

Computational Studies of Folding and Binding of Polypeptides

Eleanor Rose Turpin, BSc. (Hons.), MSc.

Thesis submitted to the University of Nottingham for the degree of Doctor
of Philosophy

JULY 2013

Abstract

In this thesis molecular dynamics simulations, in conjunction with the complementary methods of docking and QM-MM, are used, and further developed, to study two unusual polypeptide systems: the conformational preferences of isomers of an antibiotic peptide and the binding behaviour of a human transporter protein. The antibiotic peptides are analogues of a naturally occurring antibacterial called nisin which has a biological function dependent on the formation of five macrocyclic rings closed by a thioether bond between modified L-amino and D-amino residues. We propose analogues where the thioether bond is replaced by a disulfide bond between cysteine residues and the chirality of the cysteines is altered. The conformational preferences of the nisin analogues, and the dependence of ring formation on cysteine chirality, are characterised using molecular dynamics. An analogue (D-Cys3-D-Cys7-L-Cys8-L-Cys11) is identified that favours the simultaneous formation of the S₃-S₇ and S₈-S₁₁ disulfide bonds and has an RMSD of 0.6 Å to 1.7 Å between the centroids from clustering the MD trajectories and an NMR structure of *wt*-nisin. The nisin analogues contain unusual D-amino residues and using explicit solvent MD simulations of four polypeptides, it is shown that the $(\varphi, \psi) \rightarrow (-\varphi, -\psi)$ transformation of the CMAP term in the CHARMM potential energy function leads to sampling of conformations which are closest to X-ray crystallographic structures for D-amino residues and that the standard CMAP correction destabilises D-amino β -sheets and β -turns.

The ileal lipid binding protein (ILBP) shows cooperative binding comparable to haemoglobin and unusual site selectivity where one ligand will completely displace another from a binding site, despite both sites having an affinity for each ligand type and the ligands only differing by a single hydroxyl group. A probable location of the third binding site of ILBP is identified which has a role in the allosteric binding mechanism. MD simulations indicate that binding to this exterior site induces changes in the orientation of the α -helices with respect to the β -barrel by $\sim 10^\circ$. An energetic mechanism of site selectivity for

ILBP is proposed using evidence from MD simulations. The higher hydrophobicity of chenodeoxycholic acid leads it to sit deeper in the binding cavity and interact with Gln-51. This causes the cholic acid ligand to be deeper and induces the helices to move closer to the β -barrel, preventing further ligand exchange.

Refereed Publications

2. Turpin, E.R. and J.D. Hirst, *Transformation of the dihedral corrective map for D-amino residues using the CHARMM force field*. Chem. Phys. Lett., 2012. **543**: p. 142-147.
1. Turpin, E.R., B. Bonev, and J.D. Hirst, *Stereoselective Disulfide Formation Stabilizes the Local Peptide Conformation in Nisin Mimics*. Biochemistry, 2010. **49**: p. 9594-9603.

Acknowledgements

I would to like acknowledge the help, guidance and patience of my supervisor Professor Jonathan Hirst. I would also like to acknowledge my scientific collaborators Boyan Bonev and Nikos Doltsinis and my examiners Richard Wheatley and Neil Burton. I am grateful for the funding and resources provided by EPSRC, the School of Chemistry and the University of Nottingham's High Performance Computing service.

I am grateful for the computing support provided by Craig Bruce, Hainam Do and John Baker, chemistry help from David Robinson and Mark Oakley and the chance to discuss protein simulations with Clare-Louise Evans and Abrar Hussain.

Finally, I am extremely thankful for the support of my family: Isobel, Richard, Phoebe and Matt.

Contents

1	Introduction	1
1.1	Proteins and Peptides	1
1.1.1	Amino acids.....	1
1.1.2	The biological role of proteins.....	2
1.1.3	Non-covalent interactions	3
1.1.4	Structural elements	4
1.1.5	Protein-ligand interactions	6
1.2	Computational Studies of Polypeptides.....	8
1.2.1	Molecular dynamics simulations of polypeptides.....	9
1.2.2	Protein-ligand docking.....	10
1.2.3	QM-MM studies of polypeptides	11
1.3	Outline of Thesis	12
1.3.1	The molecules studied in this thesis.....	13
1.3.2	Content of Chapters	14
1.4	References	15
2	Computational Methods.....	19
2.1	Introduction	19
2.2	Molecular Docking	19
2.2.1	The AutoDock Search Strategy	20
2.2.2	Genetic Algorithms	21
2.2.3	AutoDock Scoring Function	22
2.3	Molecular Dynamics Simulations.....	24
2.3.1	Newton's laws of motion and the Verlet integrator	25
2.3.2	The CHARMM Force Field.....	27
2.3.3	Statistical Mechanics	29

2.3.4	Explicit Solvent Models and Periodic Boundary Conditions.....	31
2.3.5	The GBSW Implicit Solvent Model and Langevin Dynamics.....	33
2.4	Quantum Chemistry.....	37
2.4.1	Quantum Mechanics and the Schrödinger Equation	37
2.4.2	Hartree-Fock Methods.....	38
2.4.3	Density Functional Theory	39
2.4.4	Basis Sets.....	40
2.4.5	QM-MM and <i>Ab initio</i> Molecular Dynamics.....	42
2.4.6	Lagrangian Mechanics and CPMD	43
2.5	References	46
3	Conformational Preferences of Nisin Analogues	50
3.1	Introduction	50
3.2	The Nisin Killing Action.....	50
3.3	Disulfide Rich Peptides.....	52
3.4	Simulation Set-up.....	53
3.5	Results.....	56
3.5.1	Cysteine 3 – Cysteine 7 interactions.....	58
3.5.2	Cysteine 8 – Cysteine 11 interactions.....	61
3.6	Relative energy maps for bead and globular connectivities	64
3.7	Cluster Analysis of Nisin Analogue 12.....	67
3.8	Discussion.....	69
3.9	Conclusions	70
3.10	References.....	72
4	Extending the CHARMM Potential Energy Function for the Nisin Analogues	75
4.1	Introduction	75

4.2	The CMAP term and D-amino residues	75
4.2.1	Testing the D-CMAP term	78
4.2.2	D-CMAP Results.....	81
4.3	Impact of the CMAP term on simulations of the nisin analogues	86
4.3.1	Simulation set-up.....	86
4.3.2	Results.....	87
4.4	Reactive dynamics and CPMD simulations of the nisin analogues....	90
4.4.1	Simulation set-up.....	91
4.4.2	Results.....	93
4.5	Conclusions	96
4.6	References	97
5	Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein	100
5.1	Introduction	100
5.2	Structural and Physiological Properties of ILBP.....	100
5.2.1	Intracellular Lipid Binding Protein Family.....	100
5.2.2	ILBP and Bile Salts <i>in vivo</i>	102
5.3	Review of Experimental Evidence.....	103
5.4	Simulation Set-up.....	107
5.4.1	Initial Structures	107
5.4.2	MD Simulations.....	109
5.4.3	Docking Calculations.....	109
5.5	Results.....	111
5.5.1	Thermodynamic stability	111
5.5.2	Docking	111
5.5.3	MD Simulations from Docked Complexes	114
5.5.4	Covariance Analysis	118

5.5.5	Interaction between Gln51 and ligands	119
5.5.6	Orientation of ligands and helices	121
5.6	Discussion.....	125
5.6.1	Structural Differences of the Simulations	125
5.6.2	A Third Binding Site.....	126
5.6.3	Preferred Binding Site of Ligands and a Proposed Mechanism of Site Selectivity.....	127
5.7	Conclusions	128
5.8	References	129
6	Conclusion and Future Work	132
6.1	Conclusions	132
6.2	Major Contributions.....	137
6.3	Further Work.....	138
	Appendix - Transformed Parameter Set for Simulating D-amino Residues ...	140

1 Introduction

1.1 Proteins and Peptides

1.1.1 Amino acids

Proteins and peptides are synthesised in living organisms by connecting together a sequence of monomer units from the 20 standard amino acids that are specified by the genetic code of DNA. All of the standard amino acids have a chemical structure of a central carbon atom, C_{α} , bonded to an amino group (NH_2), a carboxyl group ($COOH$) and a third chemical group, referred to as the side chain. It is this third chemical group that distinguishes the 20 amino acids from each other and they can be grouped together according to the side chain's chemical properties: hydrophobic side chains – alanine (Ala), valine (Val), phenylalanine (Phe), proline (Pro) and methionine (Met); charged side chains – aspartic acid (Asp), glutamic acid (Glu), lysine (Lys) and arginine (Arg); polar side chains – serine (Ser), threonine (Thr), tyrosine (Tyr), histidine (His), cysteine (Cys), asparagine (Asn), glutamine (Gln) and tryptophan (Trp); and glycine (Gly), which is usually considered by itself. Amino acids are chiral molecules and only L-amino acids are selected during biosynthesis.

The amino acids are joined together via a peptide bond that forms between the carboxyl group of one amino acid and the amino group of another, releasing water in the reaction. Once in a peptide bond the monomer units are referred to as amino residues. The repeating, central chain of $N-C_{\alpha}-C-N$ is called the backbone, and because of the partial-double bond nature of the peptide bond, the backbone has limited rotational freedom. Conventionally, the amino acid sequence is counted from the N-terminus to the C-terminus. Over 700 non-standard amino residues are known to biochemistry [1]. These can either be amino acids that have a side chain with an altered chemical group or D-amino residues and both types of modification are made by enzymes following ribosomic translation. Polypeptides are usually called proteins when they have over 50 residues and specific secondary structure domains and peptide is generally used for shorter sequences that are more

1. Introduction

flexible. However, there is no set definition and peptides often have secondary structure elements.

1.1.2 The biological role of proteins

There are at least 30,000 different proteins and peptides produced by human cells alone [2] and they perform a number of essential biological roles. Important classes of proteins include [3]: structural proteins, such as keratins, that are constituent parts of hair and skin; enzymes that catalyse biological reactions so they can take place under *in vivo* conditions; antibodies that recognise foreign pathogens; regulatory enzymes that control the transcription of genes by binding to DNA; sensors that detect molecular signals in the body; transporters and pumps that control the diffusion of molecules across cell membranes and transducer proteins that convert stored chemical energy into mechanical energy. Because proteins are essential for so many biological processes they are the binding target of small drug molecules that either enhance or inhibit their action. The development of new drugs is based upon identifying protein binding sites and molecules that will interact with them.

The biosynthesis of proteins consists of three steps that are essentially the same across all classes of organism [1]. Initially a complementary sequence of RNA is transcribed from a transcription unit sequence of DNA. The RNA undergoes a number of molecular processes mediated by enzymes, such as the addition of a cap to allow recognition by the ribosome, to become messenger RNA (mRNA). The mature mRNA is read by the ribosome and a corresponding polypeptide chain is synthesised in conjunction with transfer RNA (tRNA), accessory enzymes and other proteins and energy-providing molecules like ATP. Some peptides, such as glutathione, the primary antioxidant molecule in eukaryotes [4], are not produced by translation by the ribosome, but are instead synthesised by a series of enzyme catalysed reactions.

1. Introduction

1.1.3 Non-covalent interactions

Non-covalent interactions control the shape that the linear polypeptide chain adopts and many of the biological functions that proteins and peptides perform. From Coulomb's Law, the magnitude of the attractive or repulsive force between two charged particles is inversely proportional to the square of the product of their separation and the relative dielectric constant of the medium in which the particles are embedded. In an aqueous environment the relative dielectric constant is about 80, meaning the strength of the electrostatic interactions is weak. In the solvent-excluded interior of a folded protein, the relative dielectric is 2-4, so the electrostatic interactions are strong, but it is unusual for a charged group to enter the interior region because of the energy cost associated with overcoming solvation. Salt bridges can form between the negatively charged side chains of aspartic acid or glutamic acid and the positively charged side chains of lysine and arginine.

Hydrogen bonding is important as secondary structure is stabilised by hydrogen bonds and they mediate many protein recognition events, such as the binding of ligands. A hydrogen bond is an interaction between two dipoles. An electronegative atom, usually oxygen or nitrogen in biochemistry, attracts the electrons from a covalently bonded hydrogen toward it, becoming slightly negative and the hydrogen slightly positive. The positive hydrogen is attracted to a second electronegative atom, such as the oxygen of a carbonyl group, inducing and reinforcing the dipoles in both chemical groups (Figure 1-1).

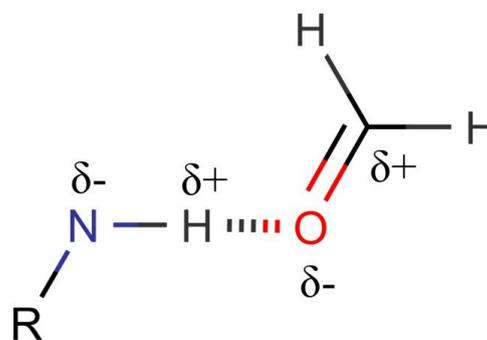


Figure 1-1 Hydrogen bonding.

1. Introduction

The polar nature of water molecules is responsible for the hydrophobic effect, which was first identified as being important to the final conformations of most proteins by Kauzmann in 1959 [5]. The peptide backbone and the side chains of polar and charged amino residues are soluble in water because they form hydrogen bonds with the water. Because the side chains of hydrophobic residues cannot form hydrogen bonds, the surrounding water molecules must form clathrate structures around these residues, which has an entropic cost. The total free energy of the structure is reduced if hydrophobic groups come together to exclude water, reducing the solvent accessible surface area and therefore the energetic cost. Hydrophobic interactions are now thought to be the dominant component of protein folding. Evidence for this includes [6]: many proteins have hydrophobic cores that exclude water; there is a 1-2 kcal mol⁻¹ cost associated with every hydrophobic residue; non-polar solvents easily denature proteins; and sequence that are mutated but retain the order of hydrophobic and polar residues still adopt the correct native state.

1.1.4 Structural elements

Protein structure is organised into four levels: the primary structure, which is the amino acid sequence; the secondary structure, such helices and sheets stabilised by hydrogen bonding; the tertiary structure, made up of arrangements of secondary structure elements; and the quaternary structure formed from the aggregation of individual polypeptide chains.

The peptide bond has a partial double bond nature and favours the *trans* conformation to residue steric interactions. The energetic cost of isomerisation to the *cis* form is ~20 kcal mol⁻¹, although this is lower for the residue proline and in some hydrophobic environments. Therefore, because of the limited rotation about the peptide bond, the conformations of polypeptide chains are characterised by the dihedral angle about the N-C_α bond, conventionally named as the φ angle, and the dihedral about the C_α-C bond, known as the ψ angle. Because of steric clashes of the side chains, most combinations of φ and ψ are disallowed. The pairs of angles that are allowed

1. Introduction

were calculated by Ramachandran [7] and plotted as a function of φ and ψ on a diagram known as the Ramachandran plot (Figure 1-2).

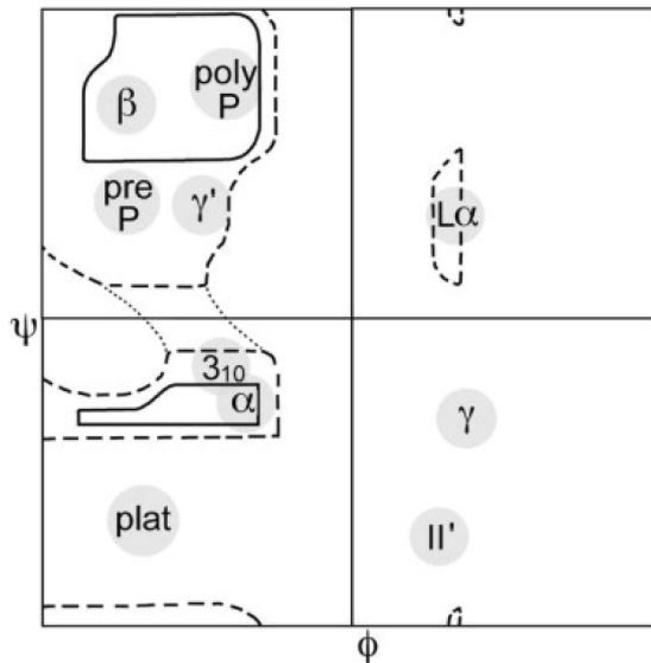


Figure 1-2 The classic Ramachandran plot from [8] after [7]. Regions marked on the plot correspond to α -helices (α), β -sheets (β), left-hand α -helices ($L\alpha$), 3_{10} helices, pre-proline residues which have more limited values (pre P), poly proline regions which also have more limited values (poly P), type-II' β -turns (II'), γ turns (γ) and γ' turns (γ').

The two primary units of protein secondary structure are the α -helix and the β -sheet, predicted by Linus Pauling in 1951 based on X-ray crystal structures of peptides and his theory of chemical bonding [9]. The α -helix is formed from consecutive residues with (φ, ψ) angles of approximately $(-60^\circ, -50^\circ)$ [10], stabilised by hydrogen bonds between the backbone carbonyl group of the i^{th} residue and the backbone amino group of the $(i + 4)^{\text{th}}$ residue. There are 3.6 residues per turn in an α -helix and each residue corresponds to a rise of approximately 1.5 Å along the central axis. Because all the hydrogen bonds are aligned in the direction of the central axis, the dipole moment of each bond is also aligned. This leads to the α -helix having a net dipole, that can attract negative ligands to bind to the N-terminus of the helix. Other types of helix are possible, such as the more loosely coiled π -helix that forms hydrogen bonds with the $(i + 5)^{\text{th}}$ residue and the tightly coiled 3_{10} -helix that forms

1. Introduction

hydrogen bonds with the $(i + 3)^{\text{th}}$ residue, but neither is as energetically favourable and structurally stable as the α -helix [10].

β -sheets are formed from extended peptide chains with (φ, ψ) values centred around $(-135^\circ, 135^\circ)$, but can adopt any angle in the large region of the upper-left hand quadrant of the Ramachandran plot. The sheets are formed by hydrogen bonds between the backbone amino groups of one strand and the carbonyl group of another. β -sheets can be described as either parallel or antiparallel, depending if they are aligned N-terminus to N-terminus or N-terminus to C-terminus. Unlike α -helices, β -sheets can be constructed from residues that are non-consecutive and in different polypeptide chains. β -turns are defined as four consecutive residues where the distance between the i^{th} residue and the $(i + 3)^{\text{th}}$ residue is less than 7 Å [11]. They are categorised into four subgroups based upon the φ and ψ angles of the two central residues and are sometimes stabilised by a hydrogen bond between the carbonyl group of the first residue and the amino group of the fourth [10]. In proteins β -turns are associated with anti-parallel β -sheets, where they frequently form the link between two adjacent strands, but in peptides they are often identified as isolated structural elements.

1.1.5 Protein-ligand interactions

The binding of certain ligands to proteins is fundamental to the biological processes that take place in organisms. Examples include the binding of enzymes to substrates, oxygen to haem, hormones to hormone receptors and protein transcription factors to nucleotide chains. The ligands can be single atoms or ions, small molecules or other macromolecules such as a peptide, another protein or a nucleotide chain. The ability of a specific ligand to bind to a protein in the cytoplasmic environment, despite the presence of many other molecules with similar properties, is called molecular recognition.

The first model of protein-ligand binding, known as the lock and key theory, was proposed by Fischer in the 1890's [12]. This model was developed to describe enzyme-substrate binding and it assumed that the enzyme binding site was rigid and the geometric complement of its substrate. However, this

1. Introduction

was an incomplete picture, as proteins are flexible molecules which lead to the induced fit model of Koshland that took account of conformation changes upon ligand binding. These changes in conformation have been shown to improve ligand specificity and are a conformational 'proofreading' mechanism that may be evolutionarily selected [13]. The current model of protein ligand binding is the population-shift theory that takes into account the dynamic nature of proteins. Developed from the 1960's MWC model of Monod, Wyman and Changeux [14], the unbound native state of a protein is not a single conformation but an ensemble of low energy states that have small energy barriers between them. The unbound protein ensemble samples some of the same conformations as the bound protein, but with a lower probability because they have a higher energy. Upon binding, the energy of the bound conformations is lowered, so they are sampled more by the ensemble [15]; hence the population of states shifts toward favouring bound conformations.

Allostery is the term used to describe the process by which a chemical event in one part of the molecule has an effect on another part of the molecule. Allostery is an important feature of protein-ligand interactions as the binding of a ligand often induces a conformational change in the protein linked to changes in biological activity. For example, the anti-cancer drug methotrexate binds to the enzyme dihydro-folate reductase (which is important to cell division), inducing a conformational change in a part of the protein that does not contain catalytic residues, but still inhibits the proteins biological function due to small changes transmitted to the substrate binding site [15]. Under the population shift paradigm, ligand binding induces conformational changes that begin at the binding site and are propagated to another part of the protein.

Cooperative binding is a special case of an allosteric interaction, where binding of a ligand promotes a conformation change that increases the affinity for binding the same ligand again. The most well known example of a cooperative binding system in biology is haemoglobin. Haemoglobin regulates the transport of O₂ in the body, changing its binding preference depending on

1. Introduction

its location in the body and showing strong positive cooperativity for oxygen molecules. It has a quaternary structure of four subunits, each with a haem group binding site that has an iron cation at its centre. Haemoglobin has two distinct functional conformations: deoxyhaemoglobin (deoxy-Hb) which has a low affinity for oxygen and oxy-haemoglobin (oxy-Hb) which has an affinity 150-300 times higher [15]. Perutz [16] suggested a model of haemoglobin that was the first to connect allostery with cooperativity: when an oxygen binds to one haem group of deoxy-Hb in the high oxygen concentration environment of the lungs, this induces conformational changes in the other three subunits so the protein becomes oxy-Hb; in the extremities of the body the oxygen concentration is low and the haemoglobin shifts back to deoxy-Hb, releasing oxygen [15] where it is needed.

1.2 Computational Studies of Polypeptides

The biological function of a protein depends on its structure and dynamics. The structure of proteins can be experimentally determined by X-ray crystallography and NMR spectroscopy. In protein X-ray crystallography a diffraction pattern is produced by scattering X-rays from a crystal. X-ray crystallography is a good method for determining the structure of a molecule because it measures the bond lengths and bond angles but has some disadvantages: hydrogen atoms cannot deflect X-rays because their electron density is too low and must be added to an X-ray structure of a protein after the model has been resolved; very high quality crystals of purified protein are required but cannot be grown for some proteins and the data collection and analysis for a protein structure often takes weeks or months. X-ray crystallography also has the disadvantage that it cannot capture the dynamics of proteins in solvent, which is often very important for biological function. Protein NMR spectroscopy uses a high strength magnetic field and radio wave pulses to assign spectral peaks to each residue and then detects the coupling of nuclear spins of ^1H , ^{13}C and ^{15}N to identify the atoms that are spatially, but not sequentially close. This coupling information is used to build a model of the structure of the protein. NMR is a faster technique than x-ray but the

1. Introduction

models produced are more open to interpretation due to the refinement process. Computational studies of polypeptides complement and extend the experimental results obtained from NMR and X-ray studies. The three computational methods used in this thesis are molecular dynamics simulations (MD), docking and QM-MM methods. An introduction to the applications of these methods is given in the rest of this section and specific details of the theory and algorithms used to perform the calculations are explained in Chapter 2.

1.2.1 Molecular dynamics simulations of polypeptides

Molecular dynamics (MD) is a method for propagating a system of many atoms through time by numerically integrating Newton's laws of motion. The first MD simulation of a polypeptide, in 1976, was of the bovine pancreatic trypsin inhibitor (BPTI), a 58 residue peptide that inhibits the action of the digestive enzyme trypsin [17] [18]. The size of the molecular systems studied has grown with available computational power and the first simulation containing over one million atoms was performed in 2007 [19]. Recently a supercomputer with architecture dedicated to MD simulations of proteins has been developed [20] to perform the first millisecond scale simulation. Two of the main advantages that MD offers in the study of polypeptides is that it can study atomic movements and interactions in detail, and on a time scale, that is not accessible to experimental methods and that non-physical processes can be modelled to reveal information about a physical system, such as the computational alchemy used to calculate the free energy difference between two similar, but different, molecular systems [21]. However, a disadvantage is that the accuracy of the simulation is dependent on the quality of the empirical force field parameter set used to calculate the changes in potential energy of the molecular system.

MD has been used extensively to study protein folding because it provides atomic-level information about equilibria and transition states. The Trp-cage miniprotein [22] is an artificial 20 residue peptide identified as being useful for testing folding simulations. It has secondary structure elements of an eight

1. Introduction

residue α -helix linked to a short, four residue, 3_{10} helix. The folding time of approximately 4 μ s was accurately predicted by implicit solvent simulations [23] and two different folding pathways were demonstrated by an explicit solvent transition state sampling method [24], in simulations that were the first to show a protein fold from a denatured state into a native conformation [25]. Other peptides that have been widely studied are the chicken villin headpiece, which was sampled during tens of thousands of independent trajectories as part of the Folding@Home project [26], and is described as '*excellent model system for the folding of small α -helical proteins*' [25] and the human Pin WW domain, which is a similar model for the folding of β -sheets [27]. These peptides are 35 and 40 residues long respectively and it remains a major challenge to simulate the folding of longer, complete proteins with over 100 residues. The implementation of GPU processors is likely to increase the timescales accessible, but folding simulations may still be limited by the quality of the empirical force fields, which have been parameterised to model folded proteins, not denatured states or intermediates [25].

MD simulations have been successfully used to determine the allosteric transitional pathway of a cooperative binding systems. One example is the protein GroEL, a chaperonin which chaperones the folding of other proteins within the cytoplasm of *Escherichia coli* [21]. It has a structure of 14-subunits, arranged into two rings and cooperatively binds ATP and a second chaperonin with each ring, but anti-cooperatively between the rings. The binding of the ligands to one ring induces changes in all parts of the protein to move from a closed, unbound conformation to an open, bound conformation. The path of this transformation is impossible to measure experimentally [28], but the transition states have been identified using targeted MD simulations [29].

1.2.2 Protein-ligand docking

Protein-ligand docking is an essential tool for computer aided drug design and is concerned with identifying the binding modes of a ligand with a protein with a known structure. Docking programs generally have two components, a search algorithm that generates the conformation of protein-ligand

1. Introduction

complexes and a scoring function that ranks the conformations, usually based on a binding energy calculated using an empirical potential energy function. The first docking program was published in 1982 [30] and treated both the protein and the ligand as rigid bodies. Modern docking algorithms (see Section 2.2) usually treat the ligand as being flexible but allowing flexibility of the protein remains a challenge due to the computational expense involved [31]. This limits the abilities of docking algorithms as, in accordance with the population shift model, proteins change conformation upon ligand binding. However, docking is still a useful technique as part of virtual screening, where it is used to rapidly identify a subset of drug-like molecules, that have favourable steric and electrostatic interactions with a target protein, from databases of compounds [32]. These leads can be studied further using higher resolution computational methods or with experimental techniques.

An example of a priority drug lead identified through virtual screening is an inhibitor of protein kinase CK2. Protein kinase CK2 processes hundreds of protein substrates and, whilst its exact function is not understood, it has a role in neoplastic growth, making it a target for anticancer drugs. Vangrevelinghe *et al.* [33] used the docking program DOCK [34] to screen a corporate library of over 400000 compounds and identified a novel lead that inhibits protein kinase CK2 which is now undergoing pharmaceutical development.

1.2.3 QM-MM studies of polypeptides

The methods discussed in the previous sections, MD simulations and docking, both use classical, Newtonian potential energy functions, calculated using empirically derived parameter sets. To accurately describe reactive processes such as the formation or breaking of covalent bonds or catalysis, electrons must be explicitly included in the simulation. However, electrons can only be accurately described by quantum mechanics (QM), which takes into account the wave-like properties of particles. QM calculations are much more computationally expensive than classical calculations so are generally used to model smaller, gas-phase molecules. However, this is unsuitable for studying

1. Introduction

polypeptides, as such model systems do not take account of electrostatic and steric interactions with the surrounding environment. One method to model reactions in polypeptides is to use *ab initio* MD where the forces for the dynamics are calculated using QM techniques. Another method is to use a hybrid calculation where QM is used to describe the reactive centre and classical molecular mechanics (MM) is used to describe the rest of the atoms in the system (see Section 2.4.5).

The QM-MM hybrid methods were originally developed specifically for studying biomolecular systems and the first example was a study of an enzymatic reaction in 1976 [35]. They have also been used to model other chemical processes in polypeptides, including calculating spectroscopic properties, excited state dynamics and electron transfer [36]. QM-MM simulations have, for example, modelled the excited state dynamics of rhodopsin, a pigment protein in the eye, showing that its isomerisation is the first event that initiates vision [37]. They have also been used to successfully explain the experimental recyclicalisation rate of a cysteine rich tetrapeptide when the stabilising disulfide bond is cleaved by a photon [38].

1.3 Outline of Thesis

In this thesis MD simulations, in conjunction with the complementary methods of docking and QM-MM, are used, and further developed, to study two unusual polypeptide systems: characterisation of the conformational preferences of isomers of an antibiotic peptide and understanding the binding behaviour of a human transporter protein. The antibiotic peptides are analogues of an antibacterial, called nisin, produced by a class of bacteria and its biological function is dependent on the formation of five unusual macrocyclic rings, each joined by a thioether bond between modified L-amino and D-amino residues. For the analogues the thioether bond is replaced by a disulfide bond between cysteine residues and the chirality of the cysteines altered. D-amino residues are rare in biology and the standard CHARMM empirical potential energy function must be modified to correctly describe them. Cysteine rich peptides are produced by all classes of organism, but the

1. Introduction

beaded connectivity of the nisin analogues is extremely unusual and has only been synthesised artificially and never observed in nature [39]. The ileal lipid binding protein (ILBP) is also an unusual polypeptide as it has been shown, in NMR experiments using isotopical labelled ligands [40], that one ligand will completely displace another from a binding site, despite both sites having an affinity for each ligand type and the ligands only differing by a single hydroxyl group. Using MD simulations to study this site selectivity behaviour makes use of one of the advantages of computational methods, because it is possible to simulate the unphysical protein-ligand combination where the incorrect ligands are in each binding site.

1.3.1 The molecules studied in this thesis

Nisin is a naturally produced antibiotic peptide belonging to a class of compounds called lantibiotics, which are secreted by a family of gram-positive bacteria and are effective against other gram-positive bacteria, as part of the microbial arms race. Nisin has two killing actions, both of which depend on its ability to bind to lipid-II, a precursor of the peptidoglycan that forms the cell walls of gram-positive bacteria. The first, at high concentration, is that eight nisin molecules and four lipid-II molecules form a pore complex that spans the cytoplasmic membrane and allows the contents of the bacterial cell to diffuse out [41]. The second, at low concentrations, is that by binding to lipid-II, nisin prevents it from performing its role in building the cell wall, stopping bacterial proliferation by cell division [42]. Nisin has a sequence of 34 amino residues and no specific secondary structure, but is conformationally stabilised by five macrocyclic rings. A complex of three enzymes modifies the chirality and side chains of several residues and closes the rings by thioether bonds between the modified residues and methionines. Nisin binds to lipid-II via hydrogen bonds between the peptide backbone of the first two rings and a pyrophosphate group [43]. Analogues of the first twelve residues of nisin, where the rings are closed by disulfide bonds between two pairs of cysteine residues, are studied in this thesis. The chirality of the four cysteines is changed from L to D in all possible combinations resulting in 16 analogues in total. The cysteine analogues are expected to have an antibiotic effect the

1. Introduction

same as the second nisin killing action, but would be easier to synthesise *in vitro*.

The second molecular system studied in this thesis is ILBP and its bile salt ligands. Bile salts are synthesised in the liver from cholesterol. The two predominant human bile salts are conjugates of cholic acid and chenodeoxycholic acid. They are released into the gut during digestion where they have a detergent effect that aids the break down of fatty foods. At the end of the digestive tract, in the ileum, they bind to ILBP, which aids their reabsorption through the intestinal wall and they are recycled back into the liver, suppressing the conversion of more cholesterol into bile salts. Any bile salts that are not reabsorbed are excreted from the body. This means that ILBP is of pharmacological interest, because if its binding action could be inhibited, more cholesterol would be converted to bile salts, lowering body cholesterol levels. ILBP belongs to the intracellular lipid binding protein family, which have the a tertiary structure of a ten strand β -barrel capped with two α -helices, with one or two ligands bound inside the β -barrel. ILBP binds at least two ligands within its binding cavity and it has been shown, using NMR experiments [40], that cholic acid and chenodeoxycholic acid bind specifically to different binding sites despite only differing by one OH group and that each binding site has an affinity for both bile salts. It has also been shown experimentally that ILBP binds bile salts cooperatively in a way that is comparable to haemoglobin [44] and that site selectivity and cooperativity are not functionally linked [45].

1.3.2 Content of Chapters

The computational methods used in the thesis are explained in detail in Chapter 2. Chapter 3 and Chapter 4 focus on the cysteine analogues of nisin. In Chapter 3, implicit solvent simulations are performed on the set of 16 analogues to examine how chirality affects the conformational preferences that lead to sulfur-sulfur interactions. In the first part of Chapter 4 the corrective term of the CHARMM potential energy function is extended to include proteins containing D-amino residues and tested on X-ray

1. Introduction

crystallographic structures from the Protein Data Bank and the nisin analogues. In the second part of the chapter, potential energy surfaces, that are a function of the inter-sulfur distances of the nisin analogues, are calculated using QM-MM methods.

It has been shown experimentally that ILBP binds cooperatively and has specific site selectivity for ligands that vary by a single OH group, despite both interior binding sites having an affinity for both ligands. In Chapter 5 MD simulations and docking calculations are used to understand the allosteric role of an unspecific exterior binding site in cooperativity and an energetic mechanism for site selectivity is proposed. Chapter 6 contains concluding remarks and summarises the major contributions of the thesis and finishes with ideas for further study arising from the work.

1.4 References

1. Zorko, M., *Amino Acids*, in *Introduction to Peptides and Proteins*, U. Langel, et al., Editors. 2010, CRC Press: Boca Raton.
2. Graslund, A., *Misfolding-based diseases*, in *Introduction to Peptides and Proteins*, U. Langel, et al., Editors. 2010, CRC Press: Boca Raton.
3. Lesk, A.M., *Introduction to protein science*. 2nd ed. 2010, Oxford: Univeristy Press.
4. Zorko, M., *The Biosynthesis of Proteins*, in *Introduction to Peptides and Proteins*, U. Langel, et al., Editors. 2010, CRC Press: Boca Raton.
5. Kauzmann, W., *Some Factors in the Interpretation of Protein Denaturation*, in *Advances in Protein Chemistry*, C.B. Anfinsen, et al., Editors. 1959, Academic Press. p. 1-63.
6. Dill, K.A., S.B. Ozkan, M.S. Shell, and T.R. Weikl, *The Protein Folding Problem*. *Ann. Rev. Biophys.*, 2008. **37**(1): p. 289-316.
7. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations*. *J. Mol. Biol.*, 1963. **7**(1): p. 95-99.
8. Lovell, S.C., I.W. Davis, W.B. Arendall, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, and D.C. Richardson, *Structure validation by $C\alpha$ geometry: ϕ, ψ and $C\beta$ deviation*. *Proteins*, 2003. **50**(3): p. 437-450.
9. Eisenberg, D., *The discovery of the α -helix and β -sheet, the principal structural features of proteins*. *Proc. Natl. Acad. Sci. USA*, 2003. **100**(20): p. 11207-11210.
10. Branden, C. and J. Tooze, *Introduction to protein structure*. 2nd ed. 1999, New York: Garland Science.

1. Introduction

11. Hutchinson, E.G. and J.M. Thornton, *PROMOTIF - A program to identify and analyze structural motifs in proteins*. Protein Sci., 1995. **5**: p. 200-212.
12. Fischer, E., *Einfluss der configuration auf die wirkung der enzyme*. Ber. Dt. Chem. Ges, 1894. **27**: p. 2985-2993.
13. Savir, Y. and T. Tlusty, *Conformational Proofreading: the impact of conformational changes on the specificity of molecular recognition*. PloS ONE, 2007. **2**(5).
14. Monod, J., J. Wyman, and J.-P. Changeux, *On the nature of allosteric transitions: A plausible model*. J. Mol. Biol., 1965. **12**(1): p. 88-118.
15. Kessel, A. and N. Ben-Tal, *Introduction to proteins; structure function and motion*. 2011, Boca Raton: CRC Press.
16. Perutz, M.F., *Stereochemistry of Cooperative Effects in Haemoglobin: Haem-Haem Interaction and the Problem of Allostery*. Nature, 1970. **228**(5273): p. 726-734.
17. McCammon, J.A., *Molecular Dynamics of the bovine pancreatic trypsin inhibitor*, in CECAM. 1976: Orsay, France. p. 137.
18. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins*. Nature, 1977. **267**: p. 585-90.
19. Chandler, D.E., J. Hisin, C.B. Harrison, J. Gumbart, and K. Schulten, *Intrinsic curvature properties of photosynthetic proteins in chromatophores*. Biophys. J., 2008. **95**: p. 2822-2836.
20. Shaw, D.E., M.M. Deneroff, R.O. Dror, J.S. Kuskin, R.H. Larson, J.K. Salmon, C. Young, B. Batson, K.J. Bowers, J.C. Chao, M.P. Eastwood, J. Gagliardo, J.P. Grossman, C.R. Ho, D.J. Ierardi, I. Kolossvary, J.L. Klepeis, T. Layman, C. McLeavey, M.A. Moraes, R. Mueller, E.C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S.C. Wang, *Anton, a special-purpose machine for molecular dynamics simulation*. SIGARCH Comput. Archit. News, 2007. **35**(2): p. 1-12.
21. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins*. Chem. Rev., 2006. **106**: p. 1589-1615.
22. Neidigh, J.W., R.M. Fesinmeyer, and N.H. Andersen, *Designing a 20-residue protein*. Nat Struct Mol Biol, 2002. **9**(6): p. 425-430.
23. Snow, C.D., B. Zagrovic, and V.S. Pande, *The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations*. J. Am. Chem. Soc., 2002. **124**(49): p. 14548-14549.
24. Juraszek, J. and P.G. Bolhuis, *Sampling the multiple folding mechanisms of Trp-cage in explicit solvent*. Proc. Natl. Acad. Sci. USA, 2006. **103**(43): p. 15859-15864.
25. Freddolino, P.L., C.B. Harrison, Y. Liu, and K. Schulten, *Challenges in protein-folding simulations*. Nat Phys, 2010. **6**(10): p. 751-758.
26. Jayachandran, G., V. Vishal, and V.S. Pande, *Using massively parallel simulations and Markovian models to study protein folding: Examining the dynamics of the villin headpiece*. J. Am. Chem. Soc., 2006. **124**: p. 164902.

1. Introduction

27. Jäger, M., H. Nguyen, J.C. Crane, J.W. Kelly, and M. Gruebele, *The folding mechanism of a β -sheet: the WW domain*. J. Mol. Biol., 2001. **311**(2): p. 373-393.
28. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules*. Nat. Struct. Biol., 2002. **9**(9): p. 646-652.
29. Ma, J., P.B. Sigler, Z. Xu, and M. Karplus, *A Dynamic Model for the Allosteric Mechanism of GroEL*. J. Mol. Biol., 2000. **302**(2): p. 303-313.
30. Kuntz, I.D., J.M. Blaney, S.J. Oatley, R. Langridge, and T.E. Ferrin, *A geometric approach to macromolecule-ligand interactions*. J. Mol. Biol., 1982. **161**(2): p. 269-288.
31. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *Protein–ligand docking: Current status and future challenges*. Proteins, 2006. **65**(1): p. 15-26.
32. Schulz-Gasch, T. and M. Stahl, *Binding site characteristics in structure-based virtual screening: evaluation of current docking tools*. J. Mol. Model, 2003. **9**(1): p. 47-57.
33. Vangrevelinghe, E., K. Zimmermann, J. Schoepfer, R. Portmann, D. Fabbro, and P. Furet, *Discovery of a Potent and Selective Protein Kinase CK2 Inhibitor by High-Throughput Docking*. J. Med. Chem., 2003. **46**(13): p. 2656-2662.
34. Kuntz, I.D., E.C. Meng, and B.K. Shoichet, *Structure-Based Molecular Design*. Accounts Chem. Res., 1994. **27**(5): p. 117-123.
35. Warshel, A. and M. Levitt, *Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*. J. Mol. Biol., 1976. **103**(2): p. 227-249.
36. Senn, H.M. and W. Thiel, *QM/MM Methods for Biomolecular Systems*. Angew. Chem. Int. Edit., 2009. **48**(7): p. 1198-1229.
37. Frutos, L.M., T. Andruniow, F. Santoro, N. Ferre, and M. Olivucci, *Tracking the excited-state time evolution of the visual pigment with multiconfigurational quantum chemistry*. Proc. Natl. Acad. Sci. USA, 2007. **104**(19): p. 7764-7769.
38. Nieber, H., A. Hellweg, and N.L. Doltsinis, *Recyclization rate of a photocleaved peptide from multiscale simulation*. J. Am. Chem. Soc., 2010. **132**: p. 1778-1779.
39. Bulaj, G. and A. Walewska, *Oxidative folding of single-stranded disulfide-rich peptides*, in *Oxidative folding of peptides and proteins*, J. Buchner and L. Moroder, Editors. 2009, RSC: Cambridge.
40. Tochtrop, G.P., G.T. DeKoster, D.F. Covey, and D.P. Cistola, *A single hydroxyl group governs ligand site selectivity in human ileal bile acid binding protein*. J. Am. Chem. Soc., 2004. **126**(35): p. 11024-11029.
41. Hasper, H.E., B. de Kruijff, and E. Breukink, *Assembly and stability of nisin-lipid II pores*. Biochemistry, 2004. **43**: p. 11567-11575.
42. Hyde, A.J., J. Parisot, A. McNichol, and B.B. Bonev, *Nisin-induced changes in Bacillus morphology suggest a paradigm of antibiotic action*. Proc. Natl. Acad. Sci. USA, 2006. **103**(52): p. 19896-19901.
43. Hsu, S.T.D., E. Breukink, E. Tischenko, M.A.G. Lutters, B. de Kruijff, R. Kaptein, A. Bonvin, and N.A.J. van Nuland, *The nisin-lipid II complex*

1. Introduction

reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. Nat. Struct. Mol. Biol., 2004. **11**(10): p. 963-967.

44. Tochtrop, G.P., J.L. Bruns, C.G. Tang, D.F. Covey, and D.P. Cistola, *Steroid ring hydroxylation patterns govern cooperativity in human bile acid binding protein.* Biochemistry, 2003. **42**(40): p. 11561-11567.
45. Toke, O., J.D. Monsey, G.T. DeKoster, G.P. Tochtrop, C.G. Tang, and D.P. Cistola, *Determinants of cooperativity and site selectivity in human ileal bile acid binding protein.* Biochemistry, 2006. **45**(3): p. 727-737.

2 Computational Methods

2.1 Introduction

Computational methods are an important group of techniques that are used to complement and extend experimental studies of polypeptides. This chapter describes the computational techniques used in this thesis.

Even a small molecular system of interest, such as a gas-phase peptide, could contain several hundred atoms and a larger system of biomolecules, such a solvated protein-DNA complex, could contain over a million atoms. To describe the behaviour of the electrons in the system correctly, they must be treated using quantum mechanics (QM), which takes into account the wave-like properties of sub-atomic particles. However, QM calculations are very computationally expensive and, until recent developments such as the linearly scalable ONETEP program [1], could not be routinely used to study systems with more than about one hundred atoms. In order to study larger systems a physical approximation can be made that assumes atoms, or small chemical groups, are charged masses joined by springs, which can be treated using classical mechanics. Computational methods that use this approximation are known as molecular mechanics (MM) methods. Hybrid methods, known as QM-MM calculations, also exist that allow a small part of a system, such as a reactive centre, to be treated using QM whilst the rest of the molecule, and the surrounding solvent, are described classically.

2.2 Molecular Docking

Docking is concerned with identifying a binding site between a target biomolecule and a ligand. The target molecule is usually a protein but can be a DNA strand or any other biomolecule, and the ligand can be a small drug molecule, a peptide or another protein. Docking is an important and widely used technique in drug discovery and is used for both lead identification (finding compounds to bind with the target protein) and lead optimisation (modifying the chemical groups of a lead molecule to improve its binding with

2. Computational Methods

the target protein) [2]. A docking program always has two components; the first searches for plausible structures of the bound complex by generating possible conformations and the second evaluates the binding energy in order to score the conformations [3]. To reduce the size of the possible search space and increase the computational speed, most docking programs treat the receptor as either a fixed body or allow only limited rotational freedom to the side chains. The ligand-protein complexes identified by docking can then be studied further by experimental techniques such as NMR or more physically rigorous computational methods such as molecular dynamics.

A recent review [2] identifies four main categories of algorithm used to generate the bound protein-ligand conformations. The first is fragment based methods where the ligand is split into fragments, important functional groups are docked independently then the entire ligand structure is built up. Popular programs that use this technique are Surflex [4], eHiTs [5] and FlexX [6]. The second are programs such as GOLD [7] and AutoDock [8] [9] that use genetic algorithms to search for docked conformations. The final two categories are Monte Carlo algorithm searches, as used by Glide [10], and shape complementarity methods as used by LigandFit [11]. In a systematic comparison [2] of these programs, AutoDock 4 [8] performed reasonably compared with other docking programs and was able to identify over 93% of protein-ligand pairs from the PDBbind database [12]. All docking results reported in this thesis were produced using AutoDock Vina [9], an update to AutoDock 4 that improves the accuracy of the predicted binding modes.

2.2.1 The AutoDock Search Strategy

The algorithm used to search the conformations in AutoDock is a genetic algorithm (GA) with local optimisation, that uses the Metropolis criterion [13] to accept a move at each step. GA's are described in the next section. The best conformations have the lowest score calculated using the scoring function described in Section 2.2.3. Local optimisation is applied at each time step of the GA to search the nearby phase-space for conformations with lower scores. The local optimisation algorithm is a quasi-Newton method [9]

2. Computational Methods

that uses the derivatives of the scoring function to decide the direction and size of the next step.

In common with other docking programs, AutoDock treats the majority of the receptor as a fixed body, but allows some movement of selected side chains. For example, if the approximate region of the binding site is known then side chains in that region are allowed to rotate around selected dihedrals chosen by the user. There is also the option of allowing flexibility of the ligand, through rotatable bonds, depending on the application of AutoDock. For high-throughput screening, where many thousands of potential ligand molecules are docked, a rigid ligand and receptor is preferable to save computational cost, but for a study involving a low number of ligands then the increased search space and physical accuracy of flexible ligands is preferable.

To allow rapid evaluation of the scoring function of a particular ligand conformation, AutoDock uses a pre-calculated grid. Before the search algorithm is run the protein is embedded in a regular three-dimensional grid and the value of the scoring function for each grid point is calculated for each atom type in the ligand. During the search the score of a ligand conformation is evaluated using an interpolation of the eight nearest grid points for each atom.

2.2.2 Genetic Algorithms

The AutoDock search strategy uses a Lamarckian GA to generate low scoring protein-ligand binding conformations. GA's were first widely used for optimisation in the 1970's [14]. They are based upon the idea of using an analogy of Darwinian evolution to create a population of solutions to a defined problem.

In the AutoDock representation [8], the chromosome is a string of real-valued numbers: three Cartesian coordinates for the ligand translation, four coordinates to specify the ligand orientation and a value for each rotatable torsion angle. The corresponding phenotype is the fitness score calculated using the scoring function. The chromosomes of the initial population of

2. Computational Methods

solutions are filled with random values from the allowed minima and maxima of the search grid.

To evolve the solutions the current population is used to produce new solutions, which are then selected based on their fitness. This is defined by their score and those with the lowest score are considered fittest and are chosen for the next generation of solutions. The new solutions are produced from current solutions using operators called crossover and mutation. In crossover, two parent solutions are selected with a probability inversely proportional to their fitness score, and their chromosomes randomly broken and the parts swapped. For example [8], two parent solutions *ABC* and *abc* produce two new offspring solutions *aBc* and *AbC*. A mutation operation is then applied where a random value, drawn from a Cauchy distribution, is added to the variables in the chromosome. A Lamarckian GA also has an additional local search operation following the crossover and mutation operators. The search is complete when all the solutions in the population are converged within a certain tolerance of the mean value or when a maximum number of generations has been reached.

GA's are considered a good choice of search algorithm for search-spaces that have many local minima or for exploring a search space where the potential energy surface is unknown. This is because the crossover and mutation operators allow for movement into new parts of the search space without having to cross barriers.

2.2.3 AutoDock Scoring Function

The AutoDock Vina scoring function is an empirical scoring function. The functional form is based upon the X-CSCORE function [15] and parameters for the atom types, T , and weighting factors, W , were fitted to reproduce docked conformations from the PDBbind dataset [12].

The conformation dependent part of the score, c , is the sum over all pairs of atoms, i and j , of the empirical energy function h_{ij} :

(2-1)

$$c = \sum_{i < j} h_{ij}(d_{ij})$$

h_{ij} is a function of the surface distance, d_{ij} , between atoms i and j which is the distance in Ångströms between the van der Waals surfaces of each atom as defined by their van der Waals radius, R_T .

(2-2)

$$d_{ij} = r_{ij} - R_{Ti} - R_{Tj}$$

h_{ij} is a weighted sum, given in Equation (2-3). The individual components of h_{ij} are given in Equations (2-4) to (2-8).

(2-3)

$$h_{ij} = W_1 G_1 + W_2 G_2 + W_3 R + W_4 Hy + W_5 Hb$$

(2-4)

$$G_1 = \exp\left(-\left(\frac{d_{ij}}{0.5}\right)^2\right)$$

(2-5)

$$G_2 = \exp\left(-\left(\frac{d_{ij} - 3}{2}\right)^2\right)$$

(2-6)

$$R = \begin{cases} d_{ij}^2 & \text{if } d_{ij} < 0 \\ 0 & \text{if } d_{ij} \geq 0 \end{cases}$$

(2-7)

$$Hy = \begin{cases} 1 & \text{if } d_{ij} < 0.5 \\ 0 & \text{if } d_{ij} > 1.5 \end{cases}$$

(2-8)

$$Hb = \begin{cases} 1 & \text{if } d_{ij} < -0.7 \\ 0 & \text{if } d_{ij} > 0 \end{cases}$$

All the components of h_{ij} have a cut-off at $r_{ij} = 8 \text{ \AA}$. H_y and H_b are the hydrophobic and hydrogen bonding terms. They are only applied if the atom types are hydrophobic or form a donor-acceptor pair. These functions linearly interpolate between $0.5 < d_{ij} < 1.5$ and $-0.7 < d_{ij} < 0$ respectively.

The final docking score, s , for a conformation is given by

(2-9)

$$s = g(c_1 - c_{intra})$$

where c_1 is the score of the best conformation in the current iteration, c_{intra} is calculated for the intramolecular interactions of the conformation and g is the function

(2-10)

$$g(c) = \frac{c}{1 + W_r N_{rot}}$$

where N_{rot} is the number of rotatable bonds and W_r is the associated weighting constant.

2.3 Molecular Dynamics Simulations

Molecular dynamics (MD) simulation is a numerical method that simulates the propagation of a system with many particles through time based on Newton's laws of motion. Several MD packages are available for simulating biochemical molecules such as proteins, small ligands, nucleic acids and sugars. Some of the most popular programs are AMBER [16], NAMD [17], LAMMPS [18], CHARMM [19] [20] and GROMACS [21] [22]. Unless otherwise stated, the work reported here uses the CHARMM and NAMD packages with the CHARMM force field [23].

2. Computational Methods

The physical simplification of MM leads to some limitations. MD simulations are dependent upon on the empirically derived potential energy function that is used to calculate the interactions between atoms. The values in the parameter sets have been carefully chosen [24], but their empirical origin may still lead to unphysical behaviour, such as the tendency of protein simulations using older CHARMM parameters to oversample π -helices [25]. Another limitation that results from the classical assumption is that without explicit electrons it is not possible to model the dynamics of reactions and the making and breaking of covalent bonds.

2.3.1 Newton's laws of motion and the Verlet integrator

It is not possible to solve a multibody system of three or more interacting objects analytically using Newton's laws of motion. However, if the initial position and velocities of each particle in the system are known, or can be assumed, numerical integration of Newton's laws can be used to propagate the system forward in time in small discretised steps, Δt .

The most commonly used integrator is based upon the equations of Verlet [26]. It is derived from two Taylor expansions about the position of the i th particle, \vec{r}_i , at time $t + \Delta t$ and time $t - \Delta t$:

(2-11)

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \frac{d\vec{r}_i(t)}{dt} \Delta t + \frac{d^2\vec{r}_i(t)}{2dt^2} \Delta t^2 + \frac{d^3\vec{r}_i(t)}{6dt^3} \Delta t^3 + O(\Delta t^4)$$

(2-12)

$$\vec{r}_i(t - \Delta t) = \vec{r}_i(t) - \frac{d\vec{r}_i(t)}{dt} \Delta t + \frac{d^2\vec{r}_i(t)}{2dt^2} \Delta t^2 - \frac{d^3\vec{r}_i(t)}{6dt^3} \Delta t^3 + O(\Delta t^4).$$

From Newton's Second law

(2-13)

$$\frac{d^2\vec{r}_i(t)}{dt^2} = \frac{\vec{F}_i(t)}{m_i}$$

2. Computational Methods

where m_i is the mass of particle i and \vec{F}_i is the force acting on it from the other particles in the system. Adding together Equations (2-11) and (2-12) and substituting Equation (2-13) gives:

(2-14)

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\vec{F}_i(t)}{m_i} \Delta t^2 + O(\Delta t^4).$$

In this algorithm velocities are not calculated explicitly, but the velocity of particle i at time t is calculated from Equation (2-15). The positions at subsequent time steps are given by Equation (2-16)

(2-15)

$$\vec{v}_i(t) = \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t)}{2\Delta t}$$

(2-16)

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t.$$

An extension of the Verlet algorithm is the “leap frog” integrator. In this algorithm the velocities are calculated at $t = \frac{1}{2}\Delta t$ (Equation (2-17)) and then used to calculate the positions at $t = \Delta t$ (Equation (2-18)) so that the positions ‘leap’ over the velocities and the velocities ‘leap’ over the positions. In this algorithm the velocities are explicitly calculated but they are not calculated at the same time as the positions. Equation (2-19) can be used to approximate the velocities at the same time as the positions.

(2-17)

$$\vec{v}_i\left(t + \frac{1}{2}\Delta t\right) = \vec{v}_i\left(t - \frac{1}{2}\Delta t\right) + \frac{\vec{F}_i(t)}{m_i} \Delta t$$

(2-18)

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i\left(t + \frac{1}{2}\Delta t\right)\Delta t$$

(2-19)

$$\vec{v}_i(t) = \frac{1}{2}\left(\vec{v}_i\left(t - \frac{1}{2}\Delta t\right) + \vec{v}_i\left(t + \frac{1}{2}\Delta t\right)\right)$$

The size of the time step of integration for MD simulations should be a tenth of the fastest mode of vibration. Hydrogen atoms vibrate with a frequency of $\sim 10^{13}$ Hz suggesting a maximum time step of 1 fs. This maximum can be increased by the use of an algorithm called SHAKE [27], which can be applied to bonds with hydrogen. SHAKE constrains bonds to their equilibrium bond length, removing the vibration of the hydrogen atoms and allowing a maximum time step of 2 fs.

2.3.2 The CHARMM Force Field

The force that acts on atom i at time t , \vec{F}_i is calculated from a potential function of the atomic positions, V , often described in the literature as a force field, using the relation

(2-20)

$$\vec{F}_i = -\nabla V.$$

The CHARMM potential function [20] is given in Equation (2-21)

(2-21)

$$\begin{aligned} V = & \sum_{bonds} K_b (b - b_0)^2 + \sum_{UB} K_{UB} (s - s_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 \\ & + \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{impropers} K_\phi (\phi - \phi_0)^2 \\ & + \sum_{nonbond} \epsilon_{ij} \left[\left(\frac{R_{minij}}{r_{ij}} \right)^{12} - \left(\frac{R_{minij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}} \\ & + \sum_{residues} V_{CMAP}(\phi, \psi) \end{aligned}$$

2. Computational Methods

The bonds term is a harmonic potential that describes the interactions of atomic pairs separated by a single bond (1,2-pairs). K_b is a parameter defined for the atom types, b is the length of the bond, a function of the position of the atoms, and b_0 is the ideal bond length. The angle term describes the energy of bond bending where K_θ is a parameter defined for the atom type that describe the angle, θ is the bond angle, a function of the atomic positions, and θ_0 is the ideal bond angle. The Urey-Bradley term describes interactions with second neighbours (1,3-pairs); it is very rarely defined for new atoms but remains in the force field for some older atom types such as protein backbone atoms. The ideal bond length and angle parameters, b_0 and θ_0 , have been optimised with respect to the values of crystallographic structures, or, if these are unavailable, QM optimised geometries. The force constants, K_b and K_θ , have been optimised to reproduce experimental or QM vibrational structure.

The dihedral potential describes the energy associated with bond twisting; it needs four atoms to define it, ABCD and the rotation is about the B-C bond. K_χ is a parameter defined by the atom types, δ is the phase difference, n is the periodicity and χ is the dihedral angle between the planes defined by ABC and BCD. The dihedral parameters are optimised with respect to energy surfaces calculated from a HF/6-31G* or higher level, torsional scan. QM methods and notation are discussed in Section 2.4. In this case QM data are preferable to experimental data because they allow for calculation of the entire energy surface. Improper dihedrals are artificial potentials that are used to restrict the conformation of a group consisting of a central atom bonded to three others. Improper terms are only defined for a few atom types, and are included when parameter optimisation of the bond, angle and dihedral terms fails to reproduce the target data sufficiently well. The group is also defined by four atoms, ABCD and the first atom, A, is the central one. φ is the angle between the plane defined by ABC and the plane defined by BCD. The force constant, K_φ is usually large to hold the atoms near the desired configuration.

2. Computational Methods

The non-bonded terms describe the interaction between two spatially near atoms, i and j . The first part of the non-bonded term is a Lennard-Jones potential that empirically models the van der Waals interaction. ϵ_{ij} and $R_{ij\min}$ are constants that depend on the atom type and their characteristic van der Waals interaction and r_{ij} is the distance between the two atoms. The van der Waals parameters are the most difficult to optimise and were originally chosen [23] to reproduce experimental properties such as heat of vaporisation and data from QM calculations of interactions with helium and neon atoms, though for new atom types they are usually chosen by analogy with values that are already defined in the force field. The second part of the non-bonded term is the Coulombic interaction that describes the electrostatic interaction between charged particles. q_i and q_j are the charges on atoms i and j and are chosen to reproduce QM geometries and interaction energies from interactions between the atoms and individual water molecules. ϵ is the dielectric constant of the surrounding medium and is set according to whether the simulation involves explicit or implicit solvent.

The final term in the CHARMM force field, the CMAP term [25], is a corrective term applied to the φ and ψ angles of the protein backbone to correct systematic errors in secondary structure. The CMAP term is discussed in depth in Section 4.2.

2.3.3 Statistical Mechanics

An atomistic simulation propagates the microscopic properties (positions and velocities or momenta of the atoms) of a molecular system. Statistical mechanics is needed to relate these microscopic properties, which define a state, to macroscopic observables. This is done using ensemble averages, where an ensemble is made up of all the possible microscopic states that can correspond to the systems observable properties. For example, the temperature (T) of a system is directly related to the ensemble average of kinetic energy, $\langle KE \rangle$, a function of momenta, as shown in Equation (2-22)

(2-22)

$$\langle KE \rangle = \frac{3Nk_B T}{2}$$

where N is the number of atoms in the system and k_B is the Boltzmann constant. The ensemble most commonly used in biomolecular simulations is the isothermal-isobaric (NPT) ensemble that maintains a constant pressure (P), total number of atoms (N) and temperature (T), as this mostly closely matches the conditions *in vivo*.

The partition function, Q , is an important concept of statistical mechanics [28]. If it is defined for an ensemble then all the macroscopic properties can be determined by applying the correct operator to it. For a classical system, where the energy difference between states is negligible so that energy, E , can be considered continuous, the partition function is given by the integral of the Boltzmann factor over all possible positions, \vec{r} , and momenta, \vec{p} , known collectively as phase space:

(2-23)

$$Q = \iint \exp\left(\frac{-E}{k_B T}\right) d\vec{r} d\vec{p}.$$

Some macroscopic properties, such as free energy are functions of Q , whilst others, such as internal energy, enthalpy and heat capacity depend on the derivative of Q . During a simulation not all parts of the phase space are visited as only low energy states are sampled and it is not possible to take an ensemble average or calculate the partition function. However, the ergodic hypothesis states that a time average is equal to an ensemble average, provided the time period is long enough that the initial conditions no longer influence the state of the system. For an MD simulation the kinetic energy-temperature relationship becomes:

(2-24)

$$\overline{KE} = \sum_{i=1}^N \frac{|\vec{p}_i|^2}{2m_i} = \frac{k_B T}{2} (3N - N_c)$$

2. Computational Methods

where \overline{KE} is the time average kinetic energy, \vec{p}_i and m_i are the total momenta and mass of the i^{th} particle and N_c is the number of constrained degrees of freedom, which is usually six for an MD simulation with rotation and translation suppressed.

For properties that are a function of the derivative of Q only the low energy states make a significant contribution, but for properties dependent on Q the high energy states also make a contribution so cannot be ignored [28]. Therefore MD simulations are able to return reasonable values through time averages of macroscopic properties that are dependent on the derivative of Q , but calculating properties such as free energy is difficult because the high energy states are not sampled.

2.3.4 Explicit Solvent Models and Periodic Boundary Conditions

In vivo proteins are surrounded by a solvent, usually water. Because water molecules are polar and can make hydrogen bonds, the stability and dynamics of a protein depends on the surrounding water molecules. The effects of water can be included in a simulation either explicitly or implicitly.

In an explicit solvent model individual water molecules are included in the simulation. This is the most physically realistic representation of water and allows for modelling of interactions between specific water molecules and residues in the protein that may be important to stability and dynamics. However, explicit solvent models require a shell of water molecules surrounding the protein at least 12 Å thick [29], leading to many thousands of extra atoms in the simulation which greatly increases computational expense.

CHARMM usually uses the TIP3P model [30], where water molecules are modelled as three point charges; the sum of the two equal positive charges corresponding to the hydrogen atoms is equal to the negative charge representing the oxygen. The van der Waals interactions are modelled by a Lennard-Jones potential centred on the oxygen; the van der Waals radii of the hydrogen atoms are present but buried within the oxygen atoms. Other water models supported by CHARMM are the SPC model [31], that uses different

2. Computational Methods

hydrogen charges and Lennard-Jones parameters than TIP3P; the TIP4P model [30] that moves the charge on the oxygen toward the hydrogens along the molecular axis of symmetry and the ST2 model [32] that places charges on the lone-pair sites of the oxygen. The TIP4P and ST2 models have a larger dipole moment than TIP3P, which are closer to the true value of liquid water but are computationally more expensive to simulate.

When simulating a system containing explicit solvent molecules, consideration must be given to the representation of the boundary of the solvent. For example, to solvate the ileal lipid binding protein (molecular weight 14.3 kDa) a cube of edge length 62 Å is required, containing 7100 solvent molecules. Interactions with a surface boundary can extend up to 10 molecular diameters into the fluid [3] and, given the diameter of water is approximately 2.8 Å, then over 99.9% of the cube volume would be in range of surface effects. In order to remove these surface effects, periodic boundary conditions (PBC) are used to model a bulk liquid. The molecular system is built so that the solvent box is a congruent shape that will pack perfectly in three dimensions. The central cell is repeated so that it is surrounded by a periodic array of image cells. Each atom in the system interacts with all the real and image atoms within a cut-off radius, which is slightly less than the unit cell length, so that it experiences forces as if it were in an infinite solution. If an atom leaves the central cell during the simulation it is replaced by an image atom entering from the opposite face, maintaining a constant number of atoms. Figure 2-1 illustrates PBC in two dimensions. The most commonly used shapes for the unit cell are cubes and truncated octahedra. Truncated octahedra have the advantage of using a smaller volume, thus less atoms, than a cube to obtain a comparable distance between images; however, calculations using a cubic cell are computationally cheap compared with many other shapes, as the position of the image atoms is calculated by adding or subtracting multiples of the box length to the coordinates of the central atoms, not requiring more computationally expensive transformations.

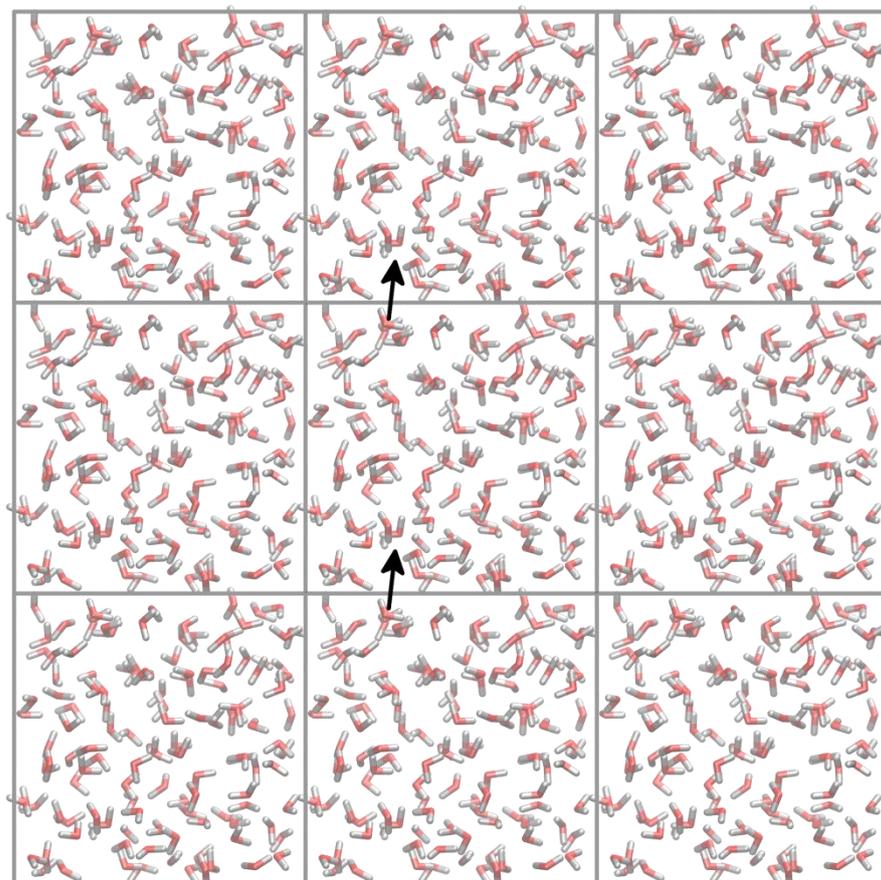


Figure 2-1 Illustration of PBC in two dimensions. When an atom leaves the central cell it is replaced by an atom entering through the opposite face from an adjacent image cell.

2.3.5 The GBSW Implicit Solvent Model and Langevin Dynamics

Implicit solvent models do not explicitly include water molecules. This decreases the size of the simulation, allowing longer timescales and increased sampling. The implicit solvent simulations in this thesis uses a generalised Born (GB) continuum solvation model [33], that incorporates the effects of the solvent as an extra term in the potential field, with Langevin dynamics, which includes the effects of frictional drag and random collisions. In a recent review on implicit solvation [34] GB models were described as '*...the prime choice for bimolecular simulations because of the favourable balance in accuracy and efficiency.*'

In a continuum model the solute (protein) is considered to be embedded in a cavity of a high dielectric medium. Equation (2-25) describes the solvation free

2. Computational Methods

energy, ΔG_{solv} , the free energy change needed to transfer a molecule from vacuum to the cavity in the solvent.

(2-25)

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

The second term, ΔG_{np} , is the non-polar solvation energy which includes the energy required to form the cavity in the solvent and solvent-solute van der Waals interaction [3]. It is approximated by Equation (2-26), where S is the solvent accessible surface area of the solute and γ is an empirical surface tension coefficient [33].

(2-26)

$$\Delta G_{np} = \gamma S$$

The first term in Equation (2-25), ΔG_{elec} , is the electrostatic term. This term can be calculated rigorously by solving the Poisson-Boltzmann equation by finite difference methods, but that is too computationally expensive for use in a bimolecular simulation [33]. Instead, the electrostatic term to describe a system of N particles with charge q_i can be approximated by the generalised Born formula. The derivation of the generalised Born formula shown here is based on Leach [3].

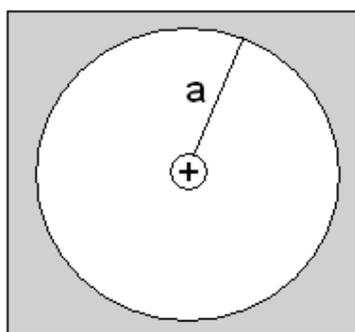


Figure 2-2 The Born model

The Born expression, Equation (2-27), describes the work done to transfer an ion from a vacuum to a spherical cavity, radius a , in a medium with dielectric constant ϵ (Figure 2-2):

(2-27)

$$\Delta G_{elec} = -\frac{q^2}{2a} \left(1 - \frac{1}{\epsilon}\right).$$

The total electrostatic free energy of a system of N particles with radii a_i and charges q_i in a medium with dielectric constant is given by the Coulombic energy plus the Born free energy of solvation

(2-28)

$$G_{elec} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{\epsilon r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \frac{q_i^2}{a_i}$$

The electrostatic free energy of solvation is the work required to move the ensemble from a vacuum to a medium of dielectric constant ϵ . This is equal to the difference between the Coulombic interaction of the charges in a vacuum and the total electrostatic free energy given in (2-28):

(2-29)

$$\Delta G_{elec} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} - \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{\epsilon r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \frac{q_i^2}{a_i}.$$

The Born radius of a particle, a_i , corresponds to the radius that would return the electrostatic energy of the system according to the Born equation if all other atoms in the molecule were uncharged [3]. The Born radii are quite complicated to calculate and are only updated with the non-bonded list during molecular dynamics. The CHARMM implementation of the GB solvent model used in this work, the GBSW module, has had the initial values for the Born radii set by optimisation of extensive folding, unfolding and equilibrium simulations for a range of peptides and mini-proteins [35].

Langevin dynamics includes the frictional drag and random collisions of water atoms as additional terms in Equation (2-13). The frictional force acting on a particle is related to its velocity, \vec{v}_i , by a constant of proportionality, ξ .

(2-30)

$$\vec{F}_{friction} = -\xi \vec{v}$$

The friction coefficient is related to the collision frequency, γ , and the diffusion constant, D , by the relationship in Equation (2-31), where m is the mass of the particle, T is the temperature and k_B is the Boltzmann constant [3].

(2-31)

$$\xi = \gamma m = \frac{k_B T}{D}$$

The collision frequency is often referred to as the friction coefficient in the literature. The term to model random collisions with solvent molecules, $\vec{R}_i(t)$, is often assumed to be independent of position, velocity and other forces, and is drawn from a Gaussian distribution with zero mean. Equation (2-32) shows the complete Langevin dynamics equation, where $\vec{F}_i(r_i(t))$ is the term calculated using the empirical force field.

(2-32)

$$m_i \frac{d^2 \vec{r}_i(t)}{dt^2} = \vec{F}_i(r_i(t)) - \gamma_i \frac{d\vec{r}_i(t)}{dt} m_i + \vec{R}_i(t)$$

Yeh and Wallqvist [36] compared the effect of GBSW, another implicit solvent model, GBMV, and explicit solvent on peptide structure and dynamics; they concluded that structural properties sampled by the explicit solvent are reasonably represented by the implicit models but dynamic properties were not as accurately represented, unless Langevin dynamics using a friction coefficient of 10 ps^{-1} was used. A Berendsen thermostat [37] maintains the temperature of a simulation by coupling the system to an external heat bath and scaling the velocities at each time step so that the rate of change of the temperature is proportional to the difference between the bath and the system [3]. Mor and Levy [38] studied proteins with long, flexible tails attached to their termini and found that using a Berendsen thermostat caused

2. Computational Methods

a temperature difference between the tails and the rigid regions of secondary structure, but using a Langevin thermostat regulated the temperature of the inhomogeneous systems reliably. Feig [39] compared GBMV with explicit solvent in simulations of an alanine dipeptide, the B1 domain of streptococcal protein G, and ubiquitin using either a Nosé–Hoover [40] thermostat or Langevin dynamics. A Nosé–Hoover thermostat includes the heat bath as an extra degree of freedom in the simulation which has a fictitious mass, so can be propagated through time with the rest of the simulation [3]. Langevin dynamics with implicit solvent matches explicit solvent, but using a Nosé–Hoover thermostat with implicit solvent reduces the system’s ability to cross energy barriers, because of the lack of stochastic collisions. Based on these studies, the GBSW simulations reported in Chapter 3 use Langevin dynamics applied to non-hydrogen atoms to regulate temperature and include the effects of friction and collisions.

2.4 Quantum Chemistry

2.4.1 Quantum Mechanics and the Schrödinger Equation

Classical Newtonian physics is unable to predict the behaviour of particles such as electrons because it does not take into account the dual particle-wave nature of moving bodies. The de Broglie hypothesis postulates that wavelength is inversely proportional to momentum with Planck’s constant as the constant of proportionality. Given that Planck’s constant is $6.62 \times 10^{-34} \text{ m}^2\text{kg s}^{-1}$, the size of the wavelength, and therefore influence of wave-like properties, is only non-negligible for particles with a very low mass, such as electrons.

Quantum mechanics is concerned with solving the Schrödinger Equation to find the wavefunction, Ψ , that describes a moving body and then using that wavefunction to calculate physical properties such as the particle’s energy, E , or probable position. Equation (2-33) gives the time-independent form of the Schrödinger Equation. \mathcal{V} is an external field that is invariant with respect to time and is often the electrostatic potential due to atomic nuclei. $\frac{\hbar}{2m} \nabla^2$ gives

2. Computational Methods

the particle's kinetic energy, where \hbar is Planck's constant over 2π , m is the particle's mass and ∇^2 is the second derivative with respect to position.

(2-33)

$$\left\{-\frac{\hbar}{2m}\nabla^2 + \mathcal{V}\right\}\Psi(\vec{r}) = E \Psi(\vec{r})$$

The left-hand side of the equation is often abbreviated to $\mathcal{H}\Psi$ where \mathcal{H} is the Hamiltonian operator which acts upon the wavefunction to give the particle's energy, E .

The Schrödinger Equation can only be solved exactly for a few idealised problems where boundary conditions are imposed on the system. Computational quantum chemistry techniques are a group of methods that compute approximate wavefunctions and energies to describe real molecular systems.

An important approximation made to solve the Schrödinger Equation for real molecular systems is the Born-Oppenheimer approximation, which states that the motion of the electrons and nuclei are de-coupled [28]. The basis of this approximation is that electrons are ~ 1840 times lighter than protons. This means that the nuclear positions and charges can be treated as fixed parameters and only the electronic Schrödinger Equation needs solving.

2.4.2 Hartree-Fock Methods

Hartree-Fock methods are a class of techniques that make a simplification that the overall wavefunction is a product of independent electronic orbitals, meaning the dynamics of each electron is independent and that each electron experiences an average interaction due the others in the system. Each electron is described by its own orbital, a function that can be related to the probability of finding an electron when the attraction of the nuclei and the average repulsion of the other electrons are included. The wavefunction is a product of these independent electronic orbitals. The orbitals are arranged in a Slater determinant which ensures that the wavefunction is antisymmetric so that if two electrons are interchanged the sign changes, which satisfies the

2. Computational Methods

Pauli exclusion principle. As the solution of each orbital depends on the others, the Hartree-Fock equations need to be solved iteratively to minimise the energy. Equation (2-34) shows the variational theorem containing the variational energy, E_{var} , which states that any approximate wavefunction will result in a higher expectation value of the energy than the exact, ground-state energy, E_{exact} . This statement can be used to adjust the electronic orbitals to minimise E_{var} and get closer to the exact energy.

(2-34)

$$\frac{\int \Psi^* \mathcal{H} \Psi}{\int \Psi^* \Psi} = E_{var} \geq E_{exact}$$

The electronic orbitals can be expanded as a basis set and the adjustment of the parameters for minimisation can be written as an eigenvalue problem. When the change in the variational energy between iterations is less than a certain tolerance the calculation is considered converged and a solution found for the wavefunction.

2.4.3 Density Functional Theory

Density Functional Theory (DFT) uses the electronic density distribution to calculate the total electronic energy. Unlike Hartree-Fock methods it does not attempt to calculate the full wavefunction, but uses the postulate that total electronic energy is a unique function of the electronic density distribution. The total sum of the electronic distribution density, $\rho(\vec{r})$, across all space is, by definition, equal to the total number of electrons, N .

(2-35)

$$\int \rho(\vec{r}) dr = N$$

For any trial density that satisfies this condition (Equation (2-35)) the electronic density calculated from it must be greater than or equal to the exact energy, leading to a statement of the variational principle for DFT.

(2-36)

$$E[\rho(\vec{r})] \geq E_{exact}$$

2. Computational Methods

The energy operator, $E[\rho(\vec{r})]$, is the energy contribution due to external potential plus the Fock operator, which is analogous to the Hamiltonian, but returns the electronic energy. Equation (2-37) shows the terms of the energy:

(2-37)

$$E[\rho] = E_{KE}[\rho] + E_H[\rho] + E_{XC}[\rho]$$

Kohn and Sham [41] suggested using non-interacting, auxiliary orbitals for each electron to calculate the kinetic energy component, E_{KE} . The auxiliary orbitals are expanded as a basis set and used with an initial guess for the electronic density distribution to iteratively minimise the changes in the total electronic energy and electronic density distribution until the calculation is converged. E_H is the Hartree electrostatic energy and is the classical electrostatic component calculated from Coulomb's Law. The final term, E_{XC} , is called the exchange-correlation functional and takes into account the quantum mechanical electron exchange and correlation interactions as well as the error in the difference between the calculated energy and the exact energy. The difference between the various types of DFT calculations depend on the functional form of the exchange-correlation energy. The simplest functionals are based on the local density approximation which assumes that the electron density is slowly varying and can be treated as a uniform gas. An improvement to this is a functional that is function of both the electron density and its derivatives, leading to gradient-corrected DFT methods. One of the most popular and successful exchange-correlation functionals is BYLP, which combines the gradient corrected exchange functional of Becke [42], shown to reduce the error in energy by two orders of magnitude compared with local density approximation methods [28], with the Lee-Yang-Parr correlation functional [43].

2.4.4 Basis Sets

A basis set is a group of basis functions that can be summed together to represent an unknown function, such as the electronic orbitals in Hartree-Fock methods or the Kohn-Sham orbitals in DFT. If the number of basis functions were infinite then the basis set would be complete and exactly

2. Computational Methods

equal the unknown function. This would be impossible to compute so a finite number of basis functions are used to approximate the unknown function [28]. The basis functions can have any form, but those commonly used in quantum chemistry programs have been chosen because they are suited to the coordinate system or are easy to perform calculations on.

Two commonly used type of basis function are Slater type orbitals (STO) and Gaussian type orbitals (GTO) [28]. In polar coordinates STO have the functional form given in equation (2-38) and GTO have the form given in equation (2-39) where n , l , m are the quantum numbers, N is a normalisation constant and $Y_{l,m}$ are spherical harmonic functions.

(2-38)

$$\chi_{\zeta,n,l,m}(r, \theta, \varphi) = NY_{l,m}(\theta, \varphi)r^{n-1}e^{-\zeta r}$$

(2-39)

$$\chi_{\zeta,n,l,m}(r, \theta, \varphi) = NY_{l,m}(\theta, \varphi)r^{2n-2-l}e^{-\zeta r^2}$$

STO result in a better representation of electronic behaviour near the nucleus and less STO are needed than GTO to achieve the same level of accuracy. However, integrals of STO are difficult to evaluate, and for some cases impossible. So usually GTO are preferable as the Gaussian function can easily be integrated.

A minimum basis set contains the number of functions to accommodate all the atomic orbitals in the neutral shell [3]. For example, for carbon, a function for 1s, 2s and 2p would be used. The STO- n G basis sets are minimum basis sets where n GTO are fitted to a STO for each atomic orbital. STO-3G is considered the absolute minimum as less than 3 GTO cannot accurately represent a STO. A double zeta (DZ) basis set has twice the number of functions than the minimal basis set. One set is then contracted close to the nucleus and the other is diffuse. This leads to higher accuracy and a better representation of each orbital but is computationally expensive. In a split-valence basis set the number of functions for the valence electrons, which are

2. Computational Methods

involved in chemical interactions, are doubled and the core electrons are represented by a minimum basis set. An example of the notation used to describe a split-valence basis set is 3-21G, where the core orbitals are described by a function of six fitted GTO and the valence orbitals are described by two functions: one a function consisting of two GTO for the contracted part and the other function a single GTO for the diffuse part. The use of extra functions for the heavy atoms to allow for polarisability is indicated with an asterisk. Therefore the HF/6-31G* calculation referred to in section 2.3.2 is a Hartree-Fock type calculation using a split valence basis set with polarisation functions.

2.4.5 QM-MM and *Ab initio* Molecular Dynamics

A major disadvantage of the classical MM approximation is that without explicit electrons it is not possible to model reactions such as the making and breaking of covalent bonds. One method for overcoming this disadvantage is to use a hybrid of QM and MM methods [44] together where a reactive centre is identified and modelled using quantum methods and the rest of the system is described classically. The quantum calculations optimise properties such as the energy, charge distribution and geometries of the atoms in the QM region and the atoms in the QM region are usually coupled to the classical region by their electrostatic interactions. Usually a linking hydrogen atom [45] is used to cap any dangling bonds in the QM region that are cut by the QM-MM boundary. The properties of the atoms in the QM region are then passed to the MD engine to calculate the next time step of dynamics.

In *ab initio* MD (AIMD) the forces used to propagate the dynamics are calculated using QM methods [46]. AIMD has the advantage that it can be used to simulate chemical reactions and un-parameterised chemical species, but because it depends on quantum calculations there is a reduction in system size compared with classical MD. There are two main AIMD schemes: Born-Oppenheimer MD and Car-Parrinello MD [47]. In Born-Oppenheimer MD the wavefunction that is used to calculate the energy and forces is optimised at every time step. This leads to accurate results and a larger time step of

2. Computational Methods

integration, but is also slow due to the computational expense. Car and Parrinello introduced a method (CPMD) that avoids wavefunction optimisation at every time step by propagating the wavefunction with the nuclei using a scheme that de-couples the nuclei and electrons into two subsystems. The CPMD method is described in more detail in the following section.

2.4.6 Lagrangian Mechanics and CPMD

The equations of CPMD are formulated using Lagrangian mechanics, a reformulation of classical mechanics that uses generalised coordinates. This means that the equations of Lagrangian mechanics can be applied to all dynamics problems, including those that take account of relativity and quantum properties. The Lagrangian, \mathcal{L} , is defined as the difference between the kinetic energy, K , and the potential energy, V

(2-40)

$$\mathcal{L} = K(\mathbf{r}, \dot{\mathbf{r}}) - V(\mathbf{r}, \dot{\mathbf{r}})$$

where \mathbf{r} are the positions and $\dot{\mathbf{r}}$ the velocities. The dot notation indicates the time derivative and the bold type that these are vectors. Provided that the forces acting on the system are conservative, the equations of motion can be derived from the Euler-Lagrange equation:

(2-41)

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_i} \right) - \frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} = 0$$

To illustrate [48], in classical mechanics K is

(2-42)

$$K = \frac{1}{2} \sum_{i=1}^N m_i \dot{\mathbf{r}}_i^2$$

where N is the number of particles and m is the mass of the i^{th} particle. Substituting this into equations (2-40) and (2-41) gives:

(2-43)

$$\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_i} = m_i \dot{\mathbf{r}}_i$$

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_i} \right) = m_i \ddot{\mathbf{r}}_i$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} = - \frac{\partial V}{\partial \mathbf{r}_i} = \mathbf{F}_i$$

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_i} \right) - \frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} = m_i \ddot{\mathbf{r}}_i - \mathbf{F}_i = 0$$

which is Newton's second law of motion.

The central idea of CPMD is that if the force on the nuclei can be obtained from the derivative of the Lagrangian with respect to the nuclear positions, this suggests that the derivative of a suitable Lagrangian with respect to the orbitals will give the forces in the orbitals. The Car-Parrinello Lagrangian [47] is

(2-44)

$$\mathcal{L}_{CP} = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2} \mu_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle - \langle \Phi_0 | \mathcal{H}_{el} | \Phi_0 \rangle + \text{constraints}$$

and the associated Euler-Lagrange equations are

(2-45)

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}_{CP}}{\partial \dot{\mathbf{R}}_I} \right) - \frac{\partial \mathcal{L}_{CP}}{\partial \mathbf{R}_I} = 0$$

(2-46)

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}_{CP}}{\partial \dot{\psi}_i^*} \right) - \frac{\partial \mathcal{L}_{CP}}{\partial \psi_i^*} = 0$$

where I is the number of nuclei, \mathbf{R} is the positions of the nuclei, μ is the 'fictitious mass' of the electrons, an inertia parameter that arises due to the

2. Computational Methods

adiabatic separation of the nuclei and the electrons, ψ are the Kohn-Sham orbitals, Φ_0 is the initial molecular wavefunction and H_{el} is the electronic Hamiltonian. Usually in CPMD simulations the ground state energy, $\langle \Phi_0 | \mathcal{H}_{el} | \Phi_0 \rangle$, that forms the potential energy term in the Lagrangian, is the Kohn-Sham energy calculated using density functional theory. This is calculated once at the start of the simulation. The constraints are Lagrange multipliers that are used to maintain orthonormality of the KS orbitals.

The nuclei evolve through time with physical temperature $\propto \sum_I M_I \dot{\mathbf{R}}_I^2$ and the electrons have a ‘fictitious temperature’ $\propto \sum_i \mu_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle$. The ‘cold’ electrons mean that the electronic subsystem will stay close to the ground state wavefunction. Forces are calculated from partial derivatives of the ground state energy with respect to independent variables (e.g. nuclear positions, Kohn-Sham orbitals).

A plane wave basis set with pseudopotentials for the core electrons is often used for CPMD calculations. Plane wave basis sets are particularly suited for systems with PBC as they have infinite range. The basis functions in a plane wave basis set have the functional form [28]:

(2-47)

$$\chi_{\mathbf{k}}(\mathbf{r}) = N \exp(i\mathbf{k}\mathbf{r})$$

The wave vector, \mathbf{k} , can be considered a frequency factor, with high \mathbf{k} values indicating rapid oscillation. The size of the basis set is determined by the largest value of \mathbf{k} included and the size of the unit cell. Core electrons experience high frequency oscillations that require a large value of \mathbf{k} , leading to a basis set with many functions that is computationally expensive to use. To overcome this problem the core electrons are often described by a pseudopotential that smoothes the high frequency oscillations whilst the valence electrons are described by the plane waves basis set.

2.5 References

1. Skylaris, C.-K., P.D. Haynes, A.A. Mostofi, and M.C. Payne, *Introducing ONETEP: Linear-scaling density functional simulations on parallel computers*. J Chem Phys, 2005. **122**(8): p. 084119.
2. Plewczynski, D., M. Lazniewski, R. Augustyniak, and K. Ginalski, *Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database*. J. Comput. Chem., 2011. **32**: p. 742-755.
3. Leach, A.R., *Molecular modelling: principles and applications*. 2nd ed. 2001, Harlow: Pearson Education.
4. Jain, A.N., *Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine*. J. Med. Chem., 2003. **46**(4): p. 499-511.
5. Zsoldos, Z., D. Reid, A. Simon, S.B. Sadjad, and A.P. Johnson, *eHiTS: A new fast, exhaustive flexible ligand docking system*. J. Mol. Graph. Model., 2007. **26**(1): p. 198-212.
6. Rarey, M., B. Kramer, T. Lengauer, and G. Klebe, *A fast flexible docking method using an incremental construction algorithm*. J. Mol. Biol., 1996. **261**(3): p. 470-489.
7. Jones, G., P. Willett, R.C. Glen, A.R. Leach, and R. Taylor, *Development and validation of a genetic algorithm for flexible docking*. J. Mol. Biol., 1997. **267**(3): p. 727-748.
8. Morris, G.M., D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, and A.J. Olson, *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy*. J. Comput. Chem., 1998. **19**(14): p. 1639-1662.
9. Trott, O. and A.J. Olson, *AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading*. J. Comput. Chem., 2010. **31**: p. 455-461.
10. Friesner, R.A., J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, and P.S. Shenkin, *Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy*. J. Med. Chem., 2004. **47**(7): p. 1739-1749.
11. Venkatachalam, C.M., X. Jiang, T. Oldfield, and M. Waldman, *LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites*. J. Mol. Graph. Model., 2003. **21**(4): p. 289-307.
12. Wang, R., X. Fang, Y. Lu, and S. Wang, *The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures*. J. Med. Chem., 2004. **47**(12): p. 2977-2980.
13. Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *Equations of state calculations by fast computing machines*. J. Chem. Phys., 1953. **99**: p. 11276-11287.
14. Holland, J.H., *Adaptation in Natural and Artificial Systems*. 1975, Ann Arbor: University of Michigan Press.

2. Computational Methods

15. Wang, R., L. Lai, and S. Wang, *Further development and validation of empirical scoring functions for structure-base binding affinity prediction*. J. Comput. Aided Mol. Des., 2002. **16**: p. 11-26.
16. Case, D.A., T.E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R. Woods, *The Amber biomolecular simulation programs*. J. Comput. Chem., 2005. **26**: p. 1668-1688.
17. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, *Scalable Molecular Dynamics with NAMD*. J. Comput. Chem., 2005. **26**(16): p. 1781-1802.
18. Plimpton, S.J., *Fast parallel algorithms for short-range molecular dynamics*. J. Comp. Phys., 1995. **117**: p. 1-19.
19. Brooks, B.R., R.E. Bruccoleri, D.J. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*. J. Comput. Chem., 1983. **4**: p. 187-217.
20. Brooks, B.R., C.L. Brooks, A.D. Mackerell Jr, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus, *CHARMM: The Biomolecular Simulation Program*. J. Comput. Chem., 2009. **30**(10): p. 1545-1614.
21. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*. J. Mol. Model, 2001. **7**(8): p. 306-317.
22. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*. J. Chem. Theory Comput., 2008. **4**(3): p. 435-447.
23. Mackerell Jr, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. J. Phys. Chem. B, 1998. **102**(18): p. 3586-3616.
24. Mackerell Jr, A.D., *Atomistic Models and Force Fields*, in *Computational Biochemistry and Biophysics*, O.M. Becker, et al., Editors. 2001, Marcel Dekker, Inc.: New York.
25. Mackerell Jr, A.D., M. Feig, and C.L. Brooks III, *Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations*. J. Comput. Chem., 2004. **25**(11): p. 1400-1415.

2. Computational Methods

26. Verlet, L., *Computer experiments on classical fluids .I. Thermodynamical properties of Lennard-Jones molecules*. Phys. Rev., 1967. **159**(1): p. 98-103.
27. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. J. Comput. Phys., 1977. **23**: p. 327-341.
28. Jensen, F., *Introduction to Computational Chemistry*. 2nd ed. 2007, Chichester: Wiley.
29. Schleif, R., *Analysis of Protein Structure and Function: A Beginner's Guide to CHARMM*. 2006, Johns Hopkins University: Baltimore.
30. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water*. J. Chem. Phys., 1983. **79**(2): p. 926-936.
31. Berendsen, H.J.C., J.R. Grigera, and T.P. Straatsma, *Interaction models for water in relation to protein hydration*. J. Phys. Chem. , 1981. **91**: p. 6269-6271.
32. Stillinger, F.H. and A. Rahman, *Improved simulation of liquid water by molecular dynamics*. J. Chem. Phys., 1974. **60**: p. 1545-1557.
33. Im, W.P., M.S. Lee, and C.L. Brooks, *Generalized Born model with a simple smoothing function*. J. Comput. Chem., 2003. **24**(14): p. 1691-1702.
34. Chen, J.H., C.L. Brooks, and J. Khandogin, *Recent advances in implicit solvent-based methods for biomolecular simulations*. Curr. Opin. Struct. Biol., 2008. **18**(2): p. 140-148.
35. Chen, J.H., W.P. Im, and C.L. Brooks, *Balancing solvation and intramolecular interactions: Toward a consistent generalized Born force field*. J. Am. Chem. Soc., 2006. **128**(11): p. 3728-3736.
36. Yeh, I. and A. Wallqvist, *Structure and dynamics of end-to-end loop formation of the penta-peptide Cys-Ala-Gly-Gln-Trp in implicit solvents*. J. Phys. Chem. B, 2009. **113**: p. 12382-12390.
37. Berendsen, H.J.C., J.P.M. Postma, W.F. van Gunsteren, A. di Nola, and J.R. Haak, *Molecular dynamics with coupling to an external bath*. J. Chem. Phys., 1984. **81**: p. 3684-3690.
38. Mor, A. and G.Z.Y. Levy, *Simulations of proteins with inhomogeneous degrees of freedom: the effect of thermostats*. J. Comput. Chem., 2008. **29**(12): p. 1992-1998.
39. Feig, M., *Kinetics from implicit solvent simulations of biomolecules as a function of viscosity*. J. Chem. Theory Comput., 2007. **3**: p. 1734-1748.
40. Hoover, W.G., *Canonical dynamics: equilibrium phase-space distributions*. Phys. Rev. A, 1985. **31**(3): p. 1695-1697.
41. Kohn, W. and L.J. Sham, *Self-consistent equations including exchange and correlation effects*. Phys. Rev. A, 1965. **140**: p. 1133-1138.
42. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behaviour*. Phys. Rev., 1988. **A38**: p. 3098-3100.

2. Computational Methods

43. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation energy formula into a functional of the electron density*. Phys. Rev., 1988. **B37**: p. 785-798.
44. Warshel, A. and M. Levitt, *Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*. J. Mol. Biol., 1976. **103**(2): p. 227-249.
45. Singh, U.C. and P.A. Kollman, *A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH₃Cl + Cl⁻ exchange reaction and gas phase protonation of polyethers*. J. Comput. Chem., 1986. **7**(6): p. 718-730.
46. Marx, D. and J. Hutter, *Ab Initio Molecular Dynamics: Theory and Implementation*, in *Modern Methods and Algorithms of Quantum Chemistry*, J. Grotendorst, Editor. 2000, John von Neumann Institute for Computing: Jülich. p. 329-477.
47. Car, R. and M. Parrinello, *Unified approach for molecular dynamics and density-functional theory*. Phys. Rev. Lett., 1985. **55**: p. 2471-2474.
48. Tuckerman, M.E., *Statistical Mechanics: Theory and Molecular Simulation*. 2010, Oxford: Oxford University Press.

3 Conformational Preferences of Nisin Analogues

3.1 Introduction

The development and introduction of penicillin in 1943 revolutionised the treatment of infectious disease. However, through the overuse and misuse of antibiotics, multi-resistant strains of pathogens have emerged, leading to an increase in human deaths from bacterial infections. One of the strategies identified by the World Health Organisation to contain resistance is the development of new drugs, with antibiotic peptides identified as one of six promising areas of research [1]. Lantibiotics are antibiotic peptides secreted by gram-positive bacteria, characterised by unusual dehydrated amino acids and thioether rings formed by post-translation modifications. Molecular dynamics simulations with implicit and explicit solvent have been performed to study analogues of the first 12 residues of nisin modified so that the thioether bonds have been replaced with disulfide bonds between cysteines. The chirality of the four cysteine residues is varied to see how this influences disulfide bond formation.

3.2 The Nisin Killing Action

Lantibiotics are a group of antibiotic peptides secreted by Gram-positive bacteria that are effective against other Gram-positive bacteria. They are characterised by unusual dehydrated amino acids and cross-linking thioether bonds [2]. One of the most studied lantibiotics is nisin. It is produced by *Lactococcus lactis* and is effective against spore-forming micro-organisms, including *Streptococci*, *Bacilli* and *Clostridia*; it is widely used as a food preservative for dairy products, fruits and vegetables [3]. Nisin (Figure 3-1a) contains five thioether rings and the sequence has 34 amino acids, including three dehydrated amino residues: dehydrobutyrine (Dhb) at position 2 and dehydroalanine (Dha) at positions 5 and 33. Nisin is synthesised by the ribosome and modified post-translation by the dehydratase NisB, which dehydrates eight of the serines and threonines. Subsequently, the cyclase NisC joins five of the dehydrated residues to five free cysteines to form the

3. Conformational Preferences of Nisin Analogues

characteristic thioether rings [4]. NMR studies of nisin in water and membrane-mimetic micelles have shown that rings A, B and C are joined to the inter-linking rings D and E by a flexible region, and that the linear tails at the N- and C-termini are also flexible [5].

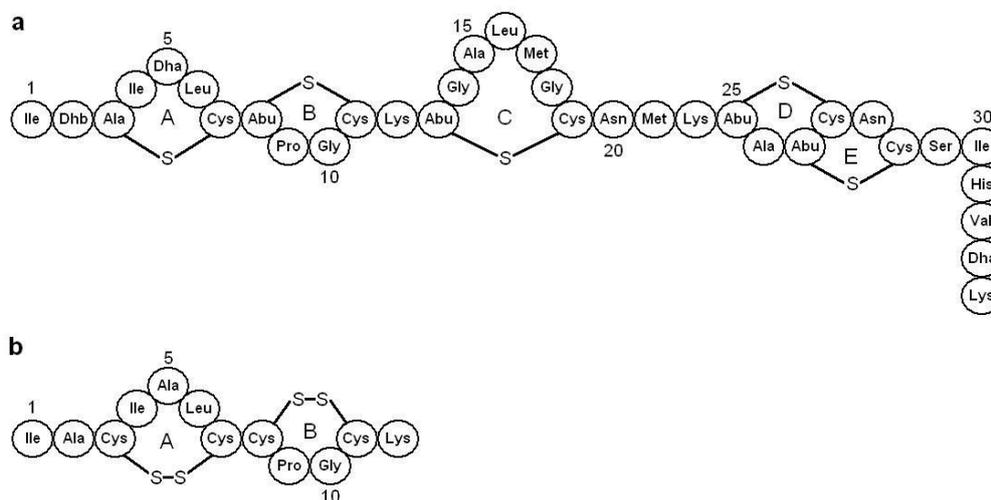


Figure 3-1 (a) The sequence of nisin showing the thioether rings and the unusual amino acids aminobutyric acid (Abu), dehydroalanine (Dha) and dehydrobutyric acid (Dhb). (b) The sequence of the proposed disulfide analogue of the first 12 residues of nisin.

The cell walls of bacteria are made of a scaffold of amino sugars called the peptidoglycan layer. Lipid II is essential for the synthesis of the cell wall by transporting peptidoglycan subunits across the cytoplasmic membrane. Nisin exerts two distinct killing actions on Gram-positive bacteria, both of which involve binding to lipid II. The first killing action is by pore formation across the cell membrane that allows the cytoplasmic contents to diffuse out. Hasper *et al.* [6] demonstrated that nisin can form stable pore complexes containing four lipid II molecules and eight nisin molecules. The second killing action is by the removal of lipid II from the site of cell wall synthesis, disrupting cell wall formation and causing aberrations [7], [8]. An NMR structure of the nisin-lipid II complex shows that the N-terminus and rings A and B form a cage around the pyrophosphate group of lipid II, bound by hydrogen bonds between the pyrophosphate and backbone amides in the nisin [9]. Vancomycin, the antibiotic of last resort, binds to the pentapeptide moiety of lipid II, where

resistance can be conferred by a few simple mutations. Nisin, and other closely related lantibiotics, are good candidates for the development of novel antibiotic compounds because they bind to a chemical entity that cannot be altered as readily, so resistance may be less likely to develop.

3.3 Disulfide Rich Peptides

There are three possible disulfide frameworks for peptides and proteins with four cysteine residues (CysI, CysII, CysIII, CysIV) and two disulfide bonds: globular connectivity (CysI-CysIII, CysII-CysIV), ribbon connectivity (CysI-CysIV, CysII-CysIII) and bead connectivity (CysI-CysII, CysIII-CysIV) [10]. Disulfide-rich peptides are produced by all classes of organism, where they perform functions such as defence against insects for plants, defence against bacteria for animals and regulatory functions [11]. An important class of disulfide-rich peptides are conotoxins, toxins and venom components evolved by spiders, scorpions and cone snails, which have been studied and characterised extensively as possible drug-leads [12]. The bead connectivity, which the nisin analogues will need to form to be active against bacteria, has not been observed in nature [11], but has been made synthetically. Gehrman *et al.* [13] synthesized all three isomers of a two-disulfide conotoxin and characterised them using NMR. The bead connectivity had a much less defined structure than the native globular and non-native ribbon connectivities, with a more stable N-terminus, a disordered C-terminus and deviations from random coil values indicating a very solvent accessible backbone.

Comprehensive searching of Web of Science has not identified any studies where thioether bonds have been substituted for disulfides, but there are examples of analogues where disulfides have been replaced with thioethers. Bondebjerg *et al.* [14] synthesised thioether analogues of a conotoxin; the analogues were significantly less potent than the native conotoxin, but changes to the orientation of the thioether bonds is expected to increase potency. Levengood and van der Donk [15] used an enzyme that forms the thioether rings in the lantibiotic lactacin, by dehydration and cyclicization, to

3. Conformational Preferences of Nisin Analogues

synthesize a thioether analogue of a snake venom conotoxin, but did not report its biological activity.

Molecular dynamics simulations on their own, and in combination with other computational techniques, have been used to study disulfide formation and bond shuffling. Schmid et al. [16] used MD simulations in explicit solvent to study disulfide bridge shuffling in bovine α -lactalbumin. To improve sampling of disulfide bond shuffling they used an unphysical representation of the cysteine residues with no constraints on bond length, bond angle or dihedral angle. They found that the simulations at 353 K and 373 K favoured a non-native disulfide bond that has been observed experimentally as a folding intermediate and at elevated temperatures. Nilsson and co-workers have used MD simulations in combination with pKa calculations, quantum chemistry calculations and experimental results to study the disulfide bond at the active site of the thioredoxin superfamily, to understand how structure and protein environment result in a difference in redox potential across members of the family [17], [18], [19] and the thiol/disulfide exchange reaction needed for dissociations of thioredoxin complexes [20].

Cysteine residue pattern, loop size, position of non-cysteine residues, backbone conformation and pre-existing disulfides all have important roles in forming native disulfide bonds [11], but how these factors interact varies between peptides. Information about the dynamics of peptides and proteins can be determined by NMR experiments, but this can be difficult and expensive and NMR cannot directly detect sulfur atoms. Therefore, MD simulations are a useful tool for studying how structure and dynamics contribute to the formation of disulfide bonds in the nisin analogues, and in peptides more generally.

3.4 Simulation Set-up

For all the simulations the peptide was built in CHARMM [21], [22] with the amino acid sequence IACIALCCPGCK (see Figure 3-1). An acetyl N-terminus and methylamine C-terminus were added so that the termini were neutral.

3. Conformational Preferences of Nisin Analogues

The lysine residue was positively charged. The chirality of the cysteine residues was altered to produce the set of 16 analogues shown in Table 3-1. The disulfide bonds between the cysteine residues were not included, in order to predict which analogues, if any, favour conformations where disulfide bonds between Cys3-Cys7 and Cys8-Cys11 can form.

Nisin analogue	Cys3	Cys7	Cys8	Cys11	Nisin Analogue	Cys3	Cys7	Cys8	Cys11
0	L	L	L	L	8	D	L	L	L
1	L	L	L	D	9	D	L	L	D
2	L	L	D	L	10	D	L	D	L
3	L	L	D	D	11	D	L	D	D
4	L	D	L	L	12	D	D	L	L
5	L	D	L	D	13	D	D	L	D
6	L	D	D	L	14	D	D	D	L
7	L	D	D	D	15	D	D	D	D

Table 3-1 Chirality of the cysteine residues in each nisin analogue. The non-cysteine residues are L-amino.

The initial conformation of the backbone was fully extended. The S_3 - S_7 and S_8 - S_{11} distances were 14.0 Å and 12.4 Å, respectively. The chirality of the cysteine residues was changed from L-form to D-form by deleting the side chain, exchanging the position of the C_β and H_α atoms and rebuilding the side chain from the C_β atom.

The final term in the CHARMM force field, the CMAP term, is a corrective term applied to the φ and ψ angles of the protein backbone. This correction was added to the parameter set in 2004 [23] after simulations of proteins and peptides in solution and lipid environments were found to be biased toward the formation of π -helices and limitations were seen in reproduction of QM energetic data for alanine dipeptide and tripeptide. QM calculations of the alanine, glycine and proline dipeptide's Ramachandran plots and optimisation to match crystallographic structures were used to produce correction maps that minimise the difference between the force field and the QM calculated energy surface. The calculations for the CMAP correction were performed

3. Conformational Preferences of Nisin Analogues

using the L-amino dipeptides to produce the L-amino Ramachandran plots. Therefore in the standard version of the CHARMM force field this term is invalid for D-amino acids and the CMAP term was not included in the simulations.

MD simulations were performed using CHARMM version 34b1 with the CHARMM22 protein force field [24]. The generalised Born with simple switching [25] (GBSW) implicit solvent model, described in the Chapter 2, was used to access a longer time scale than is usually available with explicit solvent models. The Born radii were those optimised by Nina *et al.* [26] and the GBSW model parameters were the defaults recommended in the CHARMM manual. Langevin dynamics with a friction coefficient of 10 ps^{-1} applied to non-hydrogen atoms [27] was used to regulate temperature and include the effects of friction and collisions (see Section 2.3.5). For each nisin analogue we conducted 10 independent runs of equilibration and production. The seed for the random number generator for the Langevin stochastic collisions was different for each independent run of each analogue and the peptide was equilibrated at 298 K for 2 ns. The production phase was 50 ns and the positions of the atoms were recorded every 1 ps.

To understand how the chirality of the cysteine residues affects the formation of the disulfide bridges the backbone φ and ψ dihedral angles and hydrogen bonding patterns were analysed. Linear correlation coefficients were calculated using the CORREL module in CHARMM [22]. These were calculated between the S-S distance for each cysteine pair and the distance between the backbone amino hydrogen and backbone oxygen for all possible residue combinations for each trajectory. A hydrogen – oxygen distance below 2.4 Å indicates the formation of a hydrogen bond. Linear-circular correlation coefficients were calculated between the S-S distance and each φ and ψ angle for each trajectory. Linear-circular correlation coefficients were calculated using Mardia's method [28]:

(3-1)

$$r^2 = \frac{(r_{yC}^2 + r_{yS}^2 + 2r_{yC}r_{yS}r_{CS})}{1 - r_{CS}^2}$$

where y are the linear data, χ are the circular data, r_{yC} is the correlation coefficient between y and $\cos(\chi)$, r_{yS} is the correlation coefficient between y and $\sin(\chi)$ and r_{CS} is the correlation coefficient between $\cos(\chi)$ and $\sin(\chi)$. The range of r is 0 to 1, because positive and negative correlation cannot be distinguished in the circular-linear case. If a variable (H-O distance, φ or ψ) and a S-S distance correlated during five or more trajectories, with a coefficient greater than 0.5, the correlation was investigated further.

3.5 Results

To assess the thermodynamic stability of the simulations, the temperature and total energy as a function of time were examined and were stable during all the simulations (data not shown). All six sulfur – sulfur distances (3-7, 3-8, 3-11, 7-8, 7-11, 8-11) were recorded for each independent trajectory of each nisin analogue and the associated distributions were calculated with an interval of 0.1 Å. Subject to the sufficiency of the conformational sampling, the histograms can be used to calculate the effective energy, or potential of mean force (PMF), of different conformational states as a function of sulfur-sulfur separation, using the Boltzmann relation in Equation (3-2); where N_0 is the number of conformations in the most populated, ground state, N_i is the number of conformations in the i^{th} state, ΔE_i is the effective energy difference between the ground state and the i^{th} state, k_B is the Boltzmann constant and T is the temperature.

(3-2)

$$N_i = N_0 \exp\left(\frac{-\Delta E_i}{k_B T}\right)$$

Figure 3-2 shows some typical histograms and the corresponding PMFs. The probability of the sulfur - sulfur distance being less than 5.5 Å, the cut-off for disulfide bridge formation [16], was calculated for the S₃-S₇ and S₈-S₁₁ interactions, $P(S_3-S_7)$ and $P(S_8-S_{11})$.

3. Conformational Preferences of Nisin Analogues

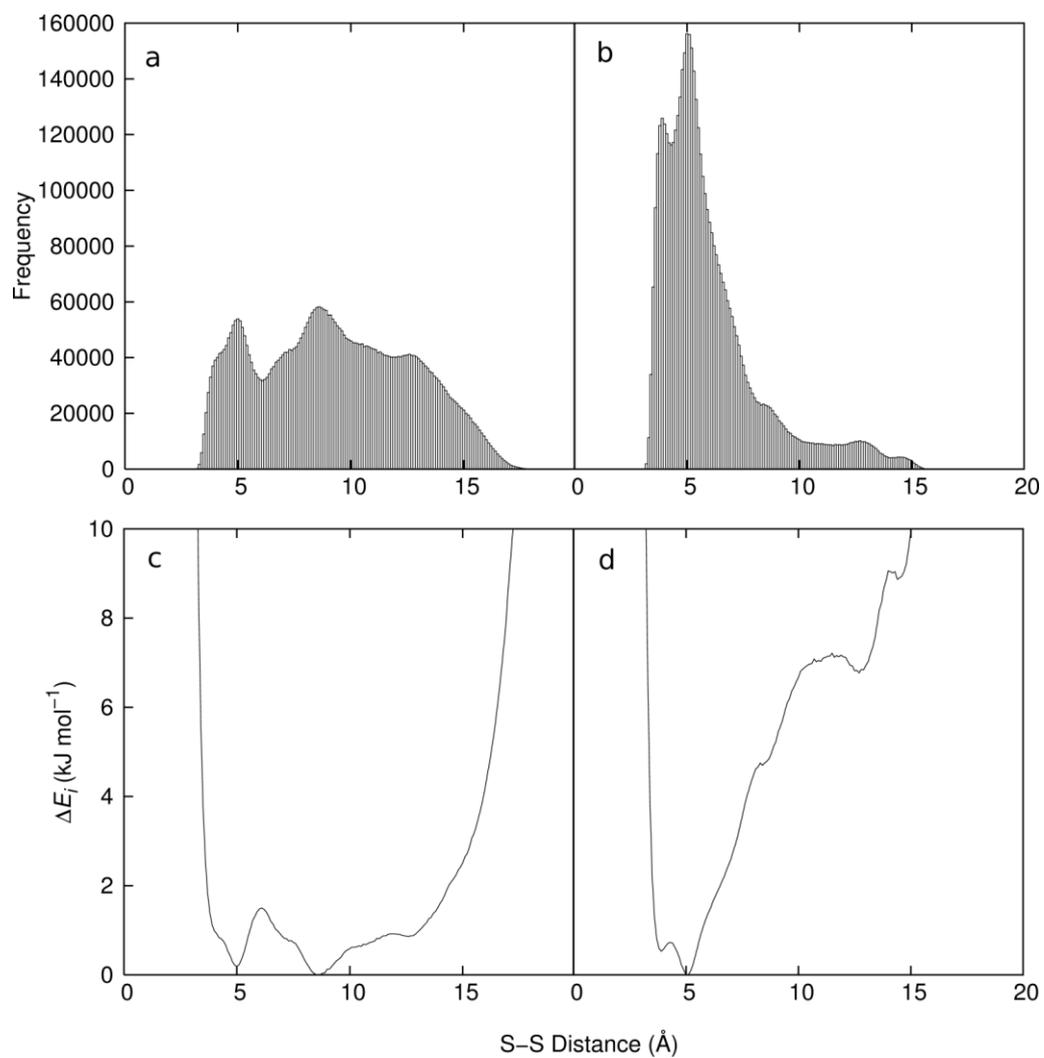


Figure 3-2 Histograms (upper panels) and corresponding potentials of mean force (lower panels) with respect to the S₃-S₇ (panels a and d) and S₈-S₁₁ (panels b and d) separations for nisin analogue 0, with four L-Cys residues.

3.5.1 Cysteine 3 – Cysteine 7 interactions

Nisin Analogue	Cysteine Chirality	ΔE at 5.5 Å (kJ mol ⁻¹)	$P(S_3-S_7)$	Location of first minimum (Å)
0	L-Cys3-L-Cys7	0.9	0.18	5.0 (0.2 kJ mol ⁻¹)
1		2.2	0.12	4.7 (1.6 kJ mol ⁻¹)
2		1.1	0.08	7.8
3		2.5	0.11	8.4
4	L-Cys3-D-Cys7	2.7	0.10	7.6
5		4.2	0.04	5.2 (3.9 kJ mol ⁻¹)
6		6.0	0.03	7.6
7		4.1	0.09	3.8 (2.8 kJ mol ⁻¹)
8	D-Cys3-L-Cys7	2.6	0.08	5.1 (2.3 kJ mol ⁻¹)
9		0.8	0.15	5.3 (0.7 kJ mol ⁻¹)
10		0.7	0.31	5.1
11		0.3	0.31	5.1
12	D-Cys3-D-Cys7	1.2	0.29	3.9 (0.6 kJ mol ⁻¹)
13		1.8	0.23	3.7 (0.3 kJ mol ⁻¹)
14		0.5	0.28	4.9
15		2.4	0.36	4.0

Table 3-2 Effect of Cys3 and Cys7 chirality on the PMF surface and the probability of S_3 and S_7 coming close enough to form a disulfide bridge. ΔE at 5.5 Å is the value of PMF when S_3 and S_7 are considered close enough to form a disulfide bridge; $P(S_3-S_7)$ is the probability of the sulfur-sulfur separation being less than 5.5 Å and location of first minimum gives the sulfur-sulfur separation of the first minimum on the PMF surface, with its value in parentheses if it is not the global minimum.

Table 3-2 summarises the effect of the chirality of Cys3 and Cys7 on the energetics of the S_3-S_7 interaction and $P(S_3-S_7)$, the probability of S_3-S_7 distance being less than 5.5 Å. For seven of the analogues where Cys3 is of the L-enantiomer (nisin analogues 1-7) there is either no minimum, or a shallow local minimum, corresponding to the possible formation of the 3-7 disulfide bridge, and $P(S_3-S_7)$ is quite low, between 0.03 and 0.12. Analogue 0 has a deep local minimum when S_3-S_7 is 5.0 Å and has a higher $P(S_3-S_7)$ of 0.18.

3. Conformational Preferences of Nisin Analogues

Changing the chirality of Cys7 from the L-enantiomer (analogues 0-3) to the D-enantiomer (analogues 4-7) decreases $P(S_3-S_7)$.

Six of the analogues with D-Cys3 favour the formation of the 3-7 disulfide bridge. Two of the analogues where Cys3 is D-enantiomer and Cys7 is L-enantiomer (analogues 10 and 11), and two analogues where both Cys3 and Cys7 are D-enantiomer (analogues 14 and 15) have a global energy minimum corresponding to the possible formation of the 3-7 disulfide bridge and $P(S_3-S_7)$ is between 0.28 and 0.36. The other two analogues where both Cys3 and Cys7 are the L-enantiomer (analogues 12 and 13) have deep local minima that correspond to the formation of the 3-7 disulfide bridge and $P(S_3-S_7)$ is 0.23 and 0.29. Analogues 8 and 9 have a lower $P(S_3-S_7)$ than the other D-Cys3 analogues (0.08 and 0.15) and the global energy minimum is at 13.4 Å and 13.5 Å, compared with between 4.0 Å and 8.9 Å for the other 14 analogues. Figure 3-3 shows two PMFs as a function of S_3-S_7 distance. Analogue 5 has L-Cys3-D-Cys7 and does not have a minimum corresponding to the S_3-S_7 interaction. Analogue 11 is D-Cys3-L-Cys7 and has a minimum corresponding to the sulfur-sulfur interaction.

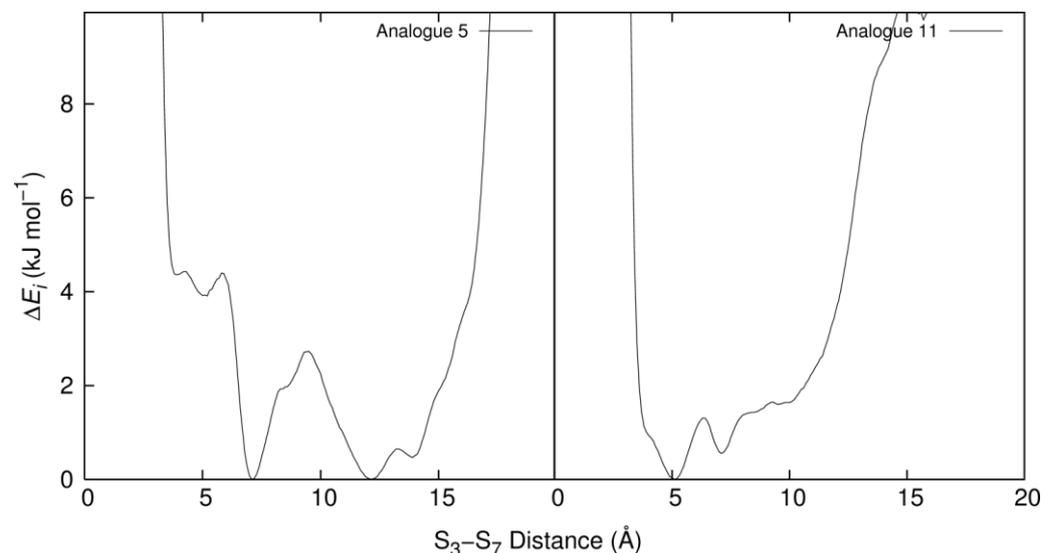


Figure 3-3 Examples of potential of mean force as a function of the S_3-S_7 separation; (left) analogue 5 (L-Cys3-D-Cys7-L-Cys8-D-Cys11), (right) analogue 11 (D-Cys3-L-Cys7-D-Cys8-D-Cys11).

3. Conformational Preferences of Nisin Analogues

The formation of the S₃-S₇ bridge in analogues 10-15 corresponds to a type IV β -turn between residues 3 to 6. A type IV β -turn is defined as four consecutive residues where the distance between the C _{α} atoms of the first and fourth residues is less than 7 Å and the ϕ , ψ dihedral angles of the central residues do not fit the criteria of the standard turn types or helices [29]. The correlation analysis did not identify any backbone hydrogen bonds or dihedral angles associated with the 3-7 disulfide bridge. For analogues 10-15, the 3-7 bridge can form in 29% of trajectory frames; 76% of these frames, or 22% of the total frames, correspond to a C _{α 3}-C _{α 6} distance of less than 7 Å. For analogues 0-9 the fraction of frames where both S₃-S₇ is less than 5.5 Å and C _{α 3}-C _{α 6} is less than 7 Å, is 5%. The mean average values of the central residues' (ϕ , ψ) angles are (-90° \pm 20°, -65° \pm 31°) for Ile4 and (-95° \pm 21°, -77° \pm 32°) for Ala5. These averages are calculated across analogues 10-15 when S₃-S₇ is less than 5.5 Å. The C _{α 3}-C _{α 6} distance and ϕ , ψ angles confirm that D-Cys3-Ile4-Ala5-Leu6 adopts a type IV β -turn when the 3-7 disulfide bridge can form.

3.5.2 Cysteine 8 – Cysteine 11 interactions

Nisin Analogue	Cysteine Chirality	ΔE at 5.5 Å (kJ mol ⁻¹)	$P(S_8-S_{11})$	Location of first minimum (Å)
0	L-Cys8-L-Cys11	0.6	0.54	5.0
4		0.4	0.49	5.2
8		0.5	0.25	5.1
12		0.4	0.32	5.2
1	L-Cys8-D-Cys11	1.3	0.19	5.1 (0.8 kJ mol ⁻¹)
5		2.4	0.07	8.7
9		2.3	0.18	5.2 (2.0 kJ mol ⁻¹)
13		1.0	0.18	5.1 (0.7 kJ mol ⁻¹)
2	D-Cys8-L-Cys11	2.2	0.07	6.9
6		5.3	0.02	7.4
10		1.1	0.11	8.0
14		4.1	0.03	10.7
3	D-Cys8-D-Cys11	1.7	0.08	9.5
7		7.7	0.01	9.6
11		2.4	0.06	9.4
15		6.9	0.01	12.0

Table 3-3 Effect of Cys8 and Cys11 chirality on the potential energy surface and the probability of Cys8(S) and Cys11(S) coming close enough to form a disulfide bridge (less than 5.5 Å). ΔE at 5.5 Å is the value of the PMF when S_8 and S_{11} are considered close enough to form a disulfide bridge; $P(S_8-S_{11})$ is the probability of the sulfur-sulfur separation being less than 5.5 Å and location of first minimum gives the sulfur-sulfur separation of the first minimum on the PMF surface, with its value in parentheses if it is not the global minimum.

The effect of the chirality of Cys8 and Cys11 on the energetics of the S_8-S_{11} interaction and on $P(S_8-S_{11})$ is summarised in Table 3-3. The nisin analogues with Cys8 and Cys11 as L-amino (0, 4, 8, 12) favour the formation of the 8-11 disulfide bridge with the global energy minimum corresponding to the possible formation of the bridge and $P(S_8-S_{11})$ between 0.32 and 0.54. When Cys11 is changed to D-amino and Cys8 remains L-amino (analogues 1, 5, 9, 13),

3. Conformational Preferences of Nisin Analogues

$P(S_8-S_{11})$ drops to between 0.07 and 0.19 and for three of the analogues there are local energy minima corresponding to the 8-11 bridge, but not a global minimum; the bridge is able to form, but it is less likely than the case of L-Cys8 and L-Cys11. A D-Cys8 stops the formation of the 8-11 disulfide bridge: analogues (2, 3, 6, 7, 10, 11, 14, 15) have lower values for $P(S_8-S_{11})$ and no energy minima corresponding to the formation of the 8-11 bridge. Figure 3-4 contains examples of PMF as a function of S_8-S_{11} distance; analogue 4 (L-Cys8-L-Cys11), with a minima corresponding to a S_8-S_{11} interaction, and analogue 7 (D-Cys8-D-Cys11) without.

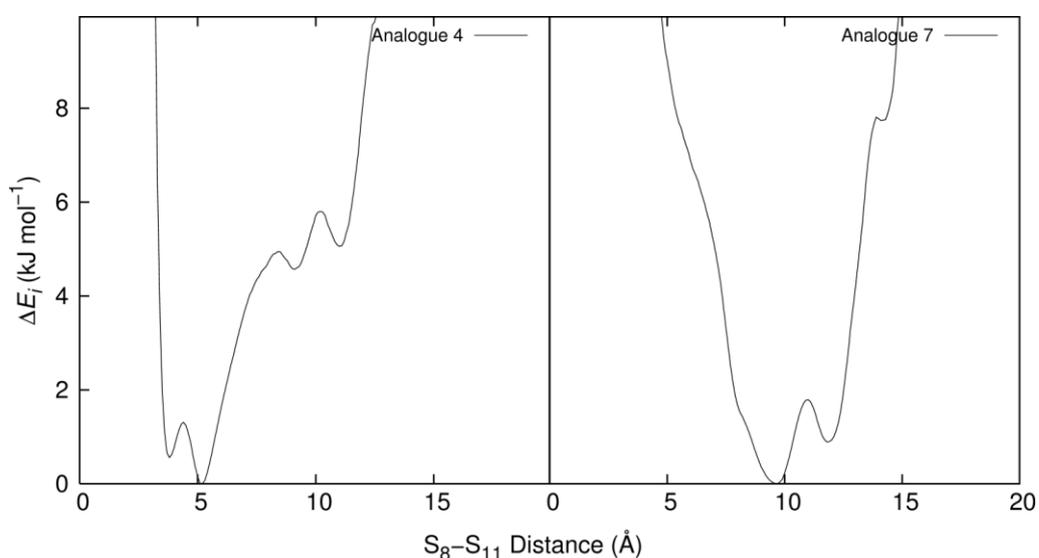


Figure 3-4 Examples of potential of mean force as a function of the S_8-S_{11} separation; (left) analogue 4 (L-Cys3-D-Cys7-L-Cys8-L-Cys11), (right) analogue 7 (L-Cys3-D-Cys7-D-Cys8-D-Cys11).

A backbone hydrogen bond between the backbone carbonyl of Cys8 and the backbone amide H of Cys11 and Lys12 was identified as being associated with the formation of the 8-11 disulfide bridge. Table 4 shows the average lifetime, average occupancy and number of trajectories where occupancy was greater than 1% for this hydrogen bond. For the analogues where both Cys8 and Cys11 are L-amino, the formation of the Cys8-Cys11/Lys12 hydrogen bond brings Cys8 and Cys11 close enough together that the S_8-S_{11} distance is less than 5.5 Å and the disulfide bridge could form. In simulations of analogues 4 and 12, the hydrogen bond and the disulfide bridge contact form simultaneously for parts of nine of the independent trajectories; during the

3. Conformational Preferences of Nisin Analogues

other trajectory neither the hydrogen bond nor the disulfide bridge form. Similarly, during simulations of analogue 8, the hydrogen bond is associated with the disulfide interaction for seven of the trajectories, and during the other three neither form. In simulations of analogue 0, the hydrogen bond and disulfide bridge contact are simultaneously formed for parts of all of the ten independent trajectories.

A hydrogen bond with lower occupancy forms between Cys8 and Cys11/Lys12 during simulations of the L-Cys8-D-Cys11 analogues 1, 5, 9 and 13; occurrences of this hydrogen bond do correlate with instances of the 8-11 disulfide bridge contact. Although the average lifetime is similar to the simulations of the analogues where both cysteines are L-amino, the lower occupancy reduces $P(S_8-S_{11})$ of the L-Cys8-D-Cys11 analogues, because there is no stabilising hydrogen bond keeping S_8 and S_{11} close together.

Analogues with a D-Cys8 are unable to form the disulfide bridge contact between S_8 and S_{11} . During the simulations of the analogues with D-Cys8 and L-Cys11 the Cys8-Cys11/Lys12 hydrogen bond was able to form, but it brought S_7 and S_{11} , rather than S_8 and S_{11} , close enough to form a disulfide bridge. For five of the trajectories of analogue 2 and four of the trajectories of analogue 14, Cys8 forms a stable hydrogen bond with Cys11 and Lys12(HN), which is associated with the S_7-S_{11} disulfide bridge contact. For three of the trajectories of analogue 6, the Cys8-Cys11/Lys12 hydrogen bond brings S_7-S_{11} together; for the other three trajectories the hydrogen bond forms between the carbonyl of Cys7 and the amine of Cys11. The interaction between S_7 and S_{11} may be blocked by the side chain of Cys8 when it is the L-enantiomer. Analogue 10 has a high occupancy (80%) and mean lifetime (21 ps) of the hydrogen bond compared with the other analogues. Although the Cys8-Cys11/Lys12(HN) hydrogen bond does correspond to the S_7-S_{11} interaction in analogue 10, the hydrogen bond is much more stable than the sulfur – sulfur interaction. Analogue 10 has a high probability for the formation of the 3-11 disulfide bridge, $P(S_3-S_{11}) = 0.59$, compared with the other 15 analogues, where $P(S_3-S_{11})$ is between 0.04 and 0.38 and the mean value is 0.16. The

3. Conformational Preferences of Nisin Analogues

Cys8-Cys11/Lys12 hydrogen bond corresponds to the 3-11 disulfide bridge for parts of nine of the ten trajectories of analogue 10. The Cys8-Cys11/Lys12 hydrogen bond brings Cys11 close to Cys7 and the d-Cys3-Ile4-Ala5-Leu6 type IV β -turn brings Cys3 close to Cys7 and therefore close to Cys11. During the simulations of the analogues with D-Cys8 and D-Cys11, the Cys8-Cys11/Lys12 hydrogen bond has low mean lifetimes, 4 ps to 7 ps, and low occupancies, 1% to 5%, compared with the other analogues and $P(S_8-S_{11})$ is low.

During the simulations of the L-Cys8, L-Cys11 analogues, residues 8 to 11 form a type IV β -turn. Across analogues 0, 4, 8 and 12 the average percentage of trajectory frames where the 8-11 bridge can form is 40%; 97% of these frames correspond to a $C_{\alpha 8}-C_{\alpha 11}$ distance of less than 7 Å. The average values of the central residues' (φ , ψ) angles, which characterise a β -turn, are $(-72^\circ \pm 9^\circ, -29^\circ \pm 70^\circ)$ for Pro9 and $(-128^\circ \pm 78^\circ, -34^\circ \pm 42^\circ)$ for Gly10. These averages are calculated across all L-Cys8, L-Cys11 analogues when S_8-S_{11} is less than 5.5 Å and the uncertainty is the angular deviation.

3.6 Relative energy maps for bead and globular connectivities

Three-dimensional histograms of S_3-S_7 and S_8-S_{11} distances with an interval of 0.1 Å were used to produce PMF surfaces using the Boltzmann relation. Figure 3-5 is the PMF for S_3-S_7 and S_8-S_{11} for analogue 12, and has a global minimum at (3.8 Å, 3.8 Å), corresponding to the bead connectivity. The S_3-S_7 and S_8-S_{11} distances during the nisin analogue 12 trajectories were checked every 2.0 ps and the atomic co-ordinates written for those that corresponded to the bead connectivity. The resulting conformations were clustered using the kclust module in MMTSB [30] and Figure 3-6 shows the centroid of the most populated cluster. Analogues 7 and 13 are not able to form the bead connectivity, but do have a global minimum corresponding to the globular connectivity at (5.0 Å, 5.0 Å) and (3.7 Å, 4.6 Å), respectively (Figure 3-7 and Figure 3-8). Figure 3-9 shows the centroid of the most populated cluster from clustering of trajectory frames, corresponding to the globular connectivity, from the simulations of nisin analogue 13.

3. Conformational Preferences of Nisin Analogues

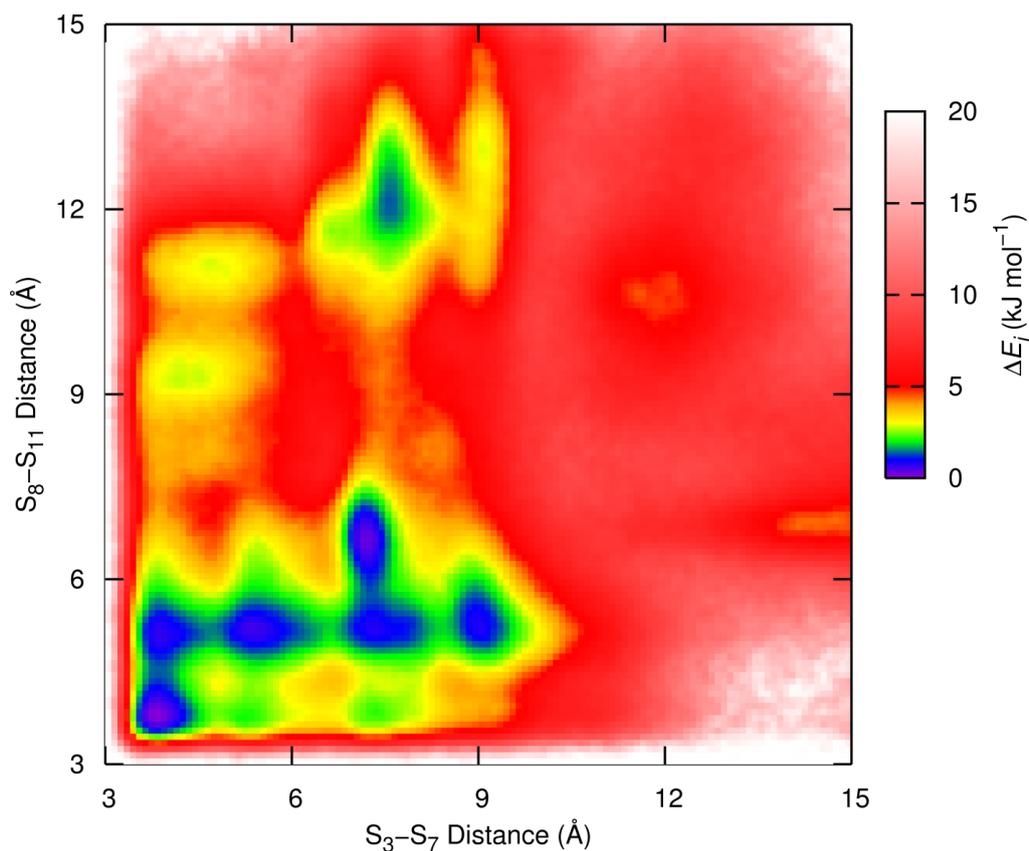


Figure 3-5 Potential of mean force as a function of S_3 - S_7 and S_8 - S_{11} separation (bead connectivity) for nisin analogue 12.

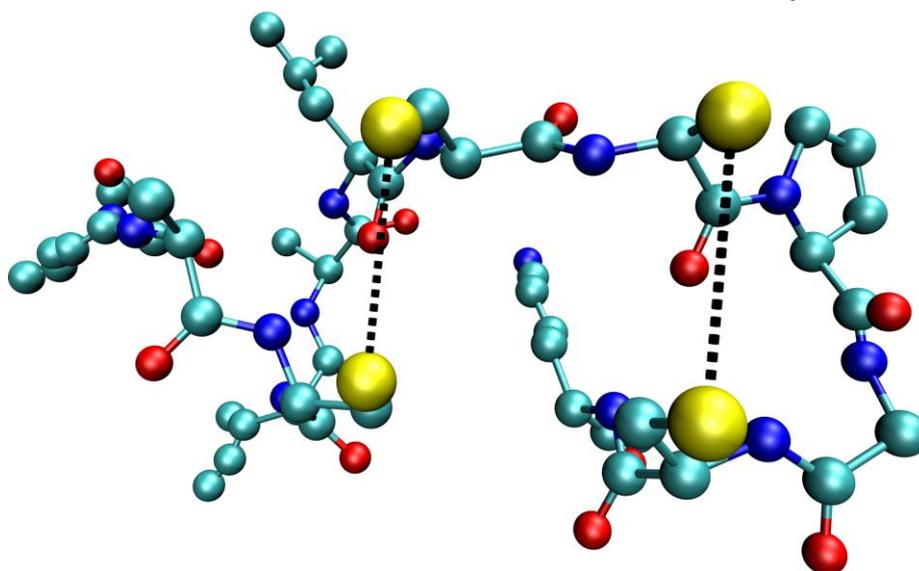


Figure 3-6 The centroid of the most populated cluster from clustering of trajectories frames from the simulations of nisin analogue 12, that corresponded to the bead connectivity. The 3-7 and 8-11 interactions are illustrated by a dashed line.

3. Conformational Preferences of Nisin Analogues

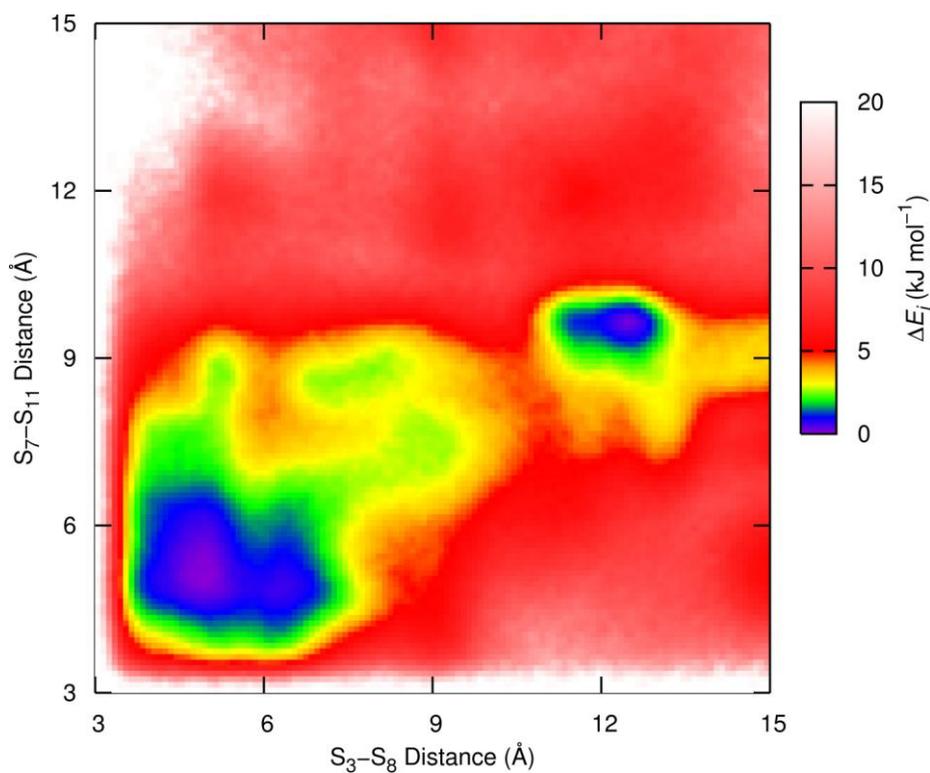


Figure 3-7 Potential of mean force as a function of S_3 - S_8 and S_7 - S_{11} separation (globular connectivity) for nisin analogue 7.

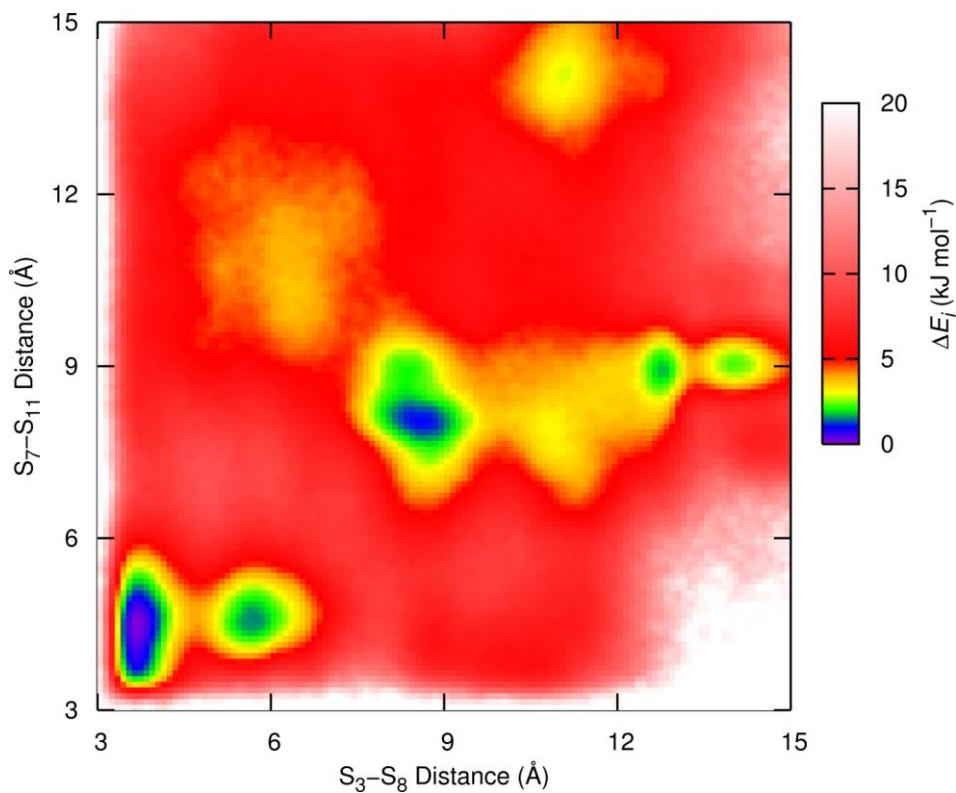


Figure 3-8 Potential of mean force as a function of S_3 - S_8 and S_7 - S_{11} separation (globular connectivity) for nisin analogue 13.

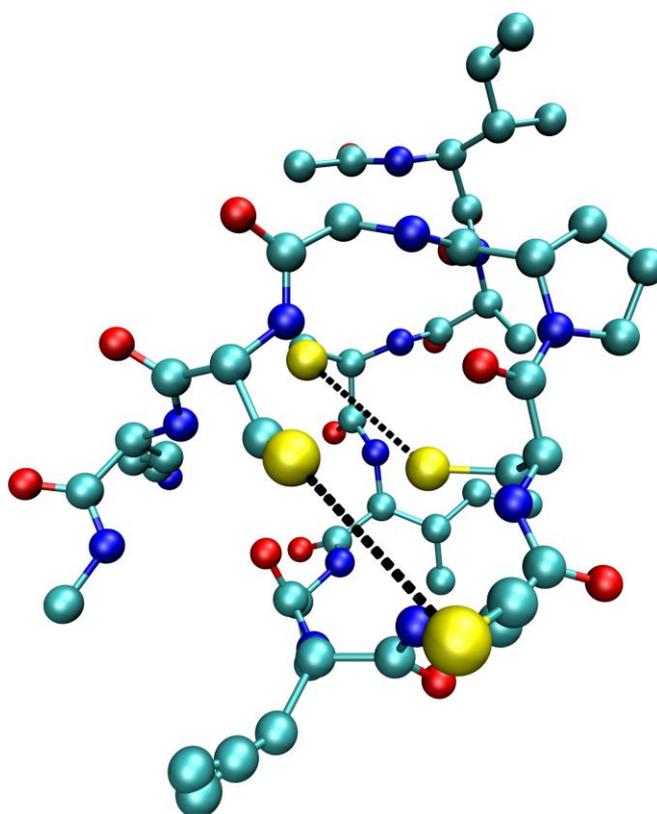


Figure 3-9 The centroid of the most populated cluster from clustering of trajectories frames from the simulations of nisin analogue 13, that corresponded to the globular connectivity. The 3-7 and 8-11 interactions are illustrated by a dashed line.

3.7 Cluster Analysis of Nisin Analogue 12

The RMS differences between the peptide backbone atoms of the first 12 residues of the NMR structure of nisin in complex with lipid II [9] (Protein Data Bank reference: 1WCO) and the four most populated clusters' centroids calculated for nisin analogue 12 are shown in Table 5.

Cluster (percentage of structures)	Residues used to calculate RMSD from 1WCO (Å)			
	3 to 7	8 to 11	2 to 11	1 to 12
1 (30.0%)	1.1	1.0	3.5	4.5
2 (18.2%)	1.6	0.6	2.6	3.1
3 (13.5%)	1.7	0.6	2.6	3.0
4 (9.4%)	1.6	0.7	2.8	3.7

Table 3-4 RMS difference between the peptide backbone atoms of the four most populated clusters' centroids of nisin analogue 12 and 1WCO. The percentage of conformations assigned to each cluster is shown in brackets.

3. Conformational Preferences of Nisin Analogues

When the RMS difference from 1WCO is calculated with respect to all the residues, 1 to 12, or excluding the termini, residues 2 to 11, the deviation is quite high. However, just using the thioether/ disulfide rings, and neglecting the flexible linking residues, RMS difference for residues 3 to 7 is less than 1.7 Å for clusters 2, 3 and 4 and is 1.1 Å for cluster 1. For residues 8 to 11 it is 1.0 Å for cluster 1 and less than 0.7 Å for clusters 2, 3 and 4. Figure 10 shows an alignment of backbone atoms for residues 3 to 7 between 1WCO and cluster 1 and Figure 11 shows an alignment of backbone atoms for residues 8 to 11 between 1WCO and clusters 2,3 and 4. The hydrogen bonding of the centroids of the four most populated clusters was also examined. Clusters 2, 3 and 4 have hydrogen bonds between the carbonyl of Cys8 and the backbone amides of Cys11 and Lys12 and between the carbonyl of Cys3 and amides of Leu6 and Cys7. Cluster 1 has hydrogen bonds between the carbonyl of Cys8 and amides of Cys11 and Lys12; the carbonyl of Ile1 and amides of Ala5 and Leu6; the carbonyl of Leu6 and amide of Cys3 and the carbonyl of Ala5 and the sidechain of Lys12. Clusters 2, 3 and 4 are very similar and only differ by the position of the N terminus. Cluster 1 is different to the other top three clusters, with the extra hydrogen bonds between the N-terminus, C-terminus and Ala5 leading to 'tucked in' termini.

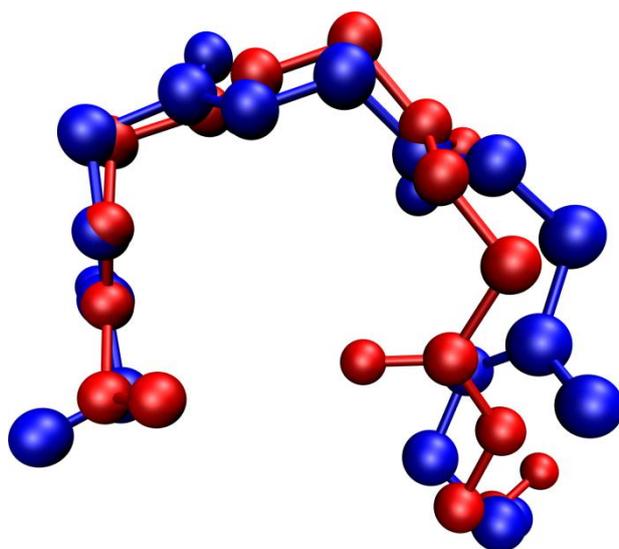


Figure 3-10 Alignment of backbone atoms for residues 3 to 7 between 1WCO (blue) and the centroid of cluster 1 (red) for nisin analogue 12.

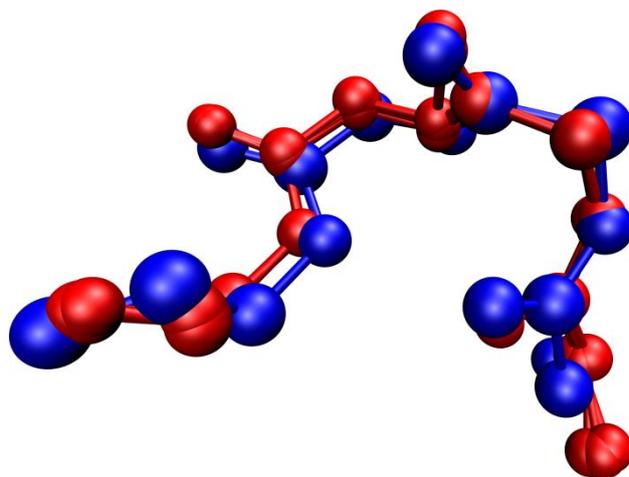


Figure 3-11 Alignment of backbone atoms for residues 8 to 11 between 1WCO (blue) and the centroids of clusters 2,3 and 4 (red) of nisin analogue 12.

3.8 Discussion

Mitchell and Smith [31] surveyed all peptide and protein entries in the Protein Data Bank for D-amino residues and observed that D-amino acid residues in L-amino acid chains have a propensity for forming β -turns. Of a representative subset of 40 D-amino acid residues in L-amino chains, where duplicate sequences had been removed, 17 of the residues were involved in a β -turn, the majority of which were classed as non-standard, or type IV. The formation of the 3-6 β -turn when Cys3 is a D-amino acid residue brings the sulfur atoms in Cys3 and Cys7 close enough to form a disulfide bridge, for two of the L-Cys7 analogues (10, 11) and all of the D-Cys7 analogues (12-15). When the Cys3 is L-amino, the 3-6 β -turn does not form and $P(S_3-S_7)$ is reduced.

Ring B is conserved across type A lantibiotics [9] and a hydrogen bond similar to the Cys8-Cys11/Lys12 hydrogen bond has been observed during NMR experiments: between D-Ala8 and Gly10 and Ala11 in nisin [5], between D-Ala8 and Ala11 in gallidermin [32] and between D-Ala8 and Gly10 and Ala11 in mutacin 1140 [33]. Ring B is described as a type II β -turn for the nisin and gallidermin structures [5, 32] and alternates between a type II and type I β -turn for the mutacin 1140 structure [33]. Hsu *et al.* [34] substituted different amino acids into a β -hairpin scaffold between cysteine residues to study their

3. Conformational Preferences of Nisin Analogues

propensity for forming β -turns and found that Cys-Pro-Gly-Cys particularly favours β -turn formation. Proline and glycine are residues 9 and 10 in the nisin analogues.

In the simulations of the nisin analogues, the 8-11 disulfide interaction is more stable than the 3-7 disulfide interaction. The 8-11 interaction is stabilised by the hydrogen bond between Cys8 and Cys11/Lys12 and by the amino acid sequence Cys-Pro-Gly-Cys, which favours a β -turn conformation. Another factor in the difference between the 8-11 and 3-7 interactions is that the stability of small disulfide loops depends on whether the number of residues between the cysteines is odd or even. Zhang and Snyder [35] measured the microscopic disulfide-exchange rate constants for C- X_m -C peptides, where m is the number of residues between the cysteines. They found that an even value of m favours disulfide bond formation more than an odd value, with $m=2$ and $m=4$ most favoured for values of m up to 5.

3.9 Conclusions

Naturally occurring antibiotic peptides are a source of new drug leads to help overcome antibiotic resistance. This work studies analogues of the first 12 residues of nisin, modified so that the thioether bonds have been replaced with disulfide bonds between cysteines. The residue sequence is the same for all the analogues with the chirality of the four cysteine residues varied through all 16 possible combinations.

50 ns Langevin dynamics simulations of the analogues, without the disulfide bonds between the cysteine residues, have been performed using the generalised Born implicit solvent model. The analysis of the simulations show that a D-Cys3 favours conformations corresponding to a S_3 - S_7 interaction and that L-Cys8 with L-Cys11 favours conformations corresponding to a S_8 - S_{11} interaction.

For six of the eight analogues with D-Cys3 there is an interaction between S_3 - S_7 . A survey of D-amino residues in L-amino chains[31] reports that D-amino residues have a propensity for forming non-standard β -turns. The sulfur

3. Conformational Preferences of Nisin Analogues

atoms in Cys3 and Cys7 are brought together by a type-IV β -turn between residues D-Cys3-Ile4-Ala5-Leu6, characterised by central residue (φ , ψ) angles of $(-90^\circ \pm 20^\circ, -65^\circ \pm 31^\circ)$ for Ile4 and $(-95^\circ \pm 21^\circ, -77^\circ \pm 32^\circ)$ for Ala5.

The interaction between S_8 - S_{11} is due a type-IV β -turn between Cys8-Pro9-Gly10-Cys11, stabilized by a hydrogen bond between 8(O) and 11(HN)/12(HN). This hydrogen bond is most stable for analogues with L-Cys8 and L-Cys11, but also forms for L-Cys8 with D-Cys11, corresponding a less stable S_8 - S_{11} interaction, and D-Cys8 with L-Cys11, corresponding to an interaction between S_7 - S_{11} , not S_8 - S_{11} . A hydrogen bond has been observed during NMR experiments between D-Ala8(CO) and Gly10(HN) and Ala11(HN) in nisin[5], between D-Ala8(CO) and Ala11(HN) in gallidermin[32] and between D-Ala8(CO) and Gly10(HN) and Ala11(HN) in mutacin 1140[33]. The Cys-Pro-Gly-Cys motif has been shown experimentally to favour β -turn formation [34]. The 8-11 interaction is more stable than the 3-7 interaction because of the 8(O) and 11(HN)/12(HN) hydrogen bond, the Cys-Pro-Gly-Cys motif which favours a β -turn and because disulfide loops with an even number of residues between the cysteines have been shown to cyclise more easily than those with an odd number.

Nisin analogue 12 (D-Cys3-D-Cys7-L-Cys8-L-Cys11) has a global minimum on the relative energy surface corresponding to the simultaneous formation of the S_3 - S_7 and S_8 - S_{11} disulfide bonds. Trajectory frames corresponding to simultaneous 3-7, 8-11 interactions were clustered and the centroids of the top four clusters compared to the first 12 residues of an NMR structure of nisin in complex with lipid II [9]. The backbone RMSD calculated for the rings only is between 0.6 and 1.7 Å, suggesting that the disulfide analogues are worth pursuing further as possible new peptide antibiotics.

3.10 References

1. WHO, *Antibiotic resistance: synthesis of recommendations by expert policy groups*. 2001, WHO.
2. Asaduzzaman, S.M. and K. Sonomoto, *Lantibiotics: Diverse activities and unique modes of action*. *J. Biosci. Bioeng.*, 2009. **107**(5): p. 475-487.
3. Belitz, H.D., W. Grosch, and P. Schieberle, *Food Chemistry*. 4th ed. 2009, Berlin: Springer-Verlag.
4. Lubelski, J., R. Khusainov, and O.P. Kuipers, *Directionality and coordination of dehydration and ring formation during biosynthesis of the lantibiotic nisin*. *J. Biol. Chem.*, 2009. **284**(38): p. 25962-25972.
5. Van Den Hooven, H.W., C.C.M. Doeland, M. Van De Kamp, R.N.H. Konings, C.W. Hilbers, and F.J.M. Van De Ven, *Three-dimensional structure of the lantibiotic nisin in the presence of membrane-mimetic micelles of dodecylphosphocholine and of sodium dodecylsulphate*. *Eur. J. Biochem.*, 1996. **235**(1-2): p. 382-393.
6. Hasper, H.E., B. de Kruijff, and E. Breukink, *Assembly and stability of nisin-lipid II pores*. *Biochemistry*, 2004. **43**: p. 11567-11575.
7. Hasper, H.E., N.E. Kramer, J.L. Smith, J.D. Hillman, C. Zachariah, O.P. Kuipers, B. de Kruijff, and E. Breukink, *An alternative bactericidal mechanism of action for lantibiotic peptides that target lipid II*. *Science*, 2006. **313**: p. 1636-1637.
8. Hyde, A.J., J. Parisot, A. McNichol, and B.B. Bonev, *Nisin-induced changes in Bacillus morphology suggest a paradigm of antibiotic action*. *Proc. Natl. Acad. Sci. USA*, 2006. **103**(52): p. 19896-19901.
9. Hsu, S.T.D., E. Breukink, E. Tischenko, M.A.G. Lutters, B. de Kruijff, R. Kaptein, A. Bonvin, and N.A.J. van Nuland, *The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics*. *Nat. Struct. Mol. Biol.*, 2004. **11**(10): p. 963-967.
10. Daly, N.L. and D.J. Craik, *Folding motifs of cysteine-rich peptides*, in *Oxidative folding of peptides and proteins*, J. Buchner and L. Moroder, Editors. 2009, RSC: Cambridge.
11. Bulaj, G. and A. Walewska, *Oxidative folding of single-stranded disulfide-rich peptides*, in *Oxidative folding of peptides and proteins*, J. Buchner and L. Moroder, Editors. 2009, RSC: Cambridge.
12. Livett, B.G., K.R. Gayler, and Z. Khalil, *Drugs from the sea: conopeptides as potential therapeutics*. *Curr. Med. Chem.*, 2004. **11**(13): p. 1715-1723.
13. Gehrman, J., P.F. Alewood, and D.J. Craik, *Structure determination of the three disulfide bond isomers of α -conotoxin G1: a model for the role of disulfide bonds in structural stability*. *J. Mol. Biol.*, 1998. **278**: p. 401-415.
14. Bondebjerg, J., M. Grunnet, T. Jespersen, and M. Meldal, *Solid-phase synthesis and biological activity of a thioether analogue of conotoxin G1*. *ChemBioChem* 2003. **4**: p. 186-194.

3. Conformational Preferences of Nisin Analogues

15. Levensgood, M.R. and W. van der Donk, *Use of lantibiotic synthetases for the preparation of bioactive constrained peptides*. *Bioorg. Med. Chem.*, 2008. **18**: p. 3025-3028.
16. Schmid, N., C. Bolliger, L.J. Smith, and W.F. van Gunsteren, *Disulfide bond shuffling in bovine alpha-Lactalbumin: MD simulation confirms experiment*. *Biochemistry*, 2008. **47**(46): p. 12104-12107.
17. Foloppe, N., J. Sagemark, K. Nordstrand, K.D. Berndt, and L. Nilsson, *Structure, dynamics and electrostatics of the active site of glutaredoxin 3 from Escherichia coli: comparison with functionally related proteins*. *J. Mol. Biol.*, 2001. **310**: p. 449-470.
18. Foloppe, N. and L. Nilsson, *The glutaredoxin -C-P-Y-C- motif: influence of peripheral residues*. *Structure*, 2004. **12**(2): p. 289-300.
19. Foloppe, N. and L. Nilsson, *Stabilization of the catalytic thiolate in a mammalian glutaredoxin: structure, dynamics and electrostatics of reduced pig glutaredoxin and its mutants*. *J. Mol. Biol.*, 2007. **372**: p. 798-816.
20. Roos, G., N. Foloppe, K. van Laer, L. Wyns, L. Nilsson, P. Geerlings, and J. Messens, *How thioredoxin dissociates its mixed disulfide*. *PLoS Comp. Biol.*, 2009. **5**(8): p. e1000461.
21. Brooks, B.R., R.E. Bruccoleri, D.J. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*. *J. Comput. Chem.*, 1983. **4**: p. 187-217.
22. Brooks, B.R., C.L. Brooks, A.D. Mackerell Jr, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus, *CHARMM: The Biomolecular Simulation Program*. *J. Comput. Chem.*, 2009. **30**(10): p. 1545-1614.
23. Mackerell Jr, A.D., M. Feig, and C.L. Brooks III, *Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations*. *J. Comput. Chem.*, 2004. **25**(11): p. 1400-1415.
24. Mackerell Jr, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. *J. Phys. Chem. B*, 1998. **102**(18): p. 3586-3616.
25. Im, W.P., M.S. Lee, and C.L. Brooks, *Generalized Born model with a simple smoothing function*. *J. Comput. Chem.*, 2003. **24**(14): p. 1691-1702.

3. Conformational Preferences of Nisin Analogues

26. Nina, M., D. Beglov, and B. Roux, *Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations*. J. Phys. Chem. B, 1997. **101**(26): p. 5239-5248.
27. Yeh, I. and A. Wallqvist, *Structure and dynamics of end-to-end loop formation of the penta-peptide Cys-Ala-Gly-Gln-Trp in implicit solvents*. J. Phys. Chem. B, 2009. **113**: p. 12382-12390.
28. Mardia, K.V., *Linear-Circular Correlation Coefficients and Rhythmometry*. Biometrika, 1976. **63**(2): p. 403-405.
29. Hutchinson, E.G. and J.M. Thornton, *PROMOTIF - A program to identify and analyze structural motifs in proteins*. Protein Sci., 1995. **5**: p. 200-212.
30. Feig, M, J. Karanicolas, and C. Brooks III, L., *MMTSB Tool Set*. 2001, MMTSB NIH Research Resource, The Scripps Research Institute.
31. Mitchell, J.B.O. and J. Smith, *D-amino acid residues in peptides and proteins*. Proteins, 2003. **50**: p. 563-571.
32. Freund, S., G. Jung, O. Gutbrod, G. Folkers, W.A. Gibbons, H. Allgaier, and R. Werner, *The solution structure of the lantibiotic gallidermin*. Biopolymers, 1991. **31**(6): p. 803-811.
33. Smith, L., C. Zachariah, R. Thirumoorthy, J. Rocca, J. Novak, J.D. Hillman, and A.S. Edison, *Structure and dynamics of the lantibiotic mutacin 1140*. Biochemistry, 2003. **42**: p. 10372-10384.
34. Hsu, H.-J., H.-J. Chang, H.-P. Peng, S.-S. Huang, M.-Y. Lin, and A.-S. Yang, *Assessing computational amino acid β -turn propensities with a phage-displayed combinatorial library and directed evolution*. Structure, 2006. **14**: p. 1499-1510.
35. Zhang, R. and G.H. Snyder, *Dependence of formation of small disulfide loops in two-cysteine peptides on the number and types of intervening amino acids*. J. Biol. Chem., 1989. **264**(31): p. 18472-18479.

4 Extending the CHARMM Potential Energy Function for the Nisin Analogues

4.1 Introduction

This chapter focuses on developing methods to improve and develop the simulations of the nisin analogues (Chapter 3), and by extension, simulations of other peptides that are cysteine-rich or contain D-amino residues. In the first section, the corrective map term of the CHARMM force field is transformed to describe D-amino residues correctly, improving modelling of β -sheets and β -turns. In the second section, explicit solvent simulations of the nisin analogues are performed using no correction and the transformed correction. The results of these simulations confirm that the transformed map is needed to simulate β -turns and agrees with the predictions of the implicit solvent simulations. The similarity between the implicit simulations without the corrective map and the explicit simulations with the corrective map suggests that explicit simulations are more sensitive to the potential energy function than implicit simulations. In the final section, potential energy surfaces (PES) are calculated as a function of the sulfur-sulfur distance for ring A and ring B of the nisin analogue 12 using Car-Parrinello molecular dynamics (CPMD) [1]. The intent was to use the PES to fit a Morse potential to describe the stretching and breaking of the disulfide bonds for use in a reactive dynamics simulation, which jumps between surfaces to model dynamics events. However, the crossing point between the thiolate PES and the CPMD PES was too high to be reached during normal dynamics, indicating that the CPMD PES is unsuitable for reactive dynamics.

4.2 The CMAP term and D-amino residues

The quality of MD simulations depends on the empirically derived parameters within the force field. One of the most successful and widely used potentials is the CHARMM force field; it was used in both the first million atom simulation [2] and recently in studies performed on the supercomputer ANTON, a

machine with dedicated hardware to enable millisecond MD simulations of solvated biomolecules [3].

Amino acids are chiral and nearly all those in the proteins and peptides of living organisms are the L-enantiomer, as specific enzymes stop D-amino residues being delivered to the ribosome by tRNA [4]. Any D-amino residues that do occur are formed by post-translation modifications to the side chain by enzymes and are mainly found in bacterial cell walls [5]. However, despite the rarity in wild-type polypeptides, the use of D-amino residues and all D-proteins is a recognised tool in biochemistry. They can be used to study the role of chirality in biological mechanisms, for example, by identifying which killing actions of human defensins depend on chiral recognition [6], or where a racemic mixture forms crystals for X-ray structure determination more readily than the all L-enantiomer solution [7]. The use of D-amino acids in biochemical experiments and the interest in bacterial peptides due to antibiotic resistance has motivated MD simulations of molecules that contain D-amino residues [8], [9], [10], [11], [12]. However, as demonstrated here, care must be taken when performing such simulations using the CHARMM force field; the dihedral corrective term must be transformed, due to its dependence on chirality.

The CHARMM potential energy function [13] is given in Equation 2-21. The bonded and non-bonded (Lennard-Jones and electrostatics) terms are well established and their form and the symbols in the equations are discussed in Chapter 2. The final term in the above expression is the CMAP correction. It is a function of the φ and ψ angles of the peptide backbone. The φ angle is the dihedral about the C_{α} -C bond and ψ angle is the dihedral about the N- C_{α} bond. The CMAP term was added to the CHARMM potential energy function in 2004 [14] to correct the limitations of the original parameter set from 1998 [15].

The CMAP term is implemented through a grid based map with points in 15° intervals of φ and ψ from -180° to 180°. Values between grid points are interpolated by a bicubic method [16], which requires the function and first-

order derivatives to be defined at a grid point and ensures that they are smooth and continuous across grid squares. The correction aims to minimise the difference between the energy surface calculated by CHARMM22 and a surface calculated at the MP2 level for gas phase alanine, glycine and proline dipeptides. The correction is empirically adjusted further to match a potential of mean force (PMF) from a survey of the Protein Data Bank (PDB) [17] to take into account condensed phase contributions.

The CMAP correction addresses two problems. The first is that in the original parameterisation dihedral parameters were adjusted to describe the right-handed α -helix (α_R) region of the Ramachandran plot, centred around $(-55^\circ, -45^\circ)$, accurately at the expense of the left-handed α -helix (α_L) region at around $(55^\circ, 45^\circ)$. This caused problems for glycine-rich proteins that can sample α_L -helical conformations. The second issue was that the energy barrier between α -helices, which are energetically favoured and very abundant in nature, and π -helices $(-55^\circ, -70^\circ)$, which are rare and energetically unfavourable, was too low, resulting in over-sampling of π -helices during simulations. Figure 4-1 illustrates how the correction overcomes these problems; there is a minimum in the α_L region, which enhances sampling of that area, and a barrier in the π -helical region to discourage sampling of that area.

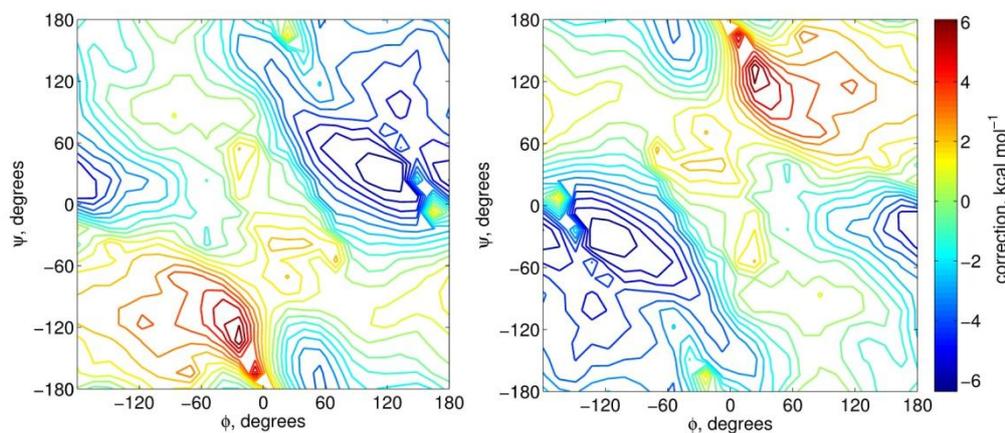


Figure 4-1 Contour map generated from the standard CMAP corrective grid (left) and the transformed map for D-amino residues (right).

The calculations for the CMAP correction were performed using L-amino dipeptides. Therefore, in the standard version of the CHARMM force field, this term is invalid for D-amino acids. The transformation $(\varphi, \psi) \rightarrow (-\varphi, -\psi)$ transforms the Ramachandran plot for L-amino residues into one that describes D-amino residues [18]. Applying the same transformation to the standard CMAP corrective grid (L-CMAP) should lead to a corrective term that improves the description of D-amino (D-CMAP) residues during MD simulations. To confirm this, we have identified from the PDB X-ray crystallographic structures of several proteins and peptides containing D-amino residues and we have performed explicit solvent simulations following the protocol used in the original CMAP paper [14]. The proteins and peptides simulated are described in Table 4-1 and Section 4.2.1.2.

4.2.1 Testing the D-CMAP term

Name	PDB code	Weight (Da)	No. residues	Edge length of water box (Å)	No. water molecules
D-Monellin	2Q33	10,692	43 (Chain A) 48 (Chain B)	79.4	7,655
D-Snowflea antifreeze protein	3BOG	6,535	81	78.8	7,600
HNP2 G17D-A mutant	1ZMK	6,953	2 x 29 (dimer)	68.1	4,628
D-HNP1	3GO0	7,011	2 x 30 (dimer)	64.7	3,905

Table 4-1 Properties of the simulated polypeptides.

4.2.1.1 Simulations

The same simulation protocol was followed as used in the explicit solvent simulations reported by MacKerell *et al.* [14]. Each polypeptide was inserted into a truncated octahedron of TIP3P [19] water molecules that has dimensions such that the solute is at least 12 Å from the edge of the solvent box. The size of each box is shown in Table 4-1. The molecular system was built and analysed in CHARMM 34b [20] [13]. The coordinates in the PDB files were used as the initial configuration; D-monellin, HNP2 G17D-A mutant and D-HNP1 were simulated as dimers in their biological configuration. Minimisation, equilibration and isothermal-isobaric (NPT) production dynamics were performed in NAMD 2.7 [21]. Periodic boundary conditions were applied to the system. The equilibration time was 1 ns at 298 K, the

production time was 5 ns for the human defensins and 100 ns for D-monellin and D-snowflea antifreeze protein. The Verlet leap frog algorithm [22] was used to propagate the system and the time step was 2 fs.

The simulation protocol was repeated three times for each polypeptide using the CHARMM22 force field and either: (a) without any correction to the potential energy field (referred to as none); (b) the D-CMAP correction applied to the D-residues and the L-CMAP correction applied to the L-residues (referred to as D-CMAP); and (c) with the L-CMAP correction applied to all the residues, irrespective of chirality (referred to as L-CMAP). A file containing the topologies and parameters for D-amino residues based upon CHARMM22 is included in the Appendix. It was expected that the simulations with the D-CMAP correction would have average backbone dihedral values closer to the original crystal structures than the simulations without the correction. In addition, it was anticipated that the simulations with the L-CMAP correction would have worse agreement with experiment.

4.2.1.2 The polypeptides studied

Monellin is a plant protein; it is a naturally occurring sweetener which is 70,000 times sweeter than sucrose. The D-variant (PDB reference 2Q33) was prepared by Hung *et al.* [23] to investigate how chirality affects sweetness and the crystal structure was resolved to confirm that D-monellin is the mirror image of L-monellin. Monellin has two chains; the A chain has 43 residues and the B chain has 48. Its secondary structure comprises a four strand anti-parallel β -sheet and an α -helix.

Snow flea antifreeze protein (sfAFP) has an unusual structure of six polyproline II helices, that have φ and ψ angles of about $(-75^\circ, 150^\circ)$, stacked to form an oblong brick-shape. One side is hydrophobic and the other polar; this relates to its biological function of disrupting the formation of ice crystals. The structure of D-sfAFP was resolved as one of the chains in a cell that was prepared from a racemic solution; Pentelute *et al.* [7] found that the mixture of enantiomers crystallised much more easily than a solution of pure L-sfAFP.

4. Extending the CHARMM Potential Energy Function for the Nisin Analogues

Both human defensins simulated have the same structure: a dimer consisting of monomers with 29 (1ZMK) and 30 (3GO0) residues and a three strand anti-parallel β -sheet. 1ZMK [24] is the human neutrophil α -defensin 2 (HNP2). It has a Gly17_D-Ala substitution, which was introduced to study the β -bulge region. 3GO0 [6] is a _D-variant of the human neutrophil α -defensin 1 (HNP1), prepared to gain insight into killing actions of the _L-variant.

4.2.2 D-CMAP Results

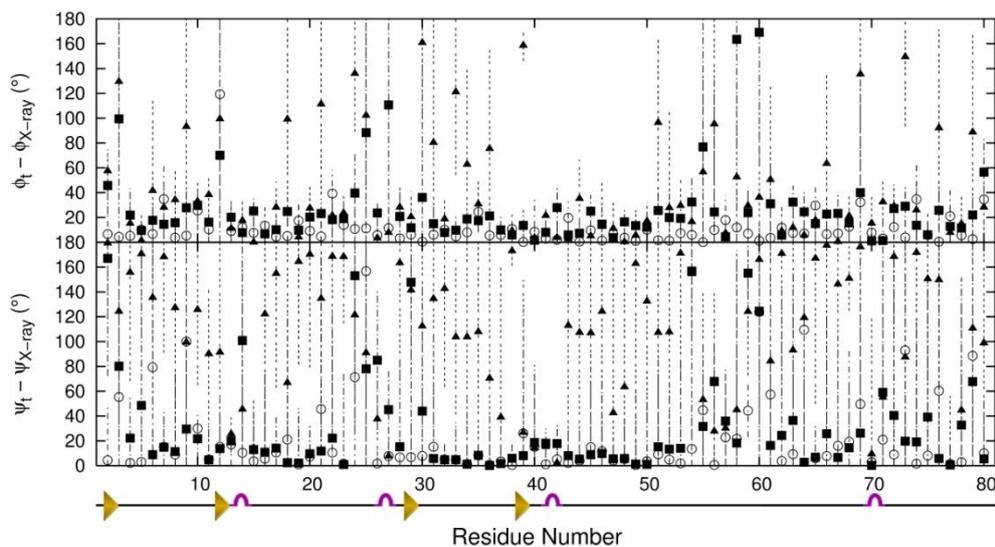


Figure 4-2 Absolute difference between MD time-average and experimental values of φ , ψ for D-sfAFP. Circles correspond to D-CMAP simulations, squares to no correction and triangles to L-CMAP. The error bars are the standard deviation of the MD time-averages, dashed lines correspond to D-CMAP simulations, alternating dots and dashes to no correction and dots to L-CMAP. Secondary structure assignment is taken from the PDB. sfAFP has an unusual structure of six poly-proline II helices, which do not have a secondary structure illustration.

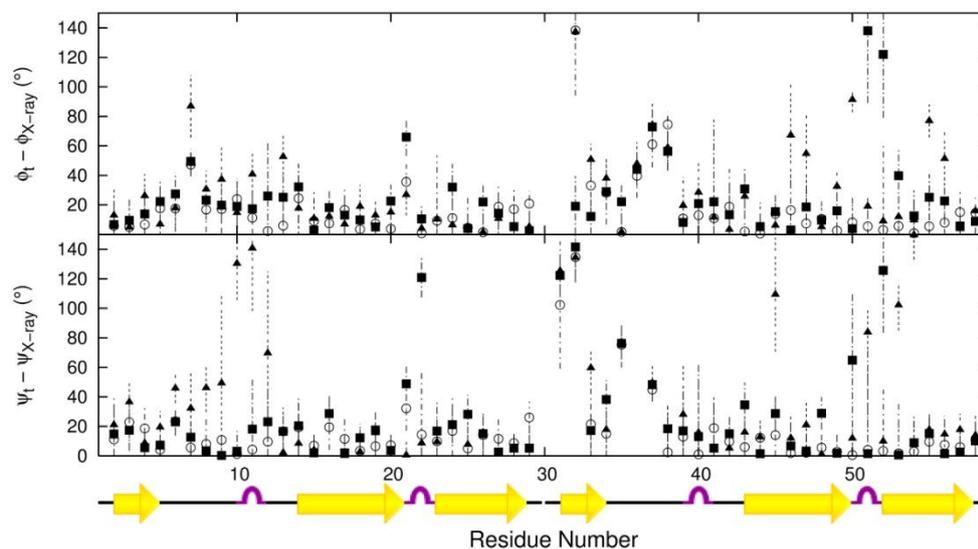


Figure 4-3 Absolute difference between MD time-average and experimental values of φ , ψ for D-HNP1. Legend the same as Figure 4-2.

4. Extending the CHARMM Potential Energy Function for the Nisin Analogues

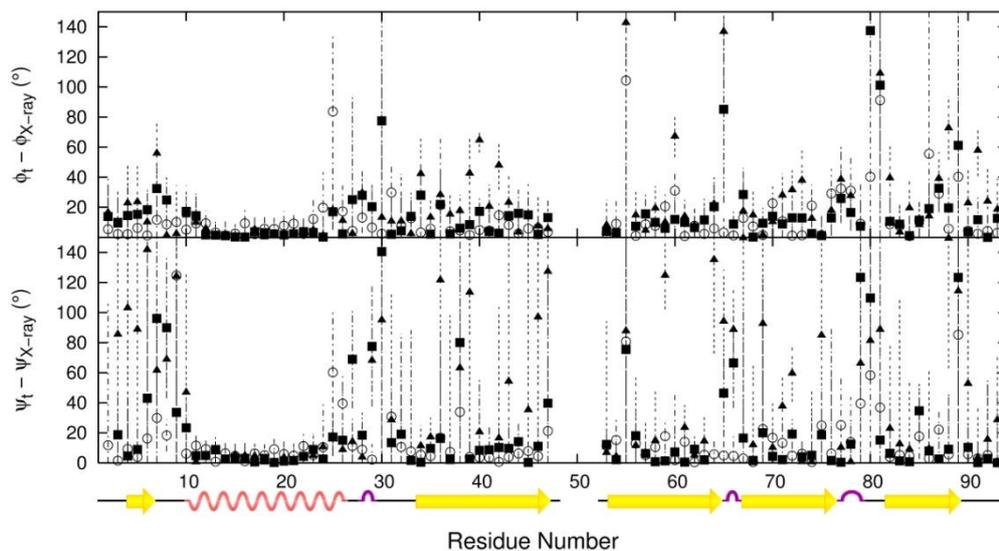


Figure 4-4 Absolute difference between MD time-average and experimental values of ϕ , ψ for D-monellin. Residue numbering from PDB file. Legend the same as Figure 4-2.

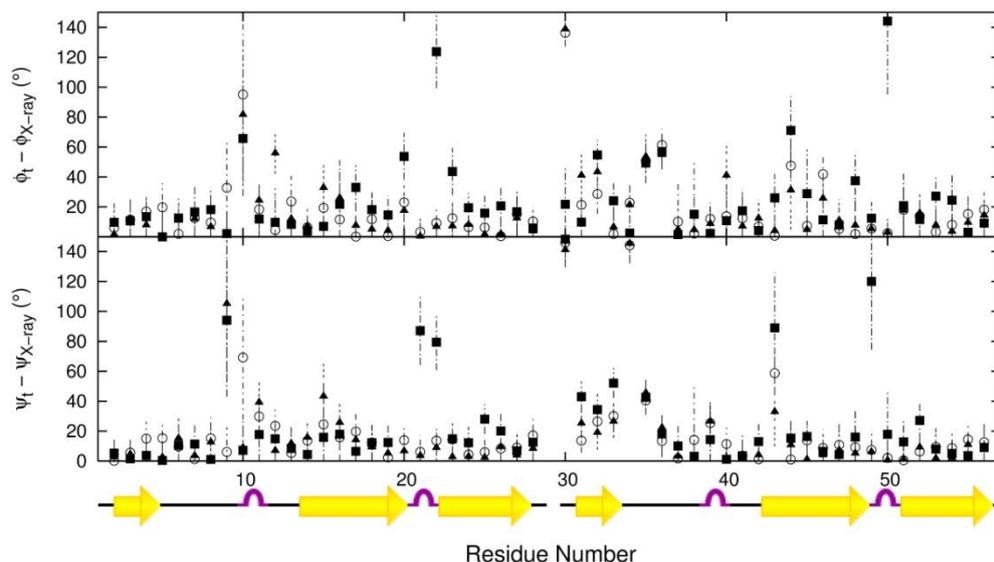


Figure 4-5 Absolute difference between MD time-average and experimental values of ϕ , ψ for HNP2. Legend the same as Figure 4-2.

The time average values of ϕ and ψ for each residue were calculated for the simulation trajectories. A comparison of the profile of a plot of residue number against time-averaged value of ϕ and ψ for the first and second half of each trajectory is the same and confirms that the values of ϕ and ψ for the simulations were converged (data not shown). The difference between the trajectory time-averages and the experimental values of ϕ and ψ for all residues, except the termini, is shown in Figures 4-2 to 4-5. The error bars show the standard deviation of the time-averages. For D-monellin, D-sfAFP

and D-HNP1, the D-CMAP correction is an improvement compared with no corrective grid and compared with the L-CMAP correction, as it reduces the difference between the trajectory time-averages and the experimental values of φ and ψ . This is particularly clear for D-sfAFP, where 84% of residues have a large difference (over 40°) between the experimental value of φ or ψ and the L-CMAP average. The reason that the effect of the correction is particularly pronounced is that D-sfAFP is 45% glycine and the correction improves the sampling of the α_L region, which is only populated by glycine residues.

For the HNP2 mutant the difference, in trajectory and experimental φ and ψ values, between L-CMAP and D-CMAP is negligible, but both are an improvement on the simulations without any corrective grid. This is because only a single residue, D-Ala17, has the D-correction applied, whilst the other residues are L-amino and are appropriately treated with the L-CMAP correction. Considering just D-Ala17, the difference in the (φ , ψ) values between the experimental values and the MD average is (6.2°, -7.7°) for L-CMAP and (3.5°, -2.9°) for D-CMAP, confirming that the D-CMAP correction improves the representation of D-amino residues in L-amino chains.

With respect to secondary structure, D-CMAP decreases the average difference between the experimental structure and the MD averages compared with no correction, for D-monellin and D-HNP1 for the residues in β -sheets. None of the corrections causes notable change in the difference or standard deviation for the α -helical residues of D-monellin. A qualitative explanation for why the L-CMAP term does not destabilise the α -helical region when used to model an all D-protein is that the minimum in the α_L region of the L-CMAP grid encourages correct sampling of the left-handed α -helix of D-amino residues. Additionally, α -helices are the most stable type of protein secondary structure and, once established, are not destabilised as easily as other types of structure by changes in the force field. For several β -turns, such as the first and second turns of D-sfAFP, the first turn of D-HNP1, the first and second turn of D-monellin and the second and final turn of HNP2, the D-CMAP correction (and the L-CMAP correction in the case of HNP2) reduces the

difference between the experimental structure and the trajectory average. A correctly treated α_L region is important for the representation of β -turns [14] as type-I' and type-II are defined as sampling that region [25]. As the CMAP correction greatly improves the treatment of the α_L region this leads to much better representation of some β -turns.

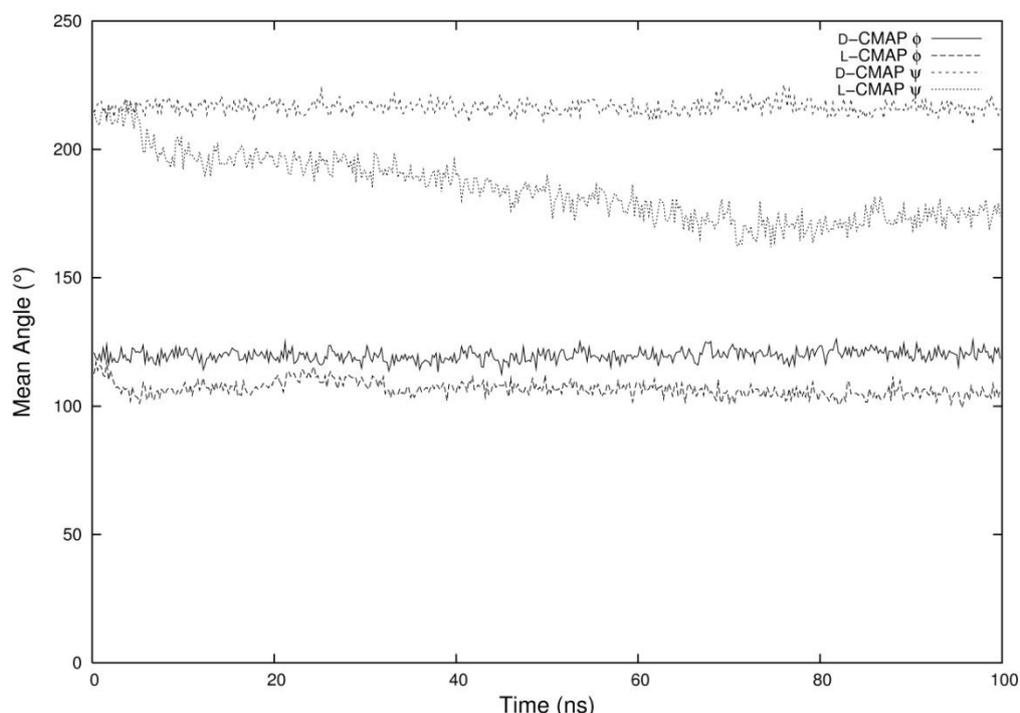


Figure 4-6 Mean value of φ and ψ for residues assigned as β -sheet in D-monellin.

Figure 4-6 shows the average value of the φ and ψ angles, of residues that are assigned as β -sheet by the PDB, for the D-monellin trajectories. For the simulation using the D-corrective map the average value of φ and ψ are converged at $120^\circ \pm 2^\circ$ and $216^\circ \pm 2^\circ$. For the simulation using the L-corrective map the value of φ is converged at $110^\circ \pm 3^\circ$ but the value of ψ shows larger fluctuations during the 100 ns trajectory and has an average value of $184^\circ \pm 12^\circ$. Figure 6 illustrates how using the incorrect map leads to different results than the correct map and introduces instabilities into secondary structure elements.

Further simulations were performed using the same protocol with the equivalent all L-amino molecules of monellin (PDB: 1KRL), sfAFP and HNP1 (PDB: 3GNY). Table 4-2 contains the average φ , ψ differences between the

4. Extending the CHARMM Potential Energy Function for the Nisin Analogues

MD average and experimental values averaged over all residues for the L-amino and D-amino variants. A similar improvement over no correction is recorded for the L-variants with the L-CMAP as for the D-variants with the D-CMAP. Again the improvement is largest for the glycine-rich sfAFP.

Protein	L-CMAP		None		D-CMAP	
	$\varphi/^\circ$	$\psi/^\circ$	$\varphi/^\circ$	$\psi/^\circ$	$\varphi/^\circ$	$\psi/^\circ$
D-Monellin	13.6	23.3	11.2	12.2	9.8	9.4
<i>L-Monellin</i>	<i>11.8</i>	<i>9.8</i>	<i>14.8</i>	<i>12.0</i>	-	-
D-sfAFP	22.5	24.5	19.5	14.0	8.5	8.2
<i>L-sfAFP</i>	<i>8.2</i>	<i>9.1</i>	<i>17.9</i>	<i>16.6</i>	-	-
D-HNP1	31.9	39.0	29.5	27.0	19.2	18.9
<i>L-HNP1</i>	<i>18.9</i>	<i>17.4</i>	<i>29.7</i>	<i>29.3</i>	-	-

Table 4-2 Differences between the MD average and experimental values of backbone dihedral angles, averaged over all residues. The results for the all L-amino molecules are shown in italics.

Yongye *et al.* [8] used 20 ns MD simulations to study the cyclisation preferences of three cyclic tetrapeptides with the chirality LLLLL, LLLDL and LDLDL, using the CHARMM force field but without the CMAP term transformed for the D-amino residues. The LLLLL and LLLDL peptides favoured the conformation identified in mass-spectrometry experiments, but only the LLLLL peptide produced relative energies for the sampled conformations consistent with QM calculations. The LDLDL simulations did not favour the conformation expected from experiment nor did they produce consistent relative energies. Furthermore, the D-amino residues occupied the experimentally less populated α_L region of the Ramachandran plot. In two recent papers, Connelly *et al.* [10] and Capone *et al.* [11] compared the ability of L- and D-enantiomers of the β -amyloid (β A) peptide to form ion channels in a lipid bilayer, part of the underlying mechanism that leads to the synaptic degeneration of Alzheimer's disease. They performed MD simulations of 80 ns [10] and 100 ns [11] of L- β A and D- β A using the CHARMM force field with the CMAP term transformed for the D-residues. They reported that L- β A and D- β A form ion channels of similar structure that were both able to transport cations through the lipid bilayer, in agreement with experimental observations from atomic force microscopy and planar lipid bilayer electrophysiological recordings. This is consistent with our observation that a transformed CMAP term is necessary

to simulate D-amino residues correctly using the CHARMM force field, as the simulations of the β A peptides were in agreement with experimental observations, in contrast to the simulations of the cyclic tetrapeptides.

4.3 Impact of the CMAP term on simulations of the nisin analogues

4.3.1 Simulation set-up

Further MD simulations were performed of the nisin analogues described in Chapter 3, this time using explicit solvent, with the simulation protocol twice repeated, once without the CMAP term included in the energy calculations and once using the D-CMAP corrective term described in the section above. The peptides were built in CHARMM [20] [13], following the same protocol used in Section 3.4. They were solvated using CHARMM in a box of 5450 TIP3P [19] water molecules. The shape of the unit cell was a truncated octahedron with maximum length vector 69.9 Å. Each system was energy minimised in NAMD [21] using the standard conjugate gradient algorithm for 10,000 steps with the protein atoms fixed, 10,000 steps with the protein backbone atoms fixed and a further 20,000 steps with the no constraints.

All heating, equilibration and production dynamics were performed using NAMD [21] with a time step of 2 fs and periodic boundary conditions. The SHAKE algorithm [26] was applied to all bonds to hydrogen atoms. Long-range electrostatics were treated using the particle-mesh Ewald method [27]. Dynamics were performed using a Langevin dynamics integrator with a friction coefficient of 5 ps⁻¹. The solvent-protein system was heated from 0 K to 298 K in increments of 30 K by temperature reassignment, where the velocities of all the atoms in the system are reassigned so that the entire system is set to the target temperature. The velocities were reassigned every 100 time steps for 10,000 time steps using the NVT ensemble. During the heating phase the protein backbone atoms were initially constrained to maintain their initial position using harmonic constraints with force constant 25 kcal mol⁻¹ Å⁻² and the force constant was gradually reduced by 2.5 kcal mol⁻¹

1 \AA^{-2} every 1000 steps to zero. The systems were equilibrated for a further 25,000 time steps using the NPT ensemble, with the pressure set to 1 atm. Production dynamics were run for 500 ns for each analogue.

4.3.2 Results

The analysis used to assess the GBSW implicit solvent simulation in Section 3.5 was repeated for the explicit solvent simulations. Table 4-3 and Table 4-4 summarise the effects of chirality on the peptides for the simulations without the corrective term included in the energy calculations. For both the Cys3(S)-Cys(7) and Cys(8)-Cys(11) interaction the probability of the sulfur atoms being closer than 5.5 \AA is very low, irrespective of cysteine chirality. This is particularly notable for the peptides with L-Cys8 (0, 1, 4, 5, 8, 9, 12 and 13) which have a high probability of a Cys8-Cys11 interaction during both the implicit solvent simulations in Chapter 3 and during the simulations reported in Table 4-6, performed with the corrective map. The amino acid sequence of the peptides is Cys-Gly-Pro-Cys in this region, and it has been shown experimentally that this sequence motif strongly favours the formation of a β -turns [28] leading to a sulfur-sulfur interaction. However, as shown in the section above, the CMAP term is needed to correctly simulate glycine residues and some β -turns, and this is the probably reason for the difference between the results in Table 4-4 and Table 4-6.

Similarly, the probability of a Cys3-Cys7 interaction for the simulations with the D-CMAP term follows the same pattern as the implicit simulations reported in Chapter 3, where nisin analogues 10 to 15 have an increased probability of interaction (with the exception of analogue 12 during the explicit simulations) compared with the other analogues and the simulations without the CMAP term. D-amino residues have been identified as being involved in non-standard type IV β -turns [18] that bring the sulfur atoms together and the CMAP correction is needed to correctly simulate some β -turns. The overall similarity of the results from the implicit solvent simulations in Chapter 3 and the simulations in this chapter with the D-CMAP term

suggested that explicit solvent simulations are more sensitive to the influence of the corrective map than implicit simulations.

Nisin Analogue	Cysteine Chirality	ΔE at 5.5 Å (kJ mol ⁻¹)	$P(S_3-S_7)$	Location of first minimum (Å)
0	L-Cys3-L-Cys7	11.0	0.00	3.8 (12.5 kJ mol ⁻¹)
1		5.7	0.03	3.8 (5.4 kJ mol ⁻¹)
2		4.2	0.07	3.7 (3.2 kJ mol ⁻¹)
3		4.3	0.06	3.8 (3.2 kJ mol ⁻¹)
4	L-Cys3-D-Cys7	4.6	0.05	3.8 (4.1 kJ mol ⁻¹)
5		5.1	0.05	3.8 (4.1 kJ mol ⁻¹)
6		3.6	0.12	3.8 (2.1 kJ mol ⁻¹)
7		3.1	0.17	3.8 (0.9 kJ mol ⁻¹)
8	D-Cys3-L-Cys7	10.0	0.02	4.1 (7.8 kJ mol ⁻¹)
9		4.4	0.05	3.7 (6.7 kJ mol ⁻¹)
10		1.3	0.19	3.8
11		5.3	0.05	3.7 (3.3 kJ mol ⁻¹)
12	D-Cys3-D-Cys7	2.5	0.11	3.8 (2.0 kJ mol ⁻¹)
13		3.0	0.12	3.8 (2.3 kJ mol ⁻¹)
14		3.7	0.07	3.7 (2.5 kJ mol ⁻¹)
15		3.0	0.12	3.9 (1.8 kJ mol ⁻¹)

Table 4-3 Effect of Cys3 and Cys7 chirality on the PMF surface and the probability of S_3 and S_7 coming close enough to form a disulfide bridge (less than 5.5 Å) for the explicit solvent simulations without the CMAP term. ΔE at 5.5 Å is the value of PMF when S_3 and S_7 are considered close enough to form a disulfide bridge; $P(S_3-S_7)$ is the probability of the sulfur-sulfur separation being less than 5.5 Å and location of first minimum gives the sulfur-sulfur separation of the first minimum on the PMF surface, with its value in parentheses if it is not the global minimum.

4. Extending the CHARMM Potential Energy Function for the Nisin Analogues

Nisin Analogue	Cysteine Chirality	ΔE at 5.5 Å (kJ mol ⁻¹)	$P(S_8-S_{11})$	Location of first minimum (Å)
0	L-Cys8-L-Cys11	5.1	0.05	3.7 (5.4 kJ mol ⁻¹)
4		3.4	0.07	3.9 (4.5 kJ mol ⁻¹)
8		5.5	0.06	3.8 (5.1 kJ mol ⁻¹)
12		2.6	0.11	3.8 (3.2 kJ mol ⁻¹)
1	L-Cys8-D-Cys11	4.5	0.06	3.8 (4.7 kJ mol ⁻¹)
5		3.8	0.07	3.9 (4.5 kJ mol ⁻¹)
9		4.9	0.06	3.7 (4.5 kJ mol ⁻¹)
13		3.7	0.06	5.1 (3.2 kJ mol ⁻¹)
2	D-Cys8-L-Cys11	6.3	0.01	7.5 (1.2 kJ mol ⁻¹)
6		6.3	0.01	10.9
10		6.7	0.01	11.9
14		6.2	0.01	7.6 (1.3 kJ mol ⁻¹)
3	D-Cys8-D-Cys11	7.0	0.01	7.9 (1.1 kJ mol ⁻¹)
7		7.3	0.01	12.3
11		7.0	0.01	12.4
15		9.8	0.00	12.4

Table 4-4 Effect of Cys8 and Cys11 chirality on the potential energy surface and the probability of Cys8(S) and Cys11(S) coming close enough to form a disulfide bridge for the explicit solvent simulations without the CMAP term. Symbols and columns the same as Table 4-3.

Nisin Analogue	Cysteine Chirality	ΔE at 5.5 Å (kJ mol ⁻¹)	$P(S_3-S_7)$	Location of first minimum (Å)
0	L-Cys3-L-Cys7			
1		11.6	0.01	3.9 (16.0 kJ mol ⁻¹)
2		5.7	0.03	12.8
3		0.9	0.14	3.9 (0.9 kJ mol ⁻¹)
4	L-Cys3-D-Cys7	8.2	0.02	9.7
5		4.1	0.09	7.2
6		7.3	0.03	4.2 (6.8 kJ mol ⁻¹)
7		1.4	0.12	3.9 (2.0 kJ mol ⁻¹)
8	D-Cys3-L-Cys7	8.5	0.02	10.1
9		8.3	0.03	3.9 (6.9 kJ mol ⁻¹)
10		0.3	0.33	3.7 (0.7 kJ mol ⁻¹)
11		1.3	0.26	3.7 (0.3 kJ mol ⁻¹)
12	D-Cys3-D-Cys7	6.3	0.02	4.0 (6.7 kJ mol ⁻¹)
13		1.2	0.29	3.8 (0.8 kJ mol ⁻¹)
14		2.2	0.16	3.8 (1.9 kJ mol ⁻¹)
15		0.3	0.44	3.8 (0.3 kJ mol ⁻¹)

Table 4-5 Effect of Cys3 and Cys7 chirality on the PMF surface and the probability of S₃ and S₇ coming close enough to form a disulfide bridge (less than 5.5 Å) for the explicit solvent simulations with the D-CMAP term. Symbols and column headings the same as Table 4-3.

Nisin Analogue	Cysteine Chirality	ΔE at 5.5 Å (kJ mol ⁻¹)	$P(S_8-S_{11})$	Location of first minimum (Å)
0	L-Cys8-L-Cys11			
4		0.4	0.24	5.3
8		0.5	0.26	5.2
12		0.5	0.35	5.2
1	L-Cys8-D-Cys11	0.8	0.30	5.2
5		1.5	0.67	5
9		1.1	0.24	5.1 (0.4 kJ mol ⁻¹)
13		2.0	0.64	4.7
2	D-Cys8-L-Cys11	4.4	0.04	4.1 (7.0 kJ mol ⁻¹)
6		3.2	0.07	7.5
10		7.3	0.03	3.9 (8.8 kJ mol ⁻¹)
14		0.3	0.38	5.1
3	D-Cys8-D-Cys11	3.9	0.06	7.4
7		2.4	0.10	7.4
11		1.8	0.15	7.3
15		3.3	0.03	7.2 (0.8 kJ mol ⁻¹)

Table 4-6 Effect of Cys8 and Cys11 chirality on the potential energy surface and the probability of Cys8(S) and Cys11(S) coming close enough to form a disulfide bridge for the explicit solvent simulations with the D-CMAP term. Symbols and columns the same as Table 4-3.

4.4 Reactive dynamics and CPMD simulations of the nisin analogues

Reactive processes, such as catalysis by enzymes, are very important in biochemistry, but these types of processes are very difficult to simulate. *Ab initio* calculations using Hartree-Fock methods or density functional theory can describe the forming and breaking of bonds, but the computational expense usually leads to short simulation times that cannot adequately describe dynamics. QM-MM methods are a better compromise between accessible time scale and level of theory, but are still limited to production runs on the picosecond to nanosecond timescale due to the computational expense of the QM part. Reactive dynamics is an extension of MD that overcomes the problem of modelling the formation of breaking of bonds by allowing the energy calculation to switch between two potential energy

surfaces (PES), one with parameters that describes the atomic interactions when the bonds are formed and one that describes them when broken.

A reactive dynamics module has recently been introduced to CHARMM by Nutt and Meuwly [29] that is suitable for describing the formation of disulfide bridges in the nisin analogues, and other cysteine rich peptides, provided that suitable PES could be defined for the unbonded sulfur atoms. Using CPMD to describe each disulfide bridge and classical mechanics for the rest of the peptide, a PES was calculated to describe the breaking of each disulfide bridge with the intention of using this to define parameters for the CHARMM reactive dynamics module. The PES can also be used to examine the effect of the influence of the different peptide environment on disulfide stability.

4.4.1 Simulation set-up

The centroid of the most populated cluster from the implicit solvent MD simulations (Section 3.7) of nisin analogue 12 (D-Cys3-D-Cys7-L-Cys8-L-Cys11) was used as the starting point for the CPMD simulations. The centroid structure was minimised using 1000 steps of steepest descent followed by 1000 steps of adopted basis Newton-Raphson in CHARMM. The description of the atom connectivity was edited to include the Cys3-Cys7 and Cys8-Cys11 disulfide bonds and then minimised further. The structure was solvated in a cube of TIP3P water with edge length of 50 Å using GROMACS [30] [31] and heated to 300 K and equilibrated for 5 ns using NAMD [21] with the OPLS force field [32]. The final coordinates from the equilibration were used as the starting coordinates for the CPMD run.

Constrained MD can be used to generate a PES for a reaction coordinate of interest [33]. A constraint can be applied to a geometric variable, ξ , that is a function of the atomic coordinates, \mathbf{R} , using a Lagrangian multiplier, λ , which is equal to the constraint force. The constraint can be written as

(4-1)

$$\sigma(\mathbf{R}) = \xi(\mathbf{R}) - \xi'$$

where $\xi(\mathbf{R})$ is the actual value of the variable and ξ' is the prescribed value. The Lagrangian of the constrained system then becomes

(4-2)

$$\mathcal{L}' = \mathcal{L} - \lambda\sigma$$

with (4-3) or (4-4) as the equations of motion

(4-3)

$$\frac{\partial}{\partial t} \frac{\partial \mathcal{L}'}{\partial \dot{\mathbf{R}}} = \frac{\partial \mathcal{L}'}{\partial \mathbf{R}}$$

(4-4)

$$M\ddot{\mathbf{R}} = -\frac{\partial V}{\partial \mathbf{R}} - \lambda \frac{\partial \sigma}{\partial \mathbf{R}}.$$

The SHAKE algorithm [26] iteratively solves for the Lagrange multiplier in combination with the Verlet algorithm [34] to propagate the dynamics. The free energy difference between two states, ξ^1 and ξ^2 , can be expressed as the reversible work, W , required to move along the reaction coordinate from ξ^1 to ξ^2 :

(4-5)

$$\Delta W = W(\xi^2) - W(\xi^1) = \int_{\xi^1}^{\xi^2} d\xi \left\langle \frac{\partial \mathcal{H}}{\partial \xi} \right\rangle$$

where \mathcal{H} is the Hamiltonian. By exploiting the fact that the derivative of the Hamiltonian with respect to the reaction coordinate is equal to the Lagrange multiplier [33]

(4-6)

$$-\frac{\partial \mathcal{H}}{\partial \xi} = \lambda$$

the free energy difference between states can be found by integrating λ with respect to ξ .

Two sets of simulations were performed, treating each disulfide bond and the cysteine C_{β} atom and its hydrogen atoms using CPMD. The rest of the peptide and the water molecules were treated using classical MD. Dummy hydrogen atoms to couple the QM region to the classical region were placed 1.1 Å from C_{β} along the C_{β} - C_{α} bond. The QM part of the simulations was performed using the CPMD program (www.cpmd.org) in Car Parrinello mode. The QM cell was a cube with sides of 21 Å, the planewave cut-off was 25 Ry and the calculations used ultra-soft pseudopotentials. The fictitious mass of the electrons was 400 a.u. and the calculation used the PBE exchange-correlation functional. The wavefunction was minimised to the Born-Oppenheimer surface at the start of each production run. The classical part of the simulation was performed with GROMOS [35] using the OPLS force field. For each S-S distance a production run of 1 ps was performed using a time step 0.096 fs and the value of λ recorded at each step.

4.4.2 Results

The constrained S-S distance is plotted against the mean force used to maintain the constraint, as shown in Figure 4-7. These functions were integrated with respect to S-S distance to produce the PES in Figure 4-8. There is no difference between the PES for ring A and ring B, indicating that either the peptide environment does not influence the strength of the disulfide bond or that by treating the solvent and all the atoms other than the disulfides using classical mechanics, the level of theory was not sensitive enough to simulate the differences.

4. Extending the CHARMM Potential Energy Function for the Nisin Analogues

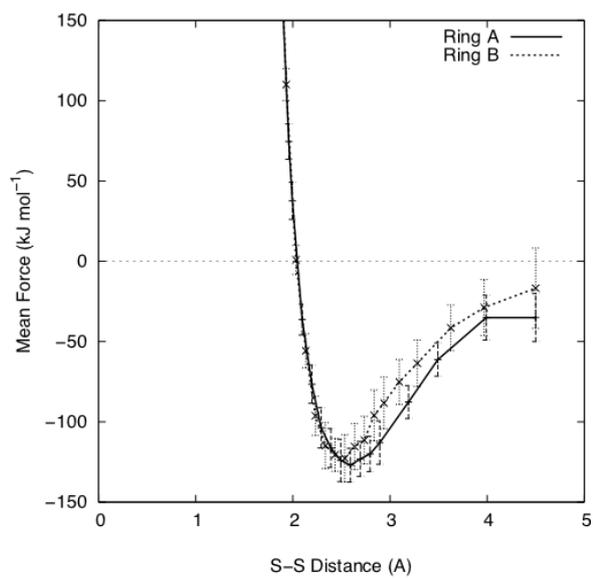


Figure 4-7 S-S distance plotted against the mean of the force used to maintain the constraint ($\langle\lambda\rangle$) for the two disulfide bridges.

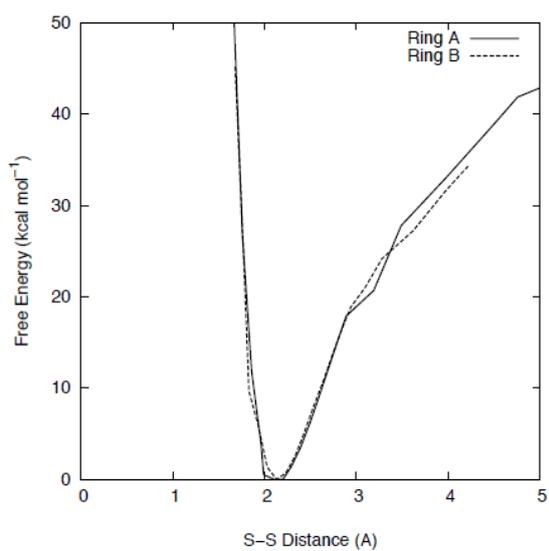


Figure 4-8 Potential energy surface for the S-S distance, calculated by integrating the curve shown in Figure 4-7.

4. Extending the CHARMM Potential Energy Function for the Nisin Analogues

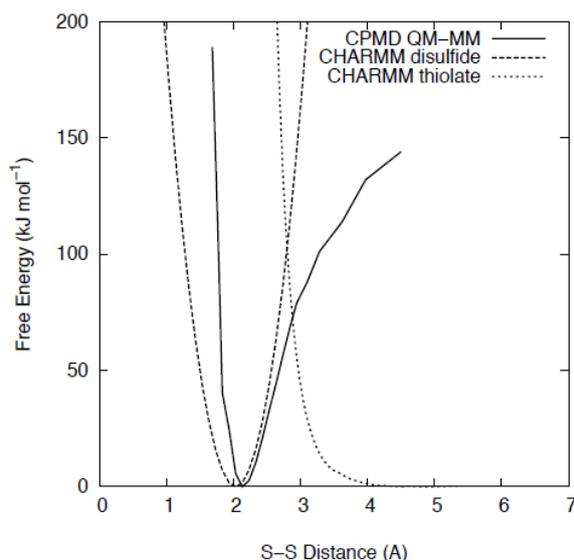


Figure 4-9 Potential energy surfaces for ring B calculated using (i) CPMD QM-MM (ii) CHARMM disulfide atoms types and (iii) CHARMM thiolate atom types.

In order to use the reactive dynamics module in CHARMM a Morse potential could be defined that is fitted to the QM-MM CPMD PES to describe the disulfide bonds whilst breaking. For the second PES, to describe the bonds when broken, the parameters in the CHARMM force field already defined for cysteine thiolate could be used [15]. Figure 4-9 shows the QM-MM PES, the CHARMM thiolate PES and the CHARMM disulfide bond PES for ring B as an example, calculated in CHARMM by constraining the sulfur-sulfur distances and running an energy minimisation. However, the crossing point between the CPMD QM-MM surface and the thiolate PES is at 72.2 kJ mol^{-1} , a barrier that could not be reached during normal MD simulations because it is too high. Therefore the CPMD QM-MM PES calculated here is not suitable for defining the Morse potential for reactive dynamics. A probable reason that it is unsuitable is that water molecules were not included in the QM region for the broken disulfide bond to interact with, so the sulfurs remained radicals with a strong attraction to each other. If a PES could be defined that can take into account the interaction with water hydrogens beyond a certain S-S distance, this would be more suitable for defining a potential for the reaction dynamics. However, to define such a PES would require simulations that were more computationally expensive: all of the peptide and the solvent would

need to be treated using QM and multiple runs would be needed to estimate the S-S distance dependence of interactions with water hydrogens.

4.5 Conclusions

In this chapter, the CMAP corrective term of the CHARMM force field is transformed to describe D-amino residues. The effect of the CMAP term is assessed using X-ray crystallographic structures containing D-amino residues and it is shown that the transformed map is necessary to correctly simulate structures containing β -turns and β -sheets and that using the standard map disrupts these secondary structure elements for D-amino residues during simulations.

Explicit solvent simulations were then performed of the nisin analogues described in Chapter 3 with and without the D-CMAP term. The simulations without the correction had a very low probability of sulfur-sulfur interactions, despite experimental evidence that suggests interactions would be likely. Such interactions are based upon the formation of β -turns and these results demonstrate again that without the CMAP term during a simulation, these structures are disrupted. The probabilities of sulfur-sulfur interaction and their dependence on cysteine chirality followed the same pattern as reported for the implicit solvent simulations reported in Chapter 3, performed without the CMAP term. This suggests that explicit solvent simulations may be more sensitive to the correction than implicit simulations.

Finally, a PES is defined for the sulfur-sulfur distance during the stretching and breaking of the disulfide bonds in the nisin analogues using QM-MM simulations with CPMD. The intention was to use this PES to define a Morse potential for use with in reactive dynamics simulations that jump between PES. However, the energy barrier between the CPMD PES and the PES calculated using the CHARMM parameters for thiolate was 72.2 kJ mol^{-1} , too high for use with reactive dynamics. A probable reason for the high barrier is that the QM part of the simulations was performed in the gas phase and

water molecules are needed to be present to reduce the attraction between the sulfur radicals and to donate hydrogen atoms.

4.6 References

1. Car, R. and M. Parrinello, *Unified approach for molecular dynamics and density-functional theory*. Phys. Rev. Lett., 1985. **55**: p. 2471-2474.
2. Chandler, D.E., J. Hisin, C.B. Harrison, J. Gumbart, and K. Schulten, *Intrinsic curvature properties of photosynthetic proteins in chromatophores*. Biophys. J., 2008. **95**: p. 2822-2836.
3. Shaw, D.E., M.M. Deneroff, R.O. Dror, J.S. Kuskin, R.H. Larson, J.K. Salmon, C. Young, B. Batson, K.J. Bowers, J.C. Chao, M.P. Eastwood, J. Gagliardo, J.P. Grossman, C.R. Ho, D.J. Ierardi, I. Kolossvary, J.L. Klepeis, T. Layman, C. McLeavey, M.A. Moraes, R. Mueller, E.C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S.C. Wang, *Anton, a special-purpose machine for molecular dynamics simulation*. SIGARCH Comput. Archit. News, 2007. **35**(2): p. 1-12.
4. Soutourina, J., P. Plateau, and S. Blanquet, *Metabolism of D-aminoacyl-tRNAs in Escherichia coli and Saccharomyces cerevisiae cells*. J. Biol. Chem., 2000. **275**(42): p. 32535-32542.
5. Lam, H., D.C. Oh, F. Cava, C.N. Takacs, J. Clardy, M.A. de Pedro, and M.K. Waldor, *D-Amino Acids Govern Stationary Phase Cell Wall Remodeling in Bacteria*. Science, 2009. **325**(5947): p. 1552-1555.
6. Wei, G., E. de Leeuw, M. Pazgier, W.R. Yuan, G.Z. Zou, J.F. Wang, B. Ericksen, W.Y. Lu, and R.I. Lehrer, *Through the Looking Glass, Mechanistic Insights from Enantiomeric Human Defensins*. J. Biol. Chem., 2009. **284**(42): p. 29180-29192.
7. Pentelute, B.L., Z.P. Gates, V. Tereshko, J.L. Dashnau, J.M. Vanderkooi, A.A. Kossiakoff, and S.B.H. Kent, *X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers*. J. Am. Chem. Soc., 2008. **130**(30): p. 9695-9701.
8. Yongye, A.B., Y. Li, M.A. Giulianotti, Y. Yu, R.A. Houghten, and K. Martínez-Mayorga, *Modeling of peptides containing D-amino acids: implications on cyclization*. J. Comput. Aided Mol. Des., 2009. **23**: p. 677-689.
9. Turpin, E.R., B. Bonev, and J.D. Hirst, *Stereoselective Disulfide Formation Stabilizes the Local Peptide Conformation in Nisin Mimics*. Biochemistry, 2010. **49**: p. 9594-9603.
10. Connelly, L., H. Jang, F.T. Arce, R. Capone, S.A. Kotler, S. Ramachandran, B.L. Kagan, R. Nussinov, and R. Lal, *Atomic Force Microscopy and MD simulations Reveal Pore-like Structures of All-D-Enantiomer of Alzheimer's β -Amyloid Peptide: Rlevance to the Ion Channel Mechanism of AD Pathology*. J. Phys. Chem. B, 2012. **116**: p. 1728-1735.
11. Capone, R., H. Jang, S.A. Kotler, L. Connelly, F.T. Arce, S. Ramachandran, B.L. Kagan, R. Nussinov, and R. Lal, *All-D-Enantiomer of*

- β -Amyloid Peptide Forms Ion Channels in Lipid Bilayers.* J. Chem. Theory. Comput., 2012. **8**(3): p. 1143-1152.
12. Nueangaudom, A., K. Lugsanangarm, S. Pianwanit, S. Kokpol, N. Nunthaboot, and F. Tanaka, *Structural basis for the temperature-induced transition of D-amino acid oxidase from pig kidney revealed by molecular dynamic simulation and photo-induced electron transfer.* Phys. Chem. Chem. Phys., 2012. **14**(8): p. 2567-2578.
 13. Brooks, B.R., C.L. Brooks, A.D. Mackerell Jr, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus, *CHARMM: The Biomolecular Simulation Program.* J. Comput. Chem., 2009. **30**(10): p. 1545-1614.
 14. Mackerell Jr, A.D., M. Feig, and C.L. Brooks III, *Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations.* J. Comput. Chem., 2004. **25**(11): p. 1400-1415.
 15. MacKerell Jr, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* J. Phys. Chem. B, 1998. **102**(18): p. 3586-3616.
 16. Teukolsky, S.A., W.T. Vetterling, and B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing.* Third ed. 1988, Cambridge: Cambridge University Press.
 17. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank.* Nucl. Acids Res., 2000. **28**(1): p. 235-242.
 18. Mitchell, J.B.O. and J. Smith, *D-amino acid residues in peptides and proteins.* Proteins, 2003. **50**: p. 563-571.
 19. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water.* J. Chem. Phys., 1983. **79**(2): p. 926-936.
 20. Brooks, B.R., R.E. Bruccoleri, D.J. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.* J. Comput. Chem., 1983. **4**: p. 187-217.
 21. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, *Scalable Molecular Dynamics with NAMD.* J. Comput. Chem., 2005. **26**(16): p. 1781-1802.
 22. Allen, M.P. and D.J. Tildesley, *Computer Simulations of Liquids.* 1987, Oxford: Oxford University Press.

23. Hung, L.W., M. Kohmura, Y. Ariyoshi, and S.H. Kim, *Structure of an enantiomeric protein, D-monellin at 1.8 angstrom resolution*. Acta Crystallogr. Sect. D-Biol. Crystallogr., 1998. **54**: p. 494-500.
24. Xie, C., A. Prahl, B. Ericksen, Z.B. Wu, P.Y. Zeng, X.Q. Li, W.Y. Lu, and J. Lubkowski, *Reconstruction of the conserved beta-bulge in mammalian defensins using D-amino acids*. J. Biol. Chem., 2005. **280**(38): p. 32921-32929.
25. Hutchinson, E.G. and J.M. Thornton, *PROMOTIF - A program to identify and analyze structural motifs in proteins*. Protein Sci., 1995. **5**: p. 200-212.
26. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. J. Comput. Phys., 1977. **23**: p. 327-341.
27. Darden, T.A., D. York, and L. Pedersen, *Particle-mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems*. J Chem Phys, 1993. **98**: p. 10089-10092.
28. Hsu, H.-J., H.-J. Chang, H.-P. Peng, S.-S. Huang, M.-Y. Lin, and A.-S. Yang, *Assessing computational amino acid β -turn propensities with a phage-displayed combinatorial library and directed evolution*. Structure, 2006. **14**: p. 1499-1510.
29. Nutt, D.R. and M. Meuwly, *Studying reactive processes with classical dynamics: Rebinding dynamics in MbNO*. Biophys. J., 2006. **90**: p. 1191-1201.
30. Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*. J. Mol. Model, 2001. **7**(8): p. 306-317.
31. Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*. J. Chem. Theory Comput., 2008. **4**(3): p. 435-447.
32. Jorgensen, W.L. and J. Tirado-Rives, *The OPLS (optimized potentials for liquid simulations) potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin*. J. Am. Chem. Soc., 1988. **110**(6): p. 1657-1666.
33. Doltsinis, N.L., *Free energy and rare events in molecular dynamics*, in *Computational Nanoscience: Do It Yourself!*, D. Marx, Editor. 2006, John von Neumann Institute for Computing: Julich. p. 375-387.
34. Verlet, L., *Computer experiments on classical fluids .I. Thermodynamical properties of Lennard-Jones molecules*. Phys. Rev., 1967. **159**(1): p. 98-103.
35. Christen, M., P.H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D.P. Geerke, T.N. Heinz, M.A. Kastenholtz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, and W.F. van Gunsteren, *The GROMOS software for biomolecular simulation: GROMOS05*. J. Comput. Chem., 2005. **26**(16): p. 1719-1751.

5 Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

5.1 Introduction

The ileal lipid binding protein (ILBP) is involved in the metabolic pathway for the regulation of cholesterol and binds at least two bile salt ligands with unusual binding behaviours. The first unusual behaviour is cooperative binding of the ligands on a level comparable with haemoglobin; the second is specific site selectivity of the ligands in the binding cavity, between cholic acid and chenodeoxycholic, and their derivatives, despite these only differing by a single hydroxyl group. In this chapter, MD simulations and docking calculations are performed with ILBP with different ligand configurations. The results agree with experimental evidence of a third binding site. A possible site on the protein exterior leads to a mechanism of allosteric interaction, where binding to the site induces changes in the apo protein conformation, in which one of the α -helices moves $\sim 10^\circ$ with respect to the β -barrel, to a conformation similar to the holo form. A mechanism is suggested for site selectivity, where the higher hydrophobicity of chenodeoxycholic acid leads it to sit deeper in the binding cavity when in site 1, inducing the cholic acid in site 2 to sit deeper in the cavity and for the helices to move closer to the β -barrel, preventing further ligand exchange.

5.2 Structural and Physiological Properties of ILBP

5.2.1 Intracellular Lipid Binding Protein Family

ILBP belongs to the intracellular lipid binding protein (iLBP) family [1]. All members have a 10-strand antiparallel β -sheet in a clam shell like structure with two, short, nearly parallel α -helices covering the opening of the β -clam. The interior of the protein contains a cavity partially filled with ordered water molecules. Protein side chains and bound water molecules are involved in hydrogen bonding that co-ordinates the ligand(s) at the binding site. A portal

region is postulated between α -helix II and the turns between β -strands C and D and E and F (see Figure 5-1 for secondary structure naming convention).

There are four subfamilies in the iLBP family [1]:

- (i) The intracellular retinoid binding proteins, cellular retinoic acid binding proteins (CRABPs) and cellular retinol binding proteins (CRBPs). The single ligand is always deeply immersed in the binding cavity.
- (ii) The liver-type fatty acid binding protein (L-FABP) and ILBP are homologous, as reflected by a significant level of sequence similarity, and both have unusual cooperative binding with two ligands. L-FABP can bind a range of ligands *e.g.* haem, bilirubin and eicosanoids (small molecules made from oxidised essential fatty acids that function as short-lived hormones). In L-FABP one ligand is located at the bottom of the protein cavity, with a bent conformation, coordinated via a hydrogen bonding network to S39, R122 and S124. The second ligand has a linear shape with the carboxylate end toward the fatty acid portal and the hydrophobic end toward the other ligand. ILBP conserves S38, R121 and S123 but binds fatty acids very weakly. It binds bile acids via hydrophobic interactions of the steroid moiety and protein side chains, with the carboxylate end pointed toward the protein solvent interface.
- (iii) Intestinal fatty acid binding protein (I-FABP) binds a single ligand with a bent conformation in the opposite direction to the first ligand in L-FABP with the carboxylate group inside cavity coordinated by R106.
- (iv) The remaining other iLBPs have an additional four-residue 3-10 helical loop at the N-terminus. These generally bind one fatty acid in a U-shaped conformation.

NMR data across the family show that protein structures stabilise upon ligand binding and that the backbone of the portal region shows greater conformational variability than the rest of the structure.

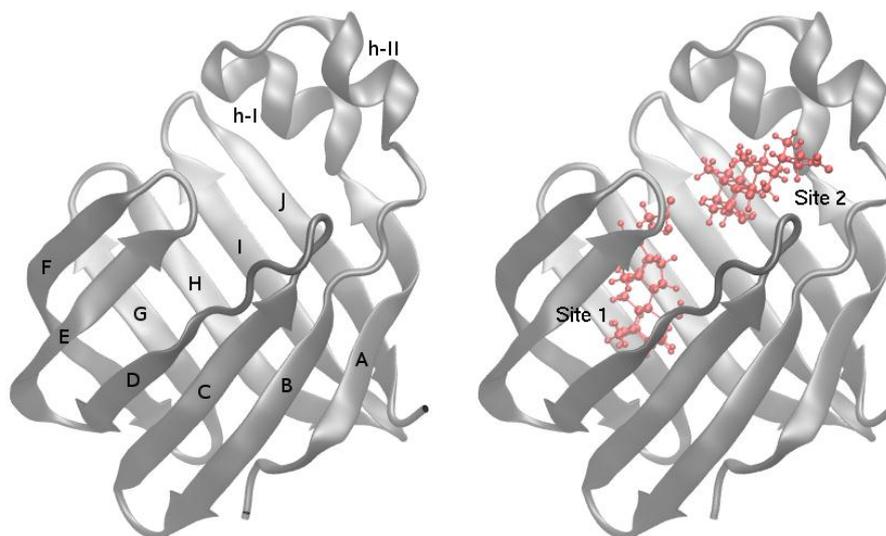


Figure 5-1 (left) Convention for naming the β -strands and α -helices of ILBP; (right) convention for naming the binding sites.

5.2.2 ILBP and Bile Salts *in vivo*

Cholic and chenodeoxycholic acid derivatives (the primary bile salts) are biosynthesised in the liver in a ratio of 2:1 and constitute 80% of the human bile acid pool [2]. The other 20% are secondary bile salts which are modified by gut flora. When synthesised *in vivo* bile salts are conjugated with glycine or taurine in a 3:1 ratio. Conjugation lowers the pKa, resulting in compounds that are fully ionised and soluble at physiological pH with increased effectiveness as detergents [2]. Bile salts need to be effective detergents, as their biological role is to aid the digestion of fats. Bile salts are synthesised from cholesterol in the liver and this synthesis accounts for about half of catabolism of body cholesterol. When bile acids are reabsorbed in the ileum they activate the nuclear farnesoid X receptor (FXR) in the enterocyte (an absorptive intestinal cell), which stimulates expression of ILBP [3]. This creates a positive feedback loop which leads to further bile acid absorption.

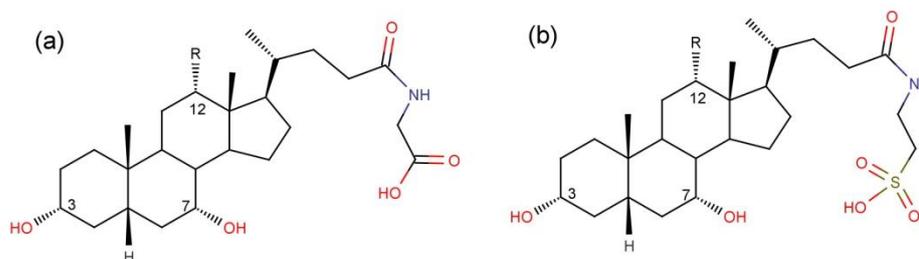


Figure 5-2 Chemical structure of the primary bile salts. For cholic acid R is OH and for chenodeoxycholic acid R is H; (a) is conjugated with glycine and (b) with taurine.

Taurocholate (TCA) is a germinant of *C.diff* spores and chenodeoxycholic acid competitively inhibits spore germination [4]. In a healthy host *C.diff* spores encounter equal numbers of cholic acid and chenodeoxycholic acid molecules. An enzyme of colonic microflora, 7 α -dehydroxylase, removes the hydroxyl group at carbon 7 to convert chenodeoxycholic acid to lithocholate, which also inhibits spore germination [4]. These conditions favour inhibition of germination. Chenodeoxycholic acid is absorbed by the colonic epithelium at a rate 10 times higher than cholic acid; so in an antibiotic treated individual, where chenodeoxycholic acid is not converted to lithochate, the conditions favour germination [4]. Sorg *et al.* tested seven analogues of chenodeoxycholic acid for spore inhibition [4]. One analogue, 37DAME, inhibits germination and is predicted to be less able to bind with FXR and ILBP because of the methyl group on C-3. This analogue would therefore remain in the gut more readily and inhibit spore germination and with further testing it could be potentially be used to treat *C.diff* infections.

5.3 Review of Experimental Evidence

Tochtrop *et al.* [2] used NMR experiments with enriched bile salts to demonstrate site selectivity in human ILBP. In the spectrum of ILBP and enriched ^{15}N -glycocholate (GCA) three peaks are resolved, one for the unbound bile salts and one for each binding site. In a spectrum of ILBP, enriched ^{15}N -GCA and un-enriched glycochenodeoxycholate (GCDA) (invisible to the NMR experiment) only two peaks were resolved – one for unbound bile salts and one at binding site 2; GCDA had completely displaced the labelled GCA from site 1. Similarly, in an experiment with ^{13}C enriched bile salts, un-

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

enriched GCDA displaced ^{13}C -GCA from binding site 1 and un-enriched GCA displaced ^{13}C -GCDA from binding site 2. GCA and GCDA are identical except for an OH at C12 in GCA. In samples containing only one species both sites were occupied; so selectivity cannot arise from steric exclusion or a lack of affinity for one site.

ILBP shows positive cooperativity comparable with haemoglobin. The Hill coefficient is a macroscopic measurement of cooperativity; it is 1 for non-cooperative systems and 2 for extremely positively cooperative systems. The Hill coefficient for ILBP was calculated as 1.94 by Tochtrop *et al.* [5].

Toke *et al.* [6] showed that cooperative binding and site selectivity are independent in human ILBP, using isothermal titration calorimetry (ITC) and NMR experiments on mutated proteins chosen from an unpublished NMR structure of ILBP in complex with GCA and GCDA. Two mutations compromised positive cooperativity in the binding of both GCA and GCDA (N61A and E110A), two mutations compromised positive cooperativity in the binding of either GCA (W49Y) or GCDA (Q99A), one mutation compromised site selectivity in bile salt human ILBP recognition (Q51A) for both bile salt species and one mutation did not affect either binding cooperativity or site selectivity within the margin of experimental error (T38A). From these results, Toke *et al.* concluded that cooperative binding and site selectivity are independent.

Stabilisation of α -II and the C-D and E-F loops of the liver bile acid binding protein and other members of the iLBP family [7] upon ligand binding has been demonstrated with NMR studies and MD simulations. A recent study of the internal backbone motions of ILBP using NMR spectroscopy by Horvath *et al.* [8] supports the hypothesis that there is an allosteric mechanism of ligand binding and that the protein shifts from sampling a closed-like state to an open state upon the first ligand binding [9].

Capaldi *et al.* [3] published the structure of zebrafish ILBP in complex with cholic acid. Two cholic acid molecules are present in the binding cavity of all

five protein models produced from the two crystal forms; for four of the five models the location of the binding site overlaps very well for both ligands; for the fifth model the location of one binding site is the same as the other four and the location of the second binding site is the same but the cholic acid molecule has been rotated by 90° about its longest axis. Direct interactions occur between the OH of Tyr97 and O12 of one cholic acid (site 1; 150); and OH of Tyr14 with O12 and Tyr53 with O7 of the other (site 2; 151). Both have interactions with O12. Contacts mediated by hydrogen-bonded water molecules are formed with Gln51, Gln99, Glu110 and Arg125 with the O7 atoms of both molecules and the carboxyl of cholic acid 150. Hydrophobic interactions occur with Ile23, Gly31, Trp49, Phe63, Met71, Val83, Leu90, Ile92 and Thr101. Comparison with human and porcine ILBPs shows ligand binding in a similar region of the binding cavity for one of the cholic acids but the precise position, interactions and orientation of the ligand are quite different. Comparison of the sequences of other iLBPs shows that residues involved in ligand binding are generally either conserved or substituted by conservative mutations.

The number of binding sites has not been conclusively determined experimentally. Tochtrop *et al.* [2] state that ILBP binds to bile salts with a 1:2 ratio, but Lucke *et al.* [10] resolve only a single cholic acid in binding cavity with NMR. Capaldi *et al.*, using X-ray crystallography, resolve two in the binding cavity and several adhered to the protein exterior. Cholates molecules are found on the surface of all five of their models, though these have a much more variable alignment than those in the interior. Two of the exterior cholic acids are present in all five models, but the position is variable because there is no hydrogen bonding at the binding site. A third exterior cholic acid is only present in two of the models and interacts with both the interior and exterior cholic acids. The region of the surface with the exterior cholic acids is almost entirely hydrophobic. Experiments on danio L-BABP and rabbit ILBP have also indicated exterior binding sites. Experimental data to support specific binding sites on hydrophobic surfaces are difficult to acquire and exterior ligands can

disrupt crystallisation. This might explain the difficulty in crystallising rabbit ILBP described by Kouvatso [11]. Capaldi *et al.* [3] speculate that the physical role of exterior binding sites could be to help guide molecules into the interior or improve binding in the presence of excess ligand. ITC data confirm the presence of interior bound ligands, but could not discriminate between a model with (i) three independent binding sites: two in the interior and a third, entropically-driven, adhesion site on the surface or (ii) three states: free, first site occupied, second site occupied, which is simpler and has been observed in other systems.

There is further experimental evidence that ILBP has three binding sites – two in the interior of the protein binding cavity mediated by hydrophobic interactions and hydrogen bonding and a third, uncharacterised site, possibly on the exterior. In Toke *et al.* [6] the binding ratio determined by NMR of ILBP:GCDA:GCA is stated as 1:1.5:1.5, but this result is not discussed further, the authors explaining that the structure will be published in a later paper. Most compellingly, Fang [12] demonstrates using electrospray mass spectroscopy that ILBP binds three ligands, even at low protein and ligand concentrations.

In this Chapter, MD simulations and docking calculations have been performed to address some of the outstanding questions about ILBP. To examine the possible position of a third binding site, docking has been performed using protein conformations generated by MD simulations; further MD has been run with the docked structures, to examine the possible binding sites for adhesion and allosteric interactions. To understand better the specific atomic and molecular interactions responsible for site selectivity, MD simulations have been performed that compared a cholic acid in both interior binding sites, cholic acid in the first and chenodeoxycholic acid in the second and chenodeoxycholic acid in the first site and cholic acid in the second. This highlights one of the strengths of using a computational method, as it would be impossible to experimentally observe the ligands within incorrect binding sites.

5.4 Simulation Set-up

5.4.1 Initial Structures

All simulations were performed using the human variant of ILBP. NMR structures are published of the apo form of human ILBP and with a single ligand of TCA resolved [13]. These structures were accessed from the Protein Data Bank (PDB) using the codes 1O1U (apo) and 1O1V (single ligand). The first conformation in each PDB file was used as the starting coordinates. For the singly ligated form, the TCA ligand was mutated in CHARMM [14] [15] to the non-conjugated cholic acid ligand and energy minimised, with harmonic constraints applied to all atoms of the protein (force constant 20 kcal mol^{-1}), for 200 steps using the steepest descent algorithm [15]. The histidine residues, H52, H57 and H98, were examined in VMD [16] for hydrogen bonding with near-by neighbours to determine the protonation state of the side chain. No hydrogen bonds were present for the single-ligand form and the neutral form with the hydrogen on N_δ (HSD), as stated in the PDB file, was used. For the apo form a hydrogen bond is between formed H98- N_δ and the backbone of Q99, confirming the use of HSD in the PDB file. The other histidines did not form any hydrogen bonds and were left as HSD.

The models for the doubly ligated human ILBP were built from the crystallographic structure of danio ILBP bound to two ligands of cholic acid (PDB 3ELZ, chain A). The exterior ligands from 3ELZ were not included in the model. Danio ILBP is three residues longer than the human form and an examination of the secondary structure elements shows that it is extended in the I-J loop region (see Figure 5-3). Sequence alignment of 1O1V and 3ELZ using ClustalW [17] shows that there are 72 conserved residues, 23 residues with strongly similar properties, 19 residues with weakly similar properties and 13 non-conserved residues.

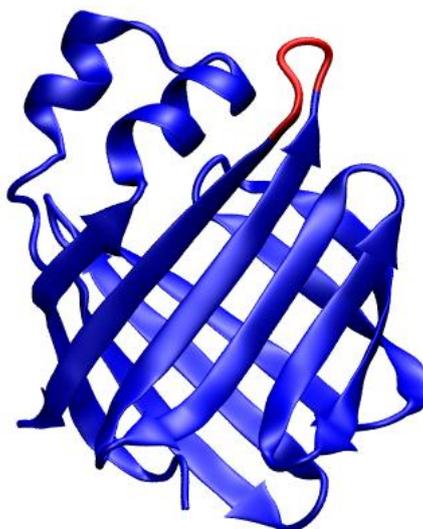


Figure 5-3 Secondary structure of danio ILBP with the residues in the I-J loop region that are additional to human ILBP highlighted in red.

To build the human double-ligand ILBP models, residues 117-119 of the danio ILBP structure were deleted. The non-conserved residues were changed in CHARMM by reading in the danio ILBP primary sequence, renaming the residues to match the human variant in each position that needed altering, deleting the side chains of the modified residues and rebuilding them from the internal coordinates in the parameter set [18]. The histidine residues were examined for hydrogen bonding using VMD [16] and all were determined to be HSD. The structures were minimised using 2000 steps of steepest descent followed by 2000 steps of adopted-basis Newton-Raphson optimisation.

Three double ligand structures of human ILBP were constructed. The first has a cholic acid molecule in both interior binding sites. The second has a cholic acid molecule in the first binding site (see Figure 5-1 for binding site naming convention) and a chenodeoxycholic acid molecule in the second site; this will be named mixed-CA in this chapter. The third has a chenodeoxycholic acid molecule in the first binding site and cholic acid in the second site and will be named mixed-CDA. These mixed ligand structures have been constructed to study site selectivity and confirm which binding site corresponds to which ligand, as no mixed ligand NMR or X-ray crystallographic structures have been published.

5.4.2 MD Simulations

The initial apo, single ligand and double ligand structures were solvated in CHARMM in a box of 4500 to 6600 TIP3P water molecules [19]. The shape of the unit cell was a truncated octahedron with maximum length vectors 68.7 Å (apo), 71.1 Å (single ligand) and 76.8 Å (double ligand). Each system was energy minimised in NAMD [20] using the standard conjugate gradient algorithm for 10,000 steps with the protein atoms fixed, 10,000 steps with the protein backbone atoms fixed and a further 20,000 steps with no constraints.

All heating, equilibration and production dynamics were performed using NAMD [20] with a time step of 2 fs, the CHARMM22 force field and periodic boundary conditions. The parameters for the bile salts were taken from the CHARMM general force field (CGenFF) of drug-like small molecules [21]. The SHAKE algorithm [22] was applied to all bonds to hydrogen atoms. Long-range electrostatics were treated using the particle-mesh Ewald method [23]. Dynamics were performed using a Langevin dynamics integrator with a friction coefficient of 5 ps^{-1} . The solvent-protein system was heated from 0 K to 298 K in increments of 30 K by temperature reassignment, where the velocities of all the atoms in the system are reassigned so that the entire system is set to the target temperature. The velocities were reassigned every 100 time steps for 10,000 time steps using the NVT ensemble. During the heating phase the protein backbone atoms were initially constrained using harmonic constraints with force constant $25 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and the force constant was gradually reduced by $2.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ every 1000 steps to zero. The systems were equilibrated for a further 25,000 time steps using the NPT ensemble, with the pressure set to 1 atm. Production dynamics were run for 300 ns for the apo, single ligand and double cholic acid ligand systems. For the mixed structures the production dynamics length was 220 ns.

5.4.3 Docking Calculations

Docking was performed between ILBP and a CA ligand. The atomic positions were taken from eleven snapshots sampled uniformly across the apo, single ligand and double CA ligand trajectories, to produce starting coordinates for

the docking receptor. For the structures bound to a single ligand, two sets of starting coordinates were generated, one with the ligand present and one with it absent. In total there were 44 sets of starting coordinates for the receptor. The ligand starting coordinates were taken from one of the interior ligands of the danio X-ray crystallographic structure. The ligand has seven rotatable bonds: C3-O3, C7-O7 and C12-O12 from the steroid rings and C17-C20, C20-C22, C22-C23 and C23-CD from the tail section, which were free to rotate during the docking calculations. The receptor structures did not have any rotatable bonds.

Calculations were performed using AutoDock Vina [24]. The input files for the ligands and receptors were prepared using AutoDock Tools [25], which automatically merges non-polar hydrogens to form united atom groups and assigns the partial charges and atom types required by the AutoDock scoring function. The atom types assigned to the ligand atoms that were part of the receptor were manually checked and they agreed with atoms types of protein side chains with the same or similar chemical groups. The origin of the search space was at the geometric centre of each receptor. The edge lengths of the cuboid that define the search space were between 36 Å and 50 Å and chosen so that the search space covered the entire exterior of the protein receptor. The exhaustiveness parameter was set to 50 and the energy range parameter to 7 kcal mol⁻¹.

5.5 Results

5.5.1 Thermodynamic stability

The stability of the MD simulations was established from the time-average of the kinetic energy, potential energy and total energy. Table 5-1 confirms that all the trajectories were stable.

	Kinetic Energy		Potential Energy		Total Energy		Temperature	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Apo	9483	75	-47338	146	-37924	135	297.2	2.3
Single Ligand	10464	79	-52614	126	-42138	155	297.2	2.2
Double CA	13336	89	-68423	132	-55201	176	297.1	2.0
Mixed CA	13588	90	-69744	159	-56278	159	297.3	2.0
Mixed CDA	13322	89	-68482	162	-55126	169	297.3	2.0

Table 5-1 Mean values and standard deviations of thermodynamic properties calculated for the production phase of each trajectory. Energies are in kcal mol⁻¹ and temperature in K.

5.5.2 Docking

The docked ligand-protein complexes predicted by AutoDock Vina were compared with the X-ray crystallographic structure of danio ILBP. To calculate the difference between ligands from a docked conformation and the experimental structure, the protein backbone atoms were aligned in VMD before the root mean square difference (RMSD) was calculated using the ligand heavy atoms. RMSD is defined as

(5-1)

$$RMSD = \sqrt{\frac{\sum_i d_i^2}{N}}$$

where d is the distance between the i^{th} atom in the two overlaid structures and N is the total number of atoms [26]. Results are reported for the first binding mode from each docking run, which is defined as the protein-ligand complex with the lowest binding affinity calculated by AutoDock Vina.

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

Receptor reference	Binding affinity (kcal mol ⁻¹)	Visualisation		RMSD from deleted ligand (Å)
0	-10.7	protein interior	tail up	1.3
1	-9.8	protein interior	tail down	3.7
2	-9.1	protein interior	tail up	1.4
3	-9.7	protein interior	tail up	1.4
4	-9.7	protein interior	tail up	1.4
5	-9	protein interior	tail down	4.0
6	-9.1	protein interior	tail up	1.8
7	-9.3	protein interior	tail up	1.4
8	-9.9	protein interior	tail up	1.5
9	-10.9	protein interior	tail up	1.4
10	-10.2	protein interior	tail up	1.4

Table 5-2 Results of binding to conformations from the single-ligand trajectories with the ligand deleted. The receptor reference 0 refers to the initial conformation at the start of the production dynamics, 1 to the conformation after 15,000,000 steps, 2 to the conformation after 30,000,000 steps etc. Tail up indicates a ligand orientation with the steroid rings pointing away from the helices.

Table 5-2 shows the results of binding to the structures taken from the single ligand trajectories after the ligand had been deleted. 9 of the 11 docking runs reproduce the original bound conformation with the same ligand orientation. The other two runs re-dock the ligand into the same site but with an incorrect orientation with the tail pointing toward the bottom of the binding cavity. The average RMSD between the docked ligand and the deleted ligand is 1.9 ± 0.9 Å.

Receptor reference	Bind affinity (kcal mol ⁻¹)	Visualisation		RMSD from danio x-ray ligands (Å)				
				site 1	site 2	site 3	site 4	site 5
0	-6.3	exterior	rings near E-F loop, tail near α -II	-	-	4.7	6.0	4.6
1	-5.9	exterior	between strands E-F at bottom of β -barrel	-	-	13.1	9.6	12.2
2	-8.7	interior	tail up	3.4	4.6	-	-	-
3	-6.6	exterior	between strand A and α -I	-	-	8.6	11.7	9.8
4	-7.1	interior	tail down	4.3	5.2	-	-	-
5	-7.9	interior	tail up	2.4	5.5	-	-	-
6	-5.8	exterior	between strand A and α -I & α -II	-	-	10.0	12.1	10.5
7	-5.8	exterior	between strands F-G and α -I- α -II loop	-	-	9.5	3.3	6.5
8	-6.4	exterior	between strands F-G and α -I- α -II loop	-	-	8.4	2.9	5.8
9	-6.0	exterior	between strand A, α -I and α -II	-	-	9.4	11.3	9.5
10	-6.1	exterior	between strand A, α -I and α -II	-	-	9.5	11.6	9.9

Table 5-3 Results of binding to conformations from the apo trajectories. Receptor reference naming convention is the same as Table 5-2. Site 1 corresponds to ligand residue 150 in the 3ELZ PDB file, site 2 to 151, site 3 to 152, site 4 to 153 and site 5 to 200.

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

Table 5-3 shows the docking results using the coordinates from the apo MD trajectories. For three of the receptor structures (2, 4 and 5) the ligand is docked in the interior of the protein in site 1. For two of these the ligand has the correct orientation with the tail pointing toward the helices; the third has the tail pointing toward the bottom of the binding cavity. The ligands docked to the protein interior have a greater binding affinity, between -7.1 to -8.7 kcal mol⁻¹, than the exterior ligands, which have a range of -5.8 to -6.6 kcal mol⁻¹. Two exterior binding sites were docked into from more than one starting conformation. The first (referred to in this chapter as site D-1) is located between β -strands F and G and the loop region connecting the two α -helices (see Figure 5-4a). Site D-1 is in a similar location to site 4 of the experimental structure and the docked ligands have an RMSD of 2.9 Å and 3.3 Å. A second exterior binding site (referred to as site D-2), between β -strand A and the helices (see Figure 5-4c), was predicted for four of the starting receptor conformations (3, 6, 9, 10). This binding site does not correspond to any of the experimentally determined exterior sites. For the other two docking runs the ligand was docked to the exterior on a site near the E-F loop and α -helix II and on a site against the E-F strand at the bottom of the β -barrel.

Receptor reference	Bind affinity (kcal mol ⁻¹)	Visualisation		RMSD from danio X-ray ligands (Å)				
				site 1	site 2	site 3	site 4	site 5
0	-5.8	exterior	near G-H strand	7.5	11.0	14.1	8.2	11.5
1	-6.6	exterior	between F-G and α -I	6.7	8.8	10.0	3.9	7.0
2	-6.3	exterior	between F-G and α -I	6.0	7.2	8.0	3.6	5.5
3	-6.0	exterior	between F-G and α -I	7.1	8.1	9.1	3.2	5.7
4	-6.2	exterior	near B-C strand	8.4	7.5	8.5	11.1	10.2
5	-7.2	exterior	between F,G and α -I- α -II loop	6.1	7.7	9.2	3.5	6.3
6	-7.2	exterior	between F,G and α -I- α -II loop	6.7	6.6	8.0	3.3	4.7
7	-6.6	exterior	between F,G and α -I- α -II loop	6.0	7.5	8.8	2.7	5.8
8	-6.9	exterior	between F,G and α -I- α -II loop	6.3	8.0	9.5	3.6	6.6
9	-6.4	exterior	between F,G and α -I- α -II loop	7.4	7.2	7.6	3.7	4.2
10	-6.1	exterior	between A, α -I and α -II	10.4	6.9	9.2	12.5	10.7

Table 5-4 Results of binding to conformations from the single ligand trajectories. Receptor reference naming the same as Table 5-2. Site 1 corresponds to ligand residue 150 in the 3ELZ PDB file, site 2 to 151, site 3 to 152, site 4 to 153 and site 5 to 200.

The results of docking to receptor structures from the single ligand trajectories with a ligand already present (in site 1) as part of the receptor are shown in Table 5-4. Eight of the docking runs predict binding to site D1 and the average RMSD from experimental site 4 for these eight complexes is 3.4 ± 0.4 Å. Of the other three docking runs, one predicts binding to the D2 site, one to a site on the G-H β -strands and one to a site on the B-C β -strands. Table 5-5 reports the results of docking to structures taken from the double CA trajectories with two ligands already present as part of the receptors. Four of the runs dock the ligand inside the binding cavity, three between the ligands in site 1 and site 2 and the fourth at the bottom of the cavity beneath the ligand in site 2. Five of the runs predict binding to site D-1 with an average RMSD from site 4 of danio ILBP of 4.2 ± 0.8 Å. The other two runs predict binding to site D-2 and a site along the C-D, E-F turns and α -helix I.

Receptor reference	Bind affinity (kcal mol ⁻¹)	Visualisation		RMSD from danio X-ray ligands (Å)				
				site 1	site 2	site 3	site 4	site 5
0	-7.3	exterior	between F-G and α -I	5.6	6.9	8.6	3.7	6.0
1	-8.3	interior	between site 1 and site 2	3.9	4.4	7.1	5.0	6.1
2	-7.9	interior	between site 1 and site 2	4.7	4.4	7.4	6.2	6.6
3	-6.8	exterior	between F, G and α -I	6.3	6.5	9.0	4.3	6.2
4	-6.9	exterior	between F, G and α -I- α -II loop	5.5	5.5	7.4	3.1	4.6
5	-8.2	exterior	between F, G and α -I- α -II loop	4.7	4.5	7.1	5.5	5.6
6	-9.1	interior	between site 1 and site 2	4.8	3.9	7.5	7.2	7.3
7	-8.7	interior	beneath site 2	4.7	4.8	8.5	8.5	7.8
8	-7.1	exterior	between C-D turn, E-F turn and α -I	7.0	4.5	4.2	5.2	3.7
9	-6.8	exterior	between A, α -I and α -II	10.3	7.1	9.9	12.3	10.8
10	-6.7	exterior	between F, G and α -I	5.6	5.8	7.4	4.3	5.0

Table 5-5 Results of binding to conformations from the double ligand trajectories. Receptor reference naming convention is the same as Table 5-2. Site 1 corresponds to ligand residue 150 in the 3ELZ PDB file, site 2 to 151, site 3 to 152, site 4 to 153 and site 5 to 200.

5.5.3 MD Simulations from Docked Complexes

Three of the docked protein-ligand complexes were selected as starting structures for further MD simulations to study binding sites D-1 and D-2. Site D-1 is of interest because it is close to the experimentally identified exterior binding site 4 and because 8 of the 11 docking runs using single ligand receptors docked into the position. Site D-2 is of interest because 4 of the 11

docking runs using apo receptors docked to it. From the results of docking to single ligand structures, the complexes corresponding to receptors 5 and 6 have the lowest binding affinity ($-7.2 \text{ kcal mol}^{-1}$); model 6 was chosen for MD simulation of D-1, because it has a smaller RMSD with respect to the experimental site 4. Models 3 and 8 were chosen from the receptors from the apo trajectories for MD simulations of sites D-2 and D-1 respectively, because they have the lowest binding affinities ($-6.6 \text{ kcal mol}^{-1}$ and $-6.4 \text{ kcal mol}^{-1}$).

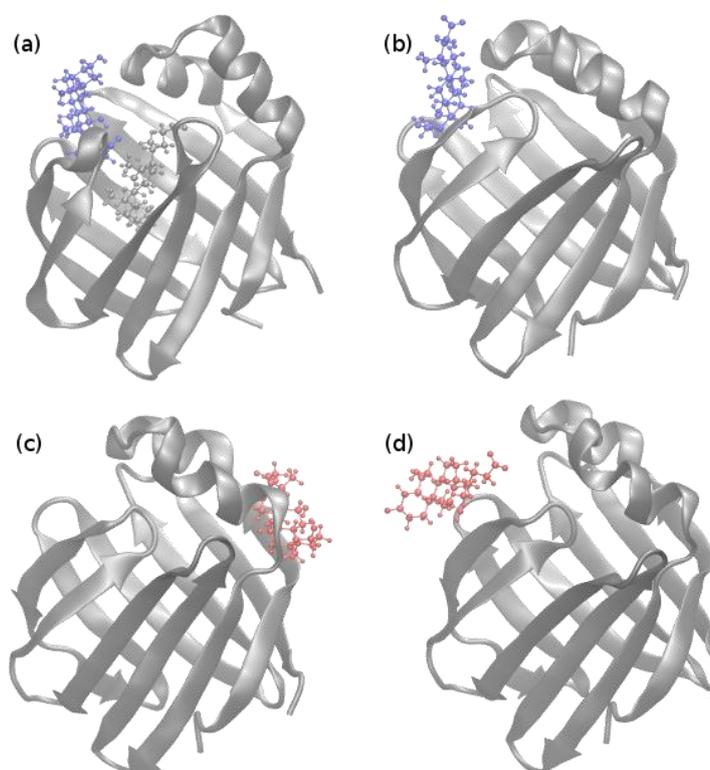


Figure 5-4 (a) cholic acid docked into binding site D-1 with ligand present in interior site 1 (single ligand receptor model 6). (b) cholic acid docked into binding site D-1 (apo receptor model 8). (c) cholic acid docked into binding site D-2 at the start of the MD trajectory and having moved near to site D-1 after 63.3 ns of simulation (d).

The MD simulations were set up and run following the same protocol as described in Section 5.4.2. The length of the production dynamics was 320 ns for the complex with a single ligand in site D-1 and 270 ns for the complex with an interior ligand in site 1 and a ligand in site D-1. Figure 5-5 shows the RMSD of the ligands from their starting position during these simulations. For the trajectory with only the exterior ligand, the RMSD is low throughout the simulation with an average of $2.5 \pm 1.1 \text{ \AA}$, indicating that the ligand remains in

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

position in the binding site throughout. For the simulation with a ligand in interior binding site 1, the average RMSD is greater, $8.8 \pm 2.8 \text{ \AA}$, with the ligand initially moving away from its starting position before moving close to the D-1 site again for the last 70 ns. Visualisation of the trajectory shows that although the ligand moves away from the binding site, it stays within the vicinity of D-1.

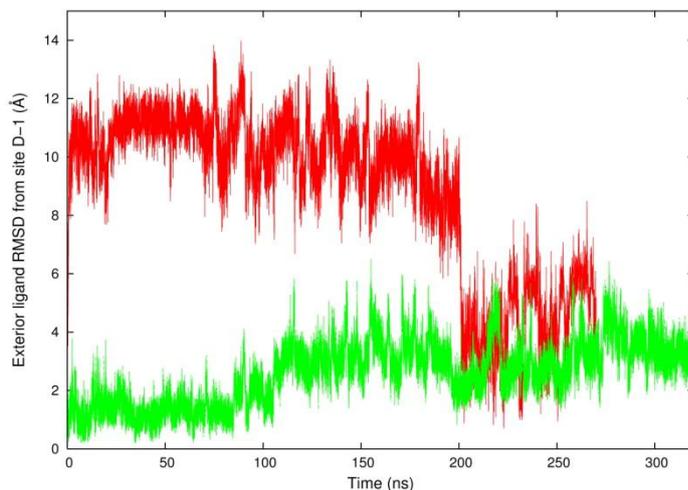


Figure 5-5 RMSD of the ligand from its initial position in site D-1. Red corresponds to cholic acid in site 1 and green to no interior ligand.

During the simulation with the ligand in site D-2 the RMSD from the starting position was $19.7 \pm 10.1 \text{ \AA}$ after 63 ns. Over the course of the trajectory, the ligand detached entirely from site D-2 and moved to site D-1. Figure 5-4c and Figure 5-4d show the start and end positions of this simulation and Figure 5-6 shows the RMSD from site D-1 during the trajectory.

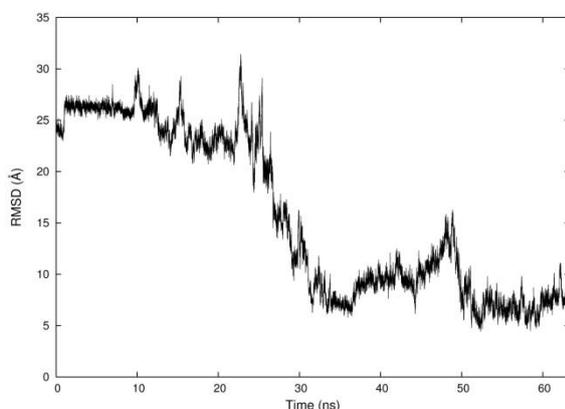


Figure 5-6 RMSD of the ligand that starts in site D-2 to the ligand in site D-1. The apo-receptor model 8 was used as the reference coordinates of D-1.

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

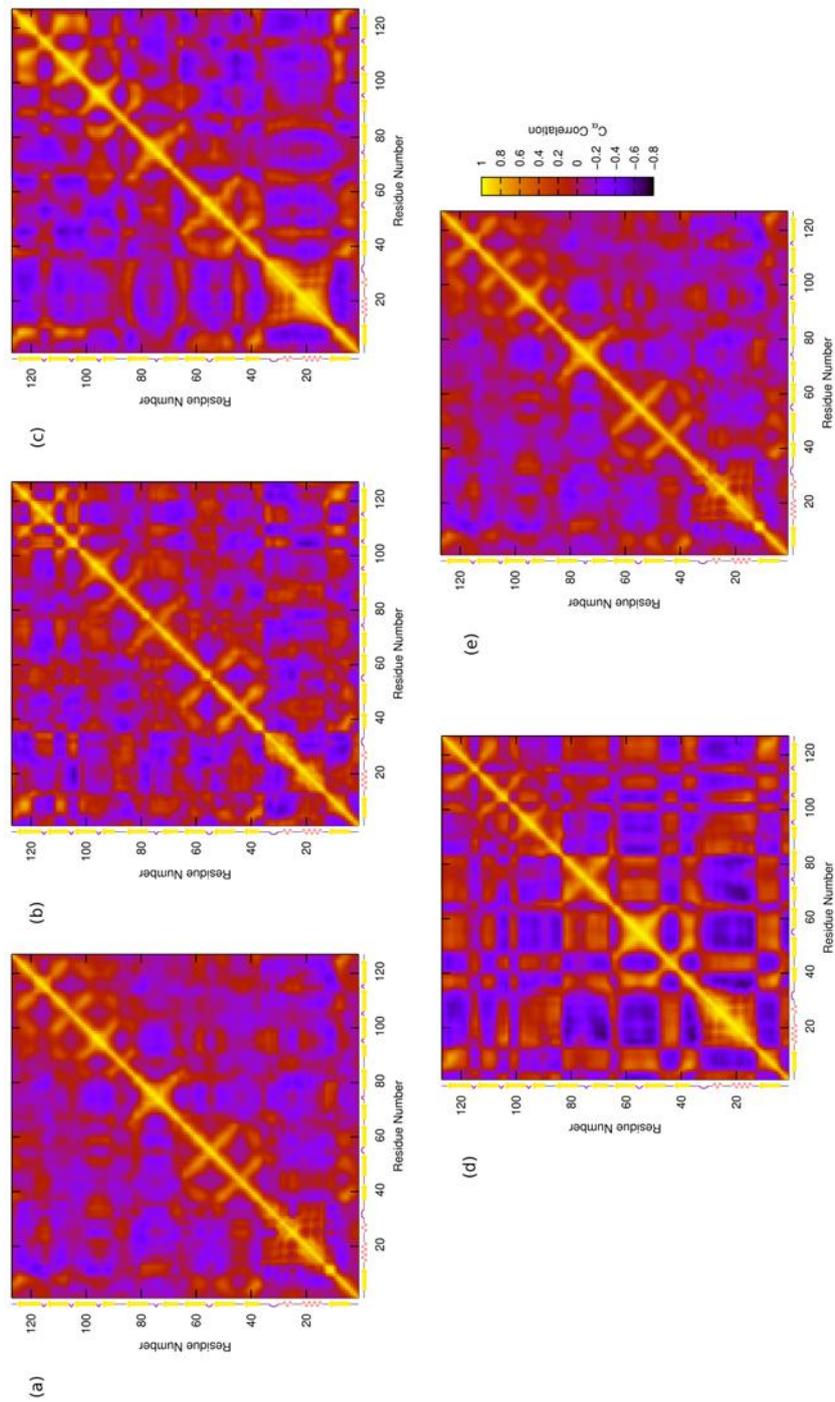


Figure 5-7 Covariance of C_{α} atoms for the (a) apo trajectory (b) single CA trajectory (c) double CA trajectory (d) mixed-CA trajectory and mixed-CDA trajectory. Secondary structure assignment from PDB.

5.5.4 Covariance Analysis

Correlation coefficients, referred to in the CHARMM literature as covariance coefficients, measure the correlation between the positions of two atoms during an MD trajectory. The correlation coefficient between two atomic position vectors, x and y , is given in Equation (5-2) [26], where M is the number of frames in the trajectory file. The value of c_{xy} ranges from -1 to 1. A value of 1 indicates positive correlation where the atoms move together; a value of 0 indicates no correlation and the movements can be considered totally independent and a value of -1 indicates negative correlation where the atomic movements are in opposite directions. By calculating covariance coefficients between atoms in different parts of the secondary structure during the simulations, it is possible to gain understanding of the internal motions and structural changes of the protein.

(5-2)

$$c_{xy} = \frac{\frac{1}{M} \sum_{i=1}^M (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\left(\frac{1}{M} \sum_{i=1}^M (x_i - \langle x \rangle)^2\right) \left(\frac{1}{M} \sum_{i=1}^M (y_i - \langle y \rangle)^2\right)}}$$

Covariance coefficients between the C_α atoms of all residues were calculated for the apo, single, double CA, mixed CA and mixed CDA trajectories (Figure 5-7). All trajectories have positive correlation between the α -helices, indicating that they move together, except the single ligand trajectory which has a region of negative correlation between α -I and α -II. Negative correlation between the E-F strands and β -turn and the G-H strands and β -turn during the apo trajectory becomes positive upon binding of a ligand into site 1 in the single trajectory but becomes negative again when two ligands are present.

The map for the double CA trajectory and the mixed CDA trajectory (Figure 5-7) are similar in terms of pattern and range of the calculated covariances. This is in contrast to mixed CA which has regions of strong negative correlation between the α -helices and strands A, B, C-D and E-F. These areas of negative correlation are indicative of a motion where the helices move away from the

front of the β -barrel, making the binding cavity more accessible for ligand exchange.

5.5.5 Interaction between Gln51 and ligands

Toke *et al.* [6] reported that the point mutation Gln51Ala disrupts the site selectivity mechanism of human ILBP. The location of Gln51 (Figure 5-8) makes it well suited as a reference point within the binding cavity to measure how 'deep' the ligands are within the binding pocket. The distance between the centre of mass (COM) of each ligand and Gln51 was calculated every 2 ps and plotted as a histogram with a bin size of 0.1 Å (shown in Figure 5-9 and Figure 5-10). The averages of these distributions were considered in terms of their modes.

When chenodeoxycholic acid is in site 1 and cholic acid in site 2, the mode of the distance between the COMs is 8.1 Å; when cholic acid is in both sites, the mode is 8.6 Å. For the trajectory with cholic acid in site 1 and chenodeoxycholic acid in site 2 there are two peaks, one at 8.6 Å and a second at 11.3 Å. The distribution of the distances between the COMs as a function of time (data not shown) shows that the first peak corresponds to the ligands position for the first ~70 ns, after which the ligand moves away further from Gln51 for the remainder of the simulation, corresponding to the second peak.

During the mixed CDA trajectory, the ligand in site 2 (cholic acid) is closer to Gln51, and therefore deeper in the binding pocket than the ligand in site 2 for the double CA trajectory and the mixed CDA trajectory. For the mixed CDA simulation the average distance (the mode of the distribution) between the COMs of the ligand and Gln51 is 8.8 Å, for mixed CA it is 10.6 Å and for double CA there are two peaks – one at 11.8 Å and a second at 13.2 Å.

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

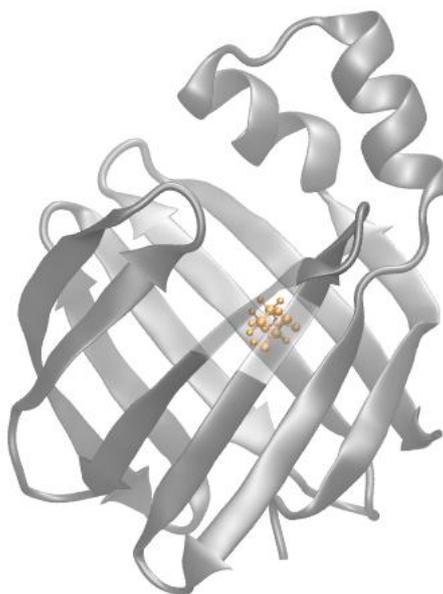


Figure 5-8 Location of Gln51 (shown in gold) inside the binding cavity.

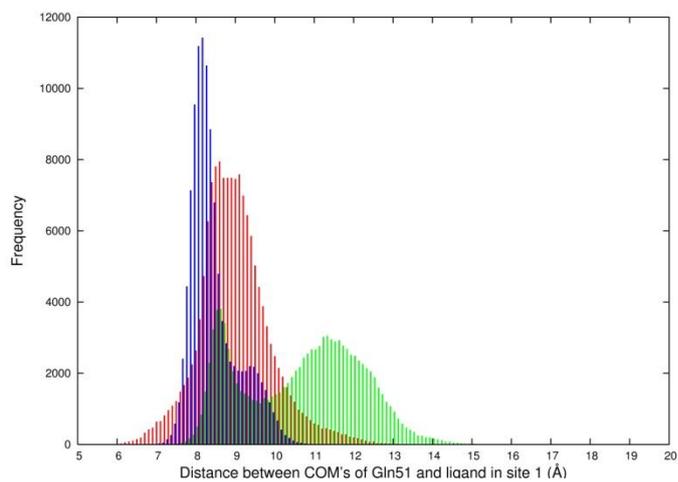


Figure 5-9 Histogram of the distance between the COM of the ligand in site 1 and Gln51 for the double CA (red), mixed CA (green) and mixed CDA (blue) trajectories.

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

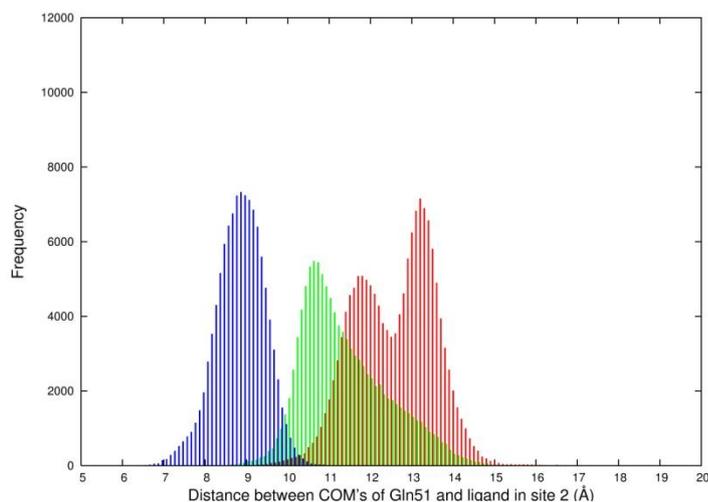


Figure 5-10 Histogram of the distance between the COM of the ligand in site 2 and Gln51 for the double (red) CA, mixed CA (green) and mixed CDA (blue) trajectories.

Hydrogen bonding between the sidechain of Gln51 and the hydroxyl groups of the steroid ring region of the ligands was identified using the CHARMM Hbond analysis module, sampling structures every 4 ps. During the double CA trajectory a hydrogen bond is present between H₇ (the hydrogen in the OH group at steroid ring position 7) and O_ε (the oxygen in the Gln-51 sidechain) for 3% of the trajectory frames. For the mixed CA trajectory this hydrogen bond is present for 17% of the frames but for the mixed CDA trajectory the hydrogen bond is present for 50% of the frames. No other direct hydrogen bonds between the ligands or between the ligands and Gln51 are present.

5.5.6 Orientation of ligands and helices

The second α -helix has been proposed to be involved in the mechanism of ligand binding for several members of the iLBP family [7]. To examine the position of the helices and ligands and to determine if the apo, single and doubly ligated proteins adopt similar conformations in this domain, three pairs of vectors were defined. To measure the angle of α -I with respect to the β -barrel, a vector was defined along the length of the helix, from C _{α} -Glu9 (strand A) to C _{α} -Gly22 (end of α -I) and along the top of the β -barrel from C _{α} -Glu9 to C _{α} -Thr78 (strand G). A pair of vectors from C _{α} -Val37 (strand B) to C _{α} -Ser25 (end of α -II) and C _{α} -Val37 to C _{α} -Gln72 (strand F) was used to measure the angle of α -II with respect to the β -barrel. For the simulations with two ligands inside the binding cavity, vectors were defined along the length of the

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

ligands from C₃ to C₂₄ (*i.e.* the carbon in the first steroid ring to the tail carboxylate group). The angle between these three pairs of vectors was calculated every 2 ps and is plotted as histograms with a bin size of 1° in Figure 5-11 to Figure 5-15.

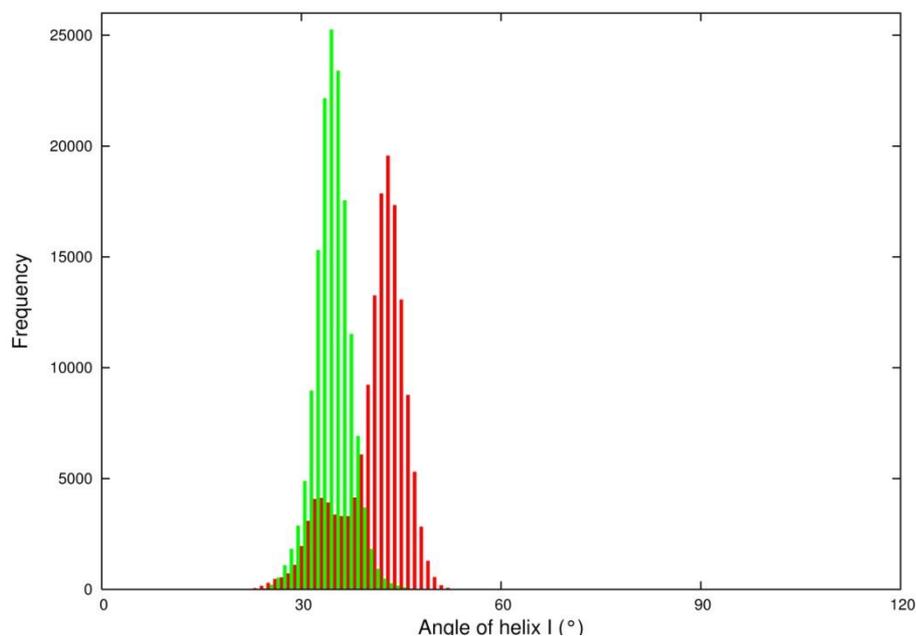


Figure 5-11 Histogram of the angle of helix I with respect to the β -barrel for the apo (green) and single CA (red) simulations

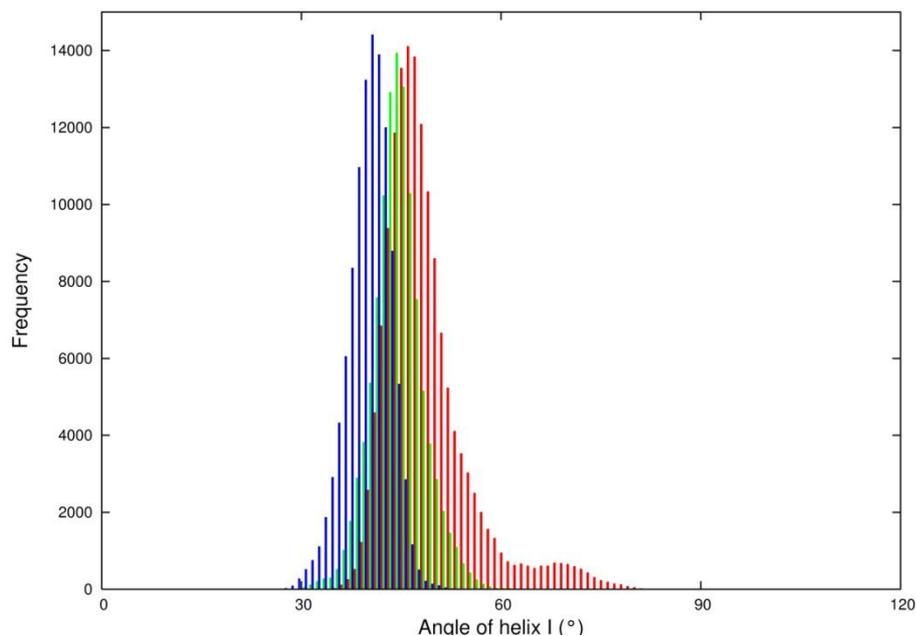


Figure 5-12 Histogram of the angle of helix I with respect to the β -barrel for the double CA (red), mixed CA (green) and mixed CDA (blue) simulations.

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

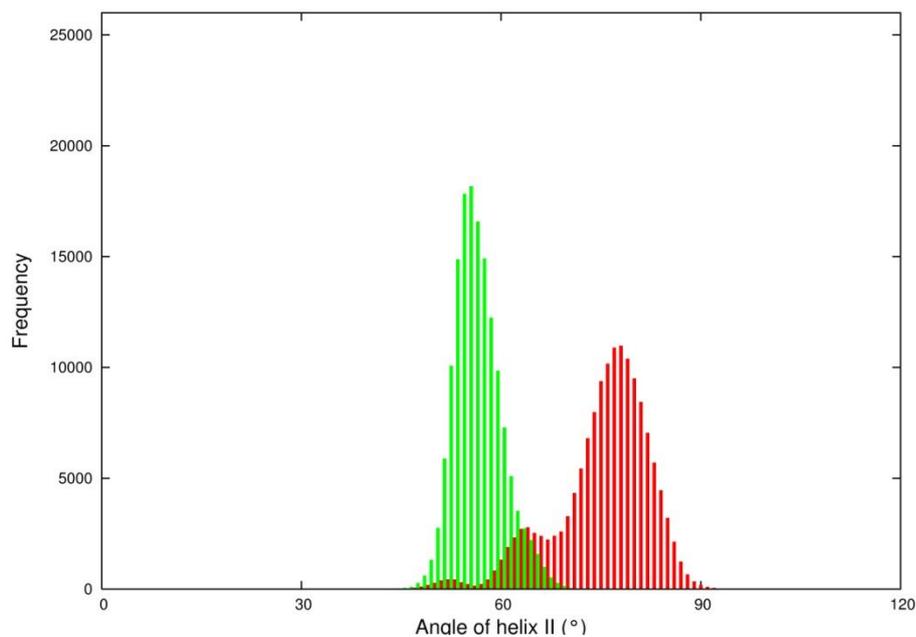


Figure 5-13 Histogram of the angle of helix II with respect to the β -barrel for the apo (green) and single CA (red) simulations

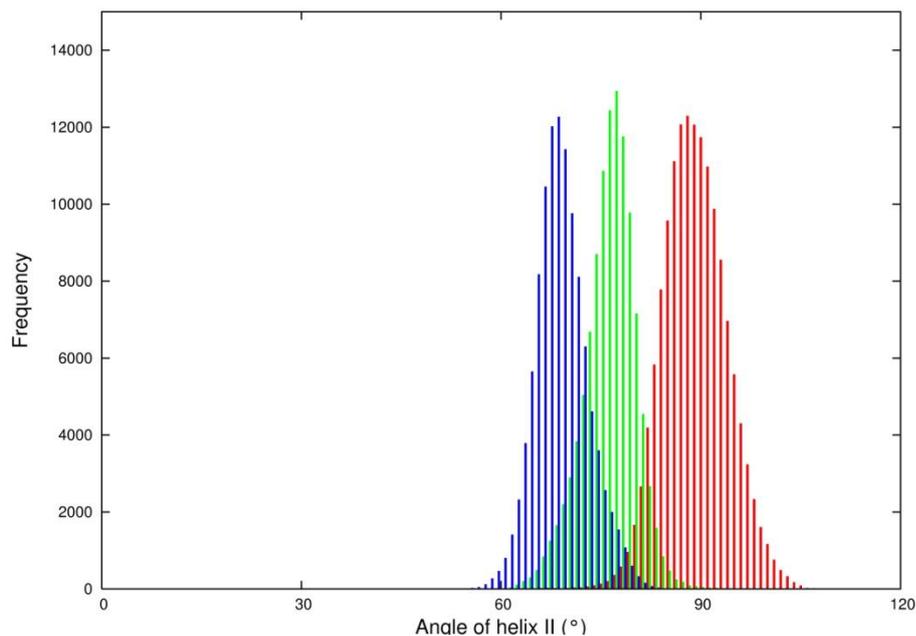


Figure 5-14 Histogram of the angle of helix II with respect to the β -barrel for the double CA (red), mixed CA (green) and mixed CDA (blue) simulations.

5. Cooperativity and Site Selectivity in the Ileal Lipid Binding Protein

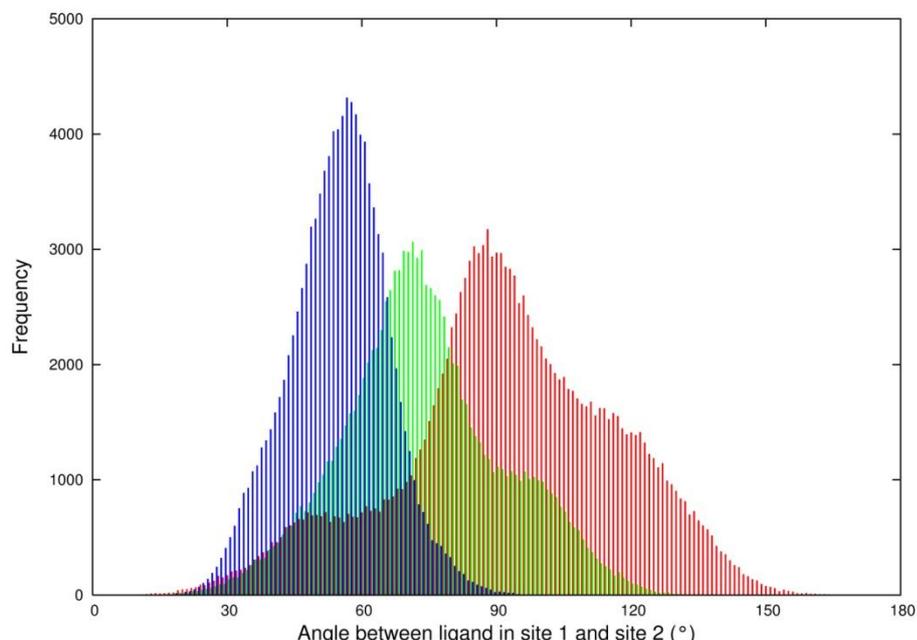


Figure 5-15 Histogram of the angle between the ligands for the double CA (red), mixed CA (green) and mixed CDA (red) simulations.

Table 5-6 reports the mean of the vectors, the standard deviation of the mean and the mode of distribution. These results are discussed below in Section 5.6.1.

Vector	α -I w.r.t β -barrel			α -II w.r.t β -barrel			site 1 w.r.t site 2		
	Mean	Std Dev	Mode	Mean	Std Dev	Mode	Mean	Std Dev	Mode
Apo	34.7	2.7	34	56.6	3.6	55	-	-	-
Single CA	41.5	4.8	43	75.8	7.1	78	-	-	-
Double CA	49.1	6.6	46	89.6	4.8	88	92.6	24.8	88
Mixed CA	44.7	3.8	44	76.5	3.9	77	73.8	18.0	71
Mixed CDA	40.1	3.2	40	69	3.9	68	54.5	11.0	56

Table 5-6 Average values of the angles, and their standard deviations, in degrees, defined to examine the angle of the α -helices with respect to the β -barrel and the angle between the ligands in sites 1 and 2.

5.6 Discussion

5.6.1 Structural Differences of the Simulations

The protein has notable variation in the tertiary structure during the different simulations. The angle between the helices and the β -barrel is smallest during the apo simulation (α -I 34° , α -II 55°). When a single ligand is present, the helices move away from the β -barrel to an angle of 46° for α -I and 78° for α -II. Figure 5-16 shows the angles of the helices with respect to the β -barrel during the simulations of the protein with cholic acid docked to the exterior binding site D-1 and no ligands inside the binding cavity. The angle of α -I increases from $\sim 36^\circ$ to $\sim 41^\circ$ and the angle of α -II increases from $\sim 52^\circ$ to $\sim 62^\circ$ and is still increasing at the end of the simulation. During the simulation of the apo structure the helices do not show this increase in angle with respect to the β -barrel (data not shown). This suggests that a ligand associated with binding site D-1 induces changes in the protein conformation to a shape that matches the conformation when ligands are bound to the protein interior.

A recent study of the internal backbone motions of IBLP using NMR spectroscopy by Horvath *et al.* [8] supports the hypothesis that there is an allosteric mechanism of ligand binding and that the protein shifts from a closed state to an open state before interior ligand binding [9]. The results of the simulation suggest that the hydrophobic adhesion of a ligand to the exterior of the protein into site D-1 initiates such a transition to an open state, where the α -helices move away from the β -barrel to allow ligands into the binding cavity.

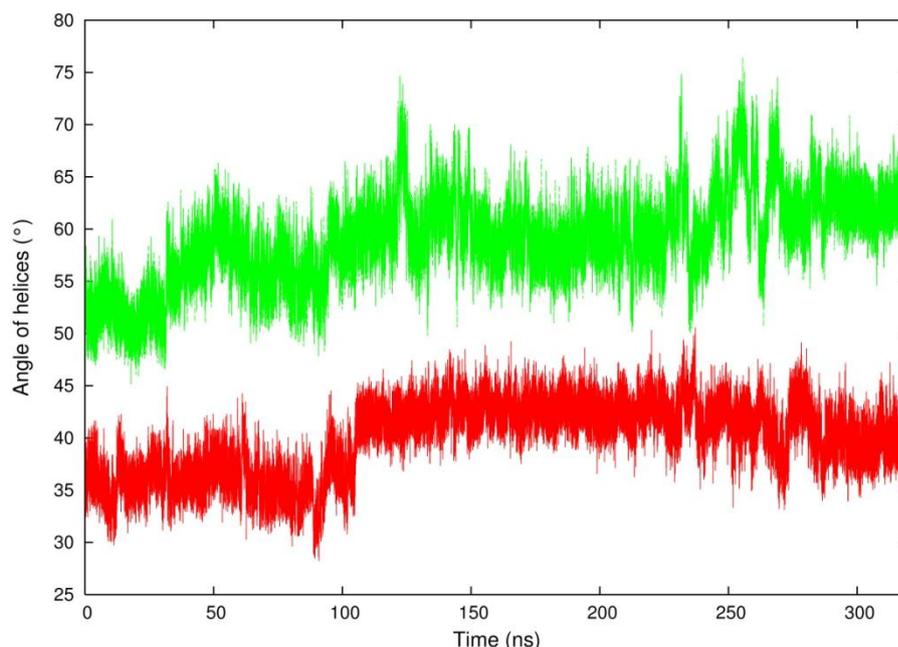


Figure 5-16 Angle of α -I with respect to the β -barrel (as defined in Section 5.5.6) as a function of time step when a ligand is bound to exterior site D-1 and the interior sites are empty (red) and a ligand is bound to site 1 (green).

5.6.2 A Third Binding Site

As discussed in Section 5.3, there is strong experimental evidence that ILBP has three binding sites. Docking to conformations from the MD trajectories identified three possible sites for the third ligand: site D-1, site D-2 or inside the binding cavity with the interior ligands. During the extended MD simulations with the docked structures as starting structures (Section 5.5.3), the ligand in D-2 moved away from its initial position and migrated toward the D-1 binding site. This suggests that the binding affinity predicted by AutoDock Vina for site D-2 was partly based on a steric interaction which is not maintained when the equilibrium conformational dynamics of the protein are considered.

Site D-1 corresponds to site 4, or site 153, of the X-ray crystallographic structure of zebrafish ILBP [3]. A ligand is resolved in this location in all models of both unit cells and is the only exterior ligand consistently identified in the same place. AutoDock Vina predicts four structures with three ligands inside the binding cavity, due to the strong hydrophobicity and lack of steric clashes.

However, the five X-ray crystallographic structures of danio ILBP [3] indicate two ligands only in the interior and several on the exterior, including the D-1 position. Furthermore, the MD simulations suggest there is a role for D-1 in an allosteric interaction and predict a migration of the ligand in D-2 to D-1. Taken together, we conclude that it is more probable that the third binding site is D-1 and not between the interior ligands.

5.6.3 Preferred Binding Site of Ligands and a Proposed Mechanism of Site Selectivity

Although no complete NMR or X-ray crystallographic structure of ILBP bound to both a conjugate of cholic acid and chenodeoxycholic acid is available, Toke *et al.* [6] propose that GCDA is bound to site 1 and GCA to site 2 based on homology modelling of porcine ILBP [10] and preliminary NMR results. The covariance analysis of the mixed CA and mixed CDA (Figure 5-7) simulations strongly supports this proposal as the correct ligand configuration. During the mixed CA simulation there is strong negative correlation between the α -helices and strands A, B, C-D and E-F, indicative of a motion where the helices are moving away the front region of the β -barrel. The covariance map of the simulations with mixed CDA has less extreme regions of positive and negative correlation, indicating a structure that is not undergoing this type of internal motion. This comparative structural stability indicates that chenodeoxycholic acid is the preferred ligand in site 1 and cholic acid is the preferred ligand in site 2.

The increased stability of the mixed CDA simulation compared with the mixed CA simulation, together with secondary structure differences, lead to a proposed mechanism of site selectivity. Chenodeoxycholic acid bile salts have one less OH group than cholic acid and are therefore more hydrophobic. When chenodeoxycholic acid is in binding site 1 this increased hydrophobicity leads it to sit deeper in the binding pocket. This is illustrated by the distribution of distances between Gln51 and the site 1 ligand (Figure 5-9) which has a mode of 8.1 Å for mixed CDA and had two peaks at 8.6 Å and 11.3 Å for mixed CA. When chenodeoxycholic acid is in site 1 it is anchored in the

binding pocket by a hydrogen bond between the OH group on carbon 7 and the side chain of Gln51. This hydrogen bond is present for 50% of the simulation when chenodeoxycholic acid is in site 1, but only 17% of the time during mixed CA simulation and for just 3% of the double CA trajectory. With the bile salt in site 1 deeper in the binding pocket, this increases the hydrophobicity of the binding cavity and encourages the ligand in site 2 to be deeper in the pocket also, as shown in Figure 5-10, where the mode of the distance between the ligands in site 2 and Gln51 is 8.8 Å for mixed CDA (with cholic acid in site 2) compared with 10.6 Å for mixed CA (with chenodeoxycholic acid in site 2) and two peaks for double CA at 11.8 Å and 13.2 Å. The mixed CDA configuration of ligands induces the protein to change to a conformation where the angle of the helices is closer to the β -barrel compared with mixed CA and double CA (Figure 5-11 to Figure 5-14) and the angle between the vectors along the bile salts is smaller for the mixed CDA trajectory than the double CA or mixed CA (Figure 5-15). This proposed mechanism of site selectivity is energetic, rather than steric, in agreement with the mechanism identified by Tochtrop *et al.* [2] using NMR methods.

5.7 Conclusions

MD simulations of ILBP were performed using various ligand configurations in combination with docking calculations to examine binding behaviour. The simulations agree with the experimental observations of Fang [12] and Toke *et al.* [6] that ILBP has three binding sites – two on the protein interior and a third uncharacterised site. Three possible locations for the third binding site are identified – D-1, D-2 and an interior site between the ligands. When a ligand was bound to the D-1 site of the protein, the helices moved away from the β -barrel, into a position similar to the conformation of the holo protein. This supports the hypothesis of an allosteric binding mechanism suggested by Toke *et al.* [9] and Horvath *et al.* [8]. Given the possible allosteric role of a ligand in D-1 and the presence of ligands in this binding site in the X-ray crystal structure of zebrafish ILBP, it is more probable that D-1 is the third binding site than D-2 or between the interior ligands.

The simulations also confirm the suggested [6] preferred ligand conformation of chenodeoxycholic acid in site 1 and cholic acid in site 2. A mechanism of site selectivity is proposed, based upon secondary structure differences between different MD trajectories, where the increased hydrophobicity of chenodeoxycholic acid, due to having one less OH group, leads it to sit deeper in the binding pocket, anchored by a hydrogen bond to Gln-51. This leads to an increase in hydrophobicity in the pocket region, causing the cholic acid ligand in site 2 to bind deeper in the pocket also and inducing a change in the protein conformation where the angle between the ligands main axis is smaller and the α -helices are closer to the β -barrel, preventing further ligand exchange.

5.8 References

1. Hanhoff, T., C. Lucke, and S. F., *Insights into binding of fatty acids by fatty acid binding proteins*. Molecular and Cellular Biochemistry, 2002. **239**: p. 45-54.
2. Tochtrop, G.P., G.T. DeKoster, D.F. Covey, and D.P. Cistola, *A single hydroxyl group governs ligand site selectivity in human ileal bile acid binding protein*. J. Am. Chem. Soc., 2004. **126**(35): p. 11024-11029.
3. Capaldi, S., G. Saccomani, D. Fessas, M. Signorelli, M. Perduca, and H.L. Monaco, *The X-Ray Structure of Zebrafish (Danio rerio) Ileal Bile Acid-Binding Protein Reveals the Presence of Binding Sites on the Surface of the Protein Molecule*. J. Mol. Biol., 2009. **385**(1): p. 99-116.
4. Sorg, J.A. and A.L. Sonenshein, *Inhibiting the Initiation of Clostridium difficile Spore Germination using Analogs of Chenodeoxycholic Acid, a Bile Acid*. J. Bacteriol., 2010. **192**(19): p. 4983-4990.
5. Tochtrop, G.P., J.L. Bruns, C.G. Tang, D.F. Covey, and D.P. Cistola, *Steroid ring hydroxylation patterns govern cooperativity in human bile acid binding protein*. Biochemistry, 2003. **42**(40): p. 11561-11567.
6. Toke, O., J.D. Monsey, G.T. DeKoster, G.P. Tochtrop, C.G. Tang, and D.P. Cistola, *Determinants of cooperativity and site selectivity in human ileal bile acid binding protein*. Biochemistry, 2006. **45**(3): p. 727-737.
7. Eberini, I., A.G. Rocco, A.R. Ientile, A.M. Baptista, E. Gianazza, S. Tomaselli, H. Molinari, and L. Ragona, *Conformational and dynamics changes induced by bile acids binding to chicken liver bile acid binding protein*. Proteins, 2008. **71**(4): p. 1889-1898.
8. Horvath, G., P. Kiraly, G. Tarkanyi, and O. Toke, *Internal motions and exchange processes in human ileal bile acid binding protein as studied by backbone ¹⁵N nuclear magnetic resonance spectroscopy*. Biochemistry, 2012. **51**: p. 1848-1861.

9. Toke, O., J.D. Monsey, and D.P. Cistola, *Kinetic mechanism of ligand binding in human ileal bile acid binding protein as determined by stopped-flow fluorescence analysis*. *Biochemistry*, 2007. **46**(18): p. 5427-5436.
10. Lucke, C., F.L. Zhang, J.A. Hamilton, J.C. Sacchettini, and H. Ruterjans, *Solution structure of ileal lipid binding protein in complex with glycocholate*. *Eur. J. Biochem.*, 2000. **267**: p. 2929-2938.
11. Kouvatso, N., *Characterisation of rabbit ileal lipid binding protein and design of new beta-scaffold proteins*, in *George Green Library*. 2006, University of Nottingham: Nottingham.
12. Fang, H.-J., *An Investigation of Ileal Bile Acid Binding Protein and its Application as a Biosensor for Performance Enhancing Drugs*, in *George Green Library*. 2011, University of Nottingham: Nottingham.
13. Kurz, M., V. Brachvogel, H. Matter, S. Stengelin, H. Thuring, and W. Kramer, *Insight into the bile acid transportation system: The human ileal lipid-binding protein-cholytaurine complex and its comparison with homologous structures*. *Proteins*, 2003. **50**(2): p. 312-328.
14. Brooks, B.R., R.E. Bruccoleri, D.J. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*. *J. Comput. Chem.*, 1983. **4**: p. 187-217.
15. Brooks, B.R., C.L. Brooks, A.D. Mackerell Jr, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, and M. Karplus, *CHARMM: The Biomolecular Simulation Program*. *J. Comput. Chem.*, 2009. **30**(10): p. 1545-1614.
16. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. *J. Mol. Graphics*, 1996. **14**(1): p. 33-38.
17. Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins, *ClustalW and ClustalX version 2*. *Bioinformatics*, 2007. **23**(21): p. 2947-2984.
18. Schleif, R., *Analysis of Protein Structure and Function: A Beginner's Guide to CHARMM*. 2006, Johns Hopkins University: Baltimore.
19. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water*. *J. Chem. Phys.*, 1983. **79**(2): p. 926-936.
20. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, *Scalable Molecular Dynamics with NAMD*. *J. Comput. Chem.*, 2005. **26**(16): p. 1781-1802.
21. Vanommeslaeghe, K., E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A.D.M. Jr., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields*. *J. Comput. Chem.*, 2009. **31**(4): p. 671-690.

22. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. J. Comput. Phys., 1977. **23**: p. 327-341.
23. Darden, T.A., D. York, and L. Pedersen, *Particle-mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems*. J Chem Phys, 1993. **98**: p. 10089-10092.
24. Trott, O. and A.J. Olson, *AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading*. J. Comput. Chem., 2010. **31**: p. 455-461.
25. Sanner, M.F., *Python: A programming Language for Software Integration and Development*. J. Mol. Graphics Mod., 1999. **17**: p. 57-61.
26. Leach, A.R., *Molecular modelling: principles and applications*. 2nd ed. 2001, Harlow: Pearson Education.

6 Conclusion and Future Work

6.1 Conclusions

In this thesis, computational methods have been used to predict the folding behaviour of an antibiotic peptide and understand the unusual binding behaviour of a transporter protein. Computational methods are a complement to experimental techniques in the study of polypeptides and often guide the research direction of more expensive experimental work. For example, docking calculations are used to screen large libraries of drug-like compounds quickly and cheaply for molecules that bind to a biomolecule of interest before synthesis and testing *in vitro*. MD simulations are used to understand specific atomic interactions and movements on timescales that cannot be easily accessed using experimental methods. MD can also be used to model unphysical processes such as computational alchemy that is used to calculate the free energy difference between two similar molecular systems. A limitation of MD simulations is that they rely on empirical potential energy functions, with parameters fitted to reproduce values and geometries from experiments and QM-level calculations, which may not transfer well to a new molecular system. A second limitation is that MD cannot simulate chemical reactions, such as catalysis and the forming and breaking of disulfide bonds. *Ab initio* MD uses QM methods to calculate the forces for dynamics and so can be used for systems where there are no parameters, or where the quality of the parameters is poor or where a reaction is taking place. However, the timescale available to this type of calculation is limited due to computational cost of calculating the forces. CPMD is an *ab initio* MD scheme that reduces the computational cost by not optimising the wavefunction at every step, but with a typical time step of less than 0.01 fs can still only simulate shorter dynamic events. To access a longer time frame, QM-MM simulations can be performed, where the reaction centre is treated using QM-level calculations and the surrounding molecular environment and solvent is treated using

classical mechanics. A combination of docking, MD simulations and QM-MM simulations have been used in this work.

Nisin is a naturally occurring antibiotic peptide, which acts against gram-positive bacteria, including so-called superbugs such as MRSA and *c.diff*. The biosynthesis of the thioether rings of nisin involves a synthesis with the steps mediated by a complex of three enzymes. This synthesis is difficult to perform *in vitro* and contains several steps. In Chapter 3, analogues of the first twelve residues of nisin are proposed, covering rings A and B, with the thioether bonds replaced by disulfide bridges between cysteine residues. Disulfide bonds form either at a slightly elevated pH or in the presence of a simple catalyst. One of the stages of the *wt*-nisin synthesis changes the chirality of some of the amino residues and the chirality of the four cysteine residues in the analogues has been changed from L to D in every possible combination to produce a set of 16 isomers. Using implicit solvent MD simulations, with a total sampling time of 0.5 μ s per analogue, the conformational preferences leading to interactions between the sulfur atoms in the cysteine residues were characterised. Analysis of the dependency of ring formation on chirality showed that D-Cys3 favours conformations corresponding to a S₃-S₇ interaction (ring A), stabilised by a type IV β -turn, and that L-Cys8 with L-Cys11 favours conformations corresponding to a S₈-S₁₁ interaction (ring B), stabilised by a type IV β -turn and a hydrogen bond between the carbonyl group of residue 8 and amines of residues 11 and 12. The D-Cys3-D-Cys7-L-Cys8-L-Cys11 analogue was shown to favour a conformation with simultaneous S₃-S₇ and S₈-S₁₁ interactions. Trajectory frames for this analogue corresponding to simultaneous 3-7, 8-11 interactions were clustered and the centroids had a backbone RMSD from the rings of *wt*-nisin in complex with lipid II of between 0.6 and 1.7 Å.

The probabilities of the predicted interactions is relatively small (0.23 to 0.36 for ring A; 0.32 to 0.54 for ring B) and the energy barriers between different conformations are one or two kJ mol⁻¹, suggesting that the system would be very sensitive to the parameterisation of the potential energy function and

the predictions could be artefacts of the simulation. However, evidence from experimental studies support the predicted conformations, such as the propensity of D-amino residues to form type IV β -turns, that the sequence Cys-Pro-Gly-Cys leads to peptide cyclisation and that the β -turn and hydrogen bond of ring B are seen in NMR studies of *wt*-nisin as well as two related peptides that are conserved across ring B. Furthermore, the conformational similarity between the centroids and the NMR structure of rings A and B indicate that the nisin analogues do warrant further investigation.

In the first part of Chapter 4, the CMAP term of the CHARMM potential energy function was extended to simulate D-amino residues correctly. The CMAP term is a function of the backbone φ and ψ angles and is based upon MP2-level calculations of the alanine dipeptide. Because of this, it is dependent on chirality and is unsuitable for simulating D-amino residues. Using the transformation $(\varphi, \psi) \rightarrow (-\varphi, -\psi)$, which changes the Ramachandran plot for L-residues to that of D-residues, a version of the CMAP term, D-CMAP, was created. X-ray crystallographic structures of two peptides and two proteins containing D-amino residues were identified from the Protein Data Bank and simulated for up to 100 ns following a protocol used in the original CMAP parameterisation paper. The simulations were repeated three times, once using the D-CMAP term to treat D-amino residues, once with no correction applied (equivalent to the original CHARMM22 potential energy function) and once using the original L-amino CMAP term irrespective of the chirality of the residues. Analysis comparing the trajectory average and crystallographic structures show that when the D-CMAP term is used the values of φ and ψ in β -turns and β -sheets are closer to the experimental structure and that using the incorrect L-CMAP correction for D-amino residues leads to instability in these structural elements. In the second part of Chapter 4 explicit solvent simulations of the nisin analogues were performed with and without the correction, using the L-CMAP for L-amino residues and D-CMAP for the D-amino residues. The probability of the S_3 - S_7 and S_8 - S_{11} interactions needed to form the macrocyclic rings was very low in the simulations without

the CMAP term because the β -turns needed to bring the sulfurs together could not form without the correction.

In the final part of Chapter 4, QM-MM simulations using CPMD and classical MD were used to define a PES to describe the stretching of the disulfide bonds in both of the macrocyclic rings of the D-Cys3-D-Cys7-L-Cys8-L-Cys11 nisin analogue. One of the motivations for this was to use the PES to define a Morse potential that could be used with the CHARMM reactive dynamics module, a surface hopping algorithm that can jump between PES to classically simulate events such as the making and breaking of covalent bonds. However, the energy barrier to cross onto the CHARMM cysteine thiolate surface was 72.2 kJ mol^{-1} , too large for dynamics under usual *in vivo* conditions. A probable reason for the high barrier is that the QM part of the simulations was performed in the gas phase and water molecules are needed to be present to shield the sulfur radicals and to donate hydrogen atoms to them. There was no difference between the PES for ring A and ring B, indicating that either the peptide environment does not influence the strength of the disulfide bond or that by treating the solvent and all the residues other than the disulfides using classical mechanics, the level of theory was not sensitive enough to characterise the differences.

Chapter 5 uses longer timescale MD simulations, in combination with docking calculations, to examine and characterise the unusual binding behaviour of ILBP. ILBP contains 127 residues and has a ten strand β -clam binding cavity capped by a pair of α -helices. The ligands of ILBP are bile salts which are amphiphilic molecules that aide the digestion of fats via a detergent action. These are biosynthesised in the liver from cholesterol and ILBP absorbs the bile salts at the end of the gut to be recycled back into the liver. This leads ILBP to be of interest as a pharmaceutical target, because if its function was suppressed the bile salts would be excreted from the body, leading to more cholesterol being catabolised. ILBP shows cooperative binding comparable to haemoglobin but the experimental literature does not agree if the protein ligand binding ratio is 1:2 or 1:3. The protein also shows unusual site

selectivity of its ligands. There are two main bile salts in humans – cholic acid and chenodeoxycholic acid. These differ by a single OH group on the steroid rings and it has been shown via labelled NMR studies that whilst two of the binding sites have an affinity for both molecules, in the presence of a mixed pool one will completely displace another from a binding site. Again it was not clear from the experimental literature which binding site corresponds to which ligand type.

Using the starting coordinates either directly from NMR studies of human ILBP or building them from an X-ray crystallographic structure of the zebra fish variant, MD simulations were performed for 300 ns of human ILBP in apo form, singly ligated with cholic acid or doubly ligated with cholic acid. Snapshots were taken from these simulations and used as the receptor coordinates for docking with an additional cholic acid. Three possible locations were identified for a third binding site – two exterior sites, D-1 and D-2 and an interior location between the ligands. Further MD simulations were performed with the ligand in sites D-1 and D-2. When the ligand was in D-2 it migrated into site D-1 after a few tens of nanoseconds. When the ligand was in site D-1 the helices moved away from the β -barrel by $\sim 10^\circ$, into a position similar to the conformation of the holo protein. This supports the hypothesis of an allosteric binding mechanism similar to haemoglobin and given this possible allosteric role, and the presence of ligands in this binding site in the X-ray crystal structure of zebra fish ILBP, it is more probable that D-1 is the third binding site than D-2 or between the interior ligands. To examine site selectivity, the X-ray structure of danio ILBP was used to produce starting coordinates of human ILBP with cholic acid in site 1 and chenodeoxycholic acid in site 2 and a structure with chenodeoxycholic acid in site 1 and cholic acid in site 2. Based upon analysis of 220 ns simulations of these structures, and comparison with the double cholic acid conformations, a mechanism of site selectivity is proposed: the increased hydrophobicity of chenodeoxycholic acid, due to having one less OH group, leads it to sit deeper in the binding pocket, anchored by a hydrogen bond to Gln-51. This leads to an increase in

hydrophobicity in the pocket region, causing the cholic acid ligand in site 2 to also bind deeper in the pocket and inducing a change in the protein conformation where the angle between the ligands' main axes is smaller and the α -helices are closer to the β -barrel preventing further ligand exchange. These results are in agreement with experimental evidence, for example the role of Gln-51 in site selectivity or ICT data supporting three binding sites, but currently do not identify any explicit atomistic interactions that induces the helices to change position or explain the specific mechanism by which the ligands enter the binding cavity. Suggestions for further work to address these issues are explained in Section 6.3.

6.2 Major Contributions

The major contributions of this thesis are:

- First characterisation of the conformational preferences of the proposed nisin analogues and the dependence of ring formation on chirality. Identification of an analogue (D-Cys3-D-Cys7-L-Cys8-L-Cys11) that favours the simultaneous formation of the S_3 - S_7 and S_8 - S_{11} disulfide bonds and has low RMSD of the rings between the centroids of clustering the MD trajectories and the NMR structure of *wt*-nisin.
- Using explicit solvent MD simulations, I have shown that the $(\varphi, \psi) \rightarrow (-\varphi, -\psi)$ transformation of the CMAP term in the CHARMM potential energy function leads to sampling of conformations which are closest to X-ray crystallographic structures for D-amino residues and that the standard CMAP correction destabilises D-amino β -sheets and β -turns. The transformed parameter set is available in the Appendix.
- Identified the probable location of the third binding site of ILBP and its role in the allosteric binding mechanism. MD simulations indicate that binding to this exterior site induces changes in the orientation of the α -helices with respect to the β -barrel by $\sim 10^\circ$.
- Proposed an energetic mechanism of site selectivity for ILBP using evidence from MD simulations. The higher hydrophobicity of one ligand leads it to sit deeper in the binding cavity and interact with a

specific residue. This causes the second ligand to be deeper and induces the helices to move closer to the β -barrel, preventing further ligand exchange.

6.3 Further Work

The biomolecular systems studied in this thesis offer many opportunities for further work. Experimental work is currently being undertaken to synthesize and characterise the nisin analogues. Further computational work could include docking and MD simulations of nisin analogues and lipid II. This may require testing and further parameterisation, because, although the pyrophosphate group is parameterised in CHARMM as part of the nucleotides set, this may not transfer well to the chemical environment of the lipid II molecule. Such simulations would also be a good application of course grain or mesoscale simulations of both wild-type nisin and the nisin analogues to gain further insights into killing action and pore formation. e.g. would the analogues be able to form pores without the final three rings and tail section of wild-type nisin. The probabilities of sulfur-sulfur interactions and their dependence on cysteine chirality for the explicit solvent simulations of the nisin analogues in Chapter 4 are very similar to those reported for the implicit solvent simulations in Chapter 3, performed without the CMAP term. This suggests that explicit solvent simulations may be more sensitive to the correction than implicit simulations, which could be investigated further. Finally, the calculation of a PES to describe stretching and breaking of disulfide bonds in a solvated environment would be valuable for use with reactive dynamics. Although this would require computationally expensive QM calculations of a relatively large system, this could be an application of recent advances such as linearly scalable DFT or an implementation of CPMD on GPU's.

Free energy calculations comparing the different ligand combinations should be used to further characterise the binding behaviour of ILBP. This molecular system is well suited to computational alchemy methods because the ligands differ by single OH group and this would be an unusual application with

6. Conclusion and Future Work

multiple ligands. Further simulations should also be considered to further study the transition of the ligand in site D-2 to site D-1. It is unclear if this was a chance event or an example of a mechanism of ligand binding where the ligand adheres to any part of the protein surface and is then guided toward the binding site. Extended MD simulations with random orientation of the ligand as the starting conformations could examine this hypothesis and help to clarify the mechanism by which the ligands enter the protein interior.

Appendix - Transformed Parameter Set for Simulating D-amino Residues

Appendix Transformed Parameter Set for Simulating D-amino Residues

* Topology for D-amino acids with D-amino CMAP correction
(transformation of L-CMAP)

* D-amino residue names from Protein Data Bank
convention, parameters based on CHARMM22

* Eleanor R. Turpin July 2011

* *****

*

read rtf card append

* Topology of D-amino residues without CMAP

*

31 1

! ERT New atom types for D-amino CMAP correction

! numbering follows on from top_all22_prot_cmap.inp

MASS 122 DC 12.01100 C ! D-amino carbonyl C, peptide
backbone

MASS 123 DCT1 12.01100 C ! D-amino aliphatic sp3 C for
CH, peptide backbone

MASS 124 DNH1 14.00700 N ! D-amino peptide nitrogen

MASS 125 DN 14.00700 N ! D-amino proline nitrogen

MASS 126 DCP1 12.01100 C ! D-amino proline alpha-
carbon

DECL -CA

DECL -C

DECL -O

DECL +N

DECL +HN

DECL +CA

RESI DAL 0.00

GROUP

ATOM N DNH1 -0.47 ! |

ATOM HN H 0.31 ! HN-N

ATOM CA DCT1 0.07 ! | HB1

ATOM HA HB 0.09 ! | /

GROUP ! HA-CA-CB-HB2

ATOM CB CT3 -0.27 ! | \

ATOM HB1 HA 0.09 ! | HB3

ATOM HB2 HA 0.09 ! O=C

ATOM HB3 HA 0.09 ! |

GROUP !

ATOM C DC 0.51

ATOM O O -0.51

BOND CB CA N HN N CA

BOND C CA C +N CA HA CB HB1 CB HB2 CB HB3

DOUBLE O C

IMPR N -C CA HN C CA +N O

CMAP -C N CA C N CA C +N

DONOR HN N

ACCEPTOR O C

RESI DAR 1.00

GROUP

ATOM N DNH1 -0.47 ! | HH11

ATOM HN H 0.31 ! HN-N |

ATOM CA DCT1 0.07 ! | HB1 HG1 HD1 HE NH1-

HH12

ATOM HA HB 0.09 ! | | | | //(+)

GROUP ! HA-CA-CB-CG-CD-NE-CZ

ATOM CB CT2 -0.18 ! | | | | \

ATOM HB1 HA 0.09 ! | HB2 HG2 HD2 NH2-HH22

ATOM HB2 HA 0.09 ! O=C |

GROUP ! | HH21

ATOM CG CT2 -0.18

ATOM HG1 HA 0.09

ATOM HG2 HA 0.09

GROUP

ATOM CD CT2 0.20

ATOM HD1 HA 0.09

ATOM HD2 HA 0.09

ATOM NE NC2 -0.70

ATOM HE HC 0.44

ATOM CZ C 0.64

ATOM NH1 NC2 -0.80

ATOM HH11 HC 0.46

ATOM HH12 HC 0.46

ATOM NH2 NC2 -0.80

ATOM HH21 HC 0.46

ATOM HH22 HC 0.46

GROUP

ATOM C DC 0.51

ATOM O O -0.51

BOND CB CA CG CB CD CG NE CD CZ NE

BOND NH2 CZ N HN N CA

BOND C CA C +N CA HA CB HB1

BOND CB HB2 CG HG1 CG HG2 CD HD1 CD HD2

BOND NE HE NH1 HH11 NH1 HH12 NH2 HH21 NH2 HH22

DOUBLE O C CZ NH1

IMPR N -C CA HN C CA +N O

CMAP -C N CA C N CA C +N

IMPR CZ NH1 NH2 NE

DONOR HN N

DONOR HE NE

DONOR HH11 NH1

DONOR HH12 NH1

DONOR HH21 NH2

DONOR HH22 NH2

ACCEPTOR O C

RESI DSG 0.00

GROUP

ATOM N DNH1 -0.47 ! |

ATOM HN H 0.31 ! HN-N

ATOM CA DCT1 0.07 ! | HB1 OD1 HD21 (cis to OD1)

ATOM HA HB 0.09 ! | | | | /

GROUP ! HA-CA-CB-CG-ND2

ATOM CB CT2 -0.18 ! | | \

ATOM HB1 HA 0.09 ! | HB2 HD22 (trans to OD1)

ATOM HB2 HA 0.09 ! O=C

GROUP ! |

ATOM CG CC 0.55

ATOM OD1 O -0.55

GROUP

ATOM ND2 NH2 -0.62

ATOM HD21 H 0.32

ATOM HD22 H 0.30

GROUP

ATOM C DC 0.51

ATOM O O -0.51

BOND CB CA CG CB ND2 CG

BOND N HN N CA C CA C +N

BOND CA HA CB HB1 CB HB2 ND2 HD21 ND2 HD22

DOUBLE C O CG OD1

IMPR N -C CA HN C CA +N O

IMPR CG ND2 CB OD1 CG CB ND2 OD1

IMPR ND2 CG HD21 HD22 ND2 CG HD22 HD21

CMAP -C N CA C N CA C +N

DONOR HN N

DONOR HD21 ND2

DONOR HD22 ND2

ACCEPTOR OD1 CG

ACCEPTOR O C

RESI DAS -1.00

GROUP

ATOM N DNH1 -0.47 ! |

ATOM HN H 0.31 ! HN-N

ATOM CA DCT1 0.07 ! | HB1 OD1

ATOM HA HB 0.09 ! | | //

GROUP ! HA-CA-CB-CG

ATOM CB CT2 -0.28 ! | | \

ATOM HB1 HA 0.09 ! | HB2 OD2(-)

ATOM HB2 HA 0.09 ! O=C

ATOM CG CC 0.62 ! |

ATOM OD1 OC -0.76

ATOM OD2 OC -0.76

GROUP

ATOM C DC 0.51

ATOM O O -0.51

BOND CB CA CG CB OD2 CG

BOND N HN N CA C CA C +N

BOND CA HA CB HB1 CB HB2

DOUBLE O C CG OD1

IMPR N -C CA HN C CA +N O

!IMPR OD1 CB OD2 CG

IMPR CG CB OD2 OD1

CMAP -C N CA C N CA C +N

DONOR HN N

ACCEPTOR OD1 CG

Appendix Transformed Parameter Set for Simulating D-amino Residues

```

ACCEPTOR OD2 CG
ACCEPTOR O C

RESI DCY 0.00
GROUP
ATOM N DNH1 -0.47 ! |
ATOM HN H 0.31 ! HN-N
ATOM CA DCT1 0.07 ! | HB1
ATOM HA HB 0.09 ! | |
GROUP ! HA-CA--CB--SG
ATOM CB CT2 -0.11 ! | | \
ATOM HB1 HA 0.09 ! | HB2 HG1
ATOM HB2 HA 0.09 ! O=C
ATOM SG S -0.23 ! |
ATOM HG1 HS 0.16
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA SG CB N HN N CA
BOND C CA C+N CA HA CB HB1
BOND CB HB2 SG HG1
DOUBLE O C
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
DONOR HG1 SG
ACCEPTOR O C

RESI DGN 0.00
GROUP
ATOM N DNH1 -0.47 ! |
ATOM HN H 0.31 ! HN-N
ATOM CA DCT1 0.07 ! | HB1 HG1 OE1 HE21 (cis to
OE1)
ATOM HA HB 0.09 ! | | | | /
GROUP ! HA-CA--CB--CG--CD--NE2
ATOM CB CT2 -0.18 ! | | | \
ATOM HB1 HA 0.09 ! | HB2 HG2 HE22 (trans to
OE1)
ATOM HB2 HA 0.09 ! O=C
GROUP ! |
ATOM CG CT2 -0.18
ATOM HG1 HA 0.09
ATOM HG2 HA 0.09
GROUP
ATOM CD CC 0.55
ATOM OE1 O -0.55
GROUP
ATOM NE2 NH2 -0.62
ATOM HE21 H 0.32
ATOM HE22 H 0.30
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG CB CD CG NE2 CD
BOND N HN N CA C CA
BOND C +N CA HA CB HB1 CB HB2 CG HG1
BOND CG HG2 NE2 HE21 NE2 HE22
DOUBLE O C CD OE1
IMPR N -C CA HN C CA +N O
IMPR CD NE2 CG OE1 CD CG NE2 OE1
IMPR NE2 CD HE21 HE22 NE2 CD HE22 HE21
CMAP -C N CA C N CA C +N
DONOR HN N
DONOR HE21 NE2
DONOR HE22 NE2
ACCEPTOR OE1 CD
ACCEPTOR O C

RESI DGL -1.00
GROUP
ATOM N DNH1 -0.47 ! |
ATOM HN H 0.31 ! HN-N
ATOM CA DCT1 0.07 ! | HB1 HG1 OE1
ATOM HA HB 0.09 ! | | //
GROUP ! HA-CA--CB--CG--CD
ATOM CB CT2 -0.18 ! | | | \
ATOM HB1 HA 0.09 ! | HB2 HG2 OE2(-)
ATOM HB2 HA 0.09 ! O=C
GROUP ! |
ATOM CG CT2 -0.28

ATOM HG1 HA 0.09
ATOM HG2 HA 0.09
ATOM CD CC 0.62
ATOM OE1 OC -0.76
ATOM OE2 OC -0.76
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG CB CD CG OE2 CD
BOND N HN N CA C CA
BOND C +N CA HA CB HB1 CB HB2 CG HG1
BOND CG HG2
DOUBLE O C CD OE1
IMPR N -C CA HN C CA +N O
IMPR CD CG OE2 OE1
CMAP -C N CA C N CA C +N
DONOR HN N
ACCEPTOR OE1 CD
ACCEPTOR OE2 CD
ACCEPTOR O C

RESI DHI 0.00 ! neutral HIS, proton on ND1
GROUP
ATOM N DNH1 -0.47 ! | HD1 HE1
ATOM HN H 0.31 ! HN-N | /
ATOM CA DCT1 0.07 ! | HB1 ND1--CE1
ATOM HA HB 0.09 ! | | / ||
GROUP ! HA-CA--CB--CG ||
ATOM CB CT2 -0.09 ! | | \ \ ||
ATOM HB1 HA 0.09 ! | HB2 CD2--NE2
ATOM HB2 HA 0.09 ! O=C |
ATOM ND1 NR1 -0.36 ! | HD2
ATOM HD1 H 0.32
ATOM CG CPH1 -0.05
GROUP
ATOM CE1 CPH2 0.25
ATOM HE1 HR1 0.13
ATOM NE2 NR2 -0.70
ATOM CD2 CPH1 0.22
ATOM HD2 HR3 0.10
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG CB ND1 CG CE1 ND1
BOND NE2 CD2 N HN N CA
BOND C CA C +N CA HA CB HB1
BOND CB HB2 ND1 HD1 CD2 HD2 CE1 HE1
DOUBLE O C CG CD2 CE1 NE2
IMPR ND1 CG CE1 HD1 CD2 CG NE2 HD2 CE1 ND1 NE2 HE1
IMPR ND1 CE1 CG HD1 CD2 NE2 CG HD2 CE1 NE2 ND1 HE1
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
DONOR HD1 ND1
ACCEPTOR NE2
ACCEPTOR O C

RESI DIL 0.00
GROUP
ATOM N DNH1 -0.47 ! | HG21 HG22
ATOM HN H 0.31 ! HN-N | /
ATOM CA DCT1 0.07 ! | CG2--HG23
ATOM HA HB 0.09 ! | /
GROUP ! HA-CA--CB--HB HD1
ATOM CB CT1 -0.09 ! | \ /
ATOM HB HA 0.09 ! | CG1--CD--HD2
GROUP ! O=C / \ \
ATOM CG2 CT3 -0.27 ! | HG11 HG12 HD3
ATOM HG21 HA 0.09
ATOM HG22 HA 0.09
ATOM HG23 HA 0.09
GROUP
ATOM CG1 CT2 -0.18
ATOM HG11 HA 0.09
ATOM HG12 HA 0.09
GROUP
ATOM CD CT3 -0.27
ATOM HD1 HA 0.09
ATOM HD2 HA 0.09
ATOM HD3 HA 0.09
GROUP

```

Appendix Transformed Parameter Set for Simulating D-amino Residues

```

ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG1 CB CG2 CB CD CG1
BOND N HN N CA C CA C +N
BOND CA HA CB HB CG1 HG11 CG1 HG12 CG2 HG21
BOND CG2 HG22 CG2 HG23 CD HD1 CD HD2 CD HD3
DOUBLE O C
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
ACCEPTOR O C

RESI DLE 0.00
GROUP
ATOM N DNH1 -0.47 ! | HD11 HD12
ATOM HN H 0.31 ! HN-N | /
ATOM CA DCT1 0.07 ! | HB1 CD1--HD13
ATOM HA HB 0.09 ! | | /
GROUP ! HA-CA--CB--CG-HG
ATOM CB CT2 -0.18 ! | | \
ATOM HB1 HA 0.09 ! | HB2 CD2--HD23
ATOM HB2 HA 0.09 ! O=C | \
GROUP ! | HD21 HD22
ATOM CG CT1 -0.09
ATOM HG HA 0.09
GROUP
ATOM CD1 CT3 -0.27
ATOM HD11 HA 0.09
ATOM HD12 HA 0.09
ATOM HD13 HA 0.09
GROUP
ATOM CD2 CT3 -0.27
ATOM HD21 HA 0.09
ATOM HD22 HA 0.09
ATOM HD23 HA 0.09
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG CB CD1 CG CD2 CG
BOND N HN N CA C CA C +N
BOND CA HA CB HB1 CB HB2 CG HG CD1 HD11
BOND CD1 HD12 CD1 HD13 CD2 HD21 CD2 HD22 CD2 HD23
DOUBLE O C
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
ACCEPTOR O C

RESI DLY 1.00
GROUP
ATOM N DNH1 -0.47 ! |
ATOM HN H 0.31 ! HN-N
ATOM CA DCT1 0.07 ! | HB1 HG1 HD1 HE1 HZ1
ATOM HA HB 0.09 ! | | | | /
GROUP ! HA-CA--CB--CG--CD--CE--NZ--HZ2
ATOM CB CT2 -0.18 ! | | | | \
ATOM HB1 HA 0.09 ! | HB2 HG2 HD2 HE2 HZ3
ATOM HB2 HA 0.09 ! O=C
GROUP ! |
ATOM CG CT2 -0.18
ATOM HG1 HA 0.09
ATOM HG2 HA 0.09
GROUP
ATOM CD CT2 -0.18
ATOM HD1 HA 0.09
ATOM HD2 HA 0.09
GROUP
ATOM CE CT2 0.21
ATOM HE1 HA 0.05
ATOM HE2 HA 0.05
ATOM NZ NH3 -0.30
ATOM HZ1 HC 0.33
ATOM HZ2 HC 0.33
ATOM HZ3 HC 0.33
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG CB CD CG CE CD NZ CE
BOND N HN N CA C CA
BOND C +N CA HA CB HB1 CB HB2 CG HG1
BOND CG HG2 CD HD1 CD HD2 CE HE1 CE HE2

DOUBLE O C
BOND NZ HZ1 NZ HZ2 NZ HZ3
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
ACCEPTOR O C

RESI MED 0.00
GROUP
ATOM N DNH1 -0.47 ! |
ATOM HN H 0.31 ! HN-N
ATOM CA DCT1 0.07 ! | HB1 HG1 HE1
ATOM HA HB 0.09 ! | | | |
GROUP ! HA-CA--CB--CG--SD--CE--HE3
ATOM CB CT2 -0.18 ! | | | |
ATOM HB1 HA 0.09 ! | HB2 HG2 HE2
ATOM HB2 HA 0.09 ! O=C
GROUP ! |
ATOM CG CT2 -0.14
ATOM HG1 HA 0.09
ATOM HG2 HA 0.09
ATOM SD S -0.09
ATOM CE CT3 -0.22
ATOM HE1 HA 0.09
ATOM HE2 HA 0.09
ATOM HE3 HA 0.09
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG CB SD CG CE SD
BOND N HN N CA C CA C +N
BOND CA HA CB HB1 CB HB2 CG HG1 CG HG2
BOND CE HE1 CE HE2 CE HE3
DOUBLE O C
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
ACCEPTOR O C

RESI DPN 0.00
GROUP
ATOM N DNH1 -0.47 ! | HD1 HE1
ATOM HN H 0.31 ! HN-N | |
ATOM CA DCT1 0.07 ! | HB1 CD1--CE1
ATOM HA HB 0.09 ! | | // \
GROUP ! HA-CA--CB--CG CZ--HZ
ATOM CB CT2 -0.18 ! | | \ _ /
ATOM HB1 HA 0.09 ! | HB2 CD2--CE2
ATOM HB2 HA 0.09 ! O=C | |
GROUP ! | HD2 HE2
ATOM CG CA 0.00
GROUP
ATOM CD1 CA -0.115
ATOM HD1 HP 0.115
GROUP
ATOM CE1 CA -0.115
ATOM HE1 HP 0.115
GROUP
ATOM CZ CA -0.115
ATOM HZ HP 0.115
GROUP
ATOM CD2 CA -0.115
ATOM HD2 HP 0.115
GROUP
ATOM CE2 CA -0.115
ATOM HE2 HP 0.115
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG CB CD2 CG CE1 CD1
BOND CZ CE2 N HN
BOND N CA C CA C +N CA HA
BOND CB HB1 CB HB2 CD1 HD1 CD2 HD2 CE1 HE1
DOUBLE O C CD1 CG CZ CE1 CE2 CD2
BOND CE2 HE2 CZ HZ
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N

```

Appendix Transformed Parameter Set for Simulating D-amino Residues

```

ACCEPTOR O C
RESI DPR      0.00
GROUP        ! HD1 HD2
ATOM N DN   -0.29 ! | \ /
ATOM CD CP3  0.00 ! N---CD HG1 ATOM CA CP1
0.02
ATOM HD1 HA  0.09 ! | \ /
ATOM HD2 HA  0.09 ! | CG
ATOM CA DCP1 0.02 ! | / \
ATOM HA HB   0.09 ! HA-CA--CB HG2
GROUP        ! | / \
ATOM CB CP2  -0.18 ! | HB1 HB2
ATOM HB1 HA  0.09 ! O=C
ATOM HB2 HA  0.09 ! |
GROUP
ATOM CG CP2  -0.18
ATOM HG1 HA  0.09
ATOM HG2 HA  0.09
GROUP
ATOM C DC    0.51
ATOM O O     -0.51
BOND C CA C +N
BOND N CA CA CB CB CG CG CD N CD
BOND HA CA HG1 CG HG2 CG HD1 CD HD2 CD HB1 CB
HB2 CB
DOUBLE O C
IMPR N -C CA CD
IMPR C CA +N O
CMAP -C N CA C N CA C +N
ACCEPTOR O C
PATCHING FIRS DPRO

RESI DSN      0.00
GROUP
ATOM N DNH1 -0.47 ! |
ATOM HN H    0.31 ! HN-N
ATOM CA DCT1 0.07 ! | HB1
ATOM HA HB   0.09 ! | |
GROUP        ! HA-CA--CB--OG
ATOM CB CT2  0.05 ! | | \
ATOM HB1 HA  0.09 ! | HB2 HG1
ATOM HB2 HA  0.09 ! O=C
ATOM OG OH1  -0.66 ! |
ATOM HG1 H    0.43
GROUP
ATOM C DC    0.51
ATOM O O     -0.51
BOND CB CA OG CB N HN N CA
BOND C CA C +N CA HA CB HB1
BOND CB HB2 OG HG1
DOUBLE O C
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
DONOR HG1 OG
ACCEPTOR OG
ACCEPTOR O C

RESI DTH      0.00
GROUP
ATOM N DNH1 -0.47 ! |
ATOM HN H    0.31 ! HN-N
ATOM CA DCT1 0.07 ! | OG1--HG1
ATOM HA HB   0.09 ! | /
GROUP        ! HA-CA--CB-HB
ATOM CB CT1  0.14 ! | \
ATOM HB HA   0.09 ! | CG2--HG21
ATOM OG1 OH1 -0.66 ! O=C / \
ATOM HG1 H    0.43 ! | HG21 HG22
GROUP
ATOM CG2 CT3 -0.27
ATOM HG21 HA  0.09
ATOM HG22 HA  0.09
ATOM HG23 HA  0.09
GROUP
ATOM C DC    0.51
ATOM O O     -0.51
BOND CB CA OG1 CB CG2 CB N HN
BOND N CA C CA C +N CA HA
BOND CB HB OG1 HG1 CG2 HG21 CG2 HG22 CG2 HG23

DOUBLE O C
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
DONOR HG1 OG1
ACCEPTOR OG1
ACCEPTOR O C

RESI DTR      0.00
GROUP
ATOM N DNH1 -0.47 ! | HE3
ATOM HN H    0.31 ! HN-N |
ATOM CA DCT1 0.07 ! | HB1 CE3
ATOM HA HB   0.09 ! | | / \ \
GROUP        ! HA-CA--CB---CG-----CD2 CZ3-HZ3
ATOM CB CT2  -0.18 ! | | | | |
ATOM HB1 HA  0.09 ! | HB2 CD1 CE2 CH2-HH2
ATOM HB2 HA  0.09 ! O=C / \ / \ //
GROUP        ! | HD1 NE1 CZ2
ATOM CG CY   -0.03 ! | |
ATOM CD1 CA  0.035 ! HE1 HZ2
ATOM HD1 HP  0.115
ATOM NE1 NY  -0.61
ATOM HE1 H    0.38
ATOM CE2 CPT 0.13
ATOM CD2 CPT -0.02
GROUP
ATOM CE3 CA  -0.115
ATOM HE3 HP  0.115
GROUP
ATOM CZ3 CA  -0.115
ATOM HZ3 HP  0.115
GROUP
ATOM CZ2 CA  -0.115
ATOM HZ2 HP  0.115
GROUP
ATOM CH2 CA  -0.115
ATOM HH2 HP  0.115
GROUP
ATOM C DC    0.51
ATOM O O     -0.51
BOND CB CA CG CB CD2 CG NE1 CD1
BOND CZ2 CE2
BOND N HN N CA C CA C +N
BOND CZ3 CH2 CD2 CE3 NE1 CE2 CA HA CB HB1
BOND CB HB2 CD1 HD1 NE1 HE1 CE3 HE3 CZ2 HZ2
BOND CZ3 HZ3 CH2 HH2
DOUBLE O C CD1 CG CE2 CD2 CZ3 CE3 CH2 CZ2
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
DONOR HE1 NE1
ACCEPTOR O C

RESI DTY      0.00
GROUP
ATOM N DNH1 -0.47 ! | HD1 HE1
ATOM HN H    0.31 ! HN-N | |
ATOM CA DCT1 0.07 ! | HB1 CD1--CE1
ATOM HA HB   0.09 ! | | // \ \
GROUP        ! HA-CA--CB--CG CZ--OH
ATOM CB CT2  -0.18 ! | | \ _ / \
ATOM HB1 HA  0.09 ! | HB2 CD2--CE2 HH
ATOM HB2 HA  0.09 ! O=C | |
GROUP        ! | HD2 HE2
ATOM CG CA   0.00
GROUP
ATOM CD1 CA  -0.115
ATOM HD1 HP  0.115
GROUP
ATOM CE1 CA  -0.115
ATOM HE1 HP  0.115
GROUP
ATOM CZ CA   0.11
ATOM OH OH1  -0.54
ATOM HH H    0.43
GROUP
ATOM CD2 CA  -0.115
ATOM HD2 HP  0.115
GROUP
ATOM CE2 CA  -0.115

```

Appendix Transformed Parameter Set for Simulating D-amino Residues

```

ATOM HE2 HP 0.115
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG CB CD2 CG CE1 CD1
BOND CZ CE2 OH CZ
BOND N HN N CA C CA C +N
BOND CA HA CB HB1 CB HB2 CD1 HD1 CD2 HD2
BOND CE1 HE1 CE2 HE2 OH HH
DOUBLE O C CD1 CG CE1 CZ CE2 CD2
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
DONOR HH OH
ACCEPTOR OH
ACCEPTOR O C

RESI DVA 0.00
GROUP
ATOM N DNH1 -0.47 ! | HG11 HG12
ATOM HN H 0.31 ! HN-N | /
ATOM CA DCT1 0.07 ! | CG1--HG13
ATOM HA HB 0.09 ! | /
GROUP ! HA-CA--CB-HB
ATOM CB CT1 -0.09 ! | \
ATOM HB HA 0.09 ! | CG2--HG21
GROUP ! O=C /\
ATOM CG1 CT3 -0.27 ! | HG21 HG22
ATOM HG11 HA 0.09
ATOM HG12 HA 0.09
ATOM HG13 HA 0.09
GROUP
ATOM CG2 CT3 -0.27
ATOM HG21 HA 0.09
ATOM HG22 HA 0.09
ATOM HG23 HA 0.09
GROUP
ATOM C DC 0.51
ATOM O O -0.51
BOND CB CA CG1 CB CG2 CB N HN
BOND N CA C CA C +N CA HA
BOND CB HB CG1 HG11 CG1 HG12 CG1 HG13 CG2 HG21
BOND CG2 HG22 CG2 HG23
DOUBLE O C
IMPR N -C CA HN C CA +N O
CMAP -C N CA C N CA C +N
DONOR HN N
ACCEPTOR O C

PRES DPRO 1.00 ! Proline N-Terminal
GROUP ! use in generate statement
ATOM N NP -0.07 ! HA
ATOM HN1 HC 0.24 ! |
ATOM HN2 HC 0.24 ! -CA HN1
ATOM CD CP3 0.16 ! / \ /
ATOM HD1 HA 0.09 ! N(+)
ATOM HD2 HA 0.09 ! /\
ATOM CA DCP1 0.16 ! -CD HN2
ATOM HA HB 0.09 ! | \
BOND HN1 N HN2 N ! HD1 HD2
DONOR HN1 N
DONOR HN2 N

END

read param card append
* append parameters
*

BONDS
DNH1 CC 370 1.3450
DC C 600 1.335
DC DC 600 1.335
CP1 DC 250 1.49
CT1 DC 250 1.49
DCT1 DC 250 1.49
CT2 DC 250 1.49
CT3 DC 250 1.49
N DC 260 1.3
NC2 DC 463 1.365

NH1 DC 370 1.345
DNH1 DC 370 1.345
O DC 620 1.23
DCT1 C 250 1.49
DCT1 CC 200 1.522
DCT1 CD 200 1.522
DCT1 CT1 222.5 1.5
DCT1 DCT1 222.5 1.5
CT2 DCT1 222.5 1.538
CT3 DCT1 222.5 1.538
HA DCT1 309 1.111
HB DCT1 330 1.08
NH1 DCT1 320 1.43
DNH1 C 370 1.345
DNH1 CT1 320 1.43
DNH1 DCT1 320 1.43
NH3 DCT1 200 1.48
OH1 DCT1 428 1.42
DNH1 CT2 320 1.43
DNH1 CT3 320 1.43
DNH1 H 440 0.997
DNH1 HC 405 0.98
DCP1 C 250 1.49
DCP1 DC 250 1.49
DCP1 CC 250 1.49
DCP1 CD 200 1.49
CP2 DCP1 222.5 1.527
HB DCP1 330 1.08
NP DCP1 320 1.485
DN C 260 1.3
DN DC 260 1.3
N DCP1 320 1.434
DN CP1 320 1.434
DN DCP1 320 1.434
DN CP3 320 1.455

ANGLES
DNH1 CC HA 44 111 50 1.98
O CC DNH1 75 122.5 50 2.37
H DNH1 CC 50 120
DCT1 DNH1 CC 50 120
CP1 DN C 60 117
CP1 DN DC 60 117
CP1 N DC 60 117
CP2 CP1 DC 52 112.3
CP2 CP2 DCP1 70 108.5
CP2 DCP1 C 52 112.3
CP2 DCP1 CC 52 112.3
CP2 DCP1 CD 50 112.3
CP2 DCP1 DC 52 112.3
CP3 DN C 60 117
CP3 DN CP1 100 114.2
CP3 DN DC 60 117
CP3 DN DCP1 100 114.2
CP3 N DC 60 117
CP3 N DCP1 100 114.2
CP3 NP DCP1 100 111
CT1 CT1 DC 52 108
CT1 DCT1 C 52 108
CT1 DCT1 CC 52 108
CT1 DCT1 CT1 53.35 111
CT1 DCT1 DC 52 108
CT1 DNH1 C 50 120
CT1 DNH1 DC 50 120
CT1 NH1 DC 50 120
CT2 CT1 DC 52 108
CT2 CT1 DCT1 53.35 111
CT2 CT2 DC 52 108
CT2 CT2 DCT1 58.35 113.5
CT2 CT3 DCT1 58.35 113.5
CT2 DCT1 C 52 108
CT2 DCT1 CC 52 108
CT2 DCT1 CD 52 108
CT2 DCT1 CT1 53.35 111
CT2 DCT1 DC 52 108
CT2 DCT1 DCT1 53.35 111
CT2 DNH1 DC 50 120
CT2 NC2 DC 62.3 120
CT2 NH1 DC 50 120
CT3 CT1 DC 52 108
CT3 CT1 DCT1 53.35 108.5

```


Appendix Transformed Parameter Set for Simulating D-amino Residues

CP3	DN	DCP1	C	0.1	3	0
CP3	DN	DCP1	CC	0.1	3	0
CP3	DN	DCP1	CP2	0.1	3	0
CP3	DN	DCP1	DC	0.1	3	0
CP3	N	C	DCP1	2.75	2	180
CP3	N	C	DCP1	0.3	4	0
CP3	N	CP1	DC	0.1	3	0
CP3	N	DC	CP1	2.75	2	180
CP3	N	DC	CP1	0.3	4	0
CP3	N	DC	DCP1	2.75	2	180
CP3	N	DC	DCP1	0.3	4	0
CP3	N	DCP1	C	0.1	3	0
CP3	N	DCP1	CC	0.1	3	0
CP3	N	DCP1	CP2	0.1	3	0
CP3	N	DCP1	DC	0.1	3	0
CP3	NP	CP1	DC	0.08	3	0
CP3	NP	DCP1	CC	0.08	3	0
CP3	NP	DCP1	CD	0.08	3	0
CP3	NP	DCP1	CP2	0.08	3	0
CT1	C	DN	CP1	2.75	2	180
CT1	C	DN	CP1	0.3	4	0
CT1	C	DN	CP3	2.75	2	180
CT1	C	DN	CP3	0.3	4	0
CT1	C	DN	DCP1	2.75	2	180
CT1	C	DN	DCP1	0.3	4	0
CT1	C	DNH1	CT1	1.6	1	0
CT1	C	DNH1	CT1	2.5	2	180
CT1	C	DNH1	DCT1	1.6	1	0
CT1	C	DNH1	DCT1	2.5	2	180
CT1	C	N	DCP1	2.75	2	180
CT1	C	N	DCP1	0.3	4	0
CT1	C	NH1	DCT1	1.6	1	0
CT1	C	NH1	DCT1	2.5	2	180
CT1	CT1	DNH1	C	1.8	1	0
CT1	CT1	DNH1	DC	1.8	1	0
CT1	CT1	NH1	DC	1.8	1	0
CT1	DC	DN	CP1	2.75	2	180
CT1	DC	DN	CP1	0.3	4	0
CT1	DC	DN	CP3	2.75	2	180
CT1	DC	DN	CP3	0.3	4	0
CT1	DC	DN	DCP1	2.75	2	180
CT1	DC	DN	DCP1	0.3	4	0
CT1	DC	DNH1	CT1	1.6	1	0
CT1	DC	DNH1	CT1	2.5	2	180
CT1	DC	DNH1	DCT1	1.6	1	0
CT1	DC	DNH1	DCT1	2.5	2	180
CT1	DC	N	CP1	2.75	2	180
CT1	DC	N	CP1	0.3	4	0
CT1	DC	N	CP3	2.75	2	180
CT1	DC	N	CP3	0.3	4	0
CT1	DC	N	DCP1	2.75	2	180
CT1	DC	N	DCP1	0.3	4	0
CT1	DC	NH1	CT1	1.6	1	0
CT1	DC	NH1	CT1	2.5	2	180
CT1	DC	NH1	DCT1	1.6	1	0
CT1	DC	NH1	DCT1	2.5	2	180
CT1	DCT1	DNH1	C	1.8	1	0
CT1	DCT1	DNH1	DC	1.8	1	0
CT1	DCT1	NH1	C	1.8	1	0
CT1	DCT1	NH1	DC	1.8	1	0
CT1	DNH1	C	CP1	1.6	1	0
CT1	DNH1	C	CP1	2.5	2	180
CT1	DNH1	DC	CP1	1.6	1	0
CT1	DNH1	DC	CP1	2.5	2	180
CT1	DNH1	DC	CP1	1.6	1	0
CT1	DNH1	DC	CP1	2.5	2	180
CT2	C	DN	CP1	2.75	2	180
CT2	C	DN	CP1	0.3	4	0
CT2	C	DN	CP3	2.75	2	180
CT2	C	DN	CP3	0.3	4	0
CT2	C	DN	DCP1	2.75	2	180
CT2	C	DN	DCP1	0.3	4	0
CT2	C	DNH1	CT1	1.6	1	0
CT2	C	DNH1	CT1	2.5	2	180
CT2	C	DNH1	CT2	1.6	1	0
CT2	C	DNH1	CT2	2.5	2	180
CT2	C	DNH1	CT3	1.6	1	0
CT2	C	DNH1	CT3	2.5	2	180
CT2	C	N	DCP1	2.75	2	180
CT2	C	N	DCP1	0.3	4	0
CT2	C	NH1	DCT1	1.6	1	0
CT2	C	NH1	DCT1	2.5	2	180
CT2	CT1	DNH1	C	1.8	1	0
CT2	CT1	DNH1	DC	1.8	1	0
CT2	CT1	NH1	DC	1.8	1	0
CT2	CT2	DNH1	C	1.8	1	0
CT2	CT2	DNH1	DC	1.8	1	0
CT2	CT2	NH1	DC	1.8	1	0
CT2	DC	DN	CP1	2.75	2	180
CT2	DC	DN	CP1	0.3	4	0
CT2	DC	DN	CP3	2.75	2	180
CT2	DC	DN	CP3	0.3	4	0
CT2	DC	DN	DCP1	2.75	2	180
CT2	DC	DN	DCP1	0.3	4	0
CT2	DC	DNH1	CT1	1.6	1	0
CT2	DC	DNH1	CT1	2.5	2	180
CT2	DC	DNH1	CT2	1.6	1	0
CT2	DC	DNH1	CT2	2.5	2	180
CT2	DC	DNH1	CT3	1.6	1	0
CT2	DC	DNH1	CT3	2.5	2	180
CT2	DC	N	CP1	2.75	2	180
CT2	DC	N	CP1	0.3	4	0
CT2	DC	N	CP3	2.75	2	180
CT2	DC	N	CP3	0.3	4	0
CT2	DC	N	DCP1	2.75	2	180
CT2	DC	N	DCP1	0.3	4	0
CT2	DC	NH1	CT1	1.6	1	0
CT2	DC	NH1	CT1	2.5	2	180
CT2	DC	NH1	CT2	1.6	1	0
CT2	DC	NH1	CT2	2.5	2	180
CT2	DC	NH1	CT3	1.6	1	0
CT2	DC	NH1	CT3	2.5	2	180
CT2	DC	NH1	DCT1	1.6	1	0
CT2	DC	NH1	DCT1	2.5	2	180
CT2	DCT1	DNH1	C	1.8	1	0
CT2	DCT1	DNH1	DC	1.8	1	0
CT2	DCT1	NH1	C	1.8	1	0
CT2	DCT1	NH1	DC	1.8	1	0
CT2	DNH1	C	CP1	1.6	1	0
CT2	DNH1	C	CP1	2.5	2	180
CT2	DNH1	C	CT1	1.6	1	0
CT2	DNH1	C	CT1	2.5	2	180
CT2	DNH1	C	DCT1	1.6	1	0
CT2	DNH1	C	DCT1	2.5	2	180
CT2	DNH1	DC	CP1	1.6	1	0
CT2	DNH1	DC	CP1	2.5	2	180
CT2	DNH1	DC	CT1	1.6	1	0
CT2	DNH1	DC	CT1	2.5	2	180
CT2	DNH1	DC	DCT1	1.6	1	0
CT2	DNH1	DC	DCT1	2.5	2	180
CT2	NH1	C	DCT1	1.6	1	0
CT2	NH1	C	DCT1	2.5	2	180
CT2	NH1	DC	CP1	1.6	1	0
CT2	NH1	DC	CP1	2.5	2	180
CT2	NH1	DC	CT1	1.6	1	0
CT2	NH1	DC	CT1	2.5	2	180
CT2	NH1	DC	CT1	1.6	1	0
CT2	NH1	DC	CT1	2.5	2	180
CT2	NH1	DC	DCP1	1.6	1	0
CT2	NH1	DC	DCP1	2.5	2	180
CT2	NH1	DC	DCT1	1.6	1	0
CT2	NH1	DC	DCT1	2.5	2	180
CT3	C	DN	CP1	2.75	2	180
CT3	C	DN	CP1	0.3	4	0
CT3	C	DN	CP3	2.75	2	180
CT3	C	DN	CP3	0.3	4	0
CT3	C	DN	DCP1	2.75	2	180
CT3	C	DN	DCP1	0.3	4	0
CT3	C	DNH1	CT1	1.6	1	0
CT3	C	DNH1	CT1	2.5	2	180
CT3	C	DNH1	CT2	1.6	1	0
CT3	C	DNH1	CT2	2.5	2	180
CT3	C	DNH1	CT3	1.6	1	0
CT3	C	DNH1	CT3	2.5	2	180
CT3	C	DNH1	DCT1	1.6	1	0
CT3	C	DNH1	DCT1	2.5	2	180
CT3	C	N	DCP1	2.75	2	180
CT3	C	N	DCP1	0.3	4	0
CT3	C	NH1	DCT1	1.6	1	0
CT3	C	NH1	DCT1	2.5	2	180
CT3	CT1	DNH1	C	1.8	1	0
CT3	CT1	DNH1	DC	1.8	1	0

Appendix Transformed Parameter Set for Simulating D-amino Residues

H	NH1	DC	CP1	2.5	2	180	HC	NP	DCP1	CP2	0.08	3	0
H	NH1	DC	CT1	2.5	2	180	HC	NP	DCP1	HB	0.08	3	0
H	NH1	DC	CT2	2.5	2	180	HS	S	CT2	DCT1	0.24	1	0
H	NH1	DC	CT3	2.5	2	180	HS	S	CT2	DCT1	0.15	2	0
H	NH1	DC	DCP1	2.5	2	180	HS	S	CT2	DCT1	0.27	3	0
H	NH1	DC	DCT1	2.5	2	180	N	C	CP1	DN	0.3	1	0
H	NH1	DCT1	C	0	1	0	N	C	CP1	DN	-0.3	4	0
H	NH1	DCT1	CC	0	1	0	N	C	CT1	DCT1	0	1	0
H	NH1	DCT1	CD	0	1	0	N	C	DCP1	CP2	0.4	1	0
H	NH1	DCT1	CT1	0	1	0	N	C	DCP1	CP2	0.6	2	0
H	NH1	DCT1	CT2	0	1	0	N	C	DCP1	DN	0.3	1	0
H	NH1	DCT1	CT3	0	1	0	N	C	DCP1	DN	-0.3	4	0
H	NH1	DCT1	DC	0	1	0	N	C	DCP1	HB	0.4	1	180
H	NH1	DCT1	DCT1	0	1	0	N	C	DCP1	HB	0.6	2	0
H	NH2	CC	DCT1	1.4	2	180	N	C	DCP1	N	0.3	1	0
H	NH2	CC	DCP1	2.5	2	180	N	C	DCP1	N	-0.3	4	0
H	OH1	CT1	DCT1	1.33	1	0	N	C	DCT1	CT1	0	1	0
H	OH1	CT1	DCT1	0.18	2	0	N	C	DCT1	CT2	0	1	0
H	OH1	CT1	DCT1	0.32	3	0	N	C	DCT1	CT3	0	1	0
H	OH1	CT2	DCT1	1.3	1	0	N	C	DCT1	DCT1	0	1	0
H	OH1	CT2	DCT1	0.3	2	0	N	C	DCT1	HB	0	1	0
H	OH1	CT2	DCT1	0.42	3	0	N	DC	CP1	CP2	0.4	1	0
H	OH1	DCT1	CT1	1.33	1	0	N	DC	CP1	CP2	0.6	2	0
H	OH1	DCT1	CT1	0.18	2	0	N	DC	CP1	HB	0.4	1	180
H	OH1	DCT1	CT1	0.32	3	0	N	DC	CP1	HB	0.6	2	0
H	OH1	DCT1	CT3	1.33	1	0	N	DC	CP1	N	0.3	1	0
H	OH1	DCT1	CT3	0.18	2	0	N	DC	CP1	N	-0.3	4	0
H	OH1	DCT1	CT3	0.32	3	0	N	DC	CT1	CT1	0	1	0
H	OH1	DCT1	DCT1	1.33	1	0	N	DC	CT1	CT2	0	1	0
H	OH1	DCT1	DCT1	0.18	2	0	N	DC	CT1	CT3	0	1	0
H	OH1	DCT1	DCT1	0.32	3	0	N	DC	CT1	DCT1	0	1	0
HA	CP3	DN	C	0	3	180	N	DC	CT1	HB	0	1	0
HA	CP3	DN	CP1	0.1	3	0	N	DC	CT2	HB	0	1	0
HA	CP3	DN	DC	0	3	180	N	DC	CT3	HA	0	1	0
HA	CP3	DN	DCP1	0.1	3	0	N	DC	DCP1	CP2	0.4	1	0
HA	CP3	N	DC	0	3	180	N	DC	DCP1	CP2	0.6	2	0
HA	CP3	N	DCP1	0.1	3	0	N	DC	DCP1	DN	0.3	1	0
HA	CP3	NP	DCP1	0.08	3	0	N	DC	DCP1	DN	-0.3	4	0
HA	CT2	DNH1	C	0	3	0	N	DC	DCP1	HB	0.4	1	180
HA	CT2	DNH1	DC	0	3	0	N	DC	DCP1	HB	0.6	2	0
HA	CT2	DNH1	H	0	3	0	N	DC	DCP1	N	0.3	1	0
HA	CT2	NH1	DC	0	3	0	N	DC	DCP1	N	-0.3	4	0
HA	CT3	DNH1	C	0	3	0	N	DC	DCT1	CT1	0	1	0
HA	CT3	DNH1	DC	0	3	0	N	DC	DCT1	CT2	0	1	0
HA	CT3	DNH1	H	0	3	0	N	DC	DCT1	CT3	0	1	0
HA	CT3	NH1	DC	0	3	0	N	DC	DCT1	DCT1	0	1	0
HA	DCT1	CT2	CA	0.04	3	0	N	DC	DCT1	HB	0	1	0
HB	CP1	DN	C	0.8	3	0	N	DCT1	CT2	CA	0.04	3	0
HB	CP1	DN	CP3	0.1	3	0	NH1	C	CP1	DN	-0.3	4	0
HB	CP1	DN	DC	0.8	3	0	NH1	C	CT1	DCT1	0	1	0
HB	CP1	N	DC	0.8	3	0	NH1	C	CT1	DNH1	0.6	1	0
HB	CT1	DNH1	C	0	1	0	NH1	C	CT2	DNH1	0.6	1	0
HB	CT1	DNH1	DC	0	1	0	NH1	C	DCP1	CP2	0.4	1	0
HB	CT1	DNH1	H	0	1	0	NH1	C	DCP1	CP2	0.6	2	0
HB	CT1	NH1	DC	0	1	0	NH1	C	DCP1	DN	0.3	1	0
HB	CT2	DNH1	C	0	1	0	NH1	C	DCP1	DN	-0.3	4	0
HB	CT2	DNH1	DC	0	1	0	NH1	C	DCP1	HB	0.4	1	180
HB	CT2	DNH1	H	0	1	0	NH1	C	DCP1	HB	0.6	2	0
HB	CT2	NH1	DC	0	1	0	NH1	C	DCP1	N	0.3	1	0
HB	CT3	DNH1	C	0	1	0	NH1	C	DCP1	N	-0.3	4	0
HB	CT3	DNH1	DC	0	1	0	NH1	C	DCT1	CT1	0	1	0
HB	CT3	DNH1	H	0	1	0	NH1	C	DCT1	DCT1	0	1	0
HB	CT3	NH1	DC	0	1	0	NH1	C	DCT1	DNH1	0.6	1	0
HB	DCP1	DN	C	0.8	3	0	NH1	C	DCT1	HB	0	1	0
HB	DCP1	DN	DC	0.8	3	0	NH1	C	DCT1	NH1	0.6	1	0
HB	DCP1	DN	CP3	0.1	3	0	NH1	CT1	C	DN	0.4	1	0
HB	DCP1	N	C	0.8	3	0	NH1	CT1	DC	N	0.4	1	0
HB	DCP1	N	DC	0.8	3	0	NH1	CT1	DC	DN	0.4	1	0
HB	DCP1	N	CP3	0.1	3	0	NH1	CT2	C	DN	0.4	1	0
HB	DCP1	NP	CP3	0.08	3	0	NH1	CT2	DC	N	0.4	1	0
HB	DCT1	DNH1	C	0	1	0	NH1	CT2	DC	DN	0.4	1	0
HB	DCT1	DNH1	DC	0	1	0	NH1	DC	CP1	CP2	0.4	1	0
HB	DCT1	DNH1	H	0	1	0	NH1	DC	CP1	CP2	0.6	2	0
HB	DCT1	NH1	C	0	1	0	NH1	DC	CP1	DN	0.3	1	0
HB	DCT1	NH1	DC	0	1	0	NH1	DC	CP1	DN	-0.3	4	0
HB	DCT1	NH1	H	0	1	0	NH1	DC	CP1	HB	0.4	1	180
HC	NP	CP1	DC	0.08	3	0	NH1	DC	CP1	HB	0.6	2	0
HC	NP	DCP1	C	0.08	3	0	NH1	DC	CP1	N	0.3	1	0
HC	NP	DCP1	DC	0.08	3	0	NH1	DC	CP1	N	-0.3	4	0
HC	NP	DCP1	CC	0.08	3	0	NH1	DC	CT1	CT1	0	1	0
HC	NP	DCP1	CD	0.08	3	0	NH1	DC	CT1	CT2	0	1	0

Appendix Transformed Parameter Set for Simulating D-amino Residues

O DC DNH1 DCT1 2.5 2 180	0 0 0 0
O DC DNH1 H 2.5 2 180	
O DC N CP1 2.75 2 180	!-120
O DC N CP1 0.3 4 0	0 0 0 0 0
O DC N CP3 2.75 2 180	0 0 0 0 0
O DC N CP3 0.3 4 0	0 0 0 0 0
O DC N DCP1 2.75 2 180	0 0 0 0 0
O DC N DCP1 0.3 4 0	0 0 0 0
O DC NH1 CT1 2.5 2 180	
O DC NH1 CT2 2.5 2 180	!-105
O DC NH1 CT3 2.5 2 180	0 0 0 0 0
O DC NH1 DCT1 2.5 2 180	0 0 0 0 0
O DC NH1 H 2.5 2 180	0 0 0 0 0
O DCT1 NH2 CC 45 0 0	0 0 0 0 0
OC CC CP1 DN 0.16 3 0	0 0 0 0
OC CC DCP1 CP2 0.16 3 0	
OC CC DCP1 DN 0.16 3 0	!-90
OC CC DCP1 HB 0.16 3 0	0 0 0 0 0
OC CC DCP1 N 0.16 3 0	0 0 0 0 0
OC CC DCP1 NP 0.16 3 0	0 0 0 0 0
OC CC DCT1 NH3 3.2 2 180	0 0 0 0 0
SM SM CT2 DCT1 0.31 3 0	0 0 0 0
X CP1 DC X 0 6 180	
X DC NC2 X 2.25 2 180	!-75
X DCP1 C X 0 6 180	0 0 0 0 0
X DCP1 CC X 0 6 180	0 0 0 0 0
X DCP1 CD X 0 6 180	0 0 0 0 0
X DCP1 CP2 X 0.14 3 0	0 0 0 0 0
X DCP1 DC X 0 6 180	0 0 0 0
X DCT1 CC X 0.05 6 180	
X DCT1 CD X 0 6 180	!-60
X DCT1 CT1 X 0.2 3 0	0 0 0 0 0
X DCT1 CT2 X 0.2 3 0	0 0 0 0 0
X DCT1 CT3 X 0.2 3 0	0 0 0 0 0
X DCT1 DCT1 X 0.2 3 0	0 0 0 0 0
X DCT1 NH3 X 0.1 3 0	0 0 0 0
X DCT1 OH1 X 0.14 3 0	
X DCT1 OS X -0.1 3 0	!-45
	0 0 0 0 0
	0 0 0 0 0
IMPROPER	0 0 0 0 0
O DNH1 HA CC 45 0 0	0 0 0 0 0
O HA DNH1 CC 45 0 0	0 0 0 0 0
CC X X DCT1 96 0 0	0 0 0 0
N DC CP1 CP3 0 0 0	
DN C DCP1 CP3 0 0 0	!-30
DN DC DCP1 CP3 0 0 0	-2.3892 -5.102 -7.102 -6.8822 -6.8375
NC2 X X DC 40 0 0	-7.2198 -7.8066 -7.664 -6.6643 -3.9961
DNH1 X X H 20 0 0	-3.814 -3.8999 -1.7562 -4.2748 -4.4979
O NH2 DCT1 CC 45 0 0	-4.3419 -6.3504 -6.2626 -7.4995 -8.4761
O X X DC 120 0 0	-7.9518 -6.6493 -4.8772 -2.9624
CMAP ! mirror of L-amino residue CMAP data	!-15
! D-, D-Pro, D-ala	-5.1926 -7.156 -7.2901 -7.8816 -7.8776
DC DN DCP1 DC DN DCP1 DC DNH1 24	-7.9583 -7.5691 -6.7542 -4.0726 -4.4798
	-4.4528 -4.3772 -4.4364 -5.3679 -5.6538
	-4.7721 -7.0024 -7.6488 -8.3212 -8.2256
	-7.2676 -6.1779 -4.6683 -3.1742
!-180	
-6.0057 -7.9186 -9.6188 -9.3338 -8.3158	
-7.7395 -7.5633 -7.9619 -7.9886 -7.9668	
-8.1024 -8.6106 -8.8889 -9.0275 -8.9783	
-9.1251 -8.4192 -7.6332 -7.1853 -6.6546	
-5.4921 -4.5142 -4.2892 -4.6066	
	!0
	-5.1204 -7.1012 -8.2774 -8.6095 -8.0409
	-7.0272 -5.7891 -3.0347 -3.5574 -3.9199
	-3.646 -3.7874 -4.2891 -5.0507 -4.984
	-7.5416 -7.8122 -7.7752 -7.0542 -6.7169
	-5.9161 -5.2696 -3.8306 -2.7319
!-165	
0 0 0 0 0	
0 0 0 0 0	
0 0 0 0 0	
0 0 0 0 0	
0 0 0 0	!15
	-6.4969 -8.6267 -9.8071 -9.8633 -8.1653
	-6.4579 -2.725 -5.7532 -1.4306 -1.4593
	-1.484 -1.7812 -2.6863 -3.4786 -6.5645
	-7.3081 -7.0445 -5.7853 -4.6496 -4.5078
	-3.8047 -0.8212 -1.1141 -2.7055
!-150	
0 0 0 0 0	
0 0 0 0 0	
0 0 0 0 0	
0 0 0 0 0	
0 0 0 0	!30
	-5.4942 -7.1248 -7.5206 -6.8194 -5.1628
	-3.1495 -0.1462 -0.2845 -0.4102 -0.5004
	-0.6017 -0.8166 -1.2421 -4.7865 -6.0834
	-6.3548 -5.0805 -3.626 -2.5999 -1.8514
	-0.7044 -0.8426 -2.0668 -3.463
!-135	
0 0 0 0 0	
0 0 0 0 0	
0 0 0 0 0	
0 0 0 0 0	!45
	-3.4792 -4.4995 -4.8337 -4.1833 -2.7114

Appendix Transformed Parameter Set for Simulating D-amino Residues

0.9667 1.2811 1.0055 0.8259 0.6445
0.8087 0.9605 -1.9933 -3.5123 -4.4661
-3.9827 -2.3982 -0.5975 0.0065 0.205
-0.0683 -0.8596 -1.3074 -2.4101

!60
-1.4684 -2.0111 -2.5938 -2.3336 0.9932
1.1616 1.4931 1.3072 1.1635 1.106
1.6319 -0.2471 -1.3394 -2.555 -2.8353
-2.1019 -0.683 0.7254 1.0599 0.5337
-0.2049 -0.2521 -0.4073 -0.4882

!75
-0.3834 -1.549 -2.1349 0.2335 0.413
0.5496 0.6535 0.5846 0.4153 0.8188
1.84 -0.3503 -1.3548 -2.2491 -2.6649
-2.1767 -0.9552 0.3332 0.2124 -0.2421
-0.4539 0.003 0.1819 0.2688

!90
-0.9566 -2.0997 -2.8554 -3.2538 -2.6412
-0.4196 -0.5913 -0.7408 -0.5754 0.383
-0.9664 -1.3135 -2.242 -3.3331 -3.57
-2.979 -1.8328 -1.1421 -1.4248 -1.8437
-1.4847 -0.9234 -0.4178 -0.274

!105
-2.1005 -3.3716 -4.2394 -4.6674 -4.2193
-4.0743 -3.7697 -3.6058 -3.3467 -2.6599
-2.4919 -2.7954 -3.8852 -4.8409 -5.0203
-4.0481 -3.0498 -2.8306 -3.1541 -3.3158
-2.8444 -2.0042 -1.4466 -1.3298

!120
-3.1676 -4.391 -5.3828 -5.4744 -4.6216
-4.162 -4.2625 -4.2134 -3.8659 -3.8321
-3.7845 -4.1609 -5.1776 -5.9866 -6.1474
-5.3773 -5.2395 -4.9503 -4.5786 -4.3332
-3.5537 -2.8511 -2.2237 -2.2061

!135
-3.3379 -4.7004 -5.6871 -5.9738 -5.3341
-4.5268 -4.0969 -3.7688 -3.9502 -4.0071
-4.1585 -4.5816 -5.5755 -6.2033 -6.4149
-5.6322 -5.7983 -5.4609 -5.02 -4.6361
-3.8824 -2.9022 -2.0201 -2.1283

!150
-3.4173 -4.8039 -6.1861 -6.0528 -5.2488
-4.3562 -4.1382 -4.3438 -4.4405 -4.5079
-4.8509 -5.3335 -6.1431 -6.5422 -6.4883
-6.1682 -5.9311 -5.3692 -5.0156 -4.6162
-3.6761 -2.5512 -2.2652 -2.2375

!165
-4.4889 -5.8441 -7.4916 -6.9351 -6.1318
-5.1884 -4.8995 -5.579 -5.9051 -5.8381
-6.2495 -6.6819 -7.3944 -7.6098 -7.595
-7.1782 -6.6249 -5.7832 -5.4179 -4.9236
-3.8372 -3.0434 -2.9035 -3.2103

!L-ala, L-ala, D-ala map
C NH1 CT1 C NH1 CT1 C DNH1 24

!-180
0.126790 0.768700 0.971260 1.250970 2.121010
2.695430 2.064440 1.764790 0.755870 -0.713470
0.976130 -2.475520 -5.455650 -5.096450 -5.305850
-3.975630 -3.088580 -2.784200 -2.677120 -2.646060
-2.335350 -2.010440 -1.608040 -0.482250

!-165
-0.802290 1.377090 1.577020 1.872290 2.398990
2.461630 2.333840 1.904070 1.061460 0.518400
-0.116320 -3.575440 -5.284480 -5.160310 -4.196010
-3.276210 -2.715340 -1.806200 -1.101780 -1.210320
-1.008810 -0.637100 -1.603360 -1.776870

!-150
-0.634810 1.156210 1.624350 2.047200 2.653910
2.691410 2.296420 1.960450 1.324930 2.038290

-1.151510 -3.148610 -4.058280 -4.531850 -3.796370
-2.572090 -1.727250 -0.961410 -0.282910 -0.479120
-1.039340 -1.618060 -1.725460 -1.376360

!-135
0.214000 1.521370 1.977440 2.377950 2.929470
2.893410 2.435810 2.162970 1.761500 1.190090
-1.218610 -2.108900 -2.976100 -3.405340 -2.768440
-1.836030 -0.957950 0.021790 -0.032760 -0.665880
-1.321170 -1.212320 -0.893170 -0.897040

!-120
0.873950 1.959160 2.508990 2.841100 3.698960
3.309330 2.614300 2.481720 2.694660 1.082440
-0.398320 -1.761800 -2.945110 -3.294690 -2.308300
-0.855480 -0.087320 0.439040 0.691880 -0.586330
-1.027210 -0.976640 -0.467580 0.104020

!-105
1.767380 2.286650 2.818030 3.065500 3.370620
3.397440 2.730310 2.878790 2.542010 1.545240
-0.092150 -1.694440 -2.812310 -2.802430 -1.856360
-0.306240 -0.122440 0.444680 0.810150 -0.058630
-0.270290 -0.178830 0.202360 0.493810

!-90
1.456010 2.743180 2.589450 3.046230 3.451510
3.319160 3.052900 3.873720 2.420650 0.949100
0.008370 -1.382980 -2.138930 -2.087380 -1.268300
-0.494370 0.267580 0.908250 0.537520 0.306260
0.069540 0.097460 0.263060 0.603220

!-75
1.396790 3.349090 2.180920 2.942960 3.814070
3.675800 3.555310 3.887290 2.101260 -0.190940
-0.732240 -1.382040 -0.673880 -0.817390 -0.826980
-0.111800 0.053710 0.379430 0.692240 0.428960
-0.036100 -0.033820 -0.194300 0.400210

!-60
0.246650 1.229980 1.716960 3.168570 4.208190
4.366860 4.251080 3.348110 0.997540 -1.287540
-1.179900 -0.684300 -0.853660 -1.158760 -0.347550
0.114810 0.242800 0.322420 0.370140 -0.374950
-0.676940 -1.323430 -1.366650 -0.218770

!-45
-1.196730 0.078060 2.347410 4.211350 5.376000
5.364940 4.355200 2.436510 0.408470 -0.590840
-0.435960 -0.501210 -0.822230 -0.607210 0.057910
0.246580 -0.070570 0.379430 0.692240 -0.571680
-1.282910 -1.715770 -1.839820 -1.987110

!-30
-1.174720 1.067030 4.180460 6.741610 6.070770
4.781470 2.758340 1.295810 0.571150 -0.196480
0.251860 -0.732140 1.289360 1.497590 1.890550
2.198490 0.169290 0.534000 0.331780 -1.276320
-2.550070 -3.312150 -3.136670 -2.642260

!-15
0.293590 5.588070 3.732620 3.217620 3.272450
2.492320 1.563700 1.356760 0.831410 0.630170
1.591970 0.821920 0.486070 0.715760 0.996020
1.591580 -0.367400 0.181770 -0.613920 -2.267900
-3.516460 -3.597700 -3.043340 -1.765020

!0
2.832310 0.787990 0.323280 0.479230 0.628600
0.976330 1.238750 1.671950 1.645480 2.520340
1.606970 0.776350 0.119780 0.070390 0.121170
-1.569230 -1.213010 -1.846360 -2.744510 -3.792530
-3.934880 -3.615930 -2.675750 -0.924170

!15
-0.778340 -1.912680 -2.052140 -1.846280 -1.047430
0.183400 1.682950 2.223500 1.358370 2.448660
1.436920 0.678570 -0.237060 -0.535320 -0.790380
-2.182580 -3.251140 -4.195110 -4.269270 -3.908210
-3.455620 -2.773970 1.755370 0.313410

Appendix Transformed Parameter Set for Simulating D-amino Residues

I30
-2.963810 -3.483730 -3.517080 -2.724860 -1.405510
0.336200 1.428450 1.394630 0.970370 2.462720
1.522430 0.553620 -0.407380 -1.482950 -3.613920
-4.159810 -4.945580 -4.784040 -3.764540 -2.959140
-1.963850 -1.071260 -1.599580 -2.445320

I45
-4.029070 -3.932660 -3.558480 -2.513980 -1.037320
0.362000 0.814380 0.754110 0.502370 1.903420
0.770220 -0.416420 -3.286310 -3.875270 -4.907800
-5.704430 -5.645660 -4.396040 -2.865450 -2.368170
-2.860490 -3.416560 -3.666490 -3.859070

I60
-3.338270 -2.960220 -2.311700 -1.272890 -0.246470
0.722610 0.668070 0.438130 2.395330 1.632470
-2.041450 -3.218100 -3.915080 -4.852510 -5.696500
-6.314370 -5.683690 -4.170620 -3.141000 -3.508820
-3.756430 -3.640810 -3.640430 -3.550690

I75
-2.244860 -1.632100 -1.000640 -0.170440 0.526440
0.823710 0.517140 -0.013120 -0.370910 -1.213720
-2.305650 -3.420580 -4.484960 -5.693140 -6.199150
-6.253870 -5.211310 -4.174380 -3.685150 -4.151360
-4.161970 -3.725150 -3.715310 -2.606760

I90
-1.720840 -1.177830 -0.428430 0.277730 0.807900
0.803260 0.482510 -0.336900 -0.786270 -1.774070
-2.793220 -3.828560 -5.211800 -6.636850 -6.989940
-6.108800 -5.452410 -3.911450 -4.321000 -4.587240
-4.102610 -3.772820 -3.157300 -2.648390

I105
-1.850640 -1.092420 -0.445020 0.128490 1.005520
0.884820 0.485850 -0.218470 -0.857670 -1.682330
-3.014400 -4.481110 -6.053510 -6.865400 -6.871130
-5.728240 -3.912230 -4.802110 -5.034640 -4.715990
-4.601080 -4.086220 -3.274630 -2.410940

I120
-1.969230 -1.116650 -0.540250 -0.150330 0.763520
1.038890 0.758480 0.313530 -0.333050 -1.872770
-3.366270 -5.008260 -6.124810 -7.034830 -6.724320
-3.700200 -4.510620 -5.185650 -5.361620 -4.847490
-4.444320 -4.004260 -3.415720 -2.751230

I135
-2.111250 -1.168960 -0.322790 -0.006920 0.316660
1.086270 0.939170 0.625340 -0.166360 -1.830310
-3.469470 -4.946030 -6.112560 -1.915580 -4.047310
-4.996740 -4.996730 -4.842690 -4.886620 -4.300540
-4.494620 -4.442210 -4.163570 -3.183510

I150
-1.757590 -0.403620 0.023920 0.362390 0.634520
1.264920 1.361360 0.948420 -0.073680 -1.483560
-3.152820 1.835120 -1.762860 -5.093660 -5.744830
-5.390070 -4.783930 -4.190630 -4.115420 -4.042280
-4.125570 -4.028550 -4.026100 -2.937910

I165
-0.810590 -0.071500 0.378890 0.543310 1.277880
1.641310 1.698840 1.519950 0.631950 -1.088670
-2.736530 -0.735240 -4.563830 -6.408350 -5.889450
-5.141750 -4.194970 -3.666490 -3.843450 -3.818830
-3.826180 -3.596820 -2.994790 -2.231020

!l-alanine, D-alanine, L-alanine map
C DNH1 DCT1 DC DNH1 DCT1 DC NH1 24

I-180
-0.48225 -1.60804 -2.01044 -2.33535 -2.64606
-2.67712 -2.7842 -3.08858 -3.97563 -5.30585
-5.09645 -5.45565 -2.47552 0.97613 -0.71347
0.75587 1.76479 2.06444 2.69543 2.12101
1.25097 0.97126 0.7687 0.12679

I-165
-2.23102 -2.99479 -3.59682 -3.82618 -3.81883
-3.84345 -3.66649 -4.19497 -5.14175 -5.88945
-6.40835 -4.56383 -0.73524 -2.73653 -1.08867
0.63195 1.51995 1.69884 1.64131 1.27788
0.54331 0.37889 -0.0715 -0.81059

I-150
-2.93791 -4.0261 -4.02855 -4.12557 -4.04228
-4.11542 -4.19063 -4.78393 -5.39007 -5.74483
-5.09366 -1.76286 1.83512 -3.15282 -1.48356
-0.07368 0.94842 1.36136 1.26492 0.63452
0.36239 0.02392 -0.40362 -1.75759

I-135
-3.18351 -4.16357 -4.44221 -4.49462 -4.30054
-4.88662 -4.84269 -4.99673 -4.99674 -4.04731
-1.91558 -6.11256 -4.94603 -3.46947 -1.83031
-0.16636 0.62534 0.93917 1.08627 0.31666
-0.00692 -0.32279 -1.16896 -2.11125

I-120
-2.75123 -3.41572 -4.00426 -4.44432 -4.84749
-5.36162 -5.18565 -4.51062 -3.7002 -6.72432
-7.03483 -6.12481 -5.00826 -3.36627 -1.87277
-0.33305 0.31353 0.75848 1.03889 0.76352
-0.15033 -0.54025 -1.11665 -1.96923

I-105
-2.41094 -3.27463 -4.08622 -4.60108 -4.71599
-5.03464 -4.80211 -3.91223 -5.72824 -6.87113
-6.8654 -6.05351 -4.48111 -3.0144 -1.68233
-0.85767 -0.21847 0.48585 0.88482 1.00552
0.12849 -0.44502 -1.09242 -1.85064

I-90
-2.64839 -3.1573 -3.77282 -4.10261 -4.58724
-4.321 -3.91145 -5.45241 -6.1088 -6.98994
-6.63685 -5.2118 -3.82856 -2.79322 -1.77407
-0.78627 -0.3369 0.48251 0.80326 0.8079
0.27773 -0.42843 -1.17783 -1.72084

I-75
-2.60676 -3.71531 -3.72515 -4.16197 -4.15136
-3.68515 -4.17438 -5.21131 -6.25387 -6.19915
-5.69314 -4.48496 -3.42058 -2.30565 -1.21372
-0.37091 -0.01312 0.51714 0.82371 0.52644
-0.17044 -1.00064 -1.6321 -2.24486

I-60
-3.55069 -3.64043 -3.64081 -3.75643 -3.50882
-3.141 -4.17062 -5.68369 -6.31437 -5.6965
-4.85251 -3.91508 -3.2181 -2.04145 1.63247
2.39533 0.43813 0.66807 0.72261 -0.24647
-1.27289 -2.3117 -2.96022 -3.33827

I-45
-3.85907 -3.66649 -3.41656 -2.86049 -2.36817
-2.86545 -4.39604 -5.64566 -5.70443 -4.9078
-3.87527 -3.28631 -0.41642 0.77022 1.90342
0.50237 0.75411 0.81438 0.362 -1.03732
-2.51398 -3.55848 -3.93266 -4.02907

I-30
-2.44532 -1.59958 -1.07126 -1.96385 -2.95914
-3.76454 -4.78404 -4.94558 -4.15981 -3.61392
-1.48295 -0.40738 0.55362 1.52243 2.46272
0.97037 1.39463 1.42845 0.3362 -1.40551
-2.72486 -3.51708 -3.48373 -2.96381

I-15
0.31341 1.75537 -2.77397 -3.45562 -3.90821
-4.26927 -4.19511 -3.25114 -2.18258 -0.79038
-0.53532 -0.23706 0.67857 1.43692 2.44866
1.35837 2.2235 1.68295 0.1834 -1.04743
-1.84628 -2.05214 -1.91268 -0.77834

I0
-0.92417 -2.67575 -3.61593 -3.93488 -3.79253
-2.74451 -1.84636 -1.21301 -1.56923 0.12117
0.07039 0.11978 0.77635 1.60697 2.52034
1.64548 1.67195 1.23875 0.97633 0.6286

Appendix Transformed Parameter Set for Simulating D-amino Residues

0.47923 0.32328 0.78799 2.83231

I15
-1.76502 -3.04334 -3.59777 -3.51646 -2.2679
-0.61392 0.18177 -0.3674 1.59158 0.99602
0.71576 0.48607 0.82192 1.59197 0.63017
0.83141 1.35676 1.5637 2.49232 3.27245
3.21762 3.73262 5.58807 0.29359

I30
-2.64226 -3.13667 -3.31215 -2.55007 -1.27632
0.33178 0.534 0.16929 2.19849 1.89055
1.49759 1.28936 -0.73214 0.25186 -0.19648
0.57115 1.29581 2.75834 4.78147 6.07077
6.74161 4.18046 1.06703 -1.17472

I45
-1.98711 -1.83982 -1.71577 -1.28291 -0.57168
0.24777 0.37943 -0.07057 0.24658 0.05791
-0.60721 -0.82223 -0.50121 -0.43596 -0.59084
0.40847 2.43651 4.3552 5.36494 5.376
4.21135 2.34741 0.07806 -1.19673

I60
-0.21877 -1.36665 -1.32343 -0.67694 -0.37495
0.37014 0.32242 0.2428 0.11481 -0.34755
-1.15876 -0.85366 -0.6843 -1.1799 -1.28754
0.99754 3.34811 4.25108 4.36686 4.20819
3.16857 1.71696 1.22998 0.24665

I75
0.40021 -0.1943 -0.03382 -0.0361 0.42896
0.69224 0.2964 0.05371 -0.1118 -0.82698
-0.81739 -0.67388 -1.38204 -0.73224 -0.19094
2.10126 3.88729 3.55531 3.6758 3.81407
2.94296 2.18092 3.34909 1.39679

I90
0.60322 0.26306 0.09746 0.06954 0.30626
0.53752 0.90825 0.26758 -0.49437 -1.2683
-2.08738 -2.13893 -1.38298 0.00837 0.9491
2.42065 3.87372 3.0529 3.31916 3.45151
3.04623 2.58945 2.74318 1.45601

I105
0.49381 0.20236 -0.17883 -0.27029 -0.05863
0.81015 0.44468 -0.12244 -0.30624 -1.85636
-2.80243 -2.81231 -1.69444 -0.09215 1.54524
2.54201 2.87879 2.73031 3.39744 3.37062
3.0655 2.81803 2.28665 1.76738

I120
0.10402 -0.46758 -0.97664 -1.02721 -0.58633
0.69188 0.43904 -0.08732 -0.85548 -2.3083
-3.29469 -2.94511 -1.7618 -0.39832 1.08244
2.69466 2.48172 2.6143 3.30933 3.69896
2.8411 2.50899 1.95916 0.87395

I135
-0.89704 -0.89317 -1.21232 -1.32117 -0.66588
-0.03276 0.02179 -0.95795 -1.83603 -2.76844
-3.40534 -2.9761 -2.1089 -1.21861 1.19009
1.7615 2.16297 2.43581 2.89341 2.92947
2.37795 1.97744 1.52137 0.214

I150
-1.37636 -1.72546 -1.61806 -1.03934 -0.47912
-0.28291 -0.96141 -1.72725 -2.57209 -3.79637
-4.53185 -4.05828 -3.14861 -1.15151 2.03829
1.32493 1.96045 2.29642 2.69141 2.65391
2.0472 1.62435 1.15621 -0.63481

I165
-1.77687 -1.60336 -0.6371 -1.00881 -1.21032
-1.10178 -1.8062 -2.71534 -3.27621 -4.19601
-5.16031 -5.28448 -3.57544 -0.11632 0.5184
1.06146 1.90407 2.33384 2.46163 2.39899
1.87229 1.57702 1.37709 -0.80229

ll-ala, D-ala, D-ala map
C DNH1 DCT1 DC DNH1 DCT1 DC DNH1 24

I-180
-0.48225 -1.60804 -2.01044 -2.33535 -2.64606
-2.67712 -2.7842 -3.08858 -3.97563 -5.30585
-5.09645 -5.45565 -2.47552 0.97613 -0.71347
0.75587 1.76479 2.06444 2.69543 2.12101
1.25097 0.97126 0.7687 0.12679

I-165
-2.23102 -2.99479 -3.59682 -3.82618 -3.81883
-3.84345 -3.66649 -4.19497 -5.14175 -5.88945
-6.40835 -4.56383 -0.73524 -2.73653 -1.08867
0.63195 1.51995 1.69884 1.64131 1.27788
0.54331 0.37889 -0.0715 -0.81059

I-150
-2.93791 -4.0261 -4.02855 -4.12557 -4.04228
-4.11542 -4.19063 -4.78393 -5.39007 -5.74483
-5.09366 -1.76286 1.83512 -3.15282 -1.48356
-0.07368 0.94842 1.36136 1.26492 0.63452
0.36239 0.02392 -0.40362 -1.75759

I-135
-3.18351 -4.16357 -4.44221 -4.49462 -4.30054
-4.88662 -4.84269 -4.99673 -4.99674 -4.04731
-1.91558 -6.11256 -4.94603 -3.46947 -1.83031
-0.16636 0.62534 0.93917 1.08627 0.31666
-0.00692 -0.32279 -1.16896 -2.11125

I-120
-2.75123 -3.41572 -4.00426 -4.44432 -4.84749
-5.36162 -5.18565 -4.51062 -3.7002 -6.72432
-7.03483 -6.12481 -5.00826 -3.36627 -1.87277
-0.33305 0.31353 0.75848 1.03889 0.76352
-0.15033 -0.54025 -1.11665 -1.96923

I-105
-2.41094 -3.27463 -4.08622 -4.60108 -4.71599
-5.03464 -4.80211 -3.91223 -5.72824 -6.87113
-6.8654 -6.05351 -4.48111 -3.0144 -1.68233
-0.85767 -0.21847 0.48585 0.88482 1.00552
0.12849 -0.44502 -1.09242 -1.85064

I-90
-2.64839 -3.1573 -3.77282 -4.10261 -4.58724
-4.321 -3.91145 -5.45241 -6.1088 -6.98994
-6.63685 -5.2118 -3.82856 -2.79322 -1.77407
-0.78627 -0.3369 0.48251 0.80326 0.8079
0.27773 -0.42843 -1.17783 -1.72084

I-75
-2.60676 -3.71531 -3.72515 -4.16197 -4.15136
-3.68515 -4.17438 -5.21131 -6.25387 -6.19915
-5.69314 -4.48496 -3.42058 -2.30565 -1.21372
-0.37091 -0.01312 0.51714 0.82371 0.52644
-0.17044 -1.00064 -1.6321 -2.24486

I-60
-3.55069 -3.64043 -3.64081 -3.75643 -3.50882
-3.141 -4.17062 -5.68369 -6.31437 -5.6965
-4.85251 -3.91508 -3.2181 -2.04145 1.63247
2.39533 0.43813 0.66807 0.72261 -0.24647
-1.27289 -2.3117 -2.96022 -3.33827

I-45
-3.85907 -3.66649 -3.41656 -2.86049 -2.36817
-2.86545 -4.39604 -5.64566 -5.70443 -4.9078
-3.87527 -3.28631 -0.41642 0.77022 1.90342
0.50237 0.75411 0.81438 0.362 -1.03732
-2.51398 -3.55848 -3.93266 -4.02907

I-30
-2.44532 -1.59958 -1.07126 -1.96385 -2.95914
-3.76454 -4.78404 -4.94558 -4.15981 -3.61392
-1.48295 -0.40738 0.55362 1.52243 2.46272
0.97037 1.39463 1.42845 0.3362 -1.40551
-2.72486 -3.51708 -3.48373 -2.96381

I-15
0.31341 1.75537 -2.77397 -3.45562 -3.90821
-4.26927 -4.19511 -3.25114 -2.18258 -0.79038

Appendix Transformed Parameter Set for Simulating D-amino Residues

-0.53532 -0.23706 0.67857 1.43692 2.44866
1.35837 2.2235 1.68295 0.1834 -1.04743
-1.84628 -2.05214 -1.91268 -0.77834

!0

-0.92417 -2.67575 -3.61593 -3.93488 -3.79253
-2.74451 -1.84636 -1.21301 -1.56923 0.12117
0.07039 0.11978 0.77635 1.60697 2.52034
1.64548 1.67195 1.23875 0.97633 0.6286
0.47923 0.32328 0.78799 2.83231

!15

-1.76502 -3.04334 -3.5977 -3.51646 -2.2679
-0.61392 0.18177 -0.3674 1.59158 0.99602
0.71576 0.48607 0.82192 1.59197 0.63017
0.83141 1.35676 1.5637 2.49232 3.27245
3.21762 3.73262 5.58807 0.29359

!30

-2.64226 -3.13667 -3.31215 -2.55007 -1.27632
0.33178 0.534 0.16929 2.19849 1.89055
1.49759 1.28936 -0.73214 0.25186 -0.19648
0.57115 1.29581 2.75834 4.78147 6.07077
6.74161 4.18046 1.06703 -1.17472

!45

-1.98711 -1.83982 -1.71577 -1.28291 -0.57168
0.24777 0.37943 -0.07057 0.24658 0.05791
-0.60721 -0.82223 -0.50121 -0.43596 -0.59084
0.40847 2.43651 4.3552 5.36494 5.376
4.21135 2.34741 0.07806 -1.19673

!60

-0.21877 -1.36665 -1.32343 -0.67694 -0.37495
0.37014 0.32242 0.2428 0.11481 -0.34755
-1.15876 -0.85366 -0.6843 -1.1799 -1.28754
0.99754 3.34811 4.25108 4.36686 4.20819
3.16857 1.71696 1.22998 0.24665

!75

0.40021 -0.1943 -0.03382 -0.0361 0.42896
0.69224 0.2964 0.05371 -0.1118 -0.82698
-0.81739 -0.67388 -1.38204 -0.73224 -0.19094
2.10126 3.88729 3.55531 3.6758 3.81407
2.94296 2.18092 3.34909 1.39679

!90

0.60322 0.26306 0.09746 0.06954 0.30626
0.53752 0.90825 0.26758 -0.49437 -1.2683
-2.08738 -2.13893 -1.38298 0.00837 0.9491
2.42065 3.87372 3.0529 3.1916 3.45151
3.04623 2.58945 2.74318 1.45601

!105

0.49381 0.20236 -0.17883 -0.27029 -0.05863
0.81015 0.44468 -0.12244 -0.30624 -1.85636
-2.80243 -2.81231 -1.69444 -0.09215 1.54524
2.54201 2.87879 2.73031 3.39744 3.37062
3.0655 2.81803 2.28665 1.76738

!120

0.10402 -0.46758 -0.97664 -1.02721 -0.58633
0.69188 0.43904 -0.08732 -0.85548 -2.3083
-3.29469 -2.94511 -1.7618 -0.39832 1.08244
2.69466 2.48172 2.6143 3.30933 3.69896
2.8411 2.50899 1.95916 0.87395

!135

-0.89704 -0.89317 -1.21232 -1.32117 -0.66588
-0.03276 0.02179 -0.95795 -1.83603 -2.76844
-3.40534 -2.9761 -2.1089 -1.21861 1.19009
1.7615 2.16297 2.43581 2.89341 2.92947
2.37795 1.97744 1.52137 0.214

!150

-1.37636 -1.72546 -1.61806 -1.03934 -0.47912
-0.28291 -0.96141 -1.72725 -2.57209 -3.79637
-4.53185 -4.05828 -3.14861 -1.15151 2.03829
1.32493 1.96045 2.29642 2.69141 2.65391
2.0472 1.62435 1.15621 -0.63481

!165

-1.77687 -1.60336 -0.6371 -1.00881 -1.21032
-1.10178 -1.8062 -2.71534 -3.27621 -4.19601
-5.16031 -5.28448 -3.57544 -0.11632 0.5184
1.06146 1.90407 2.33384 2.46163 2.39899
1.87229 1.57702 1.37709 -0.80229

!D-ala, L-ala, L-ala map

DC NH1 CT1 C NH1 CT1 C NH1 24

!-180

0.126790 0.768700 0.971260 1.250970 2.121010
2.695430 2.064440 1.764790 0.755870 -0.713470
0.976130 -2.475520 -5.455650 -5.096450 -5.305850
-3.975630 -3.088580 -2.784200 -2.677120 -2.646060
-2.335350 -2.010440 -1.608040 -0.482250

!-165

-0.802290 1.377090 1.577020 1.872290 2.398990
2.461630 2.333840 1.904070 1.061460 0.518400
-0.116320 -3.575440 -5.284480 -5.160310 -4.196010
-3.276210 -2.715340 -1.806200 -1.101780 -1.210320
-1.008810 -0.637100 -1.603360 -1.776870

!-150

-0.634810 1.156210 1.624350 2.047200 2.653910
2.691410 2.296420 1.960450 1.324930 2.038290
-1.151510 -3.148610 -4.058280 -4.531850 -3.796370
-2.572090 -1.727250 -0.961410 -0.282910 -0.479120
-1.039340 -1.618060 -1.725460 -1.376360

!-135

0.214000 1.521370 1.977440 2.377950 2.929470
2.893410 2.435810 2.162970 1.761500 1.190090
-1.218610 -2.108900 -2.976100 -3.405340 -2.768440
-1.836030 -0.957950 0.021790 -0.032760 -0.665880
-1.321170 -1.212320 -0.893170 -0.897040

!-120

0.873950 1.959160 2.508990 2.841100 3.698960
3.309330 2.614300 2.481720 2.694660 1.082440
-0.398320 -1.761800 -2.945110 -3.294690 -2.308300
-0.855480 -0.087320 0.439040 0.691880 -0.586330
-1.027210 -0.976640 -0.467580 0.104020

!-105

1.767380 2.286650 2.818030 3.065500 3.370620
3.397440 2.730310 2.878790 2.542010 1.545240
-0.092150 -1.694440 -2.812310 -2.802430 -1.856360
-0.306240 -0.122440 0.444680 0.810150 -0.058630
-0.270290 -0.178830 0.202360 0.493810

!-90

1.456010 2.743180 2.589450 3.046230 3.451510
3.319160 3.052900 3.873720 2.420650 0.949100
0.008370 -1.382980 -2.138930 -2.087380 -1.268300
-0.494370 0.267580 0.908250 0.537520 0.306260
0.069540 0.097460 0.263060 0.603220

!-75

1.396790 3.349090 2.180920 2.942960 3.814070
3.675800 3.555310 3.887290 2.101260 -0.190940
-0.732240 -1.382040 -0.673880 -0.817390 -0.826980
-0.111800 0.053710 0.296400 0.692240 0.428960
-0.036100 -0.033820 -0.194300 0.400210

!-60

0.246650 1.229980 1.716960 3.168570 4.208190
4.366860 4.251080 3.348110 0.997540 -1.287540
-1.179900 -0.684300 -0.853660 -1.158760 -0.347550
0.114810 0.242800 0.322420 0.370140 -0.374950
-0.676940 -1.323430 -1.366650 -0.218770

!-45

-1.196730 0.078060 2.347410 4.211350 5.376000
5.364940 4.355200 2.436510 0.408470 -0.590840
-0.435960 -0.501210 -0.822230 -0.607210 0.057910
0.246580 -0.070570 0.379430 0.247770 -0.571680
-1.282910 -1.715770 -1.839820 -1.987110

!-30

Appendix Transformed Parameter Set for Simulating D-amino Residues

-1.174720 1.067030 4.180460 6.741610 6.070770
4.781470 2.758340 1.295810 0.571150 -0.196480
0.251860 -0.732140 1.289360 1.497590 1.890550
2.198490 0.169290 0.534000 0.331780 -1.276320
-2.550070 -3.312150 -3.136670 -2.642260

!-15
0.293590 5.588070 3.732620 3.217620 3.272450
2.492320 1.563700 1.356760 0.831410 0.630170
1.591970 0.821920 0.486070 0.715760 0.996020
1.591580 -0.367400 0.181770 -0.613920 -2.267900
-3.516460 -3.597700 -3.043340 -1.765020

!0
2.832310 0.787990 0.323280 0.479230 0.628600
0.976330 1.238750 1.671950 1.645480 2.520340
1.606970 0.776350 0.119780 0.070390 0.121170
-1.569230 -1.213010 -1.846360 -2.744510 -3.792530
-3.934880 -3.615930 -2.675750 -0.924170

!15
-0.778340 -1.912680 -2.052140 -1.846280 -1.047430
0.183400 1.682950 2.223500 1.358370 2.448660
1.436920 0.678570 -0.237060 -0.535320 -0.790380
-2.182580 -3.251140 -4.195110 -4.269270 -3.908210
-3.455620 -2.773970 1.755370 0.313410

!30
-2.963810 -3.483730 -3.517080 -2.724860 -1.405510
0.336200 1.428450 1.394630 0.970370 2.462720
1.522430 0.553620 -0.407380 -1.482950 -3.613920
-4.159810 -4.945580 -4.784040 -3.764540 -2.959140
-1.963850 -1.071260 -1.599580 -2.445320

!45
-4.029070 -3.932660 -3.558480 -2.513980 -1.037320
0.362000 0.814380 0.754110 0.502370 1.903420
0.770220 -0.416420 -3.286310 -3.875270 -4.907800
-5.704430 -5.645660 -4.396040 -2.865450 -2.368170
-2.860490 -3.416560 -3.666490 -3.859070

!60
-3.338270 -2.960220 -2.311700 -1.272890 -0.246470
0.722610 0.668070 0.438130 2.395330 1.632470
-2.041450 -3.218100 -3.915080 -4.852510 -5.696500
-6.314370 -5.683690 -4.170620 -3.141000 -3.508820
-3.756430 -3.640810 -3.640430 -3.550690

!75
-2.244860 -1.632100 -1.000640 -0.170440 0.526440
0.823710 0.517140 -0.013120 -0.370910 -1.213720
-2.305650 -3.420580 -4.484960 -5.693140 -6.199150
-6.253870 -5.211310 -4.174380 -3.685150 -4.151360
-4.161970 -3.725150 -3.715310 -2.606760

!90
-1.720840 -1.177830 -0.428430 0.277730 0.807900
0.803260 0.482510 -0.336900 -0.786270 -1.774070
-2.793220 -3.828560 -5.211800 -6.636850 -6.989940
-6.108800 -5.452410 -3.911450 -4.321000 -4.587240
-4.102610 -3.772820 -3.157300 -2.648390

!105
-1.850640 -1.092420 -0.445020 0.128490 1.005520
0.884820 0.485850 -0.218470 -0.857670 -1.682330
-3.014400 -4.481110 -6.053510 -6.865400 -6.871130
-5.728240 -3.912230 -4.802110 -5.034640 -4.715990
-4.601080 -4.086220 -3.274630 -2.410940

!120
-1.969230 -1.116650 -0.540250 -0.150330 0.763520
1.038890 0.758480 0.313530 -0.333050 -1.872770
-3.366270 -5.008260 -6.124810 -7.034830 -6.724320
-3.700200 -4.510620 -5.185650 -5.361620 -4.847490
-4.444320 -4.004260 -3.415720 -2.751230

!135
-2.111250 -1.168960 -0.322790 -0.006920 0.316660
1.086270 0.939170 0.625340 -0.166360 -1.830310
-3.469470 -4.946030 -6.112560 -1.915580 -4.047310
-4.996740 -4.996730 -4.842690 -4.886620 -4.300540

-4.494620 -4.442210 -4.163570 -3.183510

!150
-1.757590 -0.403620 0.023920 0.362390 0.634520
1.264920 1.361360 0.948420 -0.073680 -1.483560
-3.152820 1.835120 -1.762860 -5.093660 -5.744830
-5.390070 -4.783930 -4.190630 -4.115420 -4.042280
-4.125570 -4.028550 -4.026100 -2.937910

!165
-0.810590 -0.071500 0.378890 0.543310 1.277880
1.641310 1.698840 1.519950 0.631950 -1.088670
-2.736530 -0.735240 -4.563830 -6.408350 -5.889450
-5.141750 -4.194970 -3.666490 -3.843450 -3.818830
-3.826180 -3.596820 -2.994790 -2.231020

!D-ala, L-ala, L-pro map
DC NH1 CT1 C NH1 CT1 C N 24

!-180
0.126790 0.768700 0.971260 1.250970 2.121010
2.695430 2.064440 1.764790 0.755870 -0.713470
0.976130 -2.475520 -5.455650 -5.096450 -5.305850
-3.975630 -3.088580 -2.784200 -2.677120 -2.646060
-2.335350 -2.010440 -1.608040 -0.482250

!-165
-0.802290 1.377090 1.577020 1.872290 2.398990
2.461630 2.333840 1.904070 1.061460 0.518400
-0.116320 -3.575440 -5.284480 -5.160310 -4.196010
-3.276210 -2.715340 -1.806200 -1.101780 -1.210320
-1.008810 -0.637100 -1.603360 -1.776870

!-150
-0.634810 1.156210 1.624350 2.047200 2.653910
2.691410 2.296420 1.960450 1.324930 2.038290
-1.151510 -3.148610 -4.058280 -4.531850 -3.796370
-2.572090 -1.727250 -0.961410 -0.282910 -0.479120
-1.039340 -1.618060 -1.725460 -1.376360

!-135
0.214000 1.521370 1.977440 2.377950 2.929470
2.893410 2.435810 2.162970 1.761500 1.190090
-1.218610 -2.108900 -2.976100 -3.405340 -2.768440
-1.836030 -0.957950 0.021790 -0.032760 -0.665880
-1.321170 -1.212320 -0.893170 -0.897040

!-120
0.873950 1.959160 2.508990 2.841100 3.698960
3.309330 2.614300 2.481720 2.694660 1.082440
-0.398320 -1.761800 -2.945110 -3.294690 -2.308300
-0.855480 -0.087320 0.439040 0.691880 -0.586330
-1.027210 -0.976640 -0.467580 0.104020

!-105
1.767380 2.286650 2.818030 3.065500 3.370620
3.397440 2.730310 2.878790 2.542010 1.545240
-0.092150 -1.694440 -2.812310 -2.802430 -1.856360
-0.306240 -0.122440 0.444680 0.810150 -0.058630
-0.270290 -0.178830 0.202360 0.493810

!-90
1.456010 2.743180 2.589450 3.046230 3.451510
3.319160 3.052900 3.873720 2.420650 0.949100
0.008370 -1.382980 -2.138930 -2.087380 -1.268300
-0.494370 0.267580 0.908250 0.537520 0.306260
0.069540 0.097460 0.263060 0.603220

!-75
1.396790 3.349090 2.180920 2.942960 3.814070
3.675800 3.555310 3.887290 2.101260 -0.190940
-0.732240 -1.382040 -0.673880 -0.817390 -0.826980
-0.111800 0.053710 0.296400 0.692240 0.428960
-0.036100 -0.033820 -0.194300 0.400210

!-60
0.246650 1.229980 1.716960 3.168570 4.208190
4.366860 4.251080 3.348110 0.997540 -1.287540
-1.179900 -0.684300 -0.853660 -1.158760 -0.374550
0.114810 0.242800 0.322420 0.370140 -0.374950
-0.676940 -1.323430 -1.366650 -0.218770

Appendix Transformed Parameter Set for Simulating D-amino Residues

!-45
-1.196730 0.078060 2.347410 4.211350 5.376000
5.364940 4.355200 2.436510 0.408470 -0.590840
-0.435960 -0.501210 -0.822230 -0.607210 0.057910
0.246580 -0.070570 0.379430 0.247770 -0.571680
-1.282910 -1.715770 -1.839820 -1.987110

!-30
-1.174720 1.067030 4.180460 6.741610 6.070770
4.781470 2.758340 1.295810 0.571150 -0.196480
0.251860 -0.732140 1.289360 1.497590 1.890550
2.198490 0.169290 0.534000 0.331780 -1.276320
-2.550070 -3.312150 -3.136670 -2.642260

!-15
0.293590 5.588070 3.732620 3.217620 3.272450
2.492320 1.563700 1.356760 0.831410 0.630170
1.591970 0.821920 0.486070 0.715760 0.996020
1.591580 -0.367400 0.181770 -0.613920 -2.267900
-3.516460 -3.597700 -3.043340 -1.765020

!0
2.832310 0.787990 0.323280 0.479230 0.628600
0.976330 1.238750 1.671950 1.645480 2.520340
1.606970 0.776350 0.119780 0.070390 0.121170
-1.569230 -1.213010 -1.846360 -2.744510 -3.792530
-3.934880 -3.615930 -2.675750 -0.924170

!15
-0.778340 -1.912680 -2.052140 -1.846280 -1.047430
0.183400 1.682950 2.223500 1.358370 2.448660
1.436920 0.678570 -0.237060 -0.535320 -0.790380
-2.182580 -3.251140 -4.195110 -4.269270 -3.908210
-3.455620 -2.773970 1.755370 0.313410

!30
-2.963810 -3.483730 -3.517080 -2.724860 -1.405510
0.336200 1.428450 1.394630 0.970370 2.462720
1.522430 0.553620 -0.407380 -1.482950 -3.613920
-4.159810 -4.945580 -4.784040 -3.764540 -2.959140
-1.963850 -1.071260 -1.599580 -2.445320

!45
-4.029070 -3.932660 -3.558480 -2.513980 -1.037320
0.362000 0.814380 0.754110 0.502370 1.903420
0.770220 -0.416420 -3.286310 -3.875270 -4.907800
-5.704430 -5.645660 -4.396040 -2.865450 -2.368170
-2.860490 -3.416560 -3.666490 -3.859070

!60
-3.338270 -2.960220 -2.311700 -1.272890 -0.246470
0.722610 0.668070 0.438130 2.395330 1.632470
-2.041450 -3.218100 -3.915080 -4.852510 -5.696500
-6.314370 -5.683690 -4.170620 -3.141000 -3.508820
-3.756430 -3.640810 -3.640430 -3.550690

!75
-2.244860 -1.632100 -1.000640 -0.170440 0.526440
0.823710 0.517140 -0.013120 -0.370910 -1.213720
-2.305650 -3.420580 -4.484960 -5.693140 -6.199150
-6.253870 -5.211310 -4.174380 -3.685150 -4.151360
-4.161970 -3.725150 -3.715310 -2.606760

!90
-1.720840 -1.177830 -0.428430 0.277730 0.807900
0.803260 0.482510 -0.336900 -0.786270 -1.774070
-2.793220 -3.828560 -5.211800 -6.636850 -6.989940
-6.108800 -5.452410 -3.911450 -4.321000 -4.587240
-4.102610 -3.772820 -3.157300 -2.648390

!105
-1.850640 -1.092420 -0.445020 0.128490 1.005520
0.884820 0.485850 -0.218470 -0.857670 -1.682330
-3.014400 -4.481110 -6.053510 -6.865400 -6.871130
-5.728240 -3.912230 -4.802110 -5.034640 -4.715990
-4.601080 -4.086220 -3.274630 -2.410940

!120
-1.969230 -1.116650 -0.540250 -0.150330 0.763520
1.038890 0.758480 0.313530 -0.333050 -1.872770

-3.366270 -5.008260 -6.124810 -7.034830 -6.724320
-3.700200 -4.510620 -5.185650 -5.361620 -4.847490
-4.444320 -4.004260 -3.415720 -2.751230

!135
-2.111250 -1.168960 -0.322790 -0.006920 0.316660
1.086270 0.939170 0.625340 -0.166360 -1.830310
-3.469470 -4.946030 -6.112560 -1.915580 -4.047310
-4.996740 -4.996730 -4.842690 -4.886620 -4.300540
-4.494620 -4.442210 -4.163570 -3.183510

!150
-1.757590 -0.403620 0.023920 0.362390 0.634520
1.264920 1.361360 0.948420 -0.073680 -1.483560
-3.152820 1.835120 -1.762860 -5.093660 -5.744830
-5.390070 -4.783930 -4.190630 -4.115420 -4.042280
-4.125570 -4.028550 -4.026100 -2.937910

!165
-0.810590 -0.071500 0.378890 0.543310 1.277880
1.641310 1.698840 1.519950 0.631950 -1.088670
-2.736530 -0.735240 -4.563830 -6.408350 -5.889450
-5.141750 -4.194970 -3.666490 -3.843450 -3.818830
-3.826180 -3.596820 -2.994790 -2.231020

!D-ala, L-ala, D-ala map
DC NH1 CT1 C NH1 CT1 C DNH1 24

!-180
0.126790 0.768700 0.971260 1.250970 2.121010
2.695430 2.064440 1.764790 0.755870 -0.713470
0.976130 -2.475520 -5.455650 -5.096450 -5.305850
-3.975630 -3.088580 -2.784200 -2.677120 -2.646060
-2.335350 -2.010440 -1.608040 -0.482250

!-165
-0.802290 1.377090 1.577020 1.872290 2.398990
2.461630 2.333840 1.904070 1.061460 0.518400
-0.116320 -3.575440 -5.284480 -5.160310 -4.196010
-3.276210 -2.715340 -1.806200 -1.101780 -1.210320
-1.008810 -0.637100 -1.603360 -1.776870

!-150
-0.634810 1.156210 1.624350 2.047200 2.653910
2.691410 2.296420 1.960450 1.324930 2.038290
-1.151510 -3.148610 -4.058280 -4.531850 -3.796370
-2.572090 -1.727250 -0.961410 -0.282910 -0.479120
-1.039340 -1.618060 -1.725460 -1.376360

!-135
0.214000 1.521370 1.977440 2.377950 2.929470
2.893410 2.435810 2.162970 1.761500 1.190090
-1.218610 -2.108900 -2.976100 -3.405340 -2.768440
-1.836030 -0.957950 0.021790 -0.032760 -0.665880
-1.321170 -1.212320 -0.893170 -0.897040

!-120
0.873950 1.959160 2.508990 2.841100 3.698960
3.309330 2.614300 2.481720 2.694660 1.082440
-0.398320 -1.761800 -2.945110 -3.294690 -2.308300
-0.855480 -0.087320 0.439040 0.691880 -0.586330
-1.027210 -0.976640 -0.467580 0.104020

!-105
1.767380 2.286650 2.818030 3.065500 3.370620
3.397440 2.730310 2.878790 2.542010 1.545240
-0.092150 -1.694440 -2.812310 -2.802430 -1.856360
-0.306240 -0.122440 0.444680 0.810150 -0.058630
-0.270290 -0.178830 0.202360 0.493810

!-90
1.456010 2.743180 2.589450 3.046230 3.451510
3.319160 3.052900 3.873720 2.420650 0.949100
0.008370 -1.382980 -2.138930 -2.087380 -1.268300
-0.494370 0.267580 0.908250 0.537520 0.306260
0.069540 0.097460 0.263060 0.603220

!-75
1.396790 3.349090 2.180920 2.942960 3.814070
3.675800 3.555310 3.887290 2.101260 -0.190940
-0.732240 -1.382040 -0.673880 -0.817390 -0.826980

Appendix Transformed Parameter Set for Simulating D-amino Residues

-0.111800 0.053710 0.296400 0.692240 0.428960
-0.036100 -0.033820 -0.194300 0.400210

!-60
0.246650 1.229980 1.716960 3.168570 4.208190
4.366860 4.251080 3.348110 0.997540 -1.287540
-1.179900 -0.684300 -0.853660 -1.158760 -0.347550
0.114810 0.242800 0.322420 0.370140 -0.374950
-0.676940 -1.323430 -1.366650 -0.218770

!-45
-1.196730 0.078060 2.347410 4.211350 5.376000
5.364940 4.355200 2.436510 0.408470 -0.590840
-0.435960 -0.501210 -0.822230 -0.607210 0.057910
0.246580 -0.070570 0.379430 0.247770 -0.571680
-1.282910 -1.715770 -1.839820 -1.987110

!-30
-1.174720 1.067030 4.180460 6.741610 6.070770
4.781470 2.758340 1.295810 0.571150 -0.196480
0.251860 -0.732140 1.289360 1.497590 1.890550
2.198490 0.169290 0.534000 0.331780 -1.276320
-2.550070 -3.312150 -3.136670 -2.642260

!-15
0.293590 5.588070 3.732620 3.217620 3.272450
2.492320 1.563700 1.356760 0.831410 0.630170
1.591970 0.821920 0.486070 0.715760 0.996020
1.591580 -0.367400 0.181770 -0.613920 -2.267900
-3.516460 -3.597700 -3.043340 -1.765020

!0
2.832310 0.787990 0.323280 0.479230 0.628600
0.976330 1.238750 1.671950 1.645480 2.520340
1.606970 0.776350 0.119780 0.070390 0.121170
-1.569230 -1.213010 -1.846360 -2.744510 -3.792530
-3.934880 -3.615930 -2.675750 -0.924170

!15
-0.778340 -1.912680 -2.052140 -1.846280 -1.047430
0.183400 1.682950 2.223500 1.358370 2.448660
1.436920 0.678570 -0.237060 -0.535320 -0.790380
-2.182580 -3.251140 -4.195110 -4.269270 -3.908210
-3.455620 -2.773970 1.755370 0.313410

!30
-2.963810 -3.483730 -3.517080 -2.724860 -1.405510
0.336200 1.428450 1.394630 0.970370 2.462720
1.522430 0.553620 -0.407380 -1.482950 -3.613920
-4.159810 -4.945580 -4.784040 -3.764540 -2.959140
-1.963850 -1.071260 -1.599580 -2.445320

!45
-4.029070 -3.932660 -3.558480 -2.513980 -1.037320
0.362000 0.814380 0.754110 0.502370 1.903420
0.770220 -0.416420 -3.286310 -3.875270 -4.907800
-5.704430 -5.645660 -4.396040 -2.865450 -2.368170
-2.860490 -3.416560 -3.666490 -3.859070

!60
-3.338270 -2.960220 -2.311700 -1.272890 -0.246470
0.722610 0.668070 0.438130 2.395330 1.632470
-2.041450 -3.218100 -3.915080 -4.852510 -5.696500
-6.314370 -5.683690 -4.170620 -3.141000 -3.508820
-3.756430 -3.640810 -3.640430 -3.550690

!75
-2.244860 -1.632100 -1.000640 -0.170440 0.526440
0.823710 0.517140 -0.013120 -0.370910 -1.213720
-2.305650 -3.420580 -4.484960 -5.693140 -6.199150
-6.253870 -5.211310 -4.174380 -3.685150 -4.151360
-4.161970 -3.725150 -3.715310 -2.606760

!90
-1.720840 -1.177830 -0.428430 0.277730 0.807900
0.803260 0.482510 -0.336900 -0.786270 -1.774070
-2.793220 -3.828560 -5.211800 -6.636850 -6.989940
-6.108800 -5.452410 -3.911450 -4.321000 -4.587240
-4.102610 -3.772820 -3.157300 -2.648390

!105
-1.850640 -1.092420 -0.445020 0.128490 1.005520
0.884820 0.485850 -0.218470 -0.857670 -1.682330
-3.014400 -4.481110 -6.053510 -6.865400 -6.871130
-5.728240 -3.912230 -4.802110 -5.034640 -4.715990
-4.601080 -4.086220 -3.274630 -2.410940

!120
-1.969230 -1.116650 -0.540250 -0.150330 0.763520
1.038890 0.758480 0.313530 -0.333050 -1.872770
-3.366270 -5.008260 -6.124810 -7.034830 -6.724320
-3.700200 -4.510620 -5.185650 -5.361620 -4.847490
-4.444320 -4.004260 -3.415720 -2.751230

!135
-2.111250 -1.168960 -0.322790 -0.006920 0.316660
1.086270 0.939170 0.625340 -0.166360 -1.830310
-3.469470 -4.946030 -6.112560 -1.915580 -4.047310
-4.996740 -4.996730 -4.842690 -4.886620 -4.300540
-4.494620 -4.442210 -4.163570 -3.183510

!150
-1.757590 -0.403620 0.023920 0.362390 0.634520
1.264920 1.361360 0.948420 -0.073680 -1.483560
-3.152820 1.835120 -1.762860 -5.093660 -5.744830
-5.390070 -4.783930 -4.190630 -4.115420 -4.042280
-4.125570 -4.028550 -4.026100 -2.937910

!165
-0.810590 -0.071500 0.378890 0.543310 1.277880
1.641310 1.698840 1.519950 0.631950 -1.088670
-2.736530 -0.735240 -4.563830 -6.408350 -5.889450
-5.141750 -4.194970 -3.666490 -3.843450 -3.818830
-3.826180 -3.596820 -2.994790 -2.231020

ID-ala, D-ala, L-ala map
DC DNH1 DCT1 DC DNH1 DCT1 DC NH1 24

!-180
-0.48225 -1.60804 -2.01044 -2.33535 -2.64606
-2.67712 -2.7842 -3.08858 -3.97563 -5.30585
-5.09645 -5.45565 -2.47552 0.97613 -0.71347
0.75587 1.76479 2.06444 2.69543 2.12101
1.25097 0.97126 0.7687 0.12679

!-165
-2.23102 -2.99479 -3.59682 -3.82618 -3.81883
-3.84345 -3.66649 -4.19497 -5.14175 -5.88945
-6.40835 -4.56383 -0.73524 -2.73653 -1.08867
0.63195 1.51995 1.69884 1.64131 1.27788
0.54331 0.37889 -0.0715 -0.81059

!-150
-2.93791 -4.0261 -4.02855 -4.12557 -4.04228
-4.11542 -4.19063 -4.78393 -5.39007 -5.74483
-5.09366 -1.76286 1.83512 -3.15282 -1.48356
-0.07368 0.94842 1.36136 1.26492 0.63452
0.36239 0.02392 -0.40362 -1.75759

!-135
-3.18351 -4.16357 -4.44221 -4.49462 -4.30054
-4.88662 -4.84269 -4.99673 -4.99674 -4.04731
-1.91558 -6.11256 -4.94603 -3.46947 -1.83031
-0.16636 0.62534 0.93917 1.08627 0.31666
-0.00692 -0.32279 -1.16896 -2.11125

!-120
-2.75123 -3.41572 -4.00426 -4.44432 -4.84749
-5.36162 -5.18565 -4.51062 -3.7002 -6.72432
-7.03483 -6.12481 -5.00826 -3.36627 -1.87277
-0.33305 0.31353 0.75848 1.03889 0.76352
-0.15033 -0.54025 -1.11665 -1.96923

!-105
-2.41094 -3.27463 -4.08622 -4.60108 -4.71599
-5.03464 -4.80211 -3.91223 -5.72824 -6.87113
-6.8654 -6.05351 -4.48111 -3.0144 -1.68233
-0.85767 -0.21847 0.48585 0.88482 1.00552
0.12849 -0.44502 -1.09242 -1.85064

!-90
-2.64839 -3.1573 -3.77282 -4.10261 -4.58724

Appendix Transformed Parameter Set for Simulating D-amino Residues

-4.321 -3.91145 -5.45241 -6.1088 -6.98994
-6.63685 -5.2118 -3.82856 -2.79322 -1.77407
-0.78627 -0.3369 0.48251 0.80326 0.8079
0.27773 -0.42843 -1.17783 -1.72084

!-75
-2.60676 -3.71531 -3.72515 -4.16197 -4.15136
-3.68515 -4.17438 -5.21131 -6.25387 -6.19915
-5.69314 -4.48496 -3.42058 -2.30565 -1.21372
-0.37091 -0.01312 0.51714 0.82371 0.52644
-0.17044 -1.00064 -1.6321 -2.24486

!-60
-3.55069 -3.64043 -3.64081 -3.75643 -3.50882
-3.141 -4.17062 -5.68369 -6.31437 -5.6965
-4.85251 -3.91508 -3.2181 -2.04145 1.63247
2.39533 0.43813 0.66807 0.72261 -0.24647
-1.27289 -2.3117 -2.96022 -3.33827

!-45
-3.85907 -3.66649 -3.41656 -2.86049 -2.36817
-2.86545 -4.39604 -5.64556 -5.70443 -4.9078
-3.87527 -3.28631 -0.41642 0.77022 1.90342
0.50237 0.75411 0.81438 0.362 -1.03732
-2.51398 -3.55848 -3.93266 -4.02907

!-30
-2.44532 -1.59958 -1.07126 -1.96385 -2.95914
-3.76454 -4.78404 -4.94558 -4.15981 -3.61392
-1.48295 -0.40738 0.55362 1.52243 2.46272
0.97037 1.39463 1.42845 0.3362 -1.40551
-2.72486 -3.51708 -3.48373 -2.96381

!-15
0.31341 1.75537 -2.77397 -3.45562 -3.90821
-4.26927 -4.19511 -3.25114 -2.18258 -0.79038
-0.53532 -0.23706 0.67857 1.43692 2.44866
1.35837 2.2235 1.68295 0.1834 -1.04743
-1.84628 -2.05214 -1.91268 -0.77834

!0
-0.92417 -2.67575 -3.61593 -3.93488 -3.79253
-2.74451 -1.84636 -1.21301 -1.56923 0.12117
0.07039 0.11978 0.77635 1.60697 2.52034
1.64548 1.67195 1.23875 0.97633 0.6286
0.47923 0.32328 0.78799 2.83231

!15
-1.76502 -3.04334 -3.5977 -3.51646 -2.2679
-0.61392 0.18177 -0.3674 1.59158 0.99602
0.71576 0.48607 0.82192 1.59197 0.63017
0.83141 1.35676 1.5637 2.49232 3.27245
3.21762 3.73262 5.58807 0.29359

!30
-2.64226 -3.13667 -3.31215 -2.55007 -1.27632
0.33178 0.534 0.16929 2.19849 1.89055
1.49759 1.28936 -0.73214 0.25186 -0.19648
0.57115 1.29581 2.75834 4.78147 6.07077
6.74161 4.18046 1.06703 -1.17472

!45
-1.98711 -1.83982 -1.71577 -1.28291 -0.57168
0.24777 0.37943 -0.07057 0.24658 0.05791
-0.60721 -0.82223 -0.50121 -0.43596 -0.59084
0.40847 2.43651 4.3552 5.36494 5.376
4.21135 2.34741 0.07806 -1.19673

!60
-0.21877 -1.36665 -1.32343 -0.67694 -0.37495
0.37014 0.32242 0.2428 0.11481 -0.34755
-1.15876 -0.85366 -0.6843 -1.1799 -1.28754
0.99754 3.34811 4.25108 4.36686 4.20819
3.16857 1.71696 1.22998 0.24665

!75
0.40021 -0.1943 -0.03382 -0.0361 0.42896
0.69224 0.2964 0.05371 -0.1118 -0.82698
-0.81739 -0.67388 -1.38204 -0.73224 -0.19094
2.10126 3.88729 3.55531 3.6758 3.81407
2.94296 2.18092 3.34909 1.39679

!90
0.60322 0.26306 0.09746 0.06954 0.30626
0.53752 0.90825 0.26758 -0.49437 -1.2683
-2.08738 -2.13893 -1.38298 0.00837 0.9491
2.42065 3.87372 3.0529 3.31916 3.45151
3.04623 2.58945 2.74318 1.45601

!105
0.49381 0.20236 -0.17883 -0.27029 -0.05863
0.81015 0.44468 -0.12244 -0.30624 -1.85636
-2.80243 -2.81231 -1.69444 -0.09215 1.54524
2.54201 2.87879 2.73031 3.39744 3.37062
3.0655 2.81803 2.28665 1.76738

!120
0.10402 -0.46758 -0.97664 -1.02721 -0.58633
0.69188 0.43904 -0.08732 -0.85548 -2.3083
-3.29469 -2.94511 -1.7618 -0.39832 1.08244
2.69466 2.48172 2.6143 3.30933 3.69896
2.8411 2.50899 1.95916 0.87395

!135
-0.89704 -0.89317 -1.21232 -1.32117 -0.66588
-0.03276 0.02179 -0.95795 -1.83603 -2.76844
-3.40534 -2.9761 -2.1089 -1.21861 1.19009
1.7615 2.16297 2.43581 2.89341 2.92947
2.37795 1.97744 1.52137 0.214

!150
-1.37636 -1.72546 -1.61806 -1.03934 -0.47912
-0.28291 -0.96141 -1.72725 -2.57209 -3.79637
-4.53185 -4.05828 -3.14861 -1.15151 2.03829
1.32493 1.96045 2.29642 2.69141 2.65391
2.0472 1.62435 1.15621 -0.63481

!165
-1.77687 -1.60336 -0.6371 -1.00881 -1.21032
-1.10178 -1.8062 -2.71534 -3.27621 -4.19601
-5.16031 -5.28448 -3.57544 -0.11632 0.5184
1.06146 1.90407 2.33384 2.46163 2.39899
1.87229 1.57702 1.37709 -0.80229

ID-ala, D-ala, D-ala map
DC DNH1 DCT1 DC DNH1 DCT1 DC DNH1 24

!-180
-0.48225 -1.60804 -2.01044 -2.33535 -2.64606
-2.67712 -2.7842 -3.08858 -3.97563 -5.30585
-5.09645 -5.45565 -2.47552 0.97613 -0.71347
0.75587 1.76479 2.06444 2.69543 2.12101
1.25097 0.97126 0.7687 0.12679

!-165
-2.23102 -2.99479 -3.59682 -3.82618 -3.81883
-3.84345 -3.66649 -4.19497 -5.14175 -5.88945
-6.40835 -4.56383 -0.73524 -2.73653 -1.08867
0.63195 1.51995 1.69884 1.64131 1.27788
0.54331 0.37889 -0.0715 -0.81059

!-150
-2.93791 -4.0261 -4.02855 -4.12557 -4.04228
-4.11542 -4.19063 -4.78393 -5.39007 -5.74483
-5.09366 -1.76286 1.83512 -3.15282 -1.48356
-0.07368 0.94842 1.36136 1.26492 0.63452
0.36239 0.02392 -0.40362 -1.75759

!-135
-3.18351 -4.16357 -4.44221 -4.49462 -4.30054
-4.88662 -4.84269 -4.99673 -4.99674 -4.04731
-1.91558 -6.11256 -4.94603 -3.46947 -1.83031
-0.16636 0.62534 0.93917 1.08627 0.31666
-0.00692 -0.32279 -1.16896 -2.11125

!-120
-2.75123 -3.41572 -4.00426 -4.44432 -4.84749
-5.36162 -5.18565 -4.51062 -3.7002 -6.72432
-7.03483 -6.12481 -5.00826 -3.36627 -1.87277
-0.33305 0.31353 0.75848 1.03889 0.76352
-0.15033 -0.54025 -1.11665 -1.96923

Appendix Transformed Parameter Set for Simulating D-amino Residues

!-105
-2.41094 -3.27463 -4.08622 -4.60108 -4.71599
-5.03464 -4.80211 -3.91223 -5.72824 -6.87113
-6.8654 -6.05351 -4.48111 -3.0144 -1.68233
-0.85767 -0.21847 0.48585 0.88482 1.00552
0.12849 -0.44502 -1.09242 -1.85064

!-90
-2.64839 -3.1573 -3.77282 -4.10261 -4.58724
-4.321 -3.91145 -5.45241 -6.1088 -6.98994
-6.63685 -5.2118 -3.82856 -2.79322 -1.77407
-0.78627 -0.3369 0.48251 0.80326 0.8079
0.27773 -0.42843 -1.17783 -1.72084

!-75
-2.60676 -3.71531 -3.72515 -4.16197 -4.15136
-3.68515 -4.17438 -5.21131 -6.25387 -6.19915
-5.69314 -4.48496 -3.42058 -2.30565 -1.21372
-0.37091 -0.01312 0.51714 0.82371 0.52644
-0.17044 -1.00064 -1.6321 -2.24486

!-60
-3.55069 -3.64043 -3.64081 -3.75643 -3.50882
-3.141 -4.17062 -5.68369 -6.31437 -5.6965
-4.85251 -3.91508 -3.2181 -2.04145 1.63247
2.39533 0.43813 0.66807 0.72261 -0.24647
-1.27289 -2.3117 -2.96022 -3.33827

!-45
-3.85907 -3.66649 -3.41656 -2.86049 -2.36817
-2.86545 -4.39604 -5.64566 -5.70443 -4.9078
-3.87527 -3.28631 -0.41642 0.77022 1.90342
0.50237 0.75411 0.81438 0.362 -1.03732
-2.51398 -3.55848 -3.93266 -4.02907

!-30
-2.44532 -1.59958 -1.07126 -1.96385 -2.95914
-3.76454 -4.78404 -4.94558 -4.15981 -3.61392
-1.48295 -0.40738 0.55362 1.52243 2.46272
0.97037 1.39463 1.42845 0.3362 -1.40551
-2.72486 -3.51708 -3.48373 -2.96381

!-15
0.31341 1.75537 -2.77397 -3.45562 -3.90821
-4.26927 -4.19511 -3.25114 -2.18258 -0.79038
-0.53532 -0.23706 0.67857 1.43692 2.44866
1.35837 2.2235 1.68295 0.1834 -1.04743
-1.84628 -2.05214 -1.91268 -0.77834

!0
-0.92417 -2.67575 -3.61593 -3.93488 -3.79253
-2.74451 -1.84636 -1.21301 -1.56923 0.12117
0.07039 0.11978 0.77635 1.60697 2.52034
1.64548 1.67195 1.23875 0.97633 0.6286
0.47923 0.32328 0.78799 2.83231

!15
-1.76502 -3.04334 -3.5977 -3.51646 -2.2679
-0.61392 0.18177 -0.3674 1.59158 0.99602
0.71576 0.48607 0.82192 1.59197 0.63017
0.83141 1.35676 1.5637 2.49232 3.27245
3.21762 3.73262 5.58807 0.29359

!30
-2.64226 -3.13667 -3.31215 -2.55007 -1.27632
0.33178 0.534 0.16929 2.19849 1.89055
1.49759 1.28936 -0.73214 0.25186 -0.19648
0.57115 1.29581 2.75834 4.78147 6.07077
6.74161 4.18046 1.06703 -1.17472

!45
-1.98711 -1.83982 -1.71577 -1.28291 -0.57168
0.24777 0.37943 -0.07057 0.24658 0.05791
-0.60721 -0.82223 -0.50121 -0.43596 -0.59084
0.40847 2.43651 4.3552 5.36494 5.376
4.21135 2.34741 0.07806 -1.19673

!60
-0.21877 -1.36665 -1.32343 -0.67694 -0.37495
0.37014 0.32242 0.2428 0.11481 -0.34755
-1.15876 -0.85366 -0.6843 -1.1799 -1.28754

0.99754 3.34811 4.25108 4.36686 4.20819
3.16857 1.71696 1.22998 0.24665

!75
0.40021 -0.1943 -0.03382 -0.0361 0.42896
0.69224 0.2964 0.05371 -0.1118 -0.82698
-0.81739 -0.67388 -1.38204 -0.73224 -0.19094
2.10126 3.88729 3.55531 3.6758 3.81407
2.94296 2.18092 3.34909 1.39679

!90
0.60322 0.26306 0.09746 0.06954 0.30626
0.53752 0.90825 0.26758 -0.49437 -1.2683
-2.08738 -2.13893 -1.38298 0.00837 0.9491
2.42065 3.87372 3.0529 3.31916 3.45151
3.04623 2.58945 2.74318 1.45601

!105
0.49381 0.20236 -0.17883 -0.27029 -0.05863
0.81015 0.44468 -0.12244 -0.30624 -1.85636
-2.80243 -2.81231 -1.69444 -0.09215 1.54524
2.54201 2.87879 2.73031 3.39744 3.37062
3.0655 2.81803 2.28665 1.76738

!120
0.10402 -0.46758 -0.97664 -1.02721 -0.58633
0.69188 0.43904 -0.08732 -0.85548 -2.3083
-3.29469 -2.94511 -1.7618 -0.39832 1.08244
2.69466 2.48172 2.6143 3.30933 3.69896
2.8411 2.50899 1.95916 0.87395

!135
-0.89704 -0.89317 -1.21232 -1.32117 -0.66588
-0.03276 0.02179 -0.95795 -1.83603 -2.76844
-3.40534 -2.9761 -2.1089 -1.21861 1.19009
1.7615 2.16297 2.43581 2.89341 2.92947
2.37795 1.97744 1.52137 0.214

!150
-1.37636 -1.72546 -1.61806 -1.03934 -0.47912
-0.28291 -0.96141 -1.72725 -2.57209 -3.79637
-4.53185 -4.05828 -3.14861 -1.15151 2.03829
1.32493 1.96045 2.29642 2.69141 2.65391
2.0472 1.62435 1.15621 -0.63481

!165
-1.77687 -1.60336 -0.6371 -1.00881 -1.21032
-1.10178 -1.8062 -2.71534 -3.27621 -4.19601
-5.16031 -5.28448 -3.57544 -0.11632 0.5184
1.06146 1.90407 2.33384 2.46163 2.39899
1.87229 1.57702 1.37709 -0.80229

! D-, D-Pro, D-pro
DC DN DCP1 DC DN DCP1 DC DN 24

!-180
-6.0057 -7.9186 -9.6188 -9.3338 -8.3158
-7.7395 -7.5633 -7.9619 -7.9886 -7.9668
-8.1024 -8.6106 -8.8889 -9.0275 -8.9783
-9.1251 -8.4192 -7.6332 -7.1853 -6.6546
-5.4921 -4.5142 -4.2892 -4.6066

!-165
0 0 0 0
0 0 0 0
0 0 0 0
0 0 0 0
0 0 0 0

!-150
0 0 0 0
0 0 0 0
0 0 0 0
0 0 0 0
0 0 0 0

!-135
0 0 0 0
0 0 0 0
0 0 0 0
0 0 0 0

Appendix Transformed Parameter Set for Simulating D-amino Residues

-2.165440 -1.546500 0.753400 -2.949180 -2.225630
-1.664160 -0.628990 0.000490 0.033160 -0.092820
-0.339050 -0.563330 -0.794980 -0.710760

!-135
-1.787640 -2.117750 -2.143020 -1.803720 -1.567160
-0.886880 -0.801350 -0.851590 -1.020630 -1.337360
-1.062570 0.338010 -4.372310 -2.435890 -2.220710
-1.718060 -0.758950 -0.207560 0.100910 -0.055650
-0.288370 -0.880610 -1.267450 -1.465530

!-120
-2.348270 -2.593790 -2.596140 -2.364070 -1.970070
-1.705860 -1.435540 -1.289220 -1.358170 -0.975570
-3.514390 -4.283210 -3.975820 -3.215190 -2.394430
-1.455320 -0.553910 -0.158900 -0.173830 -0.297950
-0.661220 -1.068330 -1.601800 -1.914850

!-105
-2.788800 -3.079570 -3.178150 -3.013710 -2.626630
-2.266680 -1.951490 -1.681850 -1.195390 -2.567680
-3.632800 -4.748210 -4.662850 -4.255190 -2.776760
-1.695490 -0.893140 -0.633810 -0.467320 -0.540540
-0.950190 -1.401500 -1.959970 -2.412680

!-90
-3.857170 -3.713610 -3.902110 -3.611370 -3.040850
-2.406460 -1.975250 -1.452040 -0.971860 -2.808170
-4.181160 -4.981430 -5.446890 -4.359900 -2.864390
-1.898510 -1.139090 -0.971340 -1.065550 -1.020680
-1.141350 -1.794480 -2.420970 -2.939990

!-75
-4.987770 -4.995210 -4.485310 -3.892550 -3.228630
-2.345360 -1.664160 -1.105500 -1.945510 -3.715530
-4.492140 -5.536170 -5.708500 -3.675410 -2.986660
-1.859410 -0.756620 -1.269930 -1.312730 -1.607440
-1.892510 -2.659400 -3.347950 -3.970600

!-60
-6.183650 -5.456080 -4.878940 -4.000820 -2.683230
-2.067520 -1.094850 -1.119790 -2.962910 -3.687830
-4.993340 -4.666260 -3.796280 -3.374140 -2.495430
-1.453990 -0.877560 -1.002930 -1.337310 -2.431360
-2.948140 -4.008100 -4.821040 -5.565810

!-45
-6.755760 -5.850030 -4.362190 -2.714090 -1.708710
-0.526660 -0.536700 -2.037170 -3.892650 -4.558570
-4.237410 -3.735160 -3.688580 -3.009910 -2.112940
-1.455400 -0.925490 -1.121840 -1.561900 -2.751370
-4.094860 -5.207530 -6.128530 -6.613030

!-30
-5.716250 -4.434060 -2.788600 -0.974400 -0.729200
-0.904940 -1.833540 -3.017700 -3.313450 -3.336010
-3.181640 -3.594720 -1.231370 -0.603790 0.128810
-1.222610 -0.909150 -0.837700 -1.346820 -3.040880
-4.731110 -5.844860 -6.428460 -6.424880

!-15
-3.991110 -2.046000 0.082550 -2.676110 -2.828500
-2.596640 -2.843330 -3.011480 -2.312640 -2.405980
-3.086210 -1.164620 -1.231660 -0.871900 -0.348980
-1.735900 -0.914150 -0.484520 -1.818040 -3.602550
-5.330320 -5.992270 -5.588080 -5.408360

!0
-1.147060 -3.317730 -4.305100 -4.615200 -4.533780
-3.622950 -2.832800 -1.872810 -1.144300 -1.994070
-0.741980 -1.115010 -1.229250 -1.103680 -0.742430
-1.973970 -1.070020 -1.802220 -2.712770 -3.624130
-4.537100 -4.619970 -4.310890 -3.318290

!15
-3.997710 -5.408360 -5.588080 -5.992270 -5.330320
-3.602550 -1.818040 -0.484520 -0.914150 -1.735900
-0.348980 -0.871900 -1.231660 -1.164620 -3.086210
-2.405980 -2.312640 -3.011480 -2.843330 -2.596640
-2.828500 -2.676110 0.082550 -2.046000

!30
-5.710850 -6.424880 -6.428460 -5.844860 -4.731110
-3.040880 -1.346820 -0.837700 -0.909150 -1.222610
0.128810 -0.603790 -1.231370 -3.594720 -3.181640
-3.336010 -3.313450 -3.017700 -1.833540 -0.904940
-0.729200 -0.974400 -2.788600 -4.434060

!45
-6.754940 -6.613030 -6.128530 -5.207530 -4.094860
-2.751370 -1.561900 -1.121840 -0.925490 -1.455400
-2.112940 -3.009910 -3.688580 -3.735160 -4.237410
-4.558570 -3.892650 -2.037170 -0.536700 -0.526660
-1.708710 -2.714090 -4.362190 -5.850030

!60
-6.188070 -5.565810 -4.821040 -4.008100 -2.948140
-2.431360 -1.337310 -1.002930 -0.877560 -1.453990
-2.495430 -3.374140 -3.796280 -4.666260 -4.993340
-3.687830 -2.962910 -1.119790 -1.094850 -2.067520
-2.683230 -4.000820 -4.878940 -5.456080

!75
-4.986080 -3.970600 -3.347950 -2.659400 -1.892510
-1.607440 -1.312730 -1.269930 -0.756620 -1.859410
-2.986660 -3.675410 -5.708500 -5.536170 -4.492140
-3.715530 -1.945510 -1.105500 -1.664160 -2.345360
-3.228630 -3.892550 -4.485310 -4.995210

!90
-3.879190 -2.939990 -2.420970 -1.794480 -1.141350
-1.020680 -1.065550 -0.971340 -1.139090 -1.898510
-2.864390 -4.359900 -5.446890 -4.981430 -4.181160
-2.808170 -0.971860 -1.452040 -1.975250 -2.406460
-3.040850 -3.611370 -3.902110 -3.713610

!105
-2.793280 -2.412680 -1.959970 -1.401500 -0.950190
-0.540540 -0.467320 -0.633810 -0.893140 -1.695490
-2.776760 -4.255190 -4.662850 -4.448210 -3.332800
-2.567680 -1.195390 -1.681850 -1.951490 -2.266680
-2.626630 -3.013710 -3.178150 -3.079570

!120
-2.330190 -1.914850 -1.601800 -1.068330 -0.661220
-0.297950 -0.173830 -0.158900 -0.553910 -1.455320
-2.394430 -3.215190 -3.975820 -3.783210 -3.014390
-0.975570 -1.358170 -1.289220 -1.435540 -1.705860
-1.970070 -2.364070 -2.596140 -2.593790

!135
-1.796120 -1.465530 -1.267450 -0.880610 -0.288370
-0.055650 0.100910 -0.207560 -0.758950 -1.718060
-2.220710 -2.435890 -4.372310 0.338010 -1.062570
-1.337360 -1.020630 -0.851590 -0.801350 -0.886880
-1.567160 -1.803720 -2.143020 -2.117750

!150
-1.263610 -0.710760 -0.794980 -0.563330 -0.339050
-0.092820 0.033160 0.000490 -0.628990 -1.664160
-2.225630 -2.949180 0.753400 -1.546500 -2.165440
-1.626260 -0.887080 -0.290670 -0.054220 -1.642540
-1.119780 -1.656770 -1.665900 -1.482430

!165
-0.684660 -0.477660 -0.348750 -0.462500 -0.271890
-0.000430 0.196710 0.180860 -0.410680 -1.810120
-2.675640 1.822690 -1.636020 -3.153350 -2.780320
-1.715700 -0.534910 0.258100 0.329180 -0.126160
-0.940260 -1.319240 -0.990850 -0.896700

! L-, D-PRO, D-ALA
C DN DCP1 DC DN DCP1 DC DNH1 24

!-180
-6.0057 -7.9186 -9.6188 -9.3338 -8.3158
-7.7395 -7.5633 -7.9619 -7.9886 -7.9668
-8.1024 -8.6106 -8.8889 -9.0275 -8.9783
-9.1251 -8.4192 -7.6332 -7.1853 -6.6546
-5.4921 -4.5142 -4.2892 -4.6066

!-165

Appendix Transformed Parameter Set for Simulating D-amino Residues

```

0 0 0 0 0 -5.9161 -5.2696 -3.8306 -2.7319
0 0 0 0 0 !15
0 0 0 0 0 -6.4969 -8.6267 -9.8071 -9.8633 -8.1653
0 0 0 0 0 -6.4579 -2.725 -5.7532 -1.4306 -1.4593
0 0 0 0 0 -1.484 -1.7812 -2.6863 -3.4786 -6.5645
!-150 -7.3081 -7.0445 -5.7853 -4.6496 -4.5078
0 0 0 0 0 -3.8047 -0.8212 -1.1141 -2.7055
0 0 0 0 0 !30
0 0 0 0 0 -5.4942 -7.1248 -7.5206 -6.8194 -5.1628
0 0 0 0 0 -3.1495 -0.1462 -0.2845 -0.4102 -0.5004
0 0 0 0 0 -0.6017 -0.8166 -1.2421 -4.7865 -6.0834
0 0 0 0 0 -6.3548 -5.0805 -3.626 -2.5999 -1.8514
!-135 -0.7044 -0.8426 -2.0668 -3.463
0 0 0 0 0 !45
0 0 0 0 0 -3.4792 -4.4995 -4.8337 -4.1833 -2.7114
0 0 0 0 0 0.9667 1.2811 1.0055 0.8259 0.6445
0 0 0 0 0 0.8087 0.9605 -1.9933 -3.5123 -4.4661
0 0 0 0 0 -3.9827 -2.3982 -0.5975 0.0065 0.205
!-120 -0.0683 -0.8596 -1.3074 -2.4101
0 0 0 0 0 !60
0 0 0 0 0 -1.4684 -2.0111 -2.5938 -2.3336 0.9932
0 0 0 0 0 1.1616 1.4931 1.3072 1.1635 1.106
0 0 0 0 0 1.6319 -0.2471 -1.3394 -2.555 -2.8353
0 0 0 0 0 -2.1019 -0.683 0.7254 1.0599 0.5337
!-105 -0.2049 -0.2521 -0.4073 -0.4882
0 0 0 0 0 !75
0 0 0 0 0 -0.3834 -1.549 -2.1349 0.2335 0.413
0 0 0 0 0 0.5496 0.6535 0.5846 0.4153 0.8188
0 0 0 0 0 1.84 -0.3503 -1.3548 -2.2491 -2.6649
0 0 0 0 0 -2.1767 -0.9552 0.3332 0.2124 -0.2421
!-90 -0.4539 0.003 0.1819 0.2688
0 0 0 0 0 !90
0 0 0 0 0 -0.9566 -2.0997 -2.8554 -3.2538 -2.6412
0 0 0 0 0 -0.4196 -0.5913 -0.7408 -0.5754 0.383
0 0 0 0 0 -0.9664 -1.3135 -2.242 -3.3331 -3.57
0 0 0 0 0 -2.979 -1.8328 -1.1421 -1.4248 -1.8437
!-75 -1.4847 -0.9234 -0.4178 -0.274
0 0 0 0 0 !105
0 0 0 0 0 -2.1005 -3.3716 -4.2394 -4.6674 -4.2193
0 0 0 0 0 -4.0743 -3.7697 -3.6058 -3.3467 -2.6599
0 0 0 0 0 -2.4919 -2.7954 -3.8852 -4.8409 -5.0203
0 0 0 0 0 -4.0481 -3.0498 -2.8306 -3.1541 -3.3158
!-60 -2.8444 -2.0042 -1.4466 -1.3298
0 0 0 0 0 !120
0 0 0 0 0 -3.1676 -4.391 -5.3828 -5.4744 -4.6216
0 0 0 0 0 -4.162 -4.2625 -4.2134 -3.8659 -3.8321
0 0 0 0 0 -3.7845 -4.1609 -5.1776 -5.9866 -6.1474
0 0 0 0 0 -5.3773 -5.2395 -4.9503 -4.5786 -4.3332
!-45 -3.5537 -2.8511 -2.2237 -2.2061
0 0 0 0 0 !135
0 0 0 0 0 -3.3379 -4.7004 -5.6871 -5.9738 -5.3341
0 0 0 0 0 -4.5268 -4.0969 -3.7688 -3.9502 -4.0071
0 0 0 0 0 -4.1585 -4.5816 -5.5755 -6.2033 -6.4149
0 0 0 0 0 -5.6322 -5.7983 -5.4609 -5.02 -4.6361
!-30 -3.8824 -2.9022 -2.0201 -2.1283
0 0 0 0 0 !150
0 0 0 0 0 -3.4173 -4.8039 -6.1861 -6.0528 -5.2488
0 0 0 0 0 -4.3562 -4.1382 -4.3438 -4.4405 -4.5079
0 0 0 0 0 -4.8509 -5.3335 -6.1431 -6.5422 -6.4883
0 0 0 0 0 -6.1682 -5.9311 -5.3692 -5.0156 -4.6162
!-15 -3.6761 -2.5512 -2.2652 -2.2375
0 0 0 0 0 !165
0 0 0 0 0 -4.4889 -5.8441 -7.4916 -6.9351 -6.1318
0 0 0 0 0 -5.1884 -4.8995 -5.579 -5.9051 -5.8381
0 0 0 0 0 -6.2495 -6.6819 -7.3944 -7.6098 -7.595
!0 -7.1782 -6.6249 -5.7832 -5.4179 -4.9236
0 0 0 0 0 -3.8372 -3.0434 -2.9035 -3.2103
0 0 0 0 0 !D-, D-ALA, D-PRO
0 0 0 0 0 DC DNH1 DCT1 DC DNH1 DCT1 DC DN 24

```

Appendix Transformed Parameter Set for Simulating D-amino Residues

!-180
-0.48225 -1.60804 -2.01044 -2.33535 -2.64606
-2.67712 -2.7842 -3.08858 -3.97563 -5.30585
-5.09645 -5.45565 -2.47552 0.97613 -0.71347
0.75587 1.76479 2.06444 2.69543 2.12101
1.25097 0.97126 0.7687 0.12679

!-165
-2.23102 -2.99479 -3.59682 -3.82618 -3.81883
-3.84345 -3.66649 -4.19497 -5.14175 -5.88945
-6.40835 -4.56383 -0.73524 -2.73653 -1.08867
0.63195 1.51995 1.69884 1.64131 1.27788
0.54331 0.37889 -0.0715 -0.81059

!-150
-2.93791 -4.0261 -4.02855 -4.12557 -4.04228
-4.11542 -4.19063 -4.78393 -5.39007 -5.74483
-5.09366 -1.76286 1.83512 -3.15282 -1.48356
-0.07368 0.94842 1.36136 1.26492 0.63452
0.36239 0.02392 -0.40362 -1.75759

!-135
-3.18351 -4.16357 -4.44221 -4.49462 -4.30054
-4.88662 -4.84269 -4.99673 -4.99674 -4.04731
-1.91558 -6.11256 -4.94603 -3.46947 -1.83031
-0.16636 0.62534 0.93917 1.08627 0.31666
-0.00692 -0.32279 -1.16896 -2.11125

!-120
-2.75123 -3.41572 -4.00426 -4.44432 -4.84749
-5.36162 -5.18565 -4.51062 -3.7002 -6.72432
-7.03483 -6.12481 -5.00826 -3.36627 -1.87277
-0.33305 0.31353 0.75848 1.03889 0.76352
-0.15033 -0.54025 -1.11665 -1.96923

!-105
-2.41094 -3.27463 -4.08622 -4.60108 -4.71599
-5.03464 -4.80211 -3.91223 -5.72824 -6.87113
-6.8654 -6.05351 -4.48111 -3.0144 -1.68233
-0.85767 -0.21847 0.48585 0.88482 1.00552
0.12849 -0.44502 -1.09242 -1.85064

!-90
-2.64839 -3.1573 -3.77282 -4.10261 -4.58724
-4.321 -3.91145 -5.45241 -6.1088 -6.98994
-6.63685 -5.2118 -3.82856 -2.79322 -1.77407
-0.78627 -0.3369 0.48251 0.80326 0.8079
0.27773 -0.42843 -1.17783 -1.72084

!-75
-2.60676 -3.71531 -3.72515 -4.16197 -4.15136
-3.68515 -4.17438 -5.21131 -6.25387 -6.19915
-5.69314 -4.48496 -3.42058 -2.30565 -1.21372
-0.37091 -0.01312 0.51714 0.82371 0.52644
-0.17044 -1.00064 -1.6321 -2.24486

!-60
-3.55069 -3.64043 -3.64081 -3.75643 -3.50882
-3.141 -4.17062 -5.68369 -6.31437 -5.6965
-4.85251 -3.91508 -3.2181 -2.04145 1.63247
2.39533 0.43813 0.66807 0.72261 -0.24647
-1.27289 -2.3117 -2.96022 -3.33827

!-45
-3.85907 -3.66649 -3.41656 -2.86049 -2.36817
-2.86545 -4.39604 -5.64566 -5.70443 -4.9078
-3.87527 -3.28631 -0.41642 0.77022 1.90342
0.50237 0.75411 0.81438 0.362 -1.03732
-2.51398 -3.55848 -3.93266 -4.02907

!-30
-2.44532 -1.59958 -1.07126 -1.96385 -2.95914
-3.76454 -4.78404 -4.94558 -4.15981 -3.61392
-1.48295 -0.40738 0.55362 1.52243 2.46272
0.97037 1.39463 1.42845 0.3362 -1.40551
-2.72486 -3.51708 -3.48373 -2.96381

!-15
0.31341 1.75537 -2.77397 -3.45562 -3.90821
-4.26927 -4.19511 -3.25114 -2.18258 -0.79038

-0.53532 -0.23706 0.67857 1.43692 2.44866
1.35837 2.2235 1.68295 0.1834 -1.04743
-1.84628 -2.05214 -1.91268 -0.77834

!0
-0.92417 -2.67575 -3.61593 -3.93488 -3.79253
-2.74451 -1.84636 -1.21301 -1.56923 0.12117
0.07039 0.11978 0.77635 1.60697 2.52034
1.64548 1.67195 1.23875 0.97633 0.6286
0.47923 0.32328 0.78799 2.83231

!15
-1.76502 -3.04334 -3.5977 -3.51646 -2.2679
-0.61392 0.18177 -0.3674 1.59158 0.99602
0.71576 0.48607 0.82192 1.59197 0.63017
0.83141 1.35676 1.5637 2.49232 3.27245
3.21762 3.73262 5.58807 0.29359

!30
-2.64226 -3.13667 -3.31215 -2.55007 -1.27632
0.33178 0.534 0.16929 2.19849 1.89055
1.49759 1.28936 -0.73214 0.25186 -0.19648
0.57115 1.29581 2.75834 4.78147 6.07077
6.74161 4.18046 1.06703 -1.17472

!45
-1.98711 -1.83982 -1.71577 -1.28291 -0.57168
0.24777 0.37943 -0.07057 0.24658 0.05791
-0.60721 -0.82223 -0.50121 -0.43596 -0.59084
0.40847 2.43651 4.3552 5.36494 5.376
4.21135 2.34741 0.07806 -1.19673

!60
-0.21877 -1.36665 -1.32343 -0.67694 -0.37495
0.37014 0.32242 0.2428 0.11481 -0.34755
-1.15876 -0.85366 -0.6843 -1.1799 -1.28754
0.99754 3.34811 4.25108 4.36686 4.20819
3.16857 1.71696 1.22998 0.24665

!75
0.40021 -0.1943 -0.03382 -0.0361 0.42896
0.69224 0.2964 0.05371 -0.1118 -0.82698
-0.81739 -0.67388 -1.38204 -0.73224 -0.19094
2.10126 3.88729 3.55531 3.6758 3.81407
2.94296 2.18092 3.34909 1.39679

!90
0.60322 0.26306 0.09746 0.06954 0.30626
0.53752 0.90825 0.26758 -0.49437 -1.2683
-2.08738 -2.13893 -1.38298 0.00837 0.9491
2.42065 3.87372 3.0529 3.31916 3.45151
3.04623 2.58945 2.74318 1.45601

!105
0.49381 0.20236 -0.17883 -0.27029 -0.05863
0.81015 0.44468 -0.12244 -0.30624 -1.85636
-2.80243 -2.81231 -1.69444 -0.09215 1.54524
2.54201 2.87879 2.73031 3.39744 3.37062
3.0655 2.81803 2.28665 1.76738

!120
0.10402 -0.46758 -0.97664 -1.02721 -0.58633
0.69188 0.43904 -0.08732 -0.85548 -2.3083
-3.29469 -2.94511 -1.7618 -0.39832 1.08244
2.69466 2.48172 2.6143 3.30933 3.69896
2.8411 2.50899 1.95916 0.87395

!135
-0.89704 -0.89317 -1.21232 -1.32117 -0.66588
-0.03276 0.02179 -0.95795 -1.83603 -2.76844
-3.40534 -2.9761 -2.1089 -1.21861 1.19009
1.7615 2.16297 2.43581 2.89341 2.92947
2.37795 1.97744 1.52137 0.214

!150
-1.37636 -1.72546 -1.61806 -1.03934 -0.47912
-0.28291 -0.96141 -1.72725 -2.57209 -3.79637
-4.53185 -4.05828 -3.14861 -1.15151 2.03829
1.32493 1.96045 2.29642 2.69141 2.65391
2.0472 1.62435 1.15621 -0.63481

Appendix Transformed Parameter Set for Simulating D-amino Residues

!165
-1.77687 -1.60336 -0.6371 -1.00881 -1.21032
-1.10178 -1.8062 -2.71534 -3.27621 -4.19601
-5.16031 -5.28448 -3.57544 -0.11632 0.5184
1.06146 1.90407 2.33384 2.46163 2.39899
1.87229 1.57702 1.37709 -0.80229

!D-, GLY, D-ALA
DC NH1 CT2 C NH1 CT2 C DNH1 24

!-180
-0.549160 -0.535500 -0.588110 -0.754620 -0.679290
-0.038150 0.298460 0.326040 -0.375610 -1.704360
-3.061280 -3.956460 -3.576280 -1.038930 2.012450
-1.714610 -0.377660 0.317310 0.294580 -0.042920
-0.676620 -0.744600 -0.586590 -0.554770

!-165
-0.709450 -0.896700 -0.990850 -1.319240 -0.940260
-0.126160 0.329180 0.258100 -0.534910 -1.715700
-2.780320 -3.153350 -1.636020 1.822690 -2.675640
-1.810120 -0.410680 0.180860 0.196710 -0.000430
-0.271890 -0.462500 -0.348750 -0.477660

!-150
-1.224850 -1.482430 -1.665900 -1.656770 -1.119780
-1.642540 -0.054220 -0.290670 -0.887080 -1.626260
-2.165440 -1.546500 0.753400 -2.949180 -2.225630
-1.664160 -0.628990 0.000490 0.033160 -0.092820
-0.339050 -0.563330 -0.794980 -0.710760

!-135
-1.787640 -2.117750 -2.143020 -1.803720 -1.567160
-0.886880 -0.801350 -0.851590 -1.020630 -1.337360
-1.062570 0.338010 -4.372310 -2.435890 -2.220710
-1.718060 -0.758950 -0.207560 0.100910 -0.055650
-0.288370 -0.880610 -1.267450 -1.465530

!-120
-2.348270 -2.593790 -2.596140 -2.364070 -1.970070
-1.705860 -1.435540 -1.289220 -1.358170 -0.975570
-3.514390 -4.283210 -3.975820 -3.215190 -2.394430
-1.455320 -0.553910 -0.158900 -0.173830 -0.297950
-0.661220 -1.068330 -1.601800 -1.914850

!-105
-2.788800 -3.079570 -3.178150 -3.013710 -2.626630
-2.266680 -1.951490 -1.681850 -1.195390 -2.567680
-3.632800 -4.748210 -4.662850 -4.255190 -2.776760
-1.695490 -0.893140 -0.633810 -0.467320 -0.540540
-0.950190 -1.401500 -1.959970 -2.412680

!-90
-3.857170 -3.713610 -3.902110 -3.611370 -3.040850
-2.406460 -1.975250 -1.452040 -0.971860 -2.808170
-4.181160 -4.981430 -5.446890 -4.359900 -2.864390
-1.898510 -1.139090 -0.971340 -1.065550 -1.020680
-1.141350 -1.794480 -2.420970 -2.939990

!-75
-4.987770 -4.995210 -4.485310 -3.892550 -3.228630
-2.345360 -1.664160 -1.105500 -1.945510 -3.715530
-4.492140 -5.536170 -5.708500 -3.675410 -2.986660
-1.859410 -0.756620 -1.269930 -1.312730 -1.607440
-1.892510 -2.659400 -3.347950 -3.970600

!-60
-6.183650 -5.456080 -4.878940 -4.000820 -2.683230
-2.067520 -1.094850 -1.119790 -2.962910 -3.687830
-4.993340 -4.666260 -3.796280 -3.374140 -2.495430
-1.453990 -0.877560 -1.002930 -1.337310 -2.431360
-2.948140 -4.008100 -4.821040 -5.565810

!-45
-6.755760 -5.850030 -4.362190 -2.714090 -1.708710
-0.526660 -0.536700 -2.037170 -3.892650 -4.558570
-4.237410 -3.735160 -3.688580 -3.009910 -2.112940
-1.455400 -0.925490 -1.121840 -1.561900 -2.751370
-4.094860 -5.207530 -6.128530 -6.613030

!-30
-5.716250 -4.434060 -2.788600 -0.974400 -0.729200
-0.904940 -1.833540 -3.017700 -3.313450 -3.336010
-3.181640 -3.594720 -1.231370 -0.603790 0.128810
-1.222610 -0.909150 -0.837700 -1.346820 -3.040880
-4.731110 -5.844860 -6.428460 -6.424880

!-15
-3.991110 -2.046000 0.082550 -2.676110 -2.828500
-2.596640 -2.843330 -3.011480 -2.312640 -2.405980
-3.086210 -1.164620 -1.231660 -0.871900 -0.348980
-1.735900 -0.914150 -0.484520 -1.818040 -3.602550
-5.330320 -5.992270 -5.588080 -5.408360

!0
-1.147060 -3.317730 -4.305100 -4.615200 -4.533780
-3.622950 -2.832800 -1.872810 -1.144300 -1.994070
-0.741980 -1.115010 -1.229250 -1.103680 -0.742430
-1.973970 -1.070020 -1.802220 -2.712770 -3.624130
-4.537100 -4.619970 -4.310890 -3.318290

!15
-3.997710 -5.408360 -5.588080 -5.992270 -5.330320
-3.602550 -1.818040 -0.484520 -0.914150 -1.735900
-0.348980 -0.871900 -1.231660 -1.164620 -3.086210
-2.405980 -2.312640 -3.011480 -2.843330 -2.596640
-2.828500 -2.676110 0.082550 -2.046000

!30
-5.710850 -6.424880 -6.428460 -5.844860 -4.731110
-3.040880 -1.346820 -0.837700 -0.909150 -1.222610
0.128810 -0.603790 -1.231370 -3.594720 -3.181640
-3.336010 -3.313450 -3.017700 -1.833540 -0.904940
-0.729200 -0.974400 -2.788600 -4.434060

!45
-6.754940 -6.613030 -6.128530 -5.207530 -4.094860
-2.751370 -1.561900 -1.121840 -0.925490 -1.455400
-2.112940 -3.009910 -3.688580 -3.735160 -4.237410
-4.558570 -3.892650 -2.037170 -0.536700 -0.526660
-1.708710 -2.714090 -4.362190 -5.850030

!60
-6.188070 -5.565810 -4.821040 -4.008100 -2.948140
-2.431360 -1.337310 -1.002930 -0.877560 -1.453990
-2.495430 -3.374140 -3.796280 -4.666260 -4.993340
-3.687830 -2.962910 -1.119790 -1.094850 -2.067520
-2.683230 -4.000820 -4.878940 -5.456080

!75
-4.986080 -3.970600 -3.347950 -2.659400 -1.892510
-1.607440 -1.312730 -1.269930 -0.756620 -1.859410
-2.986660 -3.675410 -5.708500 -5.536170 -4.492140
-3.715530 -1.945510 -1.105500 -1.664160 -2.345360
-3.228630 -3.892550 -4.485310 -4.995210

!90
-3.879190 -2.939990 -2.420970 -1.794480 -1.141350
-1.020680 -1.065550 -0.971340 -1.139090 -1.898510
-2.864390 -4.359900 -5.446890 -4.981430 -4.181160
-2.808170 -0.971860 -1.452040 -1.975250 -2.406460
-3.040850 -3.611370 -3.902110 -3.713610

!105
-2.793280 -2.412680 -1.959970 -1.401500 -0.950190
-0.540540 -0.467320 -0.633810 -0.893140 -1.695490
-2.776760 -4.255190 -4.662850 -4.448210 -3.332800
-2.567680 -1.195390 -1.681850 -1.951490 -2.266680
-2.626630 -3.013710 -3.178150 -3.079570

!120
-2.330190 -1.914850 -1.601800 -1.068330 -0.661220
-0.297950 -0.173830 -0.158900 -0.553910 -1.455320
-2.394430 -3.215190 -3.975820 -3.783210 -3.014390
-0.975570 -1.358170 -1.289220 -1.435540 -1.705860
-1.970070 -2.364070 -2.596140 -2.593790

!135
-1.796120 -1.465530 -1.267450 -0.880610 -0.288370
-0.055650 0.100910 -0.207560 -0.758950 -1.718060
-2.220710 -2.435890 -4.372310 0.338010 -1.062570
-1.337360 -1.020630 -0.851590 -0.801350 -0.886880

Appendix Transformed Parameter Set for Simulating D-amino Residues

-1.567160 -1.803720 -2.143020 -2.117750

I150
-1.263610 -0.710760 -0.794980 -0.563330 -0.339050
-0.092820 0.033160 0.000490 -0.628990 -1.664160
-2.225630 -2.949180 0.753400 -1.546500 -2.165440
-1.626260 -0.887080 -0.290670 -0.054220 -1.642540
-1.119780 -1.656770 -1.665900 -1.482430

I165
-0.684660 -0.477660 -0.348750 -0.462500 -0.271890
-0.000430 0.196710 0.180860 -0.410680 -1.810120
-2.675640 1.822690 -1.636020 -3.153350 -2.780320
-1.715700 -0.534910 0.258100 0.329180 -0.126160
-0.940260 -1.319240 -0.990850 -0.896700

!L-, D-ALA, D-PRO
C DNH1 DCT1 DC DNH1 DCT1 DC DN 24

I-180
-0.48225 -1.60804 -2.01044 -2.33535 -2.64606
-2.67712 -2.7842 -3.08858 -3.97563 -5.30585
-5.09645 -5.45565 -2.47552 0.97613 -0.71347
0.75587 1.76479 2.06444 2.69543 2.12101
1.25097 0.97126 0.7687 0.12679

I-165
-2.23102 -2.99479 -3.59682 -3.82618 -3.81883
-3.84345 -3.66649 -4.19497 -5.14175 -5.88945
-6.40835 -4.56383 -0.73524 -2.73653 -1.08867
0.63195 1.51995 1.69884 1.64131 1.27788
0.54331 0.37889 -0.0715 -0.81059

I-150
-2.93791 -4.0261 -4.02855 -4.12557 -4.04228
-4.11542 -4.19063 -4.78393 -5.39007 -5.74483
-5.09366 -1.76286 1.83512 -3.15282 -1.48356
-0.07368 0.94842 1.36136 1.26492 0.63452
0.36239 0.02392 -0.40362 -1.75759

I-135
-3.18351 -4.16357 -4.44221 -4.49462 -4.30054
-4.88662 -4.84269 -4.99673 -4.99674 -4.04731
-1.91558 -6.11256 -4.94603 -3.46947 -1.83031
-0.16636 0.62534 0.93917 1.08627 0.31666
-0.00692 -0.32279 -1.16896 -2.11125

I-120
-2.75123 -3.41572 -4.00426 -4.44432 -4.84749
-5.36162 -5.18565 -4.51062 -3.7002 -6.72432
-7.03483 -6.12481 -5.00826 -3.36627 -1.87277
-0.33305 0.31353 0.75848 1.03889 0.76352
-0.15033 -0.54025 -1.11665 -1.96923

I-105
-2.41094 -3.27463 -4.08622 -4.60108 -4.71599
-5.03464 -4.80211 -3.91223 -5.72824 -6.87113
-6.8654 -6.05351 -4.48111 -3.0144 -1.68233
-0.85767 -0.21847 0.48585 0.88482 1.00552
0.12849 -0.44502 -1.09242 -1.85064

I-90
-2.64839 -3.1573 -3.77282 -4.10261 -4.58724
-4.321 -3.91145 -5.45241 -6.1088 -6.98994
-6.63685 -5.2118 -3.82856 -2.79322 -1.77407
-0.78627 -0.3369 0.48251 0.80326 0.8079
0.27773 -0.42843 -1.17783 -1.72084

I-75
-2.60676 -3.71531 -3.72515 -4.16197 -4.15136
-3.68515 -4.17438 -5.21131 -6.25387 -6.19915
-5.69314 -4.48496 -3.42058 -2.30565 -1.21372
-0.37091 -0.01312 0.51714 0.82371 0.52644
-0.17044 -1.00064 -1.6321 -2.24486

I-60
-3.55069 -3.64043 -3.64081 -3.75643 -3.50882
-3.141 -4.17062 -5.68369 -6.31437 -5.6965
-4.85251 -3.91508 -3.2181 -2.04145 1.63247
2.39533 0.43813 0.66807 0.72261 -0.24647
-1.27289 -2.3117 -2.96022 -3.33827

I-45
-3.85907 -3.66649 -3.41656 -2.86049 -2.36817
-2.86545 -4.39604 -5.64566 -5.70443 -4.9078
-3.87527 -3.28631 -0.41642 0.77022 1.90342
0.50237 0.75411 0.81438 0.362 -1.03732
-2.51398 -3.55848 -3.93266 -4.02907

I-30
-2.44532 -1.59958 -1.07126 -1.96385 -2.95914
-3.76454 -4.78404 -4.94558 -4.15981 -3.61392
-1.48295 -0.40738 0.55362 1.52243 2.46272
0.97037 1.39463 1.42845 0.3362 -1.40551
-2.72486 -3.51708 -3.48373 -2.96381

I-15
0.31341 1.75537 -2.77397 -3.45562 -3.90821
-4.26927 -4.19511 -3.25114 -2.18258 -0.79038
-0.53532 -0.23706 0.67857 1.43692 2.44866
1.35837 2.2235 1.68295 0.1834 -1.04743
-1.84628 -2.05214 -1.91268 -0.77834

I0
-0.92417 -2.67575 -3.61593 -3.93488 -3.79253
-2.74451 -1.84636 -1.21301 -1.56923 0.12117
0.07039 0.11978 0.77635 1.60697 2.52034
1.64548 1.67195 1.23875 0.97633 0.6286
0.47923 0.32328 0.78799 2.83231

I15
-1.76502 -3.04334 -3.5977 -3.51646 -2.2679
-0.61392 0.18177 -0.3674 1.59158 0.99602
0.71576 0.48607 0.82192 1.59197 0.63017
0.83141 1.35676 1.5637 2.49232 3.27245
3.21762 3.73262 5.58807 0.29359

I30
-2.64226 -3.13667 -3.31215 -2.55007 -1.27632
0.33178 0.534 0.16929 2.19849 1.89055
1.49759 1.28936 -0.73214 0.25186 -0.19648
0.57115 1.29581 2.75834 4.78147 6.07077
6.74161 4.18046 1.06703 -1.17472

I45
-1.98711 -1.83982 -1.71577 -1.28291 -0.57168
0.24777 0.37943 -0.07057 0.24658 0.05791
-0.60721 -0.82223 -0.50121 -0.43596 -0.59084
0.40847 2.43651 4.3552 5.36494 5.376
4.21135 2.34741 0.07806 -1.19673

I60
-0.21877 -1.36665 -1.32343 -0.67694 -0.37495
0.37014 0.32242 0.2428 0.11481 -0.34755
-1.15876 -0.85366 -0.6843 -1.1799 -1.28754
0.99754 3.34811 4.25108 4.36686 4.20819
3.16857 1.71696 1.22998 0.24665

I75
0.40021 -0.1943 -0.03382 -0.0361 0.42896
0.69224 0.2964 0.05371 -0.1118 -0.82698
-0.81739 -0.67388 -1.38204 -0.73224 -0.19094
2.10126 3.88729 3.55531 3.6758 3.81407
2.94296 2.18092 3.34909 1.39679

I90
0.60322 0.26306 0.09746 0.06954 0.30626
0.53752 0.90825 0.26758 -0.49437 -1.2683
-2.08738 -2.13893 -1.38298 0.00837 0.9491
2.42065 3.87372 3.0529 3.31916 3.45151
3.04623 2.58945 2.74318 1.45601

I105
0.49381 0.20236 -0.17883 -0.27029 -0.05863
0.81015 0.44468 -0.12244 -0.30624 -1.85636
-2.80243 -2.81231 -1.69444 -0.09215 1.54524
2.54201 2.87879 2.73031 3.39744 3.37062
3.0655 2.81803 2.28665 1.76738

I120
0.10402 -0.46758 -0.97664 -1.02721 -0.58633
0.69188 0.43904 -0.08732 -0.85548 -2.3083

Appendix Transformed Parameter Set for Simulating D-amino Residues

-3.29469 -2.94511 -1.7618 -0.39832 1.08244	0 0 0 0
2.69466 2.48172 2.6143 3.30933 3.69896	0 0 0 0
2.8411 2.50899 1.95916 0.87395	
	I-60
!135	0 0 0 0
-0.89704 -0.89317 -1.21232 -1.32117 -0.66588	0 0 0 0
-0.03276 0.02179 -0.95795 -1.83603 -2.76844	0 0 0 0
-3.40534 -2.9761 -2.1089 -1.21861 1.19009	0 0 0 0
1.7615 2.16297 2.43581 2.89341 2.92947	0 0 0 0
2.37795 1.97744 1.52137 0.214	
	I-45
!150	0 0 0 0
-1.37636 -1.72546 -1.61806 -1.03934 -0.47912	0 0 0 0
-0.28291 -0.96141 -1.72725 -2.57209 -3.79637	0 0 0 0
-4.53185 -4.05828 -3.14861 -1.15151 2.03829	0 0 0 0
1.32493 1.96045 2.29642 2.69141 2.65391	0 0 0 0
2.0472 1.62435 1.15621 -0.63481	
	I-30
!165	-2.3892 -5.102 -7.102 -6.8822 -6.8375
-1.77687 -1.60336 -0.6371 -1.00881 -1.21032	-7.2198 -7.8066 -7.664 -6.6643 -3.9961
-1.10178 -1.8062 -2.71534 -3.27621 -4.19601	-3.814 -3.8999 -1.7562 -4.2748 -4.4979
-5.16031 -5.28448 -3.57544 -0.11632 0.5184	-4.3419 -6.3504 -6.2626 -7.4995 -8.4761
1.06146 1.90407 2.33384 2.46163 2.39899	-7.9518 -6.6493 -4.8772 -2.9624
1.87229 1.57702 1.37709 -0.80229	
	I-15
ID-, D-PRO, L-ALA	-5.1926 -7.156 -7.2901 -7.8816 -7.8776
DC DN DCP1 DC DN DCP1 DC NH1 24	-7.9583 -7.5691 -6.7542 -4.0726 -4.4798
	-4.4528 -4.3772 -4.4364 -5.3679 -5.6538
	-4.7721 -7.0024 -7.6488 -8.3212 -8.2256
	-7.2676 -6.1779 -4.6683 -3.1742
	I0
!-180	-5.1204 -7.1012 -8.2774 -8.6095 -8.0409
-6.0057 -7.9186 -9.6188 -9.3338 -8.3158	-7.0272 -5.7891 -3.0347 -3.5574 -3.9199
-7.7395 -7.5633 -7.9619 -7.9886 -7.9668	-3.646 -3.7874 -4.2891 -5.0507 -4.984
-8.1024 -8.6106 -8.8889 -9.0275 -8.9783	-7.5416 -7.8122 -7.7752 -7.0542 -6.7169
-9.1251 -8.4192 -7.6332 -7.1853 -6.6546	-5.9161 -5.2696 -3.8306 -2.7319
-5.4921 -4.5142 -4.2892 -4.6066	
	!15
!-165	-6.4969 -8.6267 -9.8071 -9.8633 -8.1653
0 0 0 0	-6.4579 -2.725 -5.7532 -1.4306 -1.4593
0 0 0 0	-1.484 -1.7812 -2.6863 -3.4786 -6.5645
0 0 0 0	-7.3081 -7.0445 -5.7853 -4.6496 -4.5078
0 0 0 0	-3.8047 -0.8212 -1.1141 -2.7055
0 0 0 0	
	!30
!-150	-5.4942 -7.1248 -7.5206 -6.8194 -5.1628
0 0 0 0	-3.1495 -0.1462 -0.2845 -0.4102 -0.5004
0 0 0 0	-0.6017 -0.8166 -1.2421 -4.7865 -6.0834
0 0 0 0	-6.3548 -5.0805 -3.626 -2.5999 -1.8514
0 0 0 0	-0.7044 -0.8426 -2.0668 -3.463
	!45
!-135	-3.4792 -4.4995 -4.8337 -4.1833 -2.7114
0 0 0 0	0.9667 1.2811 1.0055 0.8259 0.6445
0 0 0 0	0.8087 0.9605 -1.9933 -3.5123 -4.4661
0 0 0 0	-3.9827 -2.3982 -0.5975 0.0065 0.205
0 0 0 0	-0.0683 -0.8596 -1.3074 -2.4101
	!60
!-120	-1.4684 -2.0111 -2.5938 -2.3336 0.9932
0 0 0 0	1.1616 1.4931 1.3072 1.1635 1.106
0 0 0 0	1.6319 -0.2471 -1.3394 -2.555 -2.8353
0 0 0 0	-2.1019 -0.683 0.7254 1.0599 0.5337
0 0 0 0	-0.2049 -0.2521 -0.4073 -0.4882
	!75
!-105	-0.3834 -1.549 -2.1349 0.2335 0.413
0 0 0 0	0.5496 0.6535 0.5846 0.4153 0.8188
0 0 0 0	1.84 -0.3503 -1.3548 -2.2491 -2.6649
0 0 0 0	-2.1767 -0.9552 0.3332 0.2124 -0.2421
0 0 0 0	-0.4539 0.003 0.1819 0.2688
	!90
!-90	-0.9566 -2.0997 -2.8554 -3.2538 -2.6412
0 0 0 0	-0.4196 -0.5913 -0.7408 -0.5754 0.383
0 0 0 0	-0.9664 -1.3135 -2.242 -3.3331 -3.57
0 0 0 0	-2.979 -1.8328 -1.1421 -1.4248 -1.8437
0 0 0 0	-1.4847 -0.9234 -0.4178 -0.274
	!105
!-75	
0 0 0 0	
0 0 0 0	
0 0 0 0	

Appendix Transformed Parameter Set for Simulating D-amino Residues

-2.1005 -3.3716 -4.2394 -4.6674 -4.2193
-4.0743 -3.7697 -3.6058 -3.3467 -2.6599
-2.4919 -2.7954 -3.8852 -4.8409 -5.0203
-4.0481 -3.0498 -2.8306 -3.1541 -3.3158
-2.8444 -2.0042 -1.4466 -1.3298

!120

-3.1676 -4.391 -5.3828 -5.4744 -4.6216
-4.162 -4.2625 -4.2134 -3.8659 -3.8321
-3.7845 -4.1609 -5.1776 -5.9866 -6.1474
-5.3773 -5.2395 -4.9503 -4.5786 -4.3332
-3.5537 -2.8511 -2.2237 -2.2061

!135

-3.3379 -4.7004 -5.6871 -5.9738 -5.3341
-4.5268 -4.0969 -3.7688 -3.9502 -4.0071
-4.1585 -4.5816 -5.5755 -6.2033 -6.4149
-5.6322 -5.7983 -5.4609 -5.02 -4.6361
-3.8824 -2.9022 -2.0201 -2.1283

!150

-3.4173 -4.8039 -6.1861 -6.0528 -5.2488
-4.3562 -4.1382 -4.3438 -4.4405 -4.5079
-4.8509 -5.3335 -6.1431 -6.5422 -6.4883
-6.1682 -5.9311 -5.3692 -5.0156 -4.6162
-3.6761 -2.5512 -2.2652 -2.2375

!165

-4.4889 -5.8441 -7.4916 -6.9351 -6.1318
-5.1884 -4.8995 -5.579 -5.9051 -5.8381
-6.2495 -6.6819 -7.3944 -7.6098 -7.595
-7.1782 -6.6249 -5.7832 -5.4179 -4.9236
-3.8372 -3.0434 -2.9035 -3.2103

ID-, GLY, L-ALA

DC NH1 CT2 C NH1 CT2 C NH1 24

!-180

-0.549160 -0.535500 -0.588110 -0.754620 -0.679290
-0.038150 0.298460 0.326040 -0.375610 -1.704360
-3.061280 -3.956460 -3.576280 -1.038930 2.012450
-1.714610 -0.377660 0.317310 0.294580 -0.042920
-0.676620 -0.744600 -0.586590 -0.554770

!-165

-0.709450 -0.896700 -0.990850 -1.319240 -0.940260
-0.126160 0.329180 0.258100 -0.534910 -1.715700
-2.780320 -3.153350 -1.636020 1.822690 -2.675640
-1.810120 -0.410680 0.180860 0.196710 -0.000430
-0.271890 -0.462500 -0.348750 -0.477660

!-150

-1.224850 -1.482430 -1.665900 -1.656770 -1.119780
-1.642540 -0.054220 -0.290670 -0.887080 -1.626260
-2.165440 -1.546500 0.753400 -2.949180 -2.225630
-1.664160 -0.628990 0.000490 0.033160 -0.092820
-0.339050 -0.563330 -0.794980 -0.710760

!-135

-1.787640 -2.117750 -2.143020 -1.803720 -1.567160
-0.886880 -0.801350 -0.851590 -1.020630 -1.337360
-1.062570 0.338010 -4.372310 -2.435890 -2.220710
-1.718060 -0.758950 -0.207560 0.100910 -0.055650
-0.288370 -0.880610 -1.267450 -1.465530

!-120

-2.348270 -2.593790 -2.596140 -2.364070 -1.970070
-1.705860 -1.435540 -1.289220 -1.358170 -0.975570
-3.514390 -4.283210 -3.975820 -3.215190 -2.394430
-1.455320 -0.553910 -0.158900 -0.173830 -0.297950
-0.661220 -1.068330 -1.601800 -1.914850

!-105

-2.788800 -3.079570 -3.178150 -3.013710 -2.626630
-2.266680 -1.951490 -1.681850 -1.195390 -2.567680
-3.632800 -4.748210 -4.662850 -4.255190 -2.776760
-1.695490 -0.893140 -0.633810 -0.467320 -0.540540
-0.950190 -1.401500 -1.959970 -2.412680

!-90

-3.857170 -3.713610 -3.902110 -3.611370 -3.040850

-2.406460 -1.975250 -1.452040 -0.971860 -2.808170
-4.181160 -4.981430 -5.446890 -4.359900 -2.864390
-1.898510 -1.139090 -0.971340 -1.065550 -1.020680
-1.141350 -1.794480 -2.420970 -2.939990

!-75

-4.987770 -4.995210 -4.485310 -3.892550 -3.228630
-2.345360 -1.664160 -1.105500 -1.945510 -3.715530
-4.492140 -5.536170 -5.708500 -3.675410 -2.986660
-1.859410 -0.756620 -1.269930 -1.312730 -1.607440
-1.892510 -2.659400 -3.347950 -3.970600

!-60

-6.183650 -5.456080 -4.878940 -4.000820 -2.683230
-2.067520 -1.094850 -1.119790 -2.962910 -3.687830
-4.993340 -4.666260 -3.796280 -3.374140 -2.495430
-1.453990 -0.877560 -1.002930 -1.337310 -2.431360
-2.948140 -4.008100 -4.821040 -5.565810

!-45

-6.755760 -5.850030 -4.362190 -2.714090 -1.708710
-0.526660 -0.536700 -2.037170 -3.892650 -4.558570
-4.237410 -3.735160 -3.688580 -3.009910 -2.112940
-1.455400 -0.925490 -1.121840 -1.561900 -2.751370
-4.094860 -5.207530 -6.128530 -6.613030

!-30

-5.716250 -4.434060 -2.788600 -0.974400 -0.729200
-0.904940 -1.833540 -3.017700 -3.313450 -3.336010
-3.181640 -3.594720 -1.231370 -0.603790 0.128810
-1.222610 -0.909150 -0.837700 -1.346820 -3.040880
-4.731110 -5.844860 -6.428460 -6.424880

!-15

-3.991110 -2.046000 0.082550 -2.676110 -2.828500
-2.596640 -2.843330 -3.011480 -2.312640 -2.405980
-3.086210 -1.164620 -1.231660 -0.871900 -0.348980
-1.735900 -0.914150 -0.484520 -1.818040 -3.602550
-5.330320 -5.992270 -5.588080 -5.408360

!0

-1.147060 -3.317730 -4.305100 -4.615200 -4.533780
-3.622950 -2.832800 -1.872810 -1.144300 -1.994070
-0.741980 -1.115010 -1.229250 -1.103680 -0.742430
-1.973970 -1.070020 -1.802220 -2.712770 -3.624130
-4.537100 -4.619970 -4.310890 -3.318290

!15

-3.997710 -5.408360 -5.588080 -5.992270 -5.330320
-3.602550 -1.818040 -0.484520 -0.914150 -1.735900
-0.348980 -0.871900 -1.231660 -1.164620 -3.086210
-2.405980 -2.312640 -3.011480 -2.843330 -2.596640
-2.828500 -2.676110 0.082550 -2.046000

!30

-5.710850 -6.424880 -6.428460 -5.844860 -4.731110
-3.040880 -1.346820 -0.837700 -0.909150 -1.222610
0.128810 -0.603790 -1.231370 -3.594720 -3.181640
-3.336010 -3.313450 -3.017700 -1.833540 -0.904940
-0.729200 -0.974400 -2.788600 -4.434060

!45

-6.754940 -6.613030 -6.128530 -5.207530 -4.094860
-2.751370 -1.561900 -1.121840 -0.925490 -1.455400
-2.112940 -3.009910 -3.688580 -3.735160 -4.237410
-4.558570 -3.892650 -2.037170 -0.536700 -0.526660
-1.708710 -2.714090 -4.362190 -5.850030

!60

-6.188070 -5.565810 -4.821040 -4.008100 -2.948140
-2.431360 -1.337310 -1.002930 -0.877560 -1.453990
-2.495430 -3.374140 -3.796280 -4.666260 -4.993340
-3.687830 -2.962910 -1.119790 -1.094850 -2.067520
-2.683230 -4.000820 -4.878940 -5.456080

!75

-4.986080 -3.970600 -3.347950 -2.659400 -1.892510
-1.607440 -1.312730 -1.269930 -0.756620 -1.859410
-2.986660 -3.675410 -5.708500 -5.536170 -4.492140
-3.715530 -1.945510 -1.105500 -1.664160 -2.345360
-3.228630 -3.892550 -4.485310 -4.995210

Appendix Transformed Parameter Set for Simulating D-amino Residues

!90
-3.879190 -2.939990 -2.420970 -1.794480 -1.141350
-1.020680 -1.065550 -0.971340 -1.139090 -1.898510
-2.864390 -4.359900 -5.446890 -4.981430 -4.181160
-2.808170 -0.971860 -1.452040 -1.975250 -2.406460
-3.040850 -3.611370 -3.902110 -3.713610

!105
-2.793280 -2.412680 -1.959970 -1.401500 -0.950190
-0.540540 -0.467320 -0.633810 -0.893140 -1.695490
-2.776760 -4.255190 -4.662850 -4.448210 -3.332800
-2.567680 -1.195390 -1.681850 -1.951490 -2.266680
-2.626630 -3.013710 -3.178150 -3.079570

!120
-2.330190 -1.914850 -1.601800 -1.068330 -0.661220
-0.297950 -0.173830 -0.158900 -0.553910 -1.455320
-2.394430 -3.215190 -3.975820 -3.783210 -3.014390
-0.975570 -1.358170 -1.289220 -1.435540 -1.705860
-1.970070 -2.364070 -2.596140 -2.593790

!135
-1.796120 -1.465530 -1.267450 -0.880610 -0.288370
-0.055650 0.100910 -0.207560 -0.758950 -1.718060
-2.220710 -2.435890 -4.372310 0.338010 -1.062570
-1.337360 -1.020630 -0.851590 -0.801350 -0.886880
-1.567160 -1.803720 -2.143020 -2.117750

!150
-1.263610 -0.710760 -0.794980 -0.563330 -0.339050
-0.092820 0.033160 0.000490 -0.628990 -1.664160
-2.225630 -2.949180 0.753400 -1.546500 -2.165440
-1.626260 -0.887080 -0.290670 -0.054220 -1.642540
-1.119780 -1.656770 -1.665900 -1.482430

!165
-0.684660 -0.477660 -0.348750 -0.462500 -0.271890
-0.000430 0.196710 0.180860 -0.410680 -1.810120
-2.675640 1.822690 -1.636020 -3.153350 -2.780320
-1.715700 -0.534910 0.258100 0.329180 -0.126160
-0.940260 -1.319240 -0.990850 -0.896700

!L-, GLY, D-PRO
C NH1 CT2 C NH1 CT2 C DN 24

!-180
-0.549160 -0.535500 -0.588110 -0.754620 -0.679290
-0.038150 0.298460 0.326040 -0.375610 -1.704360
-3.061280 -3.956460 -3.576280 -1.038930 2.012450
-1.714610 -0.377660 0.317310 0.294580 -0.042920
-0.676620 -0.744600 -0.586590 -0.554770

!-165
-0.709450 -0.896700 -0.990850 -1.319240 -0.940260
-0.126160 0.329180 0.258100 -0.534910 -1.715700
-2.780320 -3.153350 -1.636020 1.822690 -2.675640
-1.810120 -0.410680 0.180860 0.196710 -0.000430
-0.271890 -0.462500 -0.348750 -0.477660

!-150
-1.224850 -1.482430 -1.665900 -1.656770 -1.119780
-1.642540 -0.054220 -0.290670 -0.887080 -1.626260
-2.165440 -1.546500 0.753400 -2.949180 -2.225630
-1.664160 -0.628990 0.000490 0.033160 -0.092820
-0.339050 -0.563330 -0.794980 -0.710760

!-135
-1.787640 -2.117750 -2.143020 -1.803720 -1.567160
-0.886880 -0.801350 -0.851590 -1.020630 -1.337360
-1.062570 0.338010 -4.372310 -2.435890 -2.220710
-1.718060 -0.758950 -0.207560 0.100910 -0.055650
-0.288370 -0.880610 -1.267450 -1.465530

!-120
-2.348270 -2.593790 -2.596140 -2.364070 -1.970070
-1.705860 -1.435540 -1.289220 -1.358170 -0.975570
-3.514390 -4.283210 -3.975820 -3.215190 -2.394430
-1.455320 -0.553910 -0.158900 -0.173830 -0.297950
-0.661220 -1.068330 -1.601800 -1.914850

!-105
-2.788800 -3.079570 -3.178150 -3.013710 -2.626630
-2.266680 -1.951490 -1.681850 -1.195390 -2.567680
-3.632800 -4.748210 -4.662850 -4.255190 -2.776760
-1.695490 -0.893140 -0.633810 -0.467320 -0.540540
-0.950190 -1.401500 -1.959970 -2.412680

!-90
-3.857170 -3.713610 -3.902110 -3.611370 -3.040850
-2.406460 -1.975250 -1.452040 -0.971860 -2.808170
-4.181160 -4.981430 -5.446890 -4.359900 -2.864390
-1.898510 -1.139090 -0.971340 -1.065550 -1.020680
-1.141350 -1.794480 -2.420970 -2.939990

!-75
-4.987770 -4.995210 -4.485310 -3.892550 -3.228630
-2.345360 -1.664160 -1.105500 -1.945510 -3.715530
-4.492140 -5.536170 -5.708500 -3.675410 -2.986660
-1.859410 -0.756620 -1.269930 -1.312730 -1.607440
-1.892510 -2.659400 -3.347950 -3.970600

!-60
-6.183650 -5.456080 -4.878940 -4.000820 -2.683230
-2.067520 -1.094850 -1.119790 -2.962910 -3.687830
-4.993340 -4.666260 -3.796280 -3.374140 -2.495430
-1.453990 -0.877560 -1.002930 -1.337310 -2.431360
-2.948140 -4.008100 -4.821040 -5.565810

!-45
-6.755760 -5.850030 -4.362190 -2.714090 -1.708710
-0.526660 -0.536700 -2.037170 -3.892650 -4.558570
-4.237410 -3.735160 -3.688580 -3.009910 -2.112940
-1.455400 -0.925490 -1.121840 -1.561900 -2.751370
-4.094860 -5.207530 -6.128530 -6.613030

!-30
-5.716250 -4.434060 -2.788600 -0.974400 -0.729200
-0.904940 -1.833540 -3.017700 -3.313450 -3.336010
-3.181640 -3.594720 -1.231370 -0.603790 0.128810
-1.222610 -0.909150 -0.837700 -1.346820 -3.040880
-4.731110 -5.844860 -6.428460 -6.424880

!-15
-3.991110 -2.046000 0.082550 -2.676110 -2.828500
-2.596640 -2.843330 -3.011480 -2.312640 -2.405980
-3.086210 -1.164620 -1.231660 -0.871900 -0.348980
-1.735900 -0.914150 -0.484520 -1.818040 -3.602550
-5.330320 -5.992270 -5.588080 -5.408360

!0
-1.147060 -3.317730 -4.305100 -4.615200 -4.533780
-3.622950 -2.832800 -1.872810 -1.144300 -1.994070
-0.741980 -1.115010 -1.229250 -1.103680 -0.742430
-1.973970 -1.070020 -1.802220 -2.712770 -3.624130
-4.537100 -4.619970 -4.310890 -3.318290

!15
-3.997710 -5.408360 -5.588080 -5.992270 -5.330320
-3.602550 -1.818040 -0.484520 -0.914150 -1.735900
-0.348980 -0.871900 -1.231660 -1.164620 -3.086210
-2.405980 -2.312640 -3.011480 -2.843330 -2.596640
-2.828500 -2.676110 0.082550 -2.046000

!30
-5.710850 -6.424880 -6.428460 -5.844860 -4.731110
-3.040880 -1.346820 -0.837700 -0.909150 -1.222610
0.128810 -0.603790 -1.231370 -3.594720 -3.181640
-3.336010 -3.313450 -3.017700 -1.833540 -0.904940
-0.729200 -0.974400 -2.788600 -4.434060

!45
-6.754940 -6.613030 -6.128530 -5.207530 -4.094860
-2.751370 -1.561900 -1.121840 -0.925490 -1.455400
-2.112940 -3.009910 -3.688580 -3.735160 -4.237410
-4.558570 -3.892650 -2.037170 -0.536700 -0.526660
-1.708710 -2.714090 -4.362190 -5.850030

!60
-6.188070 -5.565810 -4.821040 -4.008100 -2.948140
-2.431360 -1.337310 -1.002930 -0.877560 -1.453990
-2.495430 -3.374140 -3.796280 -4.666260 -4.993340

Appendix Transformed Parameter Set for Simulating D-amino Residues

-3.687830 -2.962910 -1.119790 -1.094850 -2.067520
-2.683230 -4.000820 -4.878940 -5.456080

!75
-4.986080 -3.970600 -3.347950 -2.659400 -1.892510
-1.607440 -1.312730 -1.269930 -0.756620 -1.859410
-2.986660 -3.675410 -5.708500 -5.536170 -4.492140
-3.715530 -1.945510 -1.105500 -1.664160 -2.345360
-3.228630 -3.892550 -4.485310 -4.995210

!90
-3.879190 -2.939990 -2.420970 -1.794480 -1.141350
-1.020680 -1.065550 -0.971340 -1.139090 -1.898510
-2.864390 -4.359900 -5.446890 -4.981430 -4.181160
-2.808170 -0.971860 -1.452040 -1.975250 -2.406460
-3.040850 -3.611370 -3.902110 -3.713610

!105
-2.793280 -2.412680 -1.959970 -1.401500 -0.950190
-0.540540 -0.467320 -0.633810 -0.893140 -1.695490
-2.776760 -4.255190 -4.662850 -4.448210 -3.332800
-2.567680 -1.195390 -1.681850 -1.951490 -2.266680
-2.626630 -3.013710 -3.178150 -3.079570

!120
-2.330190 -1.914850 -1.601800 -1.068330 -0.661220
-0.297950 -0.173830 -0.158900 -0.553910 -1.455320
-2.394430 -3.215190 -3.975820 -3.783210 -3.014390
-0.975570 -1.358170 -1.289220 -1.435540 -1.705860
-1.970070 -2.364070 -2.596140 -2.593790

!135
-1.796120 -1.465530 -1.267450 -0.880610 -0.288370
-0.055650 0.100910 -0.207560 -0.758950 -1.718060
-2.220710 -2.435890 -4.372310 0.338010 -1.062570
-1.337360 -1.020630 -0.851590 -0.801350 -0.886880
-1.567160 -1.803720 -2.143020 -2.117750

!150
-1.263610 -0.710760 -0.794980 -0.563330 -0.339050
-0.092820 0.033160 0.000490 -0.628990 -1.664160
-2.225630 -2.949180 0.753400 -1.546500 -2.165440
-1.626260 -0.887080 -0.290670 -0.054220 -1.642540
-1.119780 -1.656770 -1.665900 -1.482430

!165
-0.684660 -0.477660 -0.348750 -0.462500 -0.271890
-0.000430 0.196710 0.180860 -0.410680 -1.810120
-2.675640 1.822690 -1.636020 -3.153350 -2.780320
-1.715700 -0.534910 0.258100 0.329180 -0.126160
-0.940260 -1.319240 -0.990850 -0.896700

!L-, GLY, D-ALA
C NH1 CT2 C NH1 CT2 C DNH1 24

!-180
-0.549160 -0.535500 -0.588110 -0.754620 -0.679290
-0.038150 0.298460 0.326040 -0.375610 -1.704360
-3.061280 -3.956460 -3.576280 -1.038930 2.012450
-1.714610 -0.377660 0.317310 0.294580 -0.042920
-0.676620 -0.744600 -0.586590 -0.554770

!-165
-0.709450 -0.896700 -0.990850 -1.319240 -0.940260
-0.126160 0.329180 0.258100 -0.534910 -1.715700
-2.780320 -3.153350 -1.636020 1.822690 -2.675640
-1.810120 -0.410680 0.180860 0.196710 -0.000430
-0.271890 -0.462500 -0.348750 -0.477660

!-150
-1.224850 -1.482430 -1.665900 -1.656770 -1.119780
-1.642540 -0.054220 -0.290670 -0.887080 -1.626260
-2.165440 -1.546500 0.753400 -2.949180 -2.225630
-1.664160 -0.628990 0.000490 0.033160 -0.092820
-0.339050 -0.563330 -0.794980 -0.710760

!-135
-1.787640 -2.117750 -2.143020 -1.803720 -1.567160
-0.886880 -0.801350 -0.851590 -1.020630 -1.337360
-1.062570 0.338010 -4.372310 -2.435890 -2.220710
-1.718060 -0.758950 -0.207560 0.100910 -0.055650

-0.288370 -0.880610 -1.267450 -1.465530

!-120
-2.348270 -2.593790 -2.596140 -2.364070 -1.970070
-1.705860 -1.435540 -1.289220 -1.358170 -0.975570
-3.514390 -4.283210 -3.975820 -3.215190 -2.394430
-1.455320 -0.553910 -0.158900 -0.173830 -0.297950
-0.661220 -1.068330 -1.601800 -1.914850

!-105
-2.788800 -3.079570 -3.178150 -3.013710 -2.626630
-2.266680 -1.951490 -1.681850 -1.195390 -2.567680
-3.632800 -4.748210 -4.662850 -4.255190 -2.776760
-1.695490 -0.893140 -0.633810 -0.467320 -0.540540
-0.950190 -1.401500 -1.959970 -2.412680

!-90
-3.857170 -3.713610 -3.902110 -3.611370 -3.040850
-2.406460 -1.975250 -1.452040 -0.971860 -2.808170
-4.181160 -4.981430 -5.446890 -4.359900 -2.864390
-1.898510 -1.139090 -0.971340 -1.065550 -1.020680
-1.141350 -1.794480 -2.420970 -2.939990

!-75
-4.987770 -4.995210 -4.485310 -3.892550 -3.228630
-2.345360 -1.664160 -1.105500 -1.945510 -3.715530
-4.492140 -5.536170 -5.708500 -3.675410 -2.986660
-1.859410 -0.756620 -1.269930 -1.312730 -1.607440
-1.892510 -2.659400 -3.347950 -3.970600

!-60
-6.183650 -5.456080 -4.878940 -4.000820 -2.683230
-2.067520 -1.094850 -1.119790 -2.962910 -3.687830
-4.993340 -4.666260 -3.796280 -3.675410 -2.495430
-1.453990 -0.877560 -1.002930 -1.337310 -2.431360
-2.948140 -4.008100 -4.821040 -5.565810

!-45
-6.755760 -5.850030 -4.362190 -2.714090 -1.708710
-0.526660 -0.536700 -2.037170 -3.892650 -4.558570
-4.237410 -3.735160 -3.688580 -3.009910 -2.112940
-1.455400 -0.925490 -1.121840 -1.561900 -2.751370
-4.094860 -5.207530 -6.128530 -6.613030

!-30
-5.716250 -4.434060 -2.788600 -0.974400 -0.729200
-0.904940 -1.833540 -3.017700 -3.313450 -3.336010
-3.181640 -3.594720 -1.231370 -0.603790 0.128810
-1.222610 -0.909150 -0.837700 -1.346820 -3.040880
-4.731110 -5.844860 -6.428460 -6.424880

!-15
-3.991110 -2.046000 0.082550 -2.676110 -2.828500
-2.596640 -2.843330 -3.011480 -2.312640 -2.405980
-3.086210 -1.164620 -1.231660 -0.871900 -0.348980
-1.735900 -0.914150 -0.484520 -1.818040 -3.602550
-5.330320 -5.992270 -5.588080 -5.408360

!0
-1.147060 -3.317730 -4.305100 -4.615200 -4.533780
-3.622950 -2.832800 -1.872810 -1.144300 -1.994070
-0.741980 -1.115010 -1.229250 -1.103680 -0.742430
-1.973970 -1.070020 -1.802220 -2.712770 -3.624130
-4.537100 -4.619970 -4.310890 -3.318290

!15
-3.997710 -5.408360 -5.588080 -5.992270 -5.330320
-3.602550 -1.818040 -0.484520 -0.914150 -1.735900
-0.348980 -0.871900 -1.231660 -1.164620 -3.086210
-2.405980 -2.312640 -3.011480 -2.843330 -2.596640
-2.828500 -2.676110 0.082550 -2.046000

!30
-5.710850 -6.424880 -6.428460 -5.844860 -4.731110
-3.040880 -1.346820 -0.837700 -0.909150 -1.222610
0.128810 -0.603790 -1.231370 -3.594720 -3.181640
-3.336010 -3.313450 -3.017700 -1.833540 -0.904940
-0.729200 -0.974400 -2.788600 -4.434060

!45
-6.754940 -6.613030 -6.128530 -5.207530 -4.094860

Appendix Transformed Parameter Set for Simulating D-amino Residues

-2.751370 -1.561900 -1.121840 -0.925490 -1.455400	0 0 0 0
-2.112940 -3.009910 -3.688580 -3.735160 -4.237410	0 0 0 0
-4.558570 -3.892650 -2.037170 -0.536700 -0.526660	0 0 0 0
-1.708710 -2.714090 -4.362190 -5.850030	
!60	!-135
-6.188070 -5.565810 -4.821040 -4.008100 -2.948140	0 0 0 0
-2.431360 -1.337310 -1.002930 -0.877560 -1.453990	0 0 0 0
-2.495430 -3.374140 -3.796280 -4.666260 -4.993340	0 0 0 0
-3.687830 -2.962910 -1.119790 -1.094850 -2.067520	0 0 0 0
-2.683230 -4.000820 -4.878940 -5.456080	
!75	!-120
-4.986080 -3.970600 -3.347950 -2.659400 -1.892510	0 0 0 0
-1.607440 -1.312730 -1.269930 -0.756620 -1.859410	0 0 0 0
-2.986660 -3.675410 -5.708500 -5.536170 -4.492140	0 0 0 0
-3.715530 -1.945510 -1.105500 -1.664160 -2.345360	0 0 0 0
-3.228630 -3.892550 -4.485310 -4.995210	
!90	!-105
-3.879190 -2.939990 -2.420970 -1.794480 -1.141350	0 0 0 0
-1.020680 -1.065550 -0.971340 -1.139090 -1.898510	0 0 0 0
-2.864390 -4.359900 -5.446890 -4.981430 -4.181160	0 0 0 0
-2.808170 -0.971860 -1.452040 -1.975250 -2.406460	0 0 0 0
-3.040850 -3.611370 -3.902110 -3.713610	
!105	!-90
-2.793280 -2.412680 -1.959970 -1.401500 -0.950190	0 0 0 0
-0.540540 -0.467320 -0.633810 -0.893140 -1.695490	0 0 0 0
-2.776760 -4.255190 -4.662850 -4.448210 -3.332800	0 0 0 0
-2.567680 -1.195390 -1.681850 -1.951490 -2.266680	0 0 0 0
-2.626630 -3.013710 -3.178150 -3.079570	
!120	!-75
-2.330190 -1.914850 -1.601800 -1.068330 -0.661220	0 0 0 0
-0.297950 -0.173830 -0.158900 -0.553910 -1.455320	0 0 0 0
-2.394430 -3.215190 -3.975820 -3.783210 -3.014390	0 0 0 0
-0.975570 -1.358170 -1.289220 -1.435540 -1.705860	0 0 0 0
-1.970070 -2.364070 -2.596140 -2.593790	
!135	!-60
-1.796120 -1.465530 -1.267450 -0.880610 -0.288370	0 0 0 0
-0.055650 0.100910 -0.207560 -0.758950 -1.718060	0 0 0 0
-2.220710 -2.435890 -4.372310 0.338010 -1.062570	0 0 0 0
-1.337360 -1.020630 -0.851590 -0.801350 -0.886880	0 0 0 0
-1.567160 -1.803720 -2.143020 -2.117750	
!150	!-45
-1.263610 -0.710760 -0.794980 -0.563330 -0.339050	0 0 0 0
-0.092820 0.033160 0.000490 -0.628990 -1.664160	0 0 0 0
-2.225630 -2.949180 0.753400 -1.546500 -2.165440	0 0 0 0
-1.626260 -0.887080 -0.290670 -0.054220 -1.642540	0 0 0 0
-1.119780 -1.656770 -1.665900 -1.482430	
!165	!-30
-0.684660 -0.477660 -0.348750 -0.462500 -0.271890	-2.3892 -5.102 -7.102 -6.8822 -6.8375
-0.000430 0.196710 0.180860 -0.410680 -1.810120	-7.2198 -7.8066 -7.664 -6.6643 -3.9961
-2.675640 1.822690 -1.636020 -3.153350 -2.780320	-3.814 -3.8999 -1.7562 -4.2748 -4.4979
-1.715700 -0.534910 0.258100 0.329180 -0.126160	-4.3419 -6.3504 -6.2626 -7.4995 -8.4761
-0.940260 -1.319240 -0.990850 -0.896700	-7.9518 -6.6493 -4.8772 -2.9624
!L-, D-PRO, L-ALA	!-15
C DN DCP1 DC DN DCP1 DC NH1 24	-5.1926 -7.156 -7.2901 -7.8816 -7.8776
!-180	-7.9583 -7.5691 -6.7542 -4.0726 -4.4798
-6.0057 -7.9186 -9.6188 -9.3338 -8.3158	-4.4528 -4.3772 -4.4364 -5.3679 -5.6538
-7.7395 -7.5633 -7.9619 -7.9886 -7.9668	-4.7721 -7.0024 -7.6488 -8.3212 -8.2256
-8.1024 -8.6106 -8.8889 -9.0275 -8.9783	-7.2676 -6.1779 -4.6683 -3.1742
-9.1251 -8.4192 -7.6332 -7.1853 -6.6546	
-5.4921 -4.5142 -4.2892 -4.6066	!0
!-165	-5.1204 -7.1012 -8.2774 -8.6095 -8.0409
0 0 0 0	-7.0272 -5.7891 -3.0347 -3.5574 -3.9199
0 0 0 0	-3.646 -3.7874 -4.2891 -5.0507 -4.984
0 0 0 0	-7.5416 -7.8122 -7.7752 -7.0542 -6.7169
0 0 0 0	-5.9161 -5.2696 -3.8306 -2.7319
!-150	!15
0 0 0 0	-6.4969 -8.6267 -9.8071 -9.8633 -8.1653
0 0 0 0	-6.4579 -2.725 -5.7532 -1.4306 -1.4593
0 0 0 0	-1.484 -1.7812 -2.6863 -3.4786 -6.5645
0 0 0 0	-7.3081 -7.0445 -5.7853 -4.6496 -4.5078
	-3.8047 -0.8212 -1.1141 -2.7055

Appendix Transformed Parameter Set for Simulating D-amino Residues

```
!30
-5.4942 -7.1248 -7.5206 -6.8194 -5.1628
-3.1495 -0.1462 -0.2845 -0.4102 -0.5004
-0.6017 -0.8166 -1.2421 -4.7865 -6.0834
-6.3548 -5.0805 -3.626 -2.5999 -1.8514
-0.7044 -0.8426 -2.0668 -3.463

!45
-3.4792 -4.4995 -4.8337 -4.1833 -2.7114
0.9667 1.2811 1.0055 0.8259 0.6445
0.8087 0.9605 -1.9933 -3.5123 -4.4661
-3.9827 -2.3982 -0.5975 0.0065 0.205
-0.0683 -0.8596 -1.3074 -2.4101

!60
-1.4684 -2.0111 -2.5938 -2.3336 0.9932
1.1616 1.4931 1.3072 1.1635 1.106
1.6319 -0.2471 -1.3394 -2.555 -2.8353
-2.1019 -0.683 0.7254 1.0599 0.5337
-0.2049 -0.2521 -0.4073 -0.4882

!75
-0.3834 -1.549 -2.1349 0.2335 0.413
0.5496 0.6535 0.5846 0.4153 0.8188
1.84 -0.3503 -1.3548 -2.2491 -2.6649
-2.1767 -0.9552 0.3332 0.2124 -0.2421
-0.4539 0.003 0.1819 0.2688

!90
-0.9566 -2.0997 -2.8554 -3.2538 -2.6412
-0.4196 -0.5913 -0.7408 -0.5754 0.383
-0.9664 -1.3135 -2.242 -3.3331 -3.57
-2.979 -1.8328 -1.1421 -1.4248 -1.8437
-1.4847 -0.9234 -0.4178 -0.274

!105
-2.1005 -3.3716 -4.2394 -4.6674 -4.2193
-4.0743 -3.7697 -3.6058 -3.3467 -2.6599
-2.4919 -2.7954 -3.8852 -4.8409 -5.0203
-4.0481 -3.0498 -2.8306 -3.1541 -3.3158
-2.8444 -2.0042 -1.4466 -1.3298

!120
-3.1676 -4.391 -5.3828 -5.4744 -4.6216
-4.162 -4.2625 -4.2134 -3.8659 -3.8321
-3.7845 -4.1609 -5.1776 -5.9866 -6.1474
-5.3773 -5.2395 -4.9503 -4.5786 -4.3332
-3.5537 -2.8511 -2.2237 -2.2061

!135
-3.3379 -4.7004 -5.6871 -5.9738 -5.3341
-4.5268 -4.0969 -3.7688 -3.9502 -4.0071
-4.1585 -4.5816 -5.5755 -6.2033 -6.4149
-5.6322 -5.7983 -5.4609 -5.02 -4.6361
-3.8824 -2.9022 -2.0201 -2.1283

!150
-3.4173 -4.8039 -6.1861 -6.0528 -5.2488
-4.3562 -4.1382 -4.3438 -4.4405 -4.5079
-4.8509 -5.3335 -6.1431 -6.5422 -6.4883
-6.1682 -5.9311 -5.3692 -5.0156 -4.6162
-3.6761 -2.5512 -2.2652 -2.2375

!165
-4.4889 -5.8441 -7.4916 -6.9351 -6.1318
-5.1884 -4.8995 -5.579 -5.9051 -5.8381
-6.2495 -6.6819 -7.3944 -7.6098 -7.595
-7.1782 -6.6249 -5.7832 -5.4179 -4.9236
-3.8372 -3.0434 -2.9035 -3.2103

NONBONDED nbxmod 5 atom cdieI shift vatom vdistance
vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin
1.5
!adm jr., 5/08/91, suggested cut-off scheme

DC 0.000000 -0.110000 2.000000 ! ALLOW PEP POL
ARO
! NMA pure solvent, adm jr., 3/3/93

DCP1 0.000000 -0.020000 2.275000 0.000000 -
0.010000 1.900000 ! ALLOW ALI
! alkane update, adm jr., 3/2/92
DCT1 0.000000 -0.020000 2.275000 0.000000 -
0.010000 1.900000 ! ALLOW ALI
! isobutane pure solvent properties, adm jr, 2/3/92
DNH1 0.000000 -0.200000 1.850000 0.000000 -
0.200000 1.550000 ! ALLOW PEP POL ARO
! This 1,4 vdW allows the C5 dipeptide minimum to
exist.(LK)
DN 0.000000 -0.200000 1.850000 0.000000 -0.000100
1.850000 ! ALLOW PRO
! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g
AcProNHCH3 RLD 4/23/93

END
```