulation. For SA this is varied from 800K to 0K. For standard dynamics this is set to 300k.
ntf	2	Ignore contributions from Hydrogen when forming interaction matrix.
ntc	2	Constrain bonds involving Hydrogen.

Table 5.7.3. *Standard minimization settings*

Parameter	Setting	Explanation
Maxcycl	50 000	Maximum number of allowed minimization cycles.
Drms	0.005	The convergence criteria determined by the root mean square force (kcal/mol)
Ntpr	1000	Sample energy information every 1000 steps.
Ntmin	1	Perform steepest descend minimization for n cycles and then switch to conjugate gradient descent.
Ncycl	100	Perform steepest descent minimization for 100 cycles.
Igb	0 or 1	Default minimization in gas-phase or Born solvent environment.

References

(1) Wlodawer, A.; Nachman, J.; Gilliland, G. L.; Gallagher, W.; Woodward, C. *Journal of Molecular Biology* **1987**, *198*, 469-480.

(2) Dadlez, M.; Kim, P. S. *Nature Structural Biology* **1995**, *2*, 674-679.

(3) Darby, N. J.; Vanmierlo, C. P. M.; Creighton, T. E. *Febs Letters* **1991**, *279*, 61-64.

(4) Weissman, J. S.; Kim, P. S. *Nature* **1993**, *365*, 185-188.

(5) Weissman, J. S.; Kim, P. S. *Nature Structural Biology* **1995**, *2*, 1123-1130.

(6) Creighton, T. E.; Zapun, A.; Darby, N. J. *Trends in Biotechnology* **1995**, *13*, 18-23.

(7) Ferrer, M.; Barany, G.; Woodward, C. *Nature Structural Biology* **1995**, *2*, 211-217.

(8) Creighton, T. E. *Biological Chemistry* **1997**, *378*, 731-744.

(9) Creighton, T. E. *Bioessays* **1992**, *14*, 195-200.

(10) Darby, N. J.; Morin, P. E.; Talbo, G.; Creighton, T. E. *Journal of Molecular Biology* **1995**, *249*, 463-477.

(11) Zdanowski, K.; Dadlez, M. *Journal of Molecular Biology* **1999**, *287*, 433-445.

(12) Okamoto, Y. *International Journal of Modern Physics C* **1999**, *10*, 1571-1582.

(13) Goldenberg, D. P.; Zhang, J. X. *Proteins-Structure Function and Genetics* **1993**, *15*, 322-329.

(14) Darby, N. J.; Creighton, T. E. *Journal of Molecular Biology* **1993**, *232*, 873-896.

(15) Darby, N. J.; Vanmierlo, C. P. M.; Scott, G. H. E.; Neuhaus, D.; Creighton, T. E. *Journal of Molecular Biology* **1992**, *224*, 905-911.

(16) Vanmierlo, C. P. M.; Darby, N. J.; Keeler, J.; Neuhaus, D.; Creighton, T. E. *Journal of Molecular Biology* **1993**, *229*, 1125-1146.

(17) Tam, J. P.; Wu, C. R.; Liu, W.; Zhang, J. W. *Journal of the American Chemical Society* **1991**, *113*, 6657-6662.

(18) Wu, Z. B.; Hoover, D. M.; Yang, D.; Boulegue, C.; Santamaria, F.; Oppenheim, J. J.; Lubkowski, J.; Lu, W. Y. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 8880-8885.

(19) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863-867.

(20) Daggett, V.; Fersht, A. R. *Trends in Biochemical Sciences* **2003**, *28*, 18-25.

(21) Daggett, V.; Fersht, A. *Nature Reviews Molecular Cell Biology* **2003**, *4*, 497-502.

(22) Kazmirski, S. L.; Wong, K. B.; Freund, S. M. V.; Tan, Y. J.; Fersht, A. R.; Daggett, V. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 4349-4354.

(23) Gianni, S.; Guydosh, N. R.; Khan, F.; Caldas, T. D.; Mayor, U.; White, G. W. N.; DeMarco, M. L.; Daggett, V.; Fersht, A. R. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 13286-13291.

(24) Kazmirski, S. L.; Daggett, V. *Journal of Molecular Biology* **1998**, *277*, 487-506.

(25) Schiffer, C. A.; vanGunsteren, W. F. *Proteins-Structure Function and Genetics* **1996**, *26*, 66-71.

(26) Hao, M. H.; Pincus, M. R.; Rackovsky, S.; Scheraga, H. A. *Biochemistry* **1993**, *32*, 9614-9631.

(27) Agostini, F. P.; Soares-Pinto, D. D. O.; Moret, M. A.; Osthoff, C.; Pascutti, P. G. *Journal of Computational Chemistry* **2006**, *27*, 1142-1155.

(28) Hansmann, U. H. E.; Okamoto, Y. *Physica A* **1994**, *212*, 415-437.

(29) Collura, V.; Higo, J.; Garnier, J. *Protein Science* **1993**, *2*, 1502-1510.

(30) Carlacci, L.; Englander, S. W. *Abstracts of Papers of the American Chemical Society* **1996**, *211*, 66-Comp.

(31) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671-680.

(32) Snow, M. E. *Journal of Computational Chemistry* **1992**, *13*, 579-584.

(33) Burgess, E. M. *Abstracts of Papers of the American Chemical Society* **1988**, *195*, 41-Csec.

(34) Nolting, B. *Journal of Theoretical Biology* **1999**, *197*, 113-121.

(35) Pan, Y. P.; Daggett, V. *Biochemistry* **2001**, *40*, 2723-2731.

(36) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10428-10435.

(37) Day, R.; Daggett, V. *Protein Science* **2005**, *14*, 1242-1252.

(38) Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 5179-5184.

(39) Jackson, S. E.; Elmasry, N.; Fersht, A. R. *Biochemistry* **1993**, *32*, 11270-11278.

(40) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *Journal of Molecular Biology* **1995**, *254*, 260-288.

(41) Gay, G. D.; Ruizsanz, J.; Davis, B.; Fersht, A. R. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 10943-10946.

(42) Clarke, J.; Fersht, A. R. *Biochemistry* **1993**, *32*, 4322-4329.

(43) Dalby, P. A.; Clarke, J.; Johnson, C. M.; Fersht, A. R. *Journal of Molecular Biology* **1998**, *276*, 647-656.

(44) Khan, F.; Chuang, J. I.; Gianni, S.; Fersht, A. R. *Journal of Molecular Biology* **2003**, *333*, 169-186.

(45) Li, L.; Shakhnovich, E. I. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 13014-13018.

(46) Yoda, T.; Sugita, Y.; Okamoto, Y. *Proteins-Structure Function and Bioinformatics* **2007**, *66*, 846-859.

(47) Volk, M. *European Journal of Organic Chemistry* **2001**, 2605-2621.

(48) Yon, J. M. *Cellular and Molecular Life Sciences* **1997**, *53*, 557-567.

(49) Mu, Y. G.; Yang, Y.; Xu, W. X. *Journal of Chemical Physics* **2007**, *127*, 84119.

Chapter Six

Residue Development

6.1 Introduction

A key assumption of this work is that defensin antimicrobial activity depends on, and follows preliminary binding to key components of the bacterial cell surface. Molecular models are therefore required for both the defensin and the presumed binding partner (lipid A). The AMBER[1] residue library contains standard amino acids (for describing proteins) but no appropriate organic molecules for modelling lipid A. This Chapter discusses how the required non-standard residues are created and Chapter Seven discusses how these same residues interact with defensins.

6.1.1. Developmental background and scope

A central challenge of molecular simulation is to simulate accurately the properties and conformational behaviour of as wide a range of molecules as possible. Historically, force-field software was developed and parameterised to deal with proteins and nucleic acids[1,2], and as such, the repertoire of organic molecules available within the standard force-field libraries is comparatively limited. Increasingly, resources are being directed at developing a "generalised" set of parameters that can be applied with reproducible results to a wide range of organic molecules[3-6]. This research invokes the widely tested generalised AMBER forcefield[3] (or "gaff" as it is commonly referred to) for which abundant agreement has been observed with experimentally derived geometries, hydrogen-bond energies, and relative conformational-energies[3]. The *gaff* parameter set in conjunction with an appropriate charge derivation method enjoys wide use in research fields that span everything from predicting melting and liquid

properties of simple organic molecules[7] to molecular docking studies[8-10] to predicting dispersive X-ray diffraction spectra[11].

6.2. Method

The energy of a molecular system can be approximated with a force-field equation that consists of bonded and non-bonded interaction terms (Chapter Three). The challenge of developing a non-standard residue therefore rests with calculating (or approximating) the following force-field parameters:

1. Partial charges for each atom centre.
2. Force-constants that describe bond-angle bending and dihedral torsional variation.
3. Van der Waals radii.

All residue development is performed with the AMBER antechamber component (and its related subprograms). All residue equilibration is performed in explicit SPC[12] solvent using standard MD settings (*Section 6.5*).

6.2.1 Obtaining the starting geometry and assigning global charge

The initial coordinate geometry of LPS is taken from the *E. coli* K-12 AW740 strain of bacteria for which the crystal structure is solved in complex with the porin protein Fhua[13,14]. Missing hydrogen atoms are added to lipid A using the *xleap* facility of AMBER.

Most literature published structures of lipid A assume that each phosphate group possesses a unit negative charge[15-20]. This is a reasonable assumption because generic phosphate species with a -1

charge state display a pK_a of 6.7, whereas the neutral and -2 charge state exhibit pK_a's of 2 and 12 respectively[21]. In reality, the assumption of a -1 charge state may be incorrect because the local chemical environment (that is not *a priori* known) undoubtedly plays a role in determining the local *pH*. However, despite this potential pitfall, and in the absence of the desired experimental evidence for a definitive charge state, the global charge assignment in this work follows the assumptions of previous work[15-20].

6.2.2. Modelling strategy

The defensin-Lipid A problem is tackled in a "bottom-up" approach in which a simplistic model is gradually refined towards a more complex one. In the first level of this model a major simplifying assumption is that the phosphorylated glucosamine region of the molecule is responsible for most of the vital docking interactions. This conjecture reflects the experimental observation that cationic modifications to this moiety correlate with AMP resistance[18,22-24] and that rationally designed peptides based on properties common to many AMPs bind this same region[25]. With this rationality in mind, and for purposes of computational tractability, the KDO and lipid region of the molecule are cleaved from the central glucosamine core (*Figures 6.2.2.1*).

The charge and parameter fitting procedure, as well as the *ab initio* and MD geometry optimisation are summarised in *Figure 6.2.2.2*.

Figure 6.2.2.1. *A schematic representation of Lipid A bound to KDO. The fatty acid chains and KDO residues are cleaved at regions that intersect with the red box. Only the central tri-phosphorylated sacharide region remains. "Floating" bonds generated by the cleaving process are methylated prior to structure optimisation and charge fitting.*

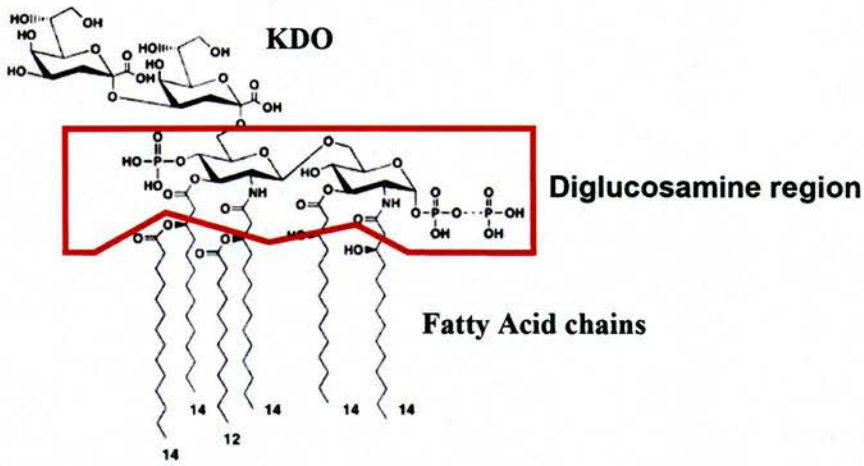
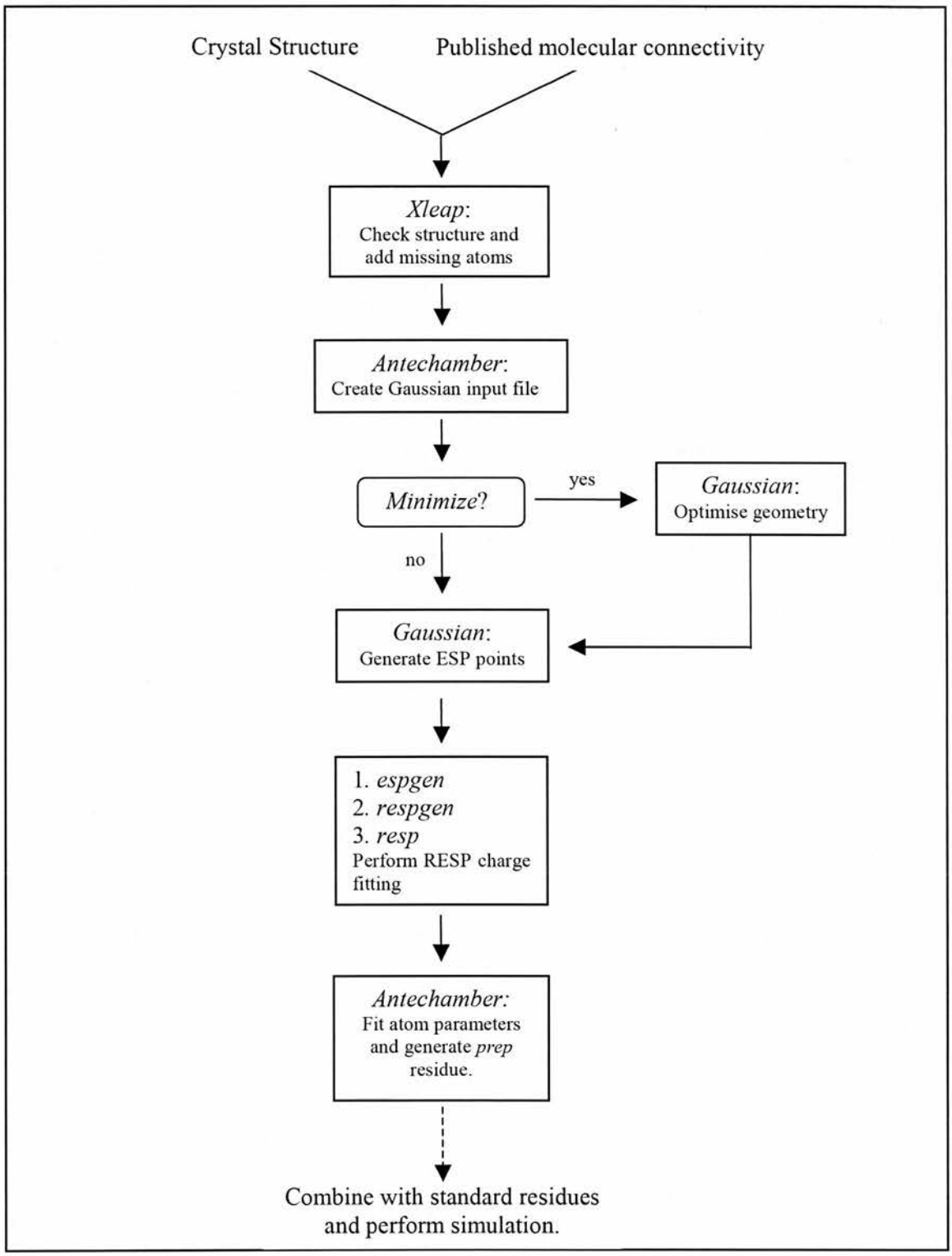


Figure 6.2.2.2. *Flow diagram of residue development strategy showing both the software used and the information flow.*



6.2.3. Optimisation

Hydrogen atoms are added to the crystal structure of the cleaved Lipid A using the *xleap* module of AMBER. The resulting structure is then optimised with *Gaussian 03*[26] using the HF/6-31G* level of theory, tight convergence, and the "iop(6/33=2) iop(6/42=6)" keyword for establishing atomic centres.

6.2.4. Charge fitting

The standard AMBER *resp*[27] procedure is used to fit partial charges to the quantum mechanically optimised lipid A structure (in which the electrostatic potential has been calculated at the HF/6-31G* level of theory). It is worth noting that more representative charges can be obtained by performing a Boltzmann weighted average of an ensemble of different MD sampled trajectories[28]. This approach, whilst offering increases in accuracy, is computationally extremely expensive and not considered in this simplified model of initial interactions, though it is a common practice for many residues that are rigorously developed[28,29].

6.2.5. Parameter fitting and MD equilibration

Parameters are fitted with the AMBER generalised force field parameter set (*gaff*)[3] using the *antechamber* module of AMBER. The bond force constants fitted to the "p5-oh-ho" connectivity are artificially strengthened by a factor of 10 in order to prevent dynamic instability that is symptomatic of AMBER force-field (over)simplifications*. The parameterised, charge fitted and geometry optimised construct is then subjected to MD equilibration for 1 nanosecond with SPC explicit solvation (The process of MD equilibration is discussed in *Section 6.5* of this Chapter).

* According to Mike Crowley (a principle architect of AMBER) the van der Waals interaction term is not computed for directly neighbouring atoms because the energy barrier to angle bending between such neighbours is assumed to be high enough to prevent atomic overlap. However, this assumption does not hold for the "p5-oh-ho" connectivity within the highly polar system investigated here. Hence, without artificially strengthening the bond, extreme (and spurious) energies arise as a result of van der Waals overlap.

Finally, the resulting geometry is used to generate a standard AMBER residue (or "prep") file using the standard antechamber procedure.

6.2.6. Mutations and variations

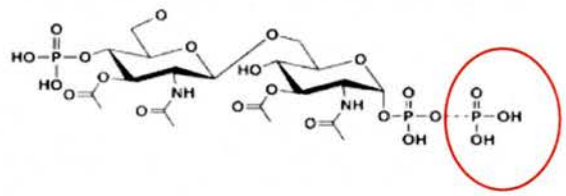
a) 4-deoxy-4-aminoarabinose (Ara4N)

Lipid A is the primary immunogenic stimulant within LPS and (unsurprisingly) a number of cationic modifications to it are associated with resistance to AMP attack. One notable mutation that is examined in this research is the addition of *Ara4N* to the diglucosamine phosphate residues[22,23]. The *Ara4N* residue is accordingly created *de novo* within *xleap* from a literature published connectivity[22]. The structure is optimised with Gaussian and subsequently attached covalently (using *xleap*) to the diglucosamine residue in place of the single phosphate group.

b) Bis-phosphorylated diglucosamine (BPDG)

There are many serum (and species) dependent lipid A variants. Some of these are confined to the lipid region[24,30,31] whilst others are confined to the glucosamine core[32]. The bis-phosphorylated variant (*Figure 6.2.6.1.*) is common and (like the tri-phosphorylated species) is investigated in Chapter Six. The necessary removal of a phosphate group from the tri-phosphorylated species is performed using *xleap* and the standard procedure followed for optimisation, charge derivation and parameter fitting (*Section 6.2.2 -5*).

Figure 6.2.6.1. *BPDG does not possess the terminal phosphate group (marked in red).*



6.3. Results and discussion

6.3.1. Equilibration and production MD

Stable MD trajectories are obtained for each of the model structures within explicit solvent[**]. An example energy and temperature output for *Ara4N* (*Figure 6.3.1.1 A* and *B*) and the other molecules follow similar behaviour (provided on CD). The standard equilibration and production procedure and parameters are given in *Section 6.5*.

6.3.2. Charge fitting and geometry optimisation

a) Tri-phosphorylated-diglucosamine (TPDG)

The MD equilibrated and minimized structure is shown in *Figure 6.3.2.1*. The regions of the molecule that would (ordinarily) attach to fatty acid chains are organised in a single plane similar to that found in the crystal structure. Partial charges for the atomic centres are shown in *Table 6.3.2.2*.

[**] Stable trajectories are not obtained for gas-phase MD, even despite artificially strengthening the *p5-oh-ho* bond. The reason for this is not clear though it is likely to result from the same AMBER (over)simplification. Many residues with "normal" levels of charge polarisation give rise to predictably stable dynamics.

b) Bis-phosphorylated diglucosamine (BPDG)

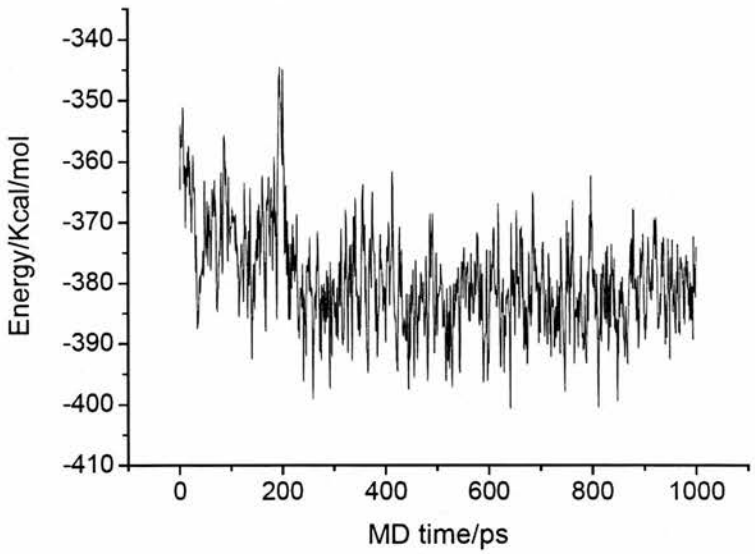
The BPDG species adopts a very similar structure to the tri-phosphorylated variant (*Figure 6.3.2.3*) and the partial charge assignments do not differ substantially for atoms outside the immediate vicinity of the cleaved phosphate residue.

c) Aminoarabinose-modified BPDG (Ara4N)

Ara4N, following MD equilibration and minimization gives rise to a geometry in which the aminoarabinose moiety sits parallel to the plane of (what would otherwise be) the extending fatty acid chains (*Figure 6.3.2.5.*). The partial charges for each of the atomic centres are shown in *Table 6.3.2.6.*

Figure 6.3.1.1. *1 nanosecond of production MD Ara4N. Production runs for other constructs follow a similar pattern and are not shown. A. Energy vs MD time and B. Temperature vs MD time.*

A



B

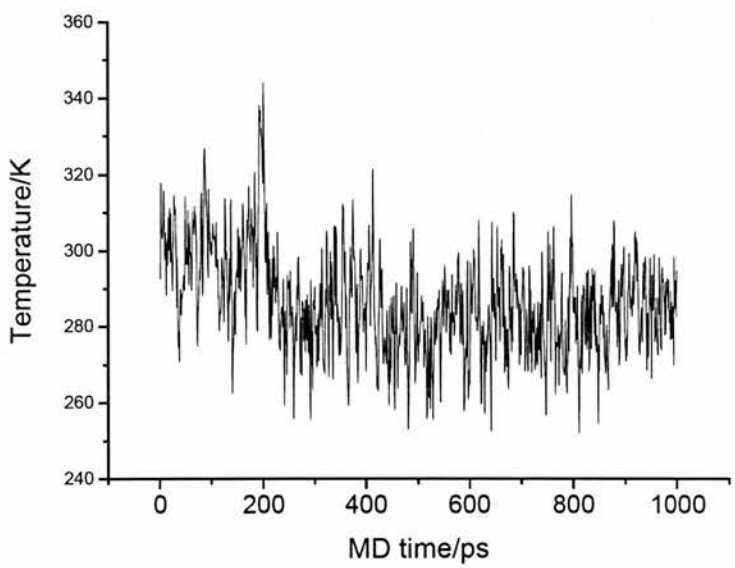


Figure 6.3.2.1. *The optimised, parameterised and charge fitted structure of TPDG. Carbon atoms are shown in turquoise, hydrogen in white, oxygen in red, phosphate in brown and nitrogen in navy blue.*

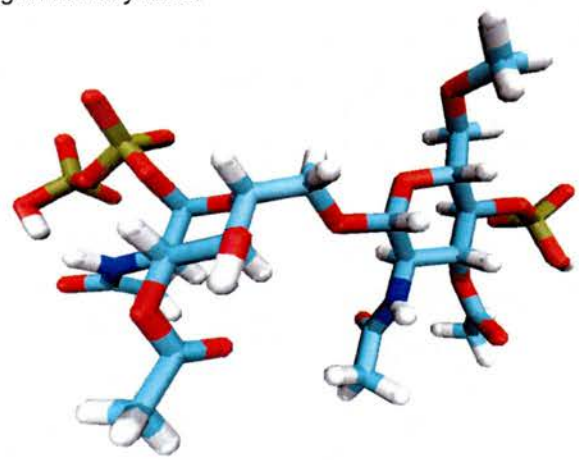


Table 6.3.2.2. *Partial charges for each of the atomic centres of TPDG. The library files for this structure are on the supplied CD.*

Atom Name	Charge	Atom Name	Charge	Atom Name	Charge	Atom Name	Charge
C13	-0.076760	**C20**	0.940720	**C5**	0.353330	**H26**	0.118830
H18	0.064640	**C21**	-0.487850	**O3**	-0.567540	**H27**	0.118830
H19	0.064640	**H31**	0.128210	**H5**	0.025520	**O12**	-0.689500
H20	0.064640	**H32**	0.128210	**C4**	-0.018210	**H2**	0.157370
O7	-0.318570	**H33**	0.128210	**O2**	-0.710260	**C1**	0.580280
C12	-0.013590	**O15**	-0.657360	**H9**	0.442330	**H1**	0.088280
H15	0.123450	**C8**	0.044970	**H4**	0.198970	**O1**	-0.537400
H17	0.123450	**N2**	-0.798980	**C3**	0.384320	**P2**	1.320280
C11	0.108890	**H16**	0.386560	**H3**	0.120440	**O17**	-0.825980
O6	-0.348900	**C14**	0.913680	**O14**	-0.539110	**O18**	-0.825980
H14	0.072470	**C15**	-0.503390	**C18**	1.017340	**O19**	-0.605700
C10	-0.116300	**H21**	0.125320	**C19**	-0.434040	**P3**	1.403300
O5	-0.433270	**H22**	0.125320	**H28**	0.107910	**O21**	-0.787990
P1	1.305310	**H23**	0.125320	**H29**	0.107910	**H34**	0.456390
O9	-0.782600	**O8**	-0.668890	**H30**	0.107910	**O22**	-0.839710
H24	0.438610	**H11**	0.174200	**O13**	-0.728490	**O20**	-0.839710
O10	-0.811080	**C7**	0.214790	**C2**	-0.434970	**H25**	0.118830
O11	-0.811080	**H10**	0.120680	**N1**	-0.423110	**H7**	0.044610
H13	0.143630	**O4**	-0.347260	**H8**	0.314200	**O16**	0.109170
C9	0.277120	**C6**	0.055190	**C16**	0.812960		
H12	0.096380	**H6**	0.044610	**C17**	-0.461670		

Figure 6.3.2.3. A liquorice depiction of *BPDG. Carbon atoms are shown in turqoiuse, hydrogen in white, oxygen in red, phosphate in brown and nitrogen in navy blue.*

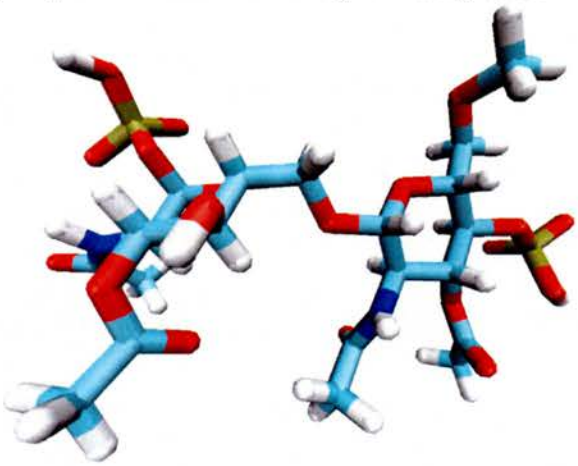


Table 6.3.2.4. *Partial charges for each of the atomic centre of BPDG. The library files for this structure are present on the supplied CD.*

Atom Name	Charge	Atom Name	Charge	Atom Name	Charge	Atom Name	Charge
C10	-0.368800	**O8**	-0.670000	**O10**	-0.311740	**C19**	0.784920
H14	0.093160	**C4**	-0.028620	**C12**	-0.073220	**C20**	-0.417070
H15	0.093160	**O3**	-0.558910	**H19**	0.093090	**H29**	0.115180
H16	0.093160	**P1**	1.330950	**H20**	0.093090	**H30**	0.115180
C9	0.780680	**O4**	-0.712390	**C13**	0.446240	**H31**	0.115180
O9	-0.627280	**H7**	0.413440	**H21**	-0.004230	**O15**	-0.667970
N1	-0.602200	**O5**	-0.825440	**C14**	0.057350	**H32**	0.173500
H13	0.356340	**O6**	-0.825440	**O12**	-0.711070	**C15**	0.429560
C8	-0.191120	**H8**	0.171650	**H22**	0.442770	**H24**	0.070980
H17	0.197310	**C3**	0.167090	**H23**	0.075940	**O13**	-0.584420
C5	0.283310	**C2**	0.019890	**O11**	-0.595660	**C16**	0.948340
H9	0.083870	**O1**	-0.425810	**C21**	0.598610	**O14**	-0.663110
O7	-0.446080	**C1**	0.036890	**H33**	0.082930	**C17**	-0.407220
C6	-0.446080	**H1**	0.050060	**O16**	-0.532290	**H25**	0.109170
C7	-0.537960	**H2**	0.050060	**P2**	1.279830	**H26**	0.109170
H10	0.143240	**H3**	0.050060	**O17**	-0.702220	**H27**	0.109170
H11	0.143240	**H4**	0.115100	**H34**	0.411820	**O2**	-0.428720
H12	0.143240	**H5**	0.115100	**O18**	-0.795460	**H18**	0.029310
H6	0.115100	**O2**	-0.428720	**C11**	0.399920	**H28**	0.297850
O19	-0.795460	**C18**	-0.389660	**N2**	-0.454160		

Figure 6.3.2.5. *The optimised, parameterised and charge fitted structure of Ara4N. Carbon atoms are shown in turquoise, hydrogen in white, oxygen in red, phosphate in brown and nitrogen in navy blue.*

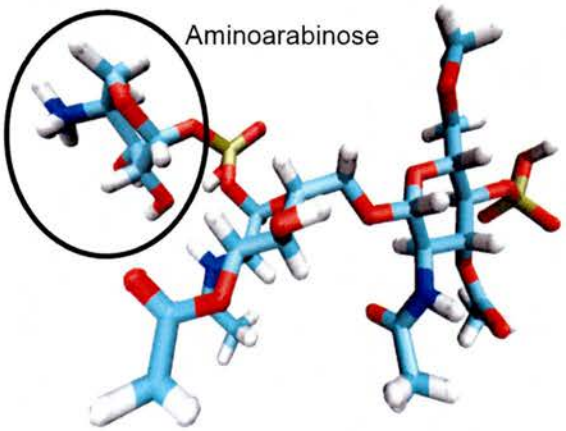


Table 6.3.2.6. *Partial charges for each of the atomic centre of Aminoarbinose modified BPDG. The library files for this structure are present on the supplied CD.*

Atom Name	Charge	Atom Name	Charge	Atom Name	Charge	Atom Name	Charge
C1	-0.48918	**O5**	-0.66466	**C13**	0.19084	**H30**	0.1291
H1	0.11905	**H10**	0.39925	**H21**	0.05255	**H31**	0.1291
H2	0.11905	**O6**	-0.84578	**O15**	-0.36864	**O14**	-0.7211
H3	0.11905	**07**	-0.84578	**C14**	0.17044	**H32**	0.12726
C2	0.85589	**H11**	0.11525	**O11**	-0.64743	**C21**	0.3291
O1	-0.67302	**C8**	0.02489	**H22**	0.39139	**H33**	0.0963
N1	-0.67360	**C9**	0.16742	**H23**	0.17408	**O16**	-0.38521
H4	0.35327	**O8**	-0.41867	**C15**	0.08649	**P2**	1.09674
C3	0.0251	**C10**	0.02777	**H24**	0.08703	**O17**	-0.77418
H5	0.18241	**H12**	0.05225	**O12**	-0.5008	**O18**	-0.77418
C4	0.07618	**H13**	0.05225	**C16**	0.93417	**O19**	-0.36496
H6	0.11916	**H14**	0.05225	**C17**	-0.57756	**C22**	0.34138
O2	-0.40563	**H15**	0.06832	**H25**	0.1594	**H34**	0.1545
C5	0.92903	**H16**	0.06832	**H26**	0.1594	**O20**	-0.46606
C6	-0.54324	**H17**	0.09303	**H27**	0.1594	**C23**	-0.04247
H7	0.1582	**O9**	-0.40918	**O13**	-0.64669	**H35**	0.13982
H8	0.1582	**C11**	0.35267	**C18**	0.00186	**H36**	0.13982
H9	0.1582	**H18**	0.07025	**N2**	-0.628	**C24**	0.34557
O3	-0.6618	**O10**	-0.4144	**H28**	0.41807	**O21**	-0.61472
C7	0.17285	**C12**	0.00308	**C19**	0.84412	**H37**	0.44816
O4	-0.56621	**H19**	0.08412	**C20**	-0.44087	**N3**	-0.64352
P1	1.31513	**H20**	0.08412	**H29**	0.1291	**H38**	0.41971
H39	0.41971	**H40**	0.41971	**C25**	0.25481	**O22**	-0.719670
H41	0.49001	**H42**	0.09725	**C26**	-0.078920		
H44	0.48300	**O23**	-0.70440	**H43**	0.115110		

6.4. Conclusion

The non-standard residues discussed in this Chapter represent a first attempt at modelling the underlying interactions of the defensin-lipid A problem. Three potential shortcomings of these (substantial) simplifications are:

1. Atomic charges in reality vary in response to changes in geometry. There is no attempt with this model to get a more representative charge via a Boltzmann weighted average of charge states obtained from many geometries.
2. The (cleaved) lipid portion and core region might, in reality, affect the structure and charge properties of the glucosamine region.
3. Defensins in reality may interact with other components of lipid A, and not merely the presumed region. (There is some evidence that modifications to the lipid A fatty acid region also direct defensin resistance[33]).

Chapter Seven explores the binding between *TPDG*, *BPDG*, *Ara4N* and defensins HBD1-3. An in-depth discussion is provided in this Chapter of the benefits and drawbacks of a simplified lipid A representation. Conclusions are also drawn about the nature and mode (if any) of binding.

6.5. Standard Settings

Equilibration in this Chapter refers to four separate equilibration steps within an explicit SPC solvent milieu (*Figure 6.5.1*). The corresponding parameter settings are also provided (*Table 6.5.2*). The four step equilibration process allows atomic overlap to be relieved (between solvent and solute), and stable values to be obtained for density, volume and temperature prior to production MD.

Figure 6.5.1. *A schematic representation of MD equilibration and production.*

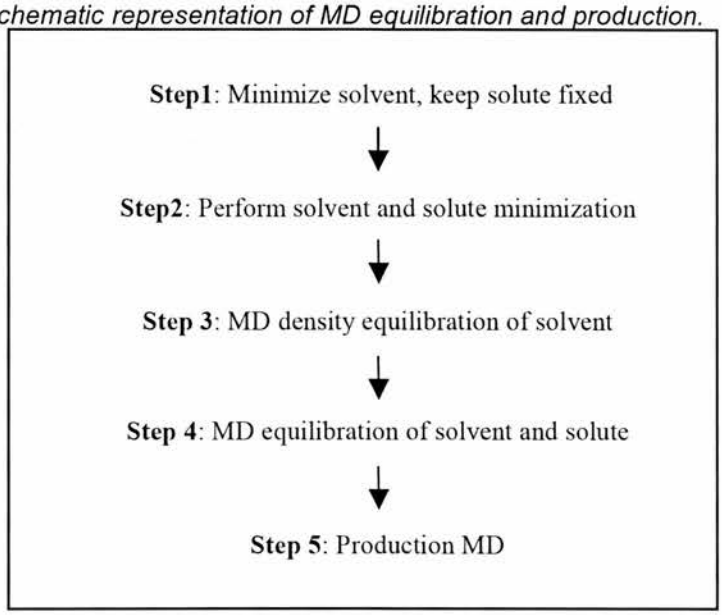


Table 6.5.2. *Parameters for steps 1 – 4 (figure 6.5.1.) using SPC explicit solvent.*

	Parameter	**Value**	**Comment**
Step 1			**solvent minimization.**
	ncyc	500	Number of steepest descent iterations.
	maxcyc	1000	Maximum number of permitted iterations.
	drms	0.005	Root mean square force threshold for stopping minimization.
	ntb	1	Maintain periodic boundary conditions.
	ntr	1	Apply harmonic restraint to solvent.
	cut	12	Non-bonded cut-off.
Step 2			**Solvent + solute minimization.**
	ncyc	1000	
	maxcyc	10000	
	ntr	0	Turn solute restraint off.
	drms	0.005	
	ntb	1	
	ntr	1	
	cut	12	
Step 3			**Density equilibration of water.**
	cut	12	
	ntr	1	Restrain solute.
	ntc	2	Ignore contributions from Hydrogen.
	ntf	2	Constrain bonds involving Hydrogen.
	tempi	0	Initial MD temperature /K.
	temp0	300	Final temperature /K.

	ntt	3	Turn on Langevin dynamics.
	gamma_ln	1	Collision frequency for Langevin equation.
	nstlim	10 000	Number of MD steps to perform.
	dt	0.02	Use a 2 femtosecond time-step.
step 4			Full equilibration of entire system.
	ntt	3	
	gamma_ln	1	
	nstlim	10 000	Number of MD steps to perform.
	dt	0.02	Use a 2 femtosecond time-step.
	cut	12	
	ntr	1	Restrain solute.
	ntc	2	Ignore contributions from Hydrogen.
	ntf	2	Constrain bonds involving Hydrogen.
	tempi	0	Initial MD temperature /K.
	temp0	300	Final temperature /K.
	ntb	1	Use periodic boundary conditions.

References

(1) Pearlman, D. A., Case, D. A., Caldwell, J. W., Cheatham, T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., Kollman, P. A. **2002**.

(2) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Computer Physics Communications* **1995**, *91*, 1-41.

(3) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157-1174.

(4) Halgren, T. A. *Journal of Computational Chemistry* **1996**, *17*, 490-519.

(5) Lii, J. H.; Allinger, N. L. *Journal of Computational Chemistry* **1991**, *12*, 186-199.

(6) Clark, M.; Cramer, R. D.; Vanopdenbosch, N. *Journal of Computational Chemistry* **1989**, *10*, 982-1012.

(7) Velardez, G. F.; Alavi, S.; Thompson, D. L. *Journal of Chemical Physics* **2003**, *119*, 6698-6708.

(8) Zhang, B.; Tan, V. B. C.; Lim, K. M.; Tay, T. E. *Journal of Computer-Aided Molecular Design* **2006**, *20*, 395-404.

(9) Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A. *Journal of the American Chemical Society* **2001**, *123*, 5221-5230.

(10) Wang, J. Y.; Deng, Y. Q.; Roux, B. *Biophysical Journal* **2006**, *91*, 2798-2814.

(11) Gontrani, L.; Rarnondo, F.; Caminiti, R. *Chemical Physics Letters* **2006**, *422*, 256-261.

(12) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *Journal of Physical Chemistry* **1987**, *91*, 6269-6271.

(13) Ferguson, A. D.; Welte, W.; Hofmann, E.; Lindner, B.; Holst, O.; Coulton, J. W.; Diederichs, K. *Structure* **2000**, *8*, 585-592.

(14) Ferguson, A. D.; Hofmann, E.; Coulton, J. W.; Diederichs, K.; Welte, W. *Science* **1998**, *282*, 2215-2220.

(15) Alexander, C.; Rietschel, E. T. *Journal of Endotoxin Research* **2001**, *7*, 167-202.

(16) Janusch, H.; Brecker, L.; Lindner, B.; Alexander, C.; Gronow, S.; Heine, H.; Ulmer, A. J.; Rietschel, E. T.; Zahringer, U. *Journal of Endotoxin Research* **2002**, *8*, 343-356.

(17) Frecer, V.; Ho, B.; Ding, J. L. *Biochimica Et Biophysica Acta-Biomembranes* **2000**, *1466*, 87-104.

(18) Peschel, A. *Trends in Microbiology* **2002**, *10*, 179-186.

(19) Kastowsky, M.; Sabisch, A.; Gutberlet, T.; Bradaczek, H. *European Journal of Biochemistry* **1991**, *197*, 707-716.

(20) Obst, S.; Kastowsky, M.; Bradaczek, H. *Biophysical Journal* **1997**, *72*, 1031-1046.

(21) Krauss, G. **2003**.

(22) Nummila, K.; Kilpelainen, I.; Zahringer, U.; Vaara, M.; Helander, I. M. *Molecular Microbiology* **1995**, *16*, 271-278.

(23) Guo, L.; Lim, K. B.; Gunn, J. S.; Bainbridge, B.; Darveau, R. P.; Hackett, M.; Miller, S. I. *Science* **1997**, *276*, 250-253.

(24) Guo, L.; Lim, K. B.; Poduje, C. M.; Daniel, M.; Gunn, J. S.; Hackett, M.; Miller, S. I. *Cell* **1998**, *95*, 189-198.

(25) Frecer, V.; Ho, B.; Ding, J. L. *Antimicrobial Agents and Chemotherapy* **2004**, *48*, 3349-3357.

(26) Gaussian 03, R. C., M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople **2004**.

(27) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *Journal of Physical Chemistry* **1993**, *97*, 10269-10280.

(28) Woods, R. J.; Chappelle, R. *Journal of Molecular Structure-Theochem* **2000**, *527*, 149-156.

(29) Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. *Journal of Computational Chemistry* **2001**, *22*, 1125-1137.

(30) Bishop, R. E.; Gibbons, H. S.; Guina, T.; Trent, M. S.; Miller, S. I.; Raetz, C. R. H. *Embo Journal* **2000**, *19*, 5071-5080.

(31) Belden, W. J.; Miller, S. I. *Infection and Immunity* **1994**, *62*, 5095-5101.

(32) Zhou, Z. M.; Ribeiro, A. A.; Lin, S. H.; Cotter, R. J.; Miller, S. I.; Raetz, C. R. H. *Journal of Biological Chemistry* **2001**, *276*, 43111-43121.

(33) Peschel, A.; Jack, R. W.; Otto, M.; Collins, L. V.; Staubitz, P.; Nicholson, G.; Kalbacher, H.; Nieuwenhuizen, W. F.; Jung, G.; Tarkowski, A.; van Kessel, K. P. M.; van Strijp, J. A. G. *Journal of Experimental Medicine* **2001**, *193*, 1067-1076.

Chapter Seven

Molecular Docking

7.1. Introduction

This Chapter investigates the interaction of the sacharide-based residues (discussed in Chapter Six) with HBD1-3. As stated in Chapter Two, mutant isolates of bacteria that possess cationic modifications to their Lipid A components are resistant to defensin attack[1-6]. This Chapter investigates the premise that β-defensins target the sacharide component of lipid A during the first (presumed) step of antimicrobial attack.

7.1.1. Previous simulation work with antimicrobial peptides

Frecer *et. al*[7] provide the only notable simulation of antimicrobial peptides interacting with surface markers of bacterial membranes. In their work in 2004 they attempt (with some success) a rational design of a novel antimicrobial agent in which key properties of common AMPs (magainins, ceprocins defensins and others)[7] are mimicked. In this approach a flexible, induced-fit docking procedure is used to generate 100 bound configurations with lipid A, the resulting structures of which are ranked on the basis of their relative interaction energies. Interestingly, the best ranked structures are found to attach to the sacharide region of the molecule[7]. Experimental synthesis of these compounds reveals a killing propensity for *E.coli, Klebsiella pneumoniae, and P. aeruginosa* that exceeds conventional antibiotics. For example, the minimum inhibitory concentration of tetracycline for these strains is in the region of 3μg/ml[8], compared to that of the most potent rational product, which possesses an MIC of 10ng/ml[7]. These tentative findings suggest a key role for lipid A in the antimicrobial mechanism of a variety of AMPs.

Unfortunately, a conspicuous absence of native NMR and crystallographic data for most cell surface markers has resulted in rather few AMP-receptor simulations. Consequently, the vast majority of membrane-AMP simulations are directed at elucidating the mechanism of bilayer permeation. An excellent example of this work is detailed by Berneche *et al*, who, in 1998 produced an all-atom model of melittin insertion into the lipid bilayer[9]. In this simulation melittin was found to partially permeate the bilayer and weaken the integrity of the membrane enough to allow a water channel to form. Similar work by Khandelia *et al.* in 2005 involved an SDS micelle binding with the antimicrobial peptide ovispirin[10,11]. Essex *at al.* investigate permeation of the bilayers by a number of small, organic molecules[12], as well as the dynamic behaviour of solutes within the bilayer itself[13].

7.1.2. The docking problem

The problem of ligand docking is two-fold; that of sampling the ligand-receptor potential energy space, and that of ranking the resulting complexes. There are an enormous (and growing) range of algorithms for tackling this problem[14-16], the choice of which depends on the nature of the investigation: Is the location of the active site known? Is the identity of the ligand known? How many ligand variations require screening? What are the available computational resources? In the case of the disaccharide-lipid A docking problem neither the binding site, nor the initial conformation (before binding), nor the final conformation (after binding) are known. The sampling technique for this problem thus employs a flexible, MD approach that bares some resemblance to the one implemented by Frecer *et al.*[7] and indeed, a variety of other small-molecule MD-docking studies[17-20]. Non-MD strategies often make use of Monte Carlo or genetic algorithm-generated structural perturbations (Software packages such as *AutoDock* often employ a combination of both "random move" and MD strategies[21-23]).

A full description of all the available docking methods is beyond the scope of this introduction and the interested reader is invited to consult the excellent reviews by Essex[14] and Halperin[15].

7.2. Method

The docking procedure makes use of a flexible MD-based method in conjunction with a relative-free-energy scoring procedure. Absolute binding free energies are computed only for the 3 best ranked structures of each of HBD1-3. These best ranked structures are then subjected to *Ara4N* and *BPDG* mutations (Chapter Six) with a view to elucidating some - or in a perfect world, all - of the binding complexities.

7.2.1. Docking strategy

Preliminary binding configurations are generated by manually placing the lipid A moiety in various (rationally intuitive) regions of the molecular surface of the peptide using *VMD*[24]. Explicit solvation of each of these docked structures with the *SPC*[25] water model is then followed by several MD simulated annealing runs (see *Section 7.5* for parameter settings). Finally, each of these annealed structures is exposed to explicitly solvated molecular dynamics for 2 ns (*Table 7.2.1.1, Figure 7.2.1.2* and *Table 7.2.1.3*). The total number of docks for a given peptide and binding partner is thus given by $N_{Docks} \times N_{SA}$, where N_{Docks} refers to the number of initial, *VMD* generated docks, and where N_{SA} refers to the number of simulated annealing cycles.

Table 7.2.1.1. *Docking and MD procedure.*

Procedure	Software/process	Comment
Initial structures	Generated using *VMD*	Manual docking proved to be most reliable form of generating initial structures.
Simulated annealing	Sander and *cshell* code	The annealing protocol employs stochastic temperature and time control for better sampling.
Final equilibriation	*Sander* and *cshell* code	Allow further relaxation and substrate surface movement.
Free energy scoring	AMBER MM-PBSA protocol	Free energy calculations permit effective comparison of binding, more so than a simple "binding energy" paradigm.

Table 7.2.1.3. *Periodic boundary-simulation set-up. The manually docked complex is solvated within xleap and neutralized with counter ions.*

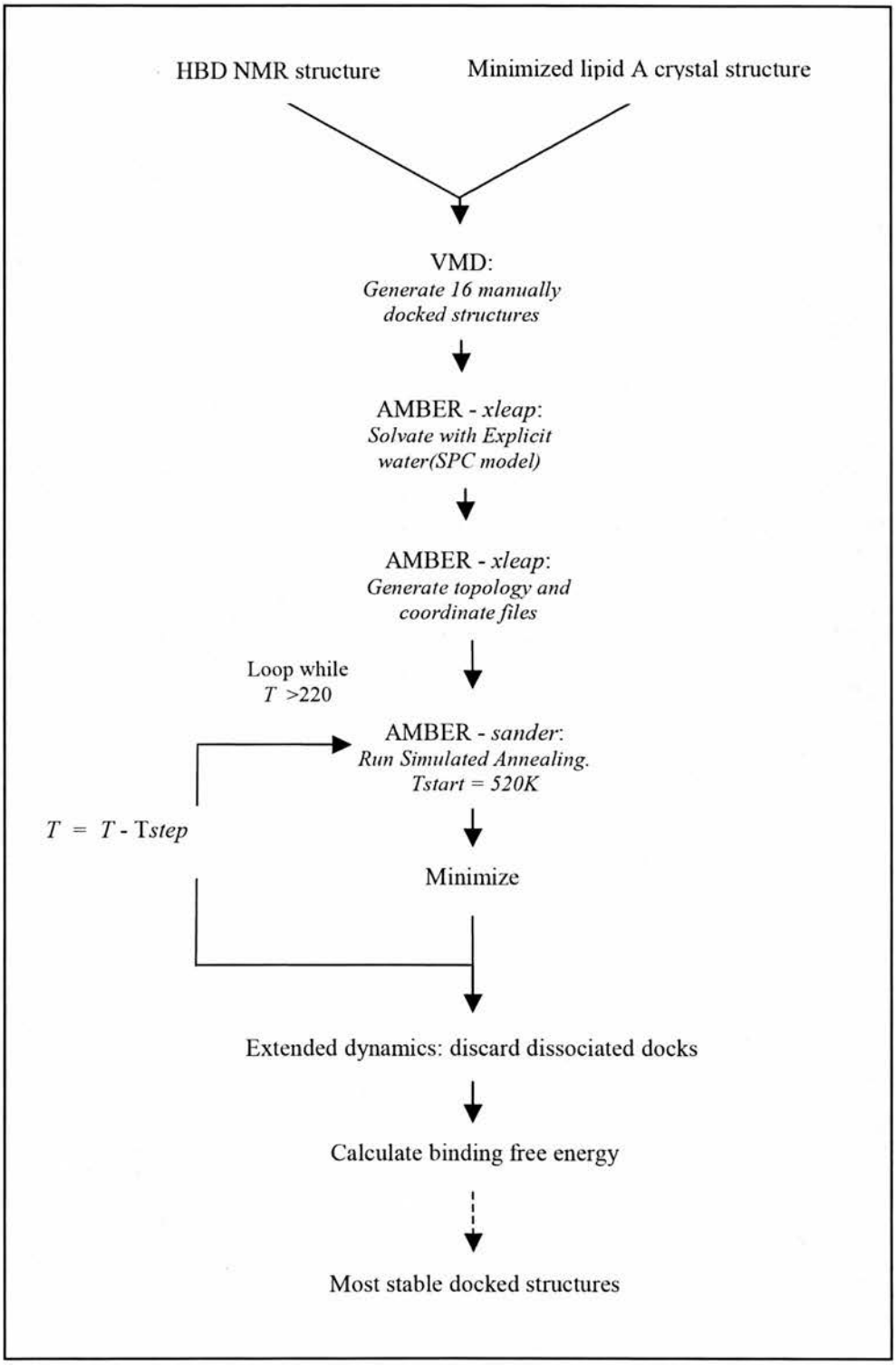
	HBD1	HBD2	HB3
Periodic box dimensions/A	*x*: 52.981 *y*: 49.027 *z*: 48.270	*x:* 46.645 *y:* 52.874 *z:* 50.124	*x:* 65.419 *y:* 62.245 *z:* 42.040
Representative water particles[2]	2740	2761	3905
Number of Cl^- counter ions	1	4	8

7.2.2. Testing for dissociation

Defensin-substrate dissociation is tested for by tracking the Euclidean distance between atoms that are buried within each of the molecular interiors of the defensin and substrate respectively. A 30 Angstrom threshold is used to distinguish between bound and unbound states.

[2] The number of water particles varies for each different orientation of ligand and defensin. The value shown is representative in general magnitude only.

Figure 7.2.1.2. *Schematic representation of docking procedure.*



7.2.3. Scoring

Structures that remain bound following the first nanosecond of extended dynamics are scored on the basis of a relative MM-PBSA[26-29] free-energy comparison (Chapter Three). With this method the free energy is expressed as a sum of individual contributions[30-32]:

$$G = E_{MM} + G_P + G_{NP} - TS_{solute} \qquad (1)$$

Where E_{MM} refers to the molecular mechanical energy, G_P refers to the polar free energy, G_{NP} to the non-polar free energy, T to the temperature of the simulation, and S to the solute entropy.

It is common practice to apply the method to a variety of trajectory snapshots so that a conformationally averaged free-energy value is obtained[30-42]. In this study the first stage of free energy scoring (for the purpose of reducing computational time) makes use of a single minimized structure. Interestingly, Kuhn *et al.* detail instances in which such single, minimized complexes give rise to comparatively more accurate free-energy values[43]. The final absolute free energies of the best ranked structures are therefore computed using both strategies for comparison sake.

The AMBER implementation of MM-PBSA used in this research estimates polar solvation free-energies with the *PBSA* procedure[26], non-polar contributions with the *Molsurf*[44] program and desolvation free energies with the *GB* component of *sander*[45-47]. The external salt concentration used for the Poisson-Boltzmann solver is set to a physiologically credible 0.15 M[48]. For each docked complex 50 trajectories are examined, each set of which is extracted from the final 500ps of MD. Details of the parameters used are given in *Table 6.2.3*.

The entropic penalty of binding is presumed to be very similar for any identical collection of docked complexes and, as such, entropic factors are ignored in the scoring model.

7.2.4. Absolute binding free-energies

In order to establish a binding hierarchy between β-defensins HBD1,2 and 3, it is necessary to consider entropic factors. Accordingly, absolute binding-free energies are computed by correcting the relative binding energy (discussed in *section 7.2.3.*) with the entropic penalty associated with docking[41]. Thus,

$$\Delta G_{Binding} = \Delta G_{relative} - T\Delta S_{solute} \qquad (2)$$

Where $\Delta G_{Binding}$ refers to the free energy of binding within a solvent environment, $\Delta G_{relative}$ refers to the relative free energy change upon binding (as set out in *equation 1*), T to temperature and ΔS_{solute} to the entropy change upon solute binding.

Entropic contributions are estimated with a normal mode analysis using the *nmode* module of AMBER as detailed in other work[30-32,40].

7.2.5. Variations to lipid A

Chapter Six discusses two mutant lipid A variants. The first, 4-deoxy-4-L-arabinose (*Ara4N*), is associated with increased resistance to AMP attack[6,49]. The second, a bis-phosphorylated species (*BPDG*), is predominant within Escherichia coli *K-12*. This variant, with one less phosphate than

the tri-phosphorylated variant (which is considered in the docking procedure), possesses both a lower net charge (-2 instead of -3) and a lower level of internal polarization. The effect of both these variants (*Ara4N* and *BPDG*) is investigated here by mutating, in *situ*, the best 3 ranked *TPDG*-defensin complexes. (The mutations are generated with the aid of a *cshell* script that superimposes the coordinates of the mutant onto the wild-type structure and redefines the AMBER atom names appropriately - see CD for code). Following the mutations, further MD equilibration is performed (with standard settings – *Section 7.5*) and a relative and absolute binding energy comparison made (as outlined in *Sections 7.2.3* and *7.2.4*).

7.3. Results and discussion

A number of assumptions are made in this docking study. These are:

1. That the lipid A region forms the first, vital docking interactions with the approaching defensin (Chapter Two).
2. That the charge distribution of a single, minimized lipid A molecule sufficiently captures the average charge distribution for many conformational trajectories (Chapter Six).
3. That the local chemical environment favours the existence of anionic phosphate moieties (Chapter Six).
4. That the most important docking interactions can be modelled with a forcefield approximation in an explicit water environment (Chapter Two and Chapter Three).
5. That the docking procedure (in conjunction with the scoring algorithm) locates a global defensin-substrate binding-minimum (this Chapter).

Where necessary these assumptions and their possible (or inherent) shortcomings are given their full consideration in the proceeding sections.

7.3.1. SA and extended MD: Is a free-energy docking minimum located?

The docking problem possesses many more degrees of freedom than a single molecule protein folding problem of equivalent size[14], and thus, the predicted receptor-ligand geometry for many classes of problem is often highly speculative. Indeed, whether or not a "true" free energy docking minimum is located for this docking problem with an SA/MD approach cannot (at this stage) be determined. Intermolecular receptor-ligand distance snapshots for each individual complex do, however, suggest effective global and local sampling (example MD equilibration and production statistics are provided in *Figure 7.3.1.2* and backbone RMSD values for each complex are given in *Figure 7.3.1.1*). It is assumed hereon that SA-sampling in conjunction with extended MD locates a good (or global) minimum for this class of problem.

Figure 7.3.1.1. *RMSD plots for initial SA docking. Large jumps in RMSD reflect manually orientated docking positions, whilst small changes reflect the existence of local, SA-generated sampling. The reference structure in each case is the first generated structure.*

HBD1

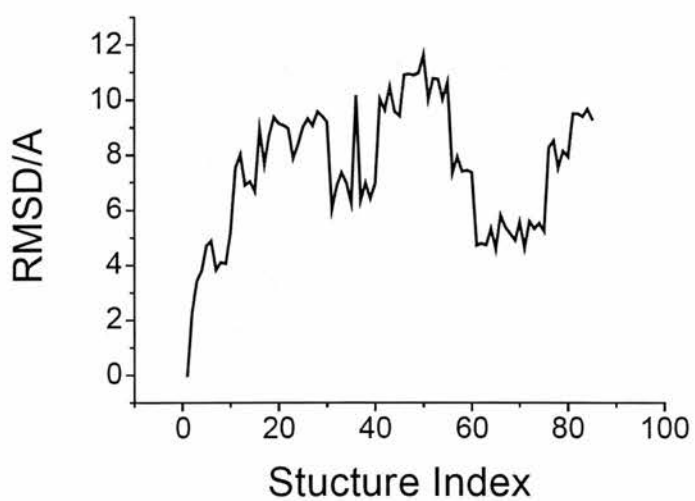
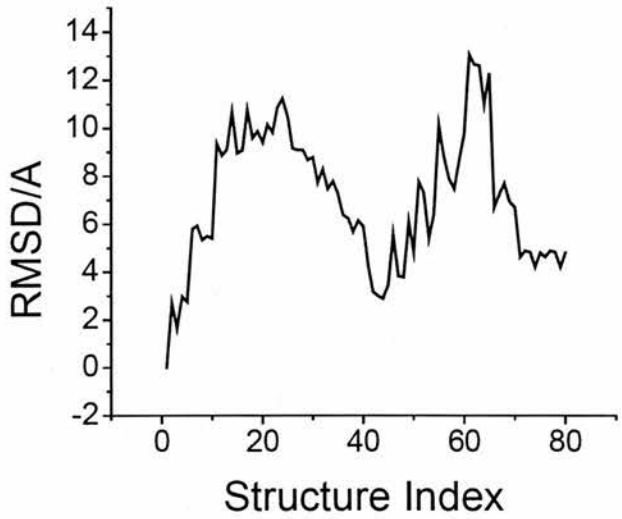
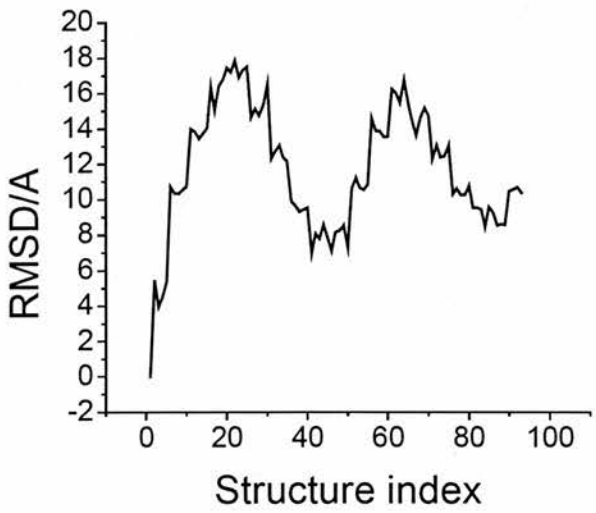


Figure 7.3.1.1 (continued). *RMSD plots for initial SA docking. Large jumps in RMSD reflect manually orientated docking positions, whilst small changes reflect the existence of local, SA-generated sampling. The reference structure in each case is the first generated structure.*

HBD2



HBD3



The extended MD runs provide, in addition to beneficial local sampling, a first level of binding discrimination. Indeed, only four HBD1 and HBD2 complexes survive the initial 1ns extended MD phase (results not shown); a result in some contrast to that of the inherently more cationic HBD3, for which there are twenty five stable complexes over the same time frame. A comparison of charge versus binding accordingly suggests (at first glance) a large electrostatic component to the observed binding stability. However, as will be shown in due course, hydrogen bonding also plays a crucial role.

Figure 7.3.1.2. *Example production statistics for a selected HBD2 run. Following initial equilibration stable production statistics are generated for: A: energy, B: temperature, C: pressure, D: density and E: volume. These statistics are representative of MD for each of HBD1-3.*

A

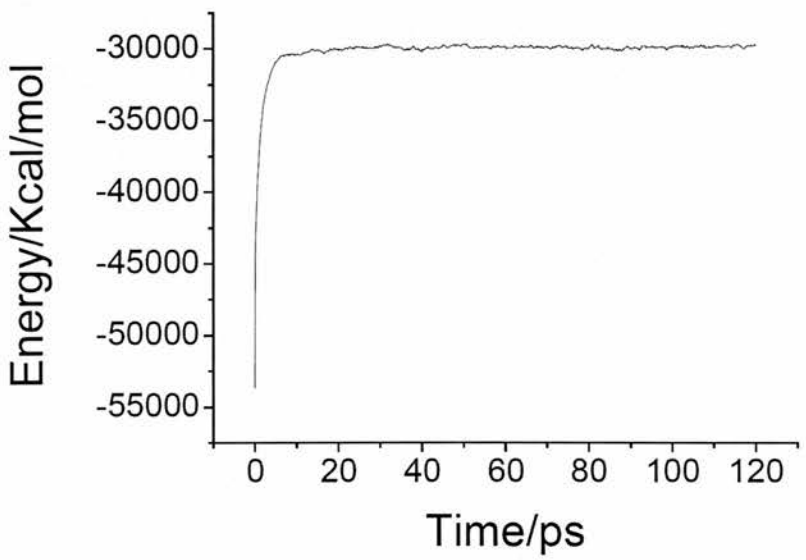
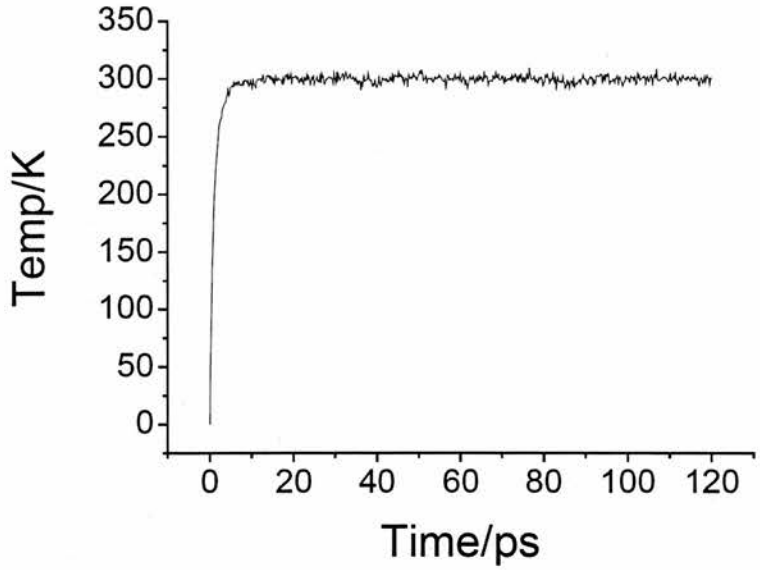


***Figure 7.3.1.2.*(continued)** *Example production statistics for a selected HBD2 run. Following initial equilibration stable production statistics are generated for: A: energy, B: temperature, C: density, D: pressure and E: volume. These statistics are representative.*

B



C

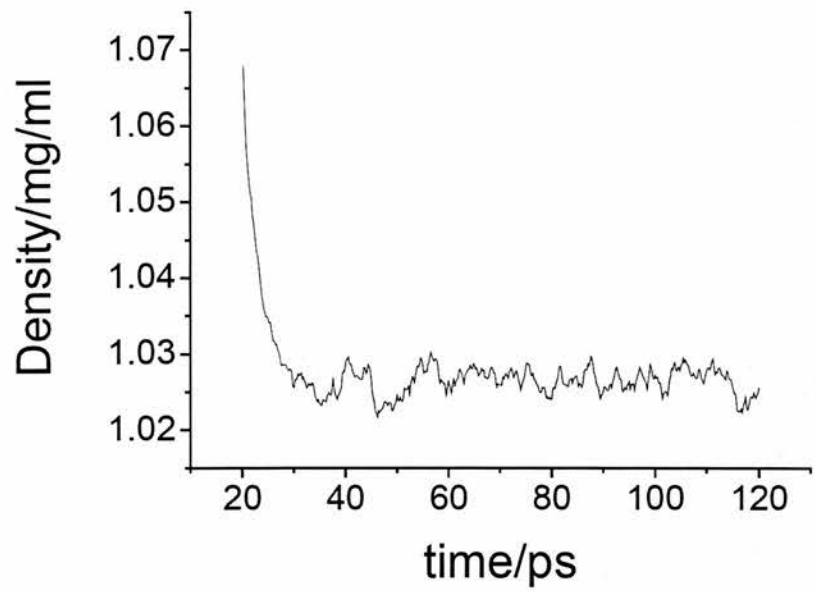
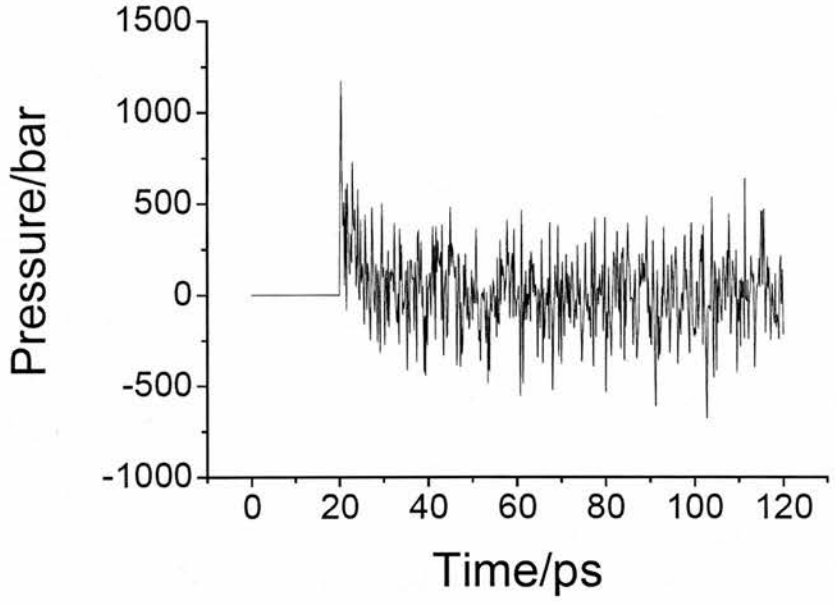
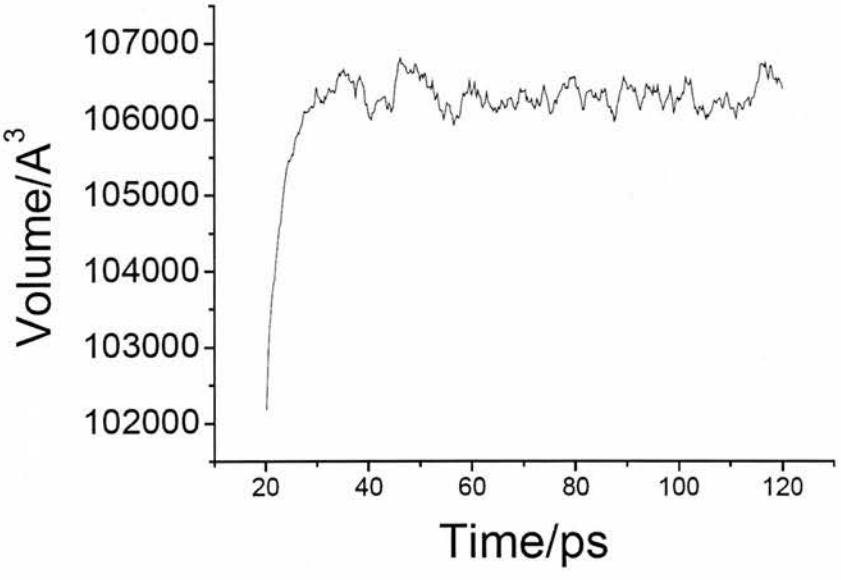


Figure 7.3.1.2.(continued) *Example production statistics for a selected HBD2 run. Following initial equilibration stable production statistics are generated for: A: energy, B: temperature, C: density, D: pressure and E: volume. These statistics are representative of equilibration for each of HBD1-3.*

D



E



7.3.2. MM-MPSA scoring

Structures that survive the first nanosecond of MD are subjected to an MM-PBSA free-energy scoring comparison. For HBD1 and HBD2 the relative free energy of binding is insubstantial, ranging from approximately +4 to -14 kcal/mol (*Figure 7.3.2.1*). This is in some contrast to the many HBD3 complexes that possess a relative free energy of binding in excess of -50 kcal/mol (*Figure 7.3.2.1).* The best three ranked structures for each of HBD1-3 are provided in *Figures 7.3.2.2 – 4* respectively.

The MM-PBSA method is approximate and suffers from two shortcomings:

1. The polar component of the free-energy term is determined with an implicit solvent approximation; does this capture adequately the inhomogeneous and complex electrostatic interactions that govern the docking landscape?

2. The non-polar free energy contribution makes use of an empirically derived solvent exposed surface area parameter that is determined by averaging the non-polar properties of many organic compounds[50,51]; does this averaging reflect the true hydrogen bonding contribution for *this* docking problem?

The latter concern is given credence by the fact that *implicitly* solvated, docked complexes of HBD1-3 are found to be stable during MD runs at 300K , 400K and 500K[3]. In each case, the number of stable complexes in existence after 2 ns of production MD greatly exceeds that of the explicitly solvated case. HBD1 and HBD2, for example, give rise to at least 12 stable

[3] Many of these initial results were regrettably deleted in order to accommodate the large explicitly solvated molecular data sets. Frustratingly, it was only following the explicit solvent study that the extent of the inaccuracy of the implicit solvent model was realized, and henceforth, the (possible) importance of hydrogen-bonding in the docking process.

complexes, whilst the explicit solvent case gives rise to just 4. This finding suggests strongly that the implicit solvent model inadequately captures the solvent milieu, and hence, that binding comparisons based on it may be inaccurate. Future work will undoubtedly require an explicit solvent free-energy binding comparison, perhaps incorporating a thermodynamic integration approach[32,52].

Figure 7.3.2.1. *MM-PBSA relative free energy of binding for HBD1-3.*

HBD1

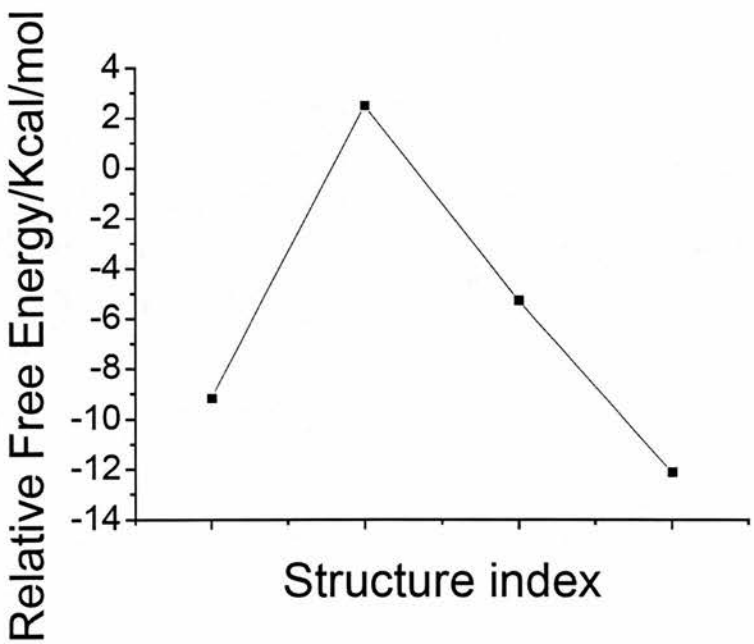
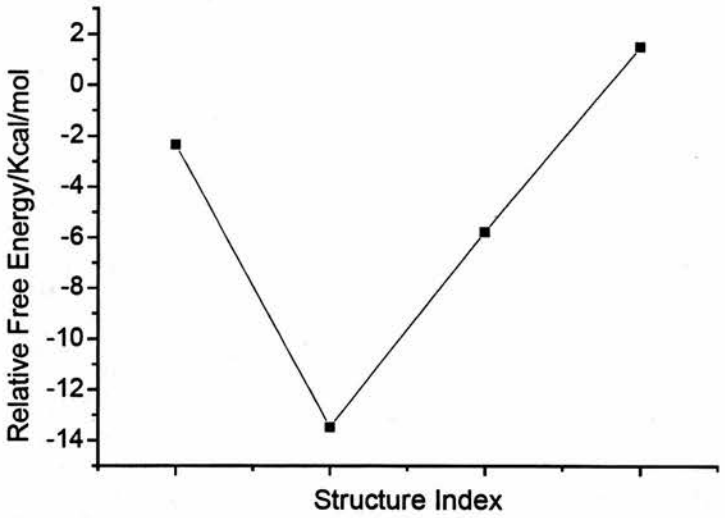


Figure 7.3.2.1 (continued). *MM-PBSA relative free-energies of binding for HBD1-3.*

HBD2



HBD3

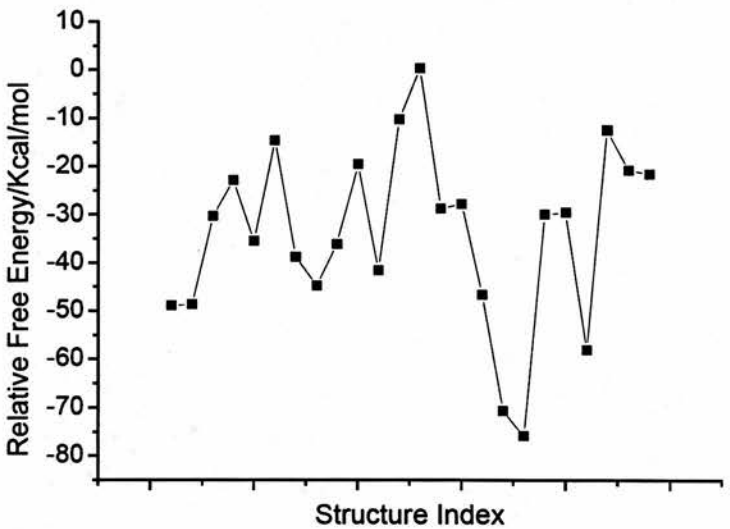
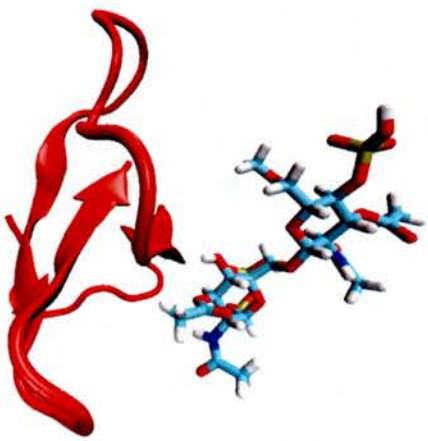


Figure 7.3.2.2. *The 3 best free-energy-ranked structures for each of HBD1. The structures are ranked A-C in order of free-energy stabilization. The defensin in each case is depicted in a red cartoon format, whilst the lipid A fragment is shown in liquorices in which phosphates appear brown, oxygen red, nitrogen blue, and carbon, turquoise.*

A

B



C

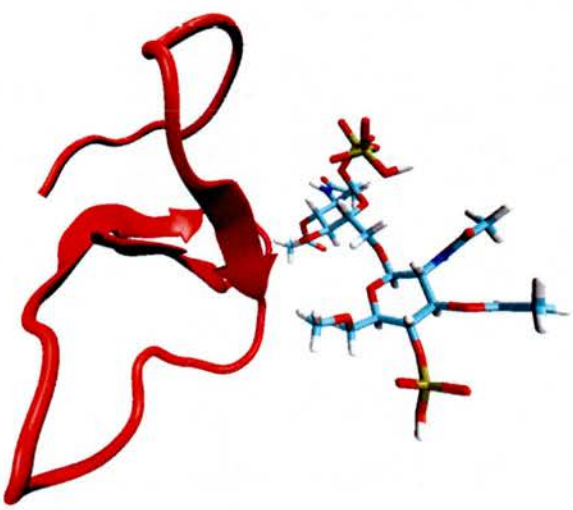


Figure 7.3.2.3. *The 3 best free-energy-ranked structures for each of HBD2. The structures are ranked A-C in order of free-energy stabilization. The defensin in each case is depicted in a red cartoon format, whilst the lipid A fragment is shown in liquorices in which phosphates appear brown, oxygen red, nitrogen blue and carbon, turquoise.*

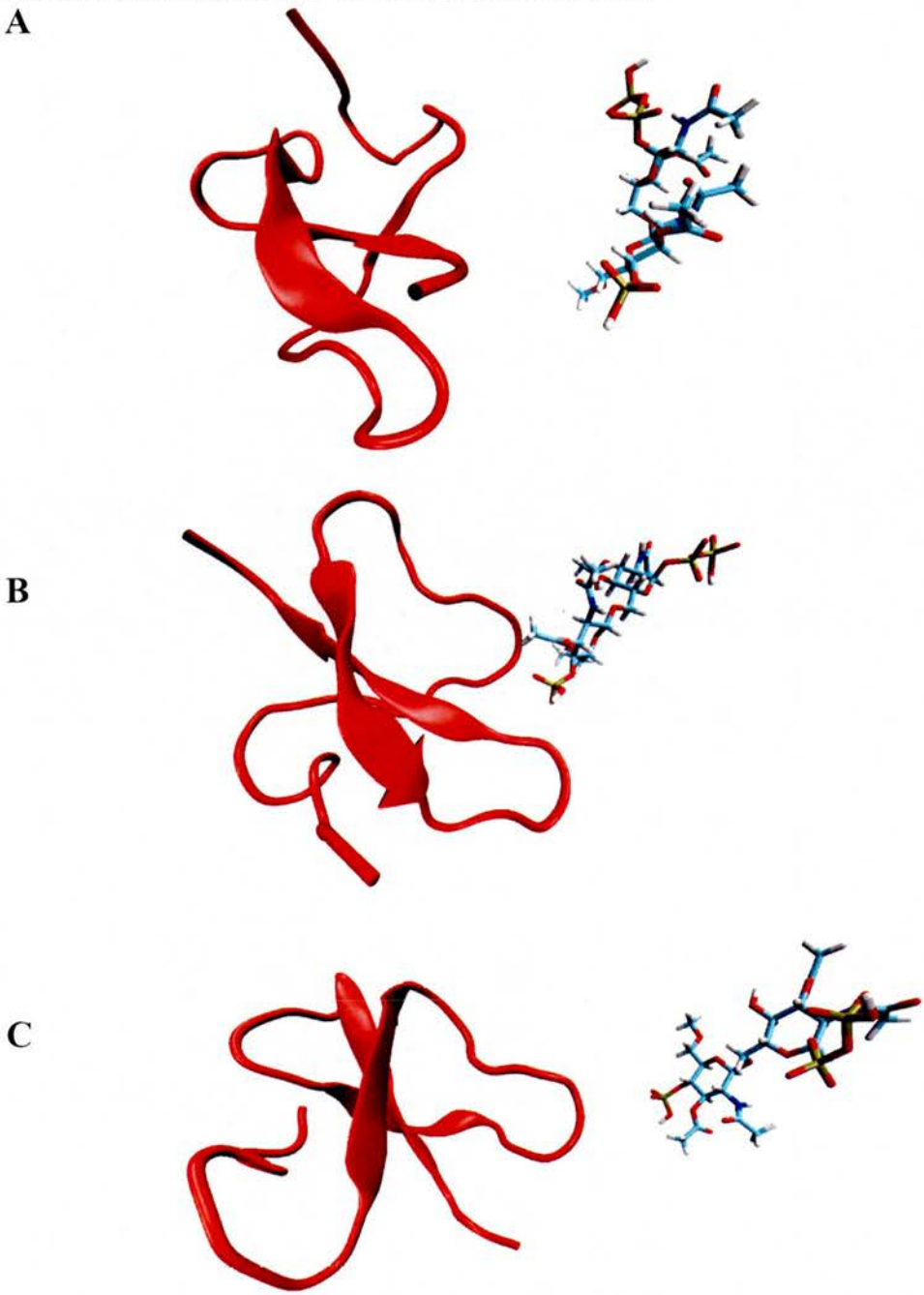
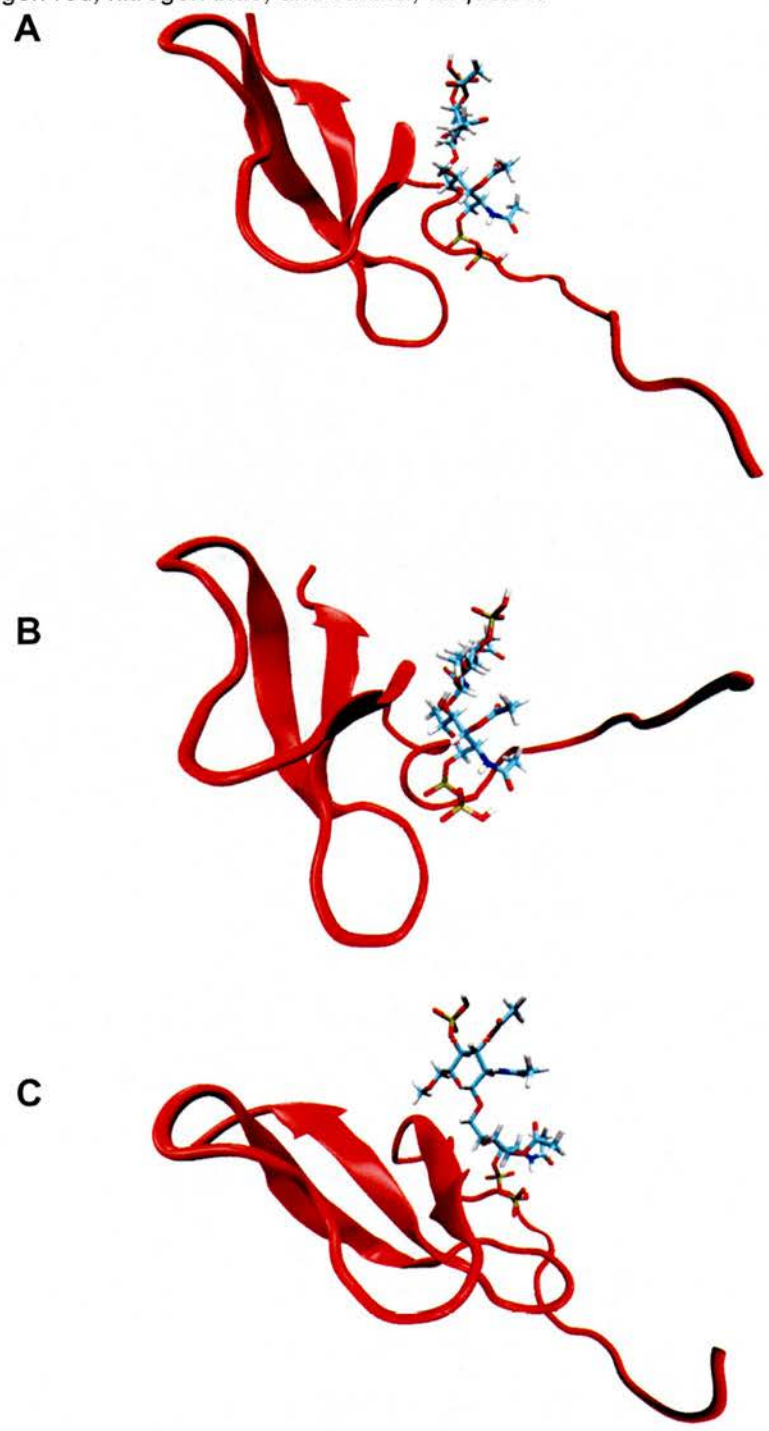


Figure 7.3.2.4. *The 3 best free-energy-ranked structures for each of HBD3. The structures are ranked A-C in order of decreasing free-energy. The defensin in each case is depicted in a red cartoon format, whilst the lipid A fragment is shown in liquorices format; phosphates are brown, oxygen red, nitrogen blue, and carbon, turquoise.*



7.3.3. Further MD of best ranked structures

Both HBD1 and HBD2 complexes dissociate within 300-400 ps and only HBD3 complexes survive a further 2ns of production MD (results not shown). It is here on assumed that only HBD3 is capable of eliciting stable binding, and as such, only HBD3 is further considered (in *section 7.3.4)* for mutational analysis.

The best ranked HBD3 complexes bind *TPDG* in a pocket that lies partway between the N-terminal region and the leading edge of the first β-sheet. This finding is consistent with the Mass spectrometric evidence of N-terminal DEFB14 defensin fragments** binding to heparin sulphate (a similar double ring to *TPDG* with an identical -3 charge state). This is a tentative finding, which, despite being a gas-phase study, supports the current *in silico* supposition that the N-terminal region is crucial in forming initial docking interactions.

7.3.4. Binding insights from residue mutations: Is hydrogen bonding important?

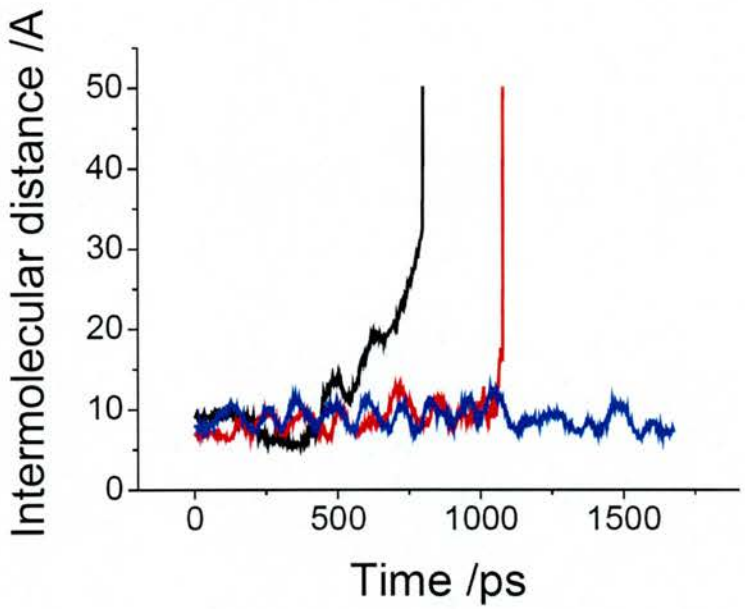
The best three ranked HBD3-*TPDG* complexes (S*ection 7.3.3*) are mutated in *situ* to incorporate *BPDG* and *Ara4N* respectively. Extended MD (of 2 ns in duration) for each of these mutants gives rise to dissociation in all but one instance (*Figure 7.3.4.1).* This finding contradicts (in part) the prevailing belief that binding results from a "predominantly" electrostatic mechanism. Indeed, whilst charge is almost certainly a factor in the observed binding efficiency of HBD3 compared to HBD1 and HBD2, it cannot be the sole factor in determining binding because, in the context of a simple charge model, changing *TPDG* to *Ara4N* should correlate with a decrease in binding stability; this is not observed – and neither, for that matter, is the expected variation on mutating *TPDG* to the less anionic *BPDG*. These findings imply that a complex interplay of electrostatic and hydrogen bonding interactions determine the binding mode, and not merely a

** DIP stands for "defensin inspired peptide".s

"simple" electrostatic mechanism, as has been postulated previously[53-57]. Indeed, a hydrogen bonding analysis of the MD trajectory for the system of *TPDG*-HBD3 reveals that interactions with solvent outweigh (in terms of time occupancy) those with substrate (*Figure 7.3.4.3* and *Table 7.3.4.4*). These findings, in addition to supporting a more complex view of β-defensin-substrate interactions, by extension challenge the assumption that resistance of *Ara4N* mutants to cationic antimicrobial peptides arises by virtue of a straightforward electrostatic mechanism in which like-like charge repulsion thwarts initial binding events.

Figure 7.3.4.1. *A graph of extended MD versus intermolecular distance for HBD3 bound to each of Ara4N and BPDG. Coordinate snapshots are sampled out of the trajectories in 5ps intervals. Blue, red and black lines indicate the free-energy ranking of the structures from which MD is seeded: blue, best ranked and black, worst ranked. Only one structure within Ara4N survives this extended MD phase.*

Ara4N



BPDG

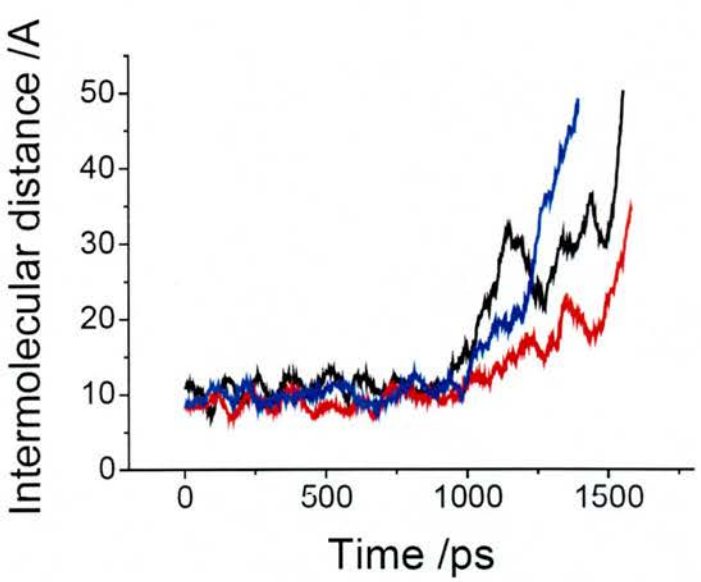


Figure 7.3.4.2. *MM-PBSA free-energy comparison of BPDG, TPDG and Ara4N. The trajectory snapshot method is used as detailed in section 7.2.3 and 7.2.4. The free-energy values do not correlate with the magnitude of ligand charge (shown in brackets, as would be expected with a dominating electrostatic mechanism.*

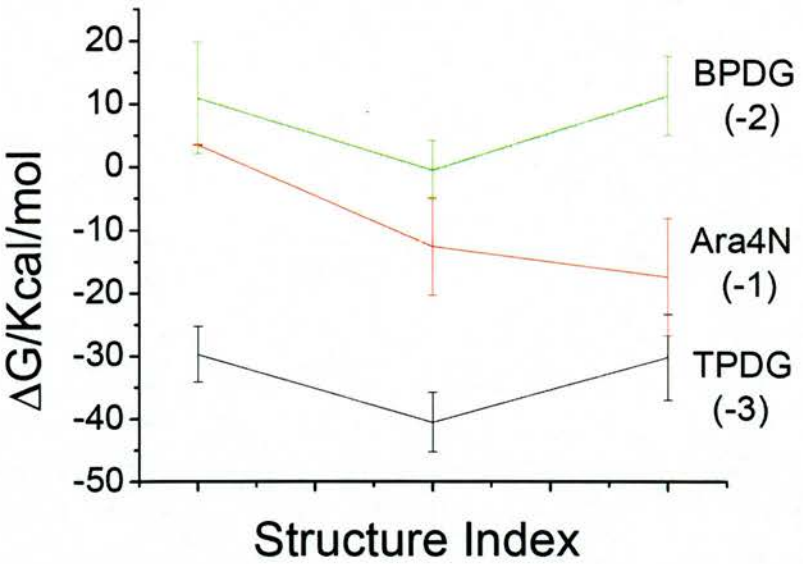


Figure 7.3.4.3. *A graph showing relative hydrogen bonding occupancy for the best ranked TPDG-HBD3 structure over a time frame of 800 ps. The red line shows hydrogen bonds formed with solvent and the black line shows hydrogen bonds formed with HBD3. The numbers label each interacting (or potentially interacting) atom. A relative occupancy of 0 indicates that no hydrogen bonding is detected for the labelled atom. The graph illustrates that hydrogen bonding with solvent outweighs hydrogen bonding with substrate.*

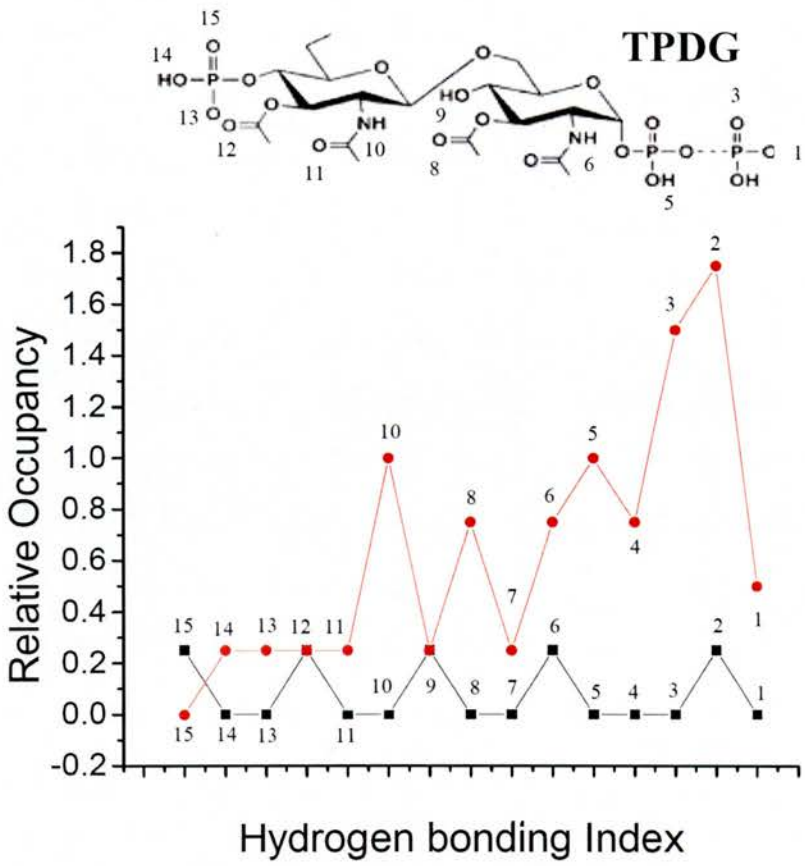


Table 7.3.4.4. *Table showing the AMBER residues and corresponding atom names involved in PDG-HBD3 hydrogen bonding. Red numbers (starting from the N-terminus) refer to HBD3 amino acid residues whilst black ones refer to TPDG residues. Numbers within the index field map to the atom references in Figure 7.3.4.3*

Index	Donor (atom Name)	Acceptor (Atom Name)	Bond Length
8	O2	18 H:N	2.6875
12	O15	34 HG:OG	2.875
1	45 OXT	H24:O9	2.859
6	26 O	H8:N1	2.897
2	24 O	H34 :O21	2.924

7.4. Conclusion

The ability of this model to elucidate the defensin structure-activity relationship depends on how representative the inherent model assumptions and simplifications are of the biological reality. As is discussed in *section 7.1* the model system of lipid A deliberately excludes many carbohydrate rings and fatty acid chains; are these components really just bulky spectators to the docking process? (Resistance in some bacterial isolates correlates with variations to the fatty-acid region of lipid A, and not to the GAG region[49]). Furthermore, are the partial atomic charges representative? Are the protonation state assignments accurate for the local chemical milieu? This model is a first, bottom-up attempt at capturing the underlying interactions that govern the structure-activity relationship and caution must be observed when interpreting such a highly simplified model. Indeed, until there is better experimental evidence and a more complete understanding of the role (if any) of the other lipid A components, definitive conclusions concerning the mode of action of this class of protein will be difficult to make. For the moment is safe to say only that:

1. More atoms need to be included in the lipid A model (Do these, in reality, shield the defensin-substrate interface from solvent interference?)
2. A more extensive sampling algorithm should be used and validated (where possible) with experiment.
3. That the MM-PBSA scoring algorithm (due to its inability to capture the hydrogen bonding contribution to binding) be improved (or exchanged for a better scoring measure).

However, assuming that the docking procedure locates a global minimum within a model framework that captures vital docking interactions then two principle conclusions (discussed below) can be drawn.

7.4.1. Failure of defensin to bind substrate is not a chance event

Most biological receptors are found to possess off rates in the region of 10^{-4} to 10^{-2} seconds^{-1} [58-61], an observation in some contrast to that of the nanosecond timescale of defensin-substrate dissociation observed in this study. Thus, assuming that the docking procedure locates a true global binding minimum, it is likely that the failure of HBD1 and HBD2 to remain bound during the simulation time scale reflects a fundamental inability of the molecules to interact, and not merely a chance *off* event.

The same rationality implies that no firm conclusions can be drawn about the ability of HBD3 to bind *TPDG*, *BPDG* and *Ara4N*. Indeed, it is conceivable that a longer simulation period in the case of *Ara4N* and *TPDG* would result in dissociation. Whilst this remains a possibility, some comfort can be drawn from the fact that the free energy of binding for *TPDG* exceeds that of other known receptor-ligand interactions. For example, HBD3-*TPDG* possesses a calculated absolute binding energy of -40 kcal/mol, which is in excess of the -13.2 kcal/mol calculated for the reverse transcription inhibitor, TIBO[31].

7.4.2. This study does not support the current resistance hypothesis

HBD3 is shown to bind to both *BPDG* and *Ara4N* with comparable binding free energies and HBD1 and HBD2 are not shown to bind BPDG, TPDG or Ara4N at all. These observations do not support the hypothesis that bacterial resistance to antimicrobial attack arises via disrupted

lipid A-defensin binding events. However, as previously discussed, this research has not examined the possibility that some (or all) defensins operate in a concerted manner. Indeed, it may well be that effective binding is observed only in dimeric or trimeric forms of HBD1 and HBD2. This remains a future investigative research area.

7.4.3. Binding is not a "simple" electrostatic process

The fact that HBD3 binds lipid A components strongly, but that HBD1 and HBD2 bind them weakly (or not at all) illustrates a strong electrostatic component to docking. However, the large observed disparity in the relative MD stability of explicitly and implicitly solvated receptor-ligand complexes, along with the hydrogen bonding occupancies, show that hydrogen bonding is also a significant factor. This inference is further supported by the observation that binding free energy differences between *Ara4N*, *BPDG* and *TPDG* do not strictly obey an electrostatic pattern (which would likely be the case if hydrogen bonding was indeed insignificant).

7.5. Standard settings

Unless otherwise stated, the parameters in *Table 7.5.1* are used for free energy calculations. Equilibration and production MD parameters used in this Chapter are identical to those discussed in Chapter Six. SA parameters are provided in *Table 7.5.2*.

***Table 7.5.1**. MM-PBSA settings for determining relative free-energies of binding for defensin-sacharide complexes.*

Paramater description	Value	Comment
PROC	2	method of solving the Poisson Boltzmann (PB) equation. This option refers to the PBSA method[26].
REFE	0	Determines reference state for PB calculation.

INDI	1.0	Dielectric constant for the solute.
EXDI	80.0	Dielectric constant for the external solvent.
SCALE	0.5	Lattice spacing in Angstroms for the number of grids. This spacing has been shown to give stable and converged values[41,62].
LINIT	1000	Number of interactions with linear PB equation.
PRBRAD	1.4	
ISTRNG	0.0	Ionic strength of the PB solvent.
RADIOPT	0	Set up radii from the prmtop file.
NPOPT	1	Use solvent accessible surface area to correlate total nonpolar solvation free energy.
CAVITY_SURFTEN	0.04352	Values used to compute the non-polar solvation free energy. Other values used by Kuhn and Kollman vary, as does that of Yan *et. al* Surften = 0.00542 kcal, b = 0.92[36,62].
CAVITY_OFFSET	-1.008	
SURFTEN	0.0072	Values used to compute the non-polar contribution to desolvation.
SURFOFF	0.00	
DIELC	1.0	Dielectric constant for electrostatic interactions.
IGB	2	Onufriev's implementation of the Born solvent model[47,63].
GBSA	1	Use LCPO method for surface area[30]
SALTCON	0.15	Concentration (in Mols) of counterions in solution. 0.1 – 1M used by Kolllman[28]. Physiological concentration is often taken to be 0.15 M[48].
EXTDIEL	80.0	Dielectric constant for the GB solvent.
INTDIEL	1.0	Dielectric constant for the GB solute.
SURFTEN	0.0072	Values used to compute the non-polar contribution of to desolvation for GB model.
SURFOFF	0.92	
PROBE	0.0	Radius of the probe sphere used to calculate Surface Area.

Table 7.5.2. *MD SA settings. Minimization settings used are identical to those outlined in Chapter Four.*

Parameter	**Setting**	**Comment**
max temperature/K	520	MD at temperatures higher than this can cause solute to denature.
cooling step/K	Chosen stochastically. Upper limit = 50, lower limit =10.	
heating time/ps	Chosen stochastically. Upper limit = 50, lower limit = 14.	Varying heating times produce better sampling than static ones.
cooling time /ps	Chosen stochastically. Upper limit = 30, lower limit = 10.	

References

(1) Ernst, R. K.; Guina, T.; Miller, S. I. *Journal of Infectious Diseases* **1999**, *179*, S326-S330.

(2) Nummila, K.; Kilpelainen, I.; Zahringer, U.; Vaara, M.; Helander, I. M. *Molecular Microbiology* **1995**, *16*, 271-278.

(3) Guo, L.; Lim, K. B.; Gunn, J. S.; Bainbridge, B.; Darveau, R. P.; Hackett, M.; Miller, S. I. *Science* **1997**, *276*, 250-253.

(4) Zhou, Z. M.; Ribeiro, A. A.; Lin, S. H.; Cotter, R. J.; Miller, S. I.; Raetz, C. R. H. *Journal of Biological Chemistry* **2001**, *276*, 43111-43121.

(5) Peschel, A.; Jack, R. W.; Otto, M.; Collins, L. V.; Staubitz, P.; Nicholson, G.; Kalbacher, H.; Nieuwenhuizen, W. F.; Jung, G.; Tarkowski, A.; van Kessel, K. P. M.; van Strijp, J. A. G. *Journal of Experimental Medicine* **2001**, *193*, 1067-1076.

(6) Peschel, A. *Trends in Microbiology* **2002**, *10*, 179-186.

(7) Frecer, V.; Ho, B.; Ding, J. L. *Antimicrobial Agents and Chemotherapy* **2004**, *48*, 3349-3357.

(8) Lancini, G.; Parenti, F.; Gallo, G. G. *Antibiotics: a multidisciplinary approach*; Plenum Press, New York, N.Y., 1995.

(9) Berneche, S.; Nina, M.; Roux, B. *Biophysical Journal* **1998**, *75*, 1603-1618.

(10) Khandelia, H.; Kaznessis, Y. N. *Peptides* **2005**, *26*, 2037-2049.

(11) Khandelia, H.; Kaznessis, Y. N. *Journal of Physical Chemistry B* **2005**, *109*, 12990-12996.

(12) Bemporad, D.; Essex, J. W.; Luttmann, C. *Journal of Physical Chemistry B* **2004**, *108*, 4875-4884.

(13) Bemporad, D.; Luttmann, C.; Essex, J. W. *Biochimica Et Biophysica Acta-Biomembranes* **2005**, *1718*, 1-21.

(14) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *Journal of Computer-Aided Molecular Design* **2002**, *16*, 151-166.

(15) Halperin, I.; Ma, B. Y.; Wolfson, H.; Nussinov, R. *Proteins-Structure Function and Genetics* **2002**, *47*, 409-443.

(16) Kroemer, R. T. *Current Protein & Peptide Science* **2007**, *8*, 312-328.

(17) Mangoni, R.; Roccatano, D.; Di Nola, A. *Proteins-Structure Function and Genetics* **1999**, *35*, 153-162.

(18) Dinola, A.; Roccatano, D.; Berendsen, H. J. C. *Proteins-Structure Function and Genetics* **1994**, *19*, 174-182.

(19) Pak, Y. S.; Wang, S. M. *Journal of Physical Chemistry B* **2000**, *104*, 354-359.

(20) Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. *Chemical Physics Letters* **1997**, *278*, 297-301.

(21) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. *Journal of Computational Chemistry* **2007**, *28*, 1145-1152.

(22) Goodsell, D. S.; Olson, A. J. *Proteins-Structure Function and Genetics* **1990**, *8*, 195-202.

(23) Goodsell, D. S.; Morris, G. M.; Olson, A. J. *Journal of Molecular Recognition* **1996**, *9*, 1-5.

(24) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33-&.

(25) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *Journal of Physical Chemistry* **1987**, *91*, 6269-6271.

(26) Luo, R.; David, L.; Gilson, M. K. *Journal of Computational Chemistry* **2002**, *23*, 1244-1253.

(27) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *Journal of the American Chemical Society* **1998**, *120*, 9401-9409.

(28) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Accounts of Chemical Research* **2000**, *33*, 889-897.

(29) Pearlman, D. A., Case, D. A., Caldwell, J. W., Cheatham, T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., Kollman, P. A. **2002**.

(30) Massova, I.; Kollman, P. A. *Perspectives in Drug Discovery and Design* **2000**, *18*, 113-135.

(31) Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A. *Journal of the American Chemical Society* **2001**, *123*, 5221-5230.

(32) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. *Biopolymers* **2003**, *68*, 16-34.

(33) Yan, S. X.; Geacintov, N. E.; Broyde, S. *Abstracts of Papers of the American Chemical Society* **2001**, *222*, U397-U398.

(34) Pricl, S.; Ferrone, M.; Fermeglia, M.; Lodi, A.; Mazzurco, A. *Abstracts of Papers of the American Chemical Society* **2003**, *225*, U260-U260.

(35) Laitinen, T.; Rouvinen, J.; Perakyla, M. *Organic & Biomolecular Chemistry* **2003**, *1*, 3535-3540.

(36) Yan, S. X.; Wu, M.; Patel, D. J.; Geacintov, N. E.; Broyde, S. *Biophysical Journal* **2003**, *84*, 2137-2148.

(37) Tan, J. J.; Kong, R.; Chen, W. Z.; Wang, C. X. *Molecular Simulation* **2005**, *31*, 1051-1056.

(38) Weinzinger, P.; Hannongbua, S.; Wolschann, P. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2005**, *20*, 129-134.

(39) Wojciechowski, M.; Fogolari, F.; Baginski, M. *Biophysical Journal* **2005**, *88*, 61a-61a.

(40) Page, C. S.; Bates, P. A. *Journal of Computational Chemistry* **2006**, *27*, 1990-2007.

(41) Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. *Journal of Medicinal Chemistry* **2006**, *49*, 6596-6606.

(42) Zhou, Z.; Bates, M.; Madura, J. D. *Proteins-Structure Function and Bioinformatics* **2006**, *65*, 580-592.

(43) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. *Journal of Medicinal Chemistry* **2005**, *48*, 4040-4048.

(44) Connolly, M. L. *Journal of Applied Crystallography* **1983**, *16*, 548-558.

(45) Bashford, D.; Case, D. A. *Annual Review of Physical Chemistry* **2000**, *51*, 129-152.

(46) Tsui, V.; Case, D. A. *Journal of the American Chemical Society* **2000**, *122*, 2489-2498.

(47) Onufriev, A.; Bashford, D.; Case, D. A. *Journal of Physical Chemistry B* **2000**, *104*, 3712-3720.

(48) Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson **1994**

(49) Peschel, A.; Collins, L. V. *Peptides* **2001**, *22*, 1651-1659.

(50) Sitkoff, D.; Sharp, K. A.; Honig, B. *Journal of Physical Chemistry* **1994**, *98*, 1978-1988.

(51) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305-320.

(52) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *Journal of Chemical Physics* **2003**, *119*, 5740-5761.

(53) Ganz, T. *Nature Reviews Immunology* **2003**, *3*, 710-720.

(54) Hwang, P. M.; Vogel, H. J. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* **1998**, *76*, 235-246.

(55) Niyonsaba, F.; Nagaoka, I.; Ogawa, H. *Critical Reviews in Immunology* **2006**, *26*, 545-575.

(56) Raj, P. A.; Dentino, A. R. *Fems Microbiology Letters* **2002**, *206*, 9-18.

(57) Selsted, M. E.; Ouellette, A. J. *Nature Immunology* **2005**, *6*, 551-557.

(58) Maerkl, S. J.; Quake, S. R. *Science* **2007**, *315*, 233-237.

(59) Lima, L. M. T. R.; Foguel, D.; Silva, J. L. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 14289-14294.

(60) Mukherjee, S. *Nature Genetics* **2004**, *36*, 1331-1339.

(61) M. L. Bulyk, X. H., Y. Choo, G. M. Church *Proceedings of the National Academy of Sciences*
2001 *98*, 7158.

(62) Kuhn, B.; Kollman, P. A. *Journal of Medicinal Chemistry* **2000**, *43*, 3786-3791.

(63) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins-Structure Function and Bioinformatics* **2004**, *55*, 383-394.

Chapter Eight

Conclusion

8.1. Introduction

The ultimate goal of this research is to elucidate the mechanism of defensin antimicrobial activity so that new, effective and safe antibiotics can be developed to meet the growing challenge of antibiotic resistance. A secondary aim is to develop a robust, *in silico* methodology for predicting the native structure of β-defensins.

When considering an in silico approach to a biological problem it is important to realise the full complexity of the reality that is being modelled. The prevailing evidence indicates that biological systems balance on a thermodynamic and kinetic knife-edge. Mice, for example, are known to become ill and die if they habitually ingest water with a deuterium content in excess of 50%[1]. The root of this toxicity lies in subtle changes in the quantum properties of hydrogen bonding that presumably disrupt one or more vital biochemical reactions[1]. Experiments of this ilk hint that there can be limitations to force-field approximations within a biological framework. Indeed, force-field approximations, by their very nature, fail to capture the true characteristics of the potential energy surface. Attempting to locate a global minimum on a potential energy surface that does not correspond to the native structure remains a significant problem for all force-field methods.

8.2. Rational Structure prediction

There are currently no correct, published rationally predicted native structures of any disulfide cross-linked systems. The in *silico* work presented here suggests that this is (in part) because the

single-Hamiltonian representations that current forcefield-based folding strategies invoke are highly susceptible to local trapping on the potential energy landscape. The dynamic-bond folding algorithm presented here (as part of a sampling algorithm) alleviates this trapping problem and (in some cases) gives rise to comparatively better native secondary structures.

Failure to locate the native geometry might be because:

1. Too much time is spent sampling extended geometries in which no cross-linking is defined; the folding problem in this instance is then much like an ordinary linear protein-folding problem (with many more degrees of freedom than the restricted case).
2. Interchanging between global and local sampling gives rise to structural collapse that is too rapid for secondary structure formation to occur.
3. The current folding algorithm only models canonical connectivities and correct folding may depend critically on the existence of one or more non-canonical folding intermediates (as is experimentally observed in the system of BPTI[2-14]).
4. The forcefield approximation does not capture the true potential energy minimum for the system.

Future work on the algorithm will therefore focus on two aspects:

1. *Improving the criterion for interchanging global and local sampling.*

   The algorithm presented in this thesis, by design, attempts to take advantage of both global and local sampling by intelligently interchanging between dynamic and static descriptions of bonding. Indeed, the upper limit of protein folding is 20 amino acids and

solving defensin structures (consisting of 35-40 amino acids) will undoubtedly require a procedure that utilises restricted conformational space. Future work will focus on improving the criterion for interchanging between local and global sampling.

2. *Introducing non-canonical pairings*

   The dynamic bond algorithm currently samples canonically connected states only. Future work will extend the Hamiltonian repertoire to include non-canonical pairings.

8.3. Lipid A-Defensin docking

Bacteria that possess cationic modifications to lipid A are resistant to defensin (and other antimicrobial attack)[15-20].This thesis tests the hypothesis that this resistance arises by virtue of a simple electrostatic mechanism in which like-charge repulsion disrupts initial (and vital) defensin-lipid A binding events. Preliminary docking calculations with a simplified representation of Lipid A - triphosphorylated diglucosamine (*TPDG*), bis-phsopshrylated diglucosamine (*BPDG*) and the cationic mutant aminoarabinose-modified-TPDG (*Ara4N*) – do not support a simple, electrostatic repulsion model of antimicrobial resistance. Indeed, only HBD3 in this model shows any appreciable binding, HBD1 and HBD2 do not. Furthermore, no strict electrostatic behaviour is observed for HBD3, implying that the observed binding results from a complex interplay of hydrogen bonding and electrostatic interactions and not merely a "simple" charge process as previously postulated)[21-25].

Future work will focus on:

1. Expanding the basic model of lipid A to include more regions of the molecule.

2. Improving docking sampling and applying a more accurate free-energy procedure to the final free-energy calculation.
3. Performing docking with oligomeric defensins.
4. Obtaining better experimental binding studies to support the *in silico* modelling.

8.4. β-defensins – A new generation of antibiotic?

Bacterial drug resistance remains a serious problem in the world and there is an urgent requirement for new and better antibiotics to counter this problem. Explaining antimicrobial resistance in mutant lipid A bacterial isolates may provide the first insight to the antimicrobial process. Regrettably, the speculation that preliminary docking between defensin and bacteria begins with attachments to lipid A is not definitively supported (or rejected) with this model. Better experimental evidence and more inclusive molecular models will be required to get a clearer picture of whether or not this interaction occurs.

Assuming that the first, preliminary interactions are elucidated then further questions are immediately raised: What are the next subsequent events? Do defensins always act as monomers? Or do they form higher order oligomers that act cooperatively (as is the case with the complement system[26], magainins[27-30] and alamethicin[31,32])? Do defensins permeate the bilayer or merely destabilise surface components? How, ultimately, do these small, cationic peptides create massive macromolecular destabilisation?

The questions posed above require significant resources at both a computational and experimental level. Answering them may be a matter of life and death for a great many people.

References

(1) Kushner, D. J.; Baker, A.; Dunstall, T. G. *Can. J. Physiol. Pharmacol.* **1999**, **77** (2), 79–88.

(2) Dadlez, M.; Kim, P. S. *Nature Structural Biology* **1995**, *2*, 674-679.

(3) Darby, N. J.; Vanmierlo, C. P. M.; Creighton, T. E. *Febs Letters* **1991**, *279*, 61-64.

(4) Weissman, J. S.; Kim, P. S. *Nature* **1993**, *365*, 185-188.

(5) Weissman, J. S.; Kim, P. S. *Nature Structural Biology* **1995**, *2*, 1123-1130.

(6) Creighton, T. E.; Zapun, A.; Darby, N. J. *Trends in Biotechnology* **1995**, *13*, 18-23.

(7) Ferrer, M.; Barany, G.; Woodward, C. *Nature Structural Biology* **1995**, *2*, 211-217.

(8) Creighton, T. E. *Biological Chemistry* **1997**, *378*, 731-744.

(9) Creighton, T. E. *Bioessays* **1992**, *14*, 195-200.

(10) Darby, N. J.; Morin, P. E.; Talbo, G.; Creighton, T. E. *Journal of Molecular Biology* **1995**, *249*, 463-477.

(11) Zdanowski, K.; Dadlez, M. *Journal of Molecular Biology* **1999**, *287*, 433-445.

(12) Okamoto, Y. *International Journal of Modern Physics C* **1999**, *10*, 1571-1582.

(13) Goldenberg, D. P.; Zhang, J. X. *Proteins-Structure Function and Genetics* **1993**, *15*, 322-329.

(14) Darby, N. J.; Creighton, T. E. *Journal of Molecular Biology* **1993**, *232*, 873-896

(15) Ernst, R. K.; Guina, T.; Miller, S. I. *Journal of Infectious Diseases* **1999**, *179*, S326-S330.

(16) Nummila, K.; Kilpelainen, I.; Zahringer, U.; Vaara, M.; Helander, I. M. *Molecular Microbiology* **1995**, *16*, 271-278.

(17) Guo, L.; Lim, K. B.; Gunn, J. S.; Bainbridge, B.; Darveau, R. P.; Hackett, M.; Miller, S. I. *Science* **1997**, *276*, 250-253.

(18) Zhou, Z. M.; Ribeiro, A. A.; Lin, S. H.; Cotter, R. J.; Miller, S. I.; Raetz, C. R. H. *Journal of Biological Chemistry* **2001**, *276*, 43111-43121.

(19) Peschel, A.; Jack, R. W.; Otto, M.; Collins, L. V.; Staubitz, P.; Nicholson, G.; Kalbacher, H.; Nieuwenhuizen, W. F.; Jung, G.; Tarkowski, A.; van Kessel, K. P. M.; van Strijp, J. A. G. *Journal of Experimental Medicine* **2001**, *193*, 1067-1076.

(20) Peschel, A. *Trends in Microbiology* **2002**, *10*, 179-186.

(21) Ganz, T. *Nature Reviews Immunology* **2003**, *3*, 710-720.

(22) Hwang, P. M.; Vogel, H. J. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* **1998**, *76*, 235-246.

(23) Niyonsaba, F.; Nagaoka, I.; Ogawa, H. *Critical Reviews in Immunology* **2006**, *26*, 545-575.

(24) Raj, P. A.; Dentino, A. R. *Fems Microbiology Letters* **2002**, *206*, 9-18.

(25) Selsted, M. E.; Ouellette, A. J. *Nature Immunology* **2005**, *6*, 551-557.

(26) Janeway, C. A. Jr.; Travers, P.; Walport, M.; Shlomchik, M. J. **2001**. *Immunobiology.*, 5th ed., Garland Publishing.

(27) Cruciani, R. A.; Stanley, E. F.; Zasloff, M.; Lewis, D. L.; Barker, J. L. *Biophysical Journal* **1988**, *53*, A9-A9.

(28) Matsuzaki, K.; Harada, M.; Funakoshi, S.; Fujii, N.; Miyajima, K. *Biochimica Et Biophysica Acta* **1991**, *1063*, 162-170.

(29) Matsuzaki, K.; Murase, O.; Miyajima, K. *Biochemistry* **1995**, *34*, 12553-12559.

(30) Westerhoff, H. V.; Zasloff, M.; Rosner, J. L.; Hendler, R. W.; Dewaal, A.; Gomes, A.

(31) Bechinger, B. *Journal of Membrane Biology* **1997**, *156*, 197-211.

(32) Fox, R. O.; Richards, F. M. *Nature* **1982**, *300*, 325-330.