

Structural Investigations of Peptide Folding and Unfolding in Solution



Nicola Steinke

Department of Biochemistry

Corpus Christi College, University of Oxford

This thesis was submitted to the Board of the Medical Sciences Division
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Michaelmas Term 2017

Abstract

Proteins are the essential biological working molecules in all living beings. Their function depends on their ability to adopt a specific three-dimensional structure in aqueous solution. The mechanism of this folding process, and conversely the loss of protein structure, unfolding, is still not well understood, especially on the atomic scale where these interactions occur *in vivo*. In this thesis, the structures of small model peptides have been investigated in aqueous solutions in order to understand how water, urea and other co-solutes influence protein folding and unfolding. Using a combination of neutron diffraction enhanced by isotopic substitution and computer simulations, it was found that urea, generally at any probed concentration, decreases ionic interactions with the peptides in solution. When the peptide backbone is solvent accessible, urea interacts directly with the polar peptide backbone and replaces water molecules at higher urea concentrations. Further, the results indicate an ambiguous influence of water on peptide folding *vs.* conformational stability. The formation of β -turn like structures, and possibly intra-peptide hydrogen bonds in general, appears to be dependent on a specific backbone hydration pattern, facilitated by the local amino acid sequence. Conversely, hydrophobic residues in folded protein structures are particularly sensitive, in terms of conformation and hydration, towards an increased peptide backbone hydration.

Publications

Below is a list of publications arising from the work presented in this thesis and associated collaborations.

On the interplay between hydration and conformation in peptide folding

Thomas P. E. Dixon, **N. Steinke**, C. D. Lorenz and S. E. McLain; *Phys. Chem. Chem. Phys.*, (submitted)

On the hydration structure of the pro-drug GPG-NH₂ and its derivatives

Paul Smith, **N. Steinke**, A. Galina, J.F.C. Turner, S. E. McLain and C. D. Lorenz; *J. Chem. Phys.*, (submitted)

Insights into the interplay between peptide conformation and peptide hydration

N. Steinke, R. J. Gillams, A. Galina, C. D. Lorenz and S. E. McLain; *J. Am. Chem. Soc.*, (To be resubmitted)

On the structure of prilocaine in aqueous and amphiphilic solutions

A. Silva-Santisteban, **N. Steinke**, A. J. Johnston, G. Ruiz, L. C. Pardo and S. E. McLain; *Phys. Chem. Chem. Phys.*, 2017

Salt Interactions in Solution Prevent Urea from Direct Association with a Peptide Backbone

N. Steinke, A. Galina, C. D. Lorenz and S. E. McLain; *J. Phys. Chem. B*, 2017

Atomic scale insights into urea-peptide interactions in solution

N. Steinke, R. J. Gillams, S. Busch, C. D. Lorenz, S. E. McLain; *Phys. Chem. Chem. Phys.*, 2016

Acknowledgements

I would like to thank everyone in the McLain group for the friendly atmosphere and helping me to settle into my life in Oxford. First, I thank Sylvia McLain for her help and support, the opportunity to pursue my DPhil project in her group and her unequivocally excellent supervision. I thank Sebastian Busch for his help and patience in teaching me. His scripts have been used to create many of the figures in this thesis. Thanks to Richard Gillams who helped me with kind and insightful advice especially with the computer setup and neutron scattering experiments. I thank Natasha Rhys for providing the Latex template for this thesis, her advice and contribution to the KLL-project. I also thank fellow DPhil students Andrew Johnston for helpful discussions and particularly Tom Dixon for his help in writing python scripts. I also thank Samvid Kurlekar for his contribution to my work during his summer project.

Over the course of my doctorate, I have also had the good fortune to work with many people outside the group. Thanks to Sam Callear and Silvia Imberti who have both helped with the neutron scattering experiments. I thank Luis Carlos Pardo for providing the ANGULA programme, adjusting it to my increasingly large peptides, and his enthusiastic help. I also thank Chris Lorenz for his kindness and patience in teaching me MD, the additional MD simulations he performed on GPG and KGXGK and Anna Genina for performing the MD simulations on KGPGK. I thank Christina Redfield for her help using NMR and for being my secondary supervisor.

Finally, I thank my family for supporting me throughout my studies.

Nicola Steinke

Contents

Abstract	iii
Publications	v
Acknowledgements	vii
Abbreviations	xv
1 Introduction	1
1.1 Protein Folding and Unfolding	1
1.1.1 Motivation for Understanding Protein Folding and Unfolding . .	1
1.1.2 Protein Structure	3
1.1.3 Fundamental Thermodynamics and Kinetics of Protein Folding .	8
1.1.4 Non-Covalent Interactions in Biological Molecules	10
1.1.4.1 Coulomb and Lennard-Jones Potential	10
1.1.4.2 The Hydrophobic Effect	13
1.1.5 The Protein Folding Problem	14
1.1.6 Protein Stability and Protein Solute-Induced Unfolding	17
1.2 Aims and Objectives	20
1.3 Structure and Scope of the Thesis	21
2 Methodical Background	23
2.1 The Structure of Biomolecules in the Liquid Phase	23
2.1.1 Local Order in the Liquid Phase	23
2.1.2 Techniques to Determine the Structure of Biomolecules in the Liq-	
uid Phase	25
2.1.2.1 Nuclear Magnetic Resonance	26
2.1.2.2 Circular Dichroism	27
2.1.2.3 Molecular Dynamics	29
2.2 Diffraction	31
2.2.1 Fundamentals	31

2.2.2	Scattering Probes	33
2.2.3	Scattering by a Single Fixed Atom	33
2.2.4	Scattering Cross Section	37
2.2.5	Extracting Structural Information from Scattering Data	38
2.2.6	Small-Angle Scattering	41
2.2.7	Neutron Scattering Experiments	42
2.2.7.1	The SANDALS and NIMROD Instruments	42
2.2.7.2	Measurements	44
2.2.7.3	Data Processing	46
2.3	Empirical Potential Structure Refinement	48
2.3.1	The Reference Potential	49
2.3.2	The Monte Carlo Simulation	51
2.3.3	The Empirical Potential	52
2.3.4	Refining the Empirical Potential	53
2.4	ANGULA	54
2.4.1	Assigning Coordinate Systems	55
2.4.2	Site Specific Analysis	56
2.4.3	Whole Molecule Analysis	59
2.4.4	Relative Molecule Orientations	59
2.4.5	Further Applications	62
3	Influence of Urea on the GPG-NH₂ Peptide	63
3.1	Introduction	63
3.2	Methods	65
3.2.1	NMR Spectroscopy Sample Preparation and Measurement	65
3.2.2	Neutron Diffraction and Sample Preparation	66
3.2.3	Empirical Potential Structure Refinement	67
3.2.4	Molecular Dynamics Simulations	68
3.2.5	ANGULA	69
3.2.6	Preference ratios	70
3.3	Results	71
3.3.1	NMR Spectroscopy	71
3.3.2	Diffraction Data and EPSR Fits	73
3.3.3	Solvent Structure	76
3.3.4	Co-Solute Structure	79

3.3.4.1	Urea-Water Interactions	79
3.3.4.2	Chloride-Water Interactions	80
3.3.4.3	Urea-Chloride Interactions	82
3.3.4.4	Urea-Urea Interactions	82
3.3.5	Peptide Structure and intramolecular distance	83
3.3.6	Peptide Solvation	85
3.3.6.1	Peptide Backbone Oxygen Solvation	85
3.3.6.2	Peptide Backbone Nitrogen Solvation	89
3.3.6.3	Apolar Solvation	97
3.3.7	Whole molecule analysis	99
3.4	Conclusions	103
4	Conformation of KGXGK Peptides and Influence of Urea on the KGPGK Peptide	107
4.1	Introduction	107
4.2	Methods	110
4.2.1	NMR Spectroscopy Sample Preparation and Measurement	110
4.2.2	Neutron and X-ray Diffraction and Sample Preparation	110
4.2.3	Empirical Potential Structure Refinement	112
4.2.4	Molecular Dynamics Simulations	113
4.2.5	ANGULA	113
4.2.5.1	Bridging Water Analysis	116
4.3	Results	117
4.3.1	NMR Spectroscopy and Elemental Analysis	117
4.3.2	Diffraction Data and EPSR Fits	118
4.3.3	Peptide Backbone Interactions	119
4.3.3.1	Peptide Backbone Nitrogen Interactions	119
4.3.3.2	Peptide Backbone Oxygens Interactions	128
4.3.4	Whole Molecule Analysis	130
4.3.5	Solvent and Co-Solute Interactions	139
4.3.5.1	Water-Water Interactions	139
4.3.5.2	TFA-Water Interactions	139
4.3.5.3	TFA-TFA Interactions	141
4.3.6	Conformation Independent Peptide Hydration	141
4.3.7	Peptide Conformation Analysis	145

4.3.7.1	Time Dependent Conformation Analysis	145
4.3.7.2	Dihedral Angles	145
4.3.7.3	Intramolecular distances	149
4.3.8	Conformation Dependent Peptide Hydration	151
4.3.9	Shared nearest neighbour water molecules	153
4.4	Conclusions	155
5	Solvent and Urea Influence on the Amphiphilic KLL-Peptides	161
5.1	Introduction	161
5.2	Methods	163
5.2.1	NMR Spectroscopy Preparation and Measurement	163
5.2.2	CD Spectroscopy Sample Preparation and Measurement	164
5.2.3	Neutron Diffraction and Sample Preparation	164
5.2.4	Empirical Potential Structure Refinement and ANGULA	166
5.3	Results	166
5.3.1	NMR Spectroscopy and Elemental Analysis	166
5.3.2	CD Spectroscopy	168
5.3.3	Diffraction Data and EPSR Fits	171
5.3.4	Small Angle Scattering	174
5.3.5	Peptide Conformations	176
5.3.5.1	Residues most affected by a change in solvation	176
5.3.6	KLL ₁₅ : Amphiphilicity	179
5.3.6.1	Peptide Water Interaction	179
5.3.6.2	Peptide Chloride Interaction	180
5.3.6.3	Peptide Methanol Oxygen Interaction	180
5.3.6.4	Peptide Methanol Carbon Interaction	182
5.3.7	KLL ₂₂ : Urea Influence	184
5.3.7.1	Peptide Water Interaction	184
5.3.7.2	Peptide Chloride Interaction	185
5.3.7.3	Peptide Urea Interaction	185
5.4	Conclusions	188
6	Conclusions and Further Work	193
6.1	Conclusions	193
6.1.1	Urea Influence on Peptides	194
6.1.2	Peptide Folding and Stability	195

6.2	Further Work	196
A	EPSR Input Parameters	199
A.1	Water EPSR Input Parameters	199
A.2	Methanol EPSR Input Parameters	200
A.3	Urea EPSR Input Parameters	201
A.4	TFA EPSR Input Parameters	202
A.5	GPG-NH ₂ EPSR Input Parameters	204
A.6	KGPGK EPSR Input Parameters	208
B	KGPGK Trajectory Analysis	217

Abbreviations

AMBER	Assisted Model Building with Energy Refinement
ATLAS	Analysis of Time-of-Flight Diffraction Data from Liquid and Amorphous Samples
CD	Circular dichroism
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CN	Coordination number
DCS	Differential cross section
DNA	Deoxyribonucleic acid
EPSR	Empirical Potential Structure Refinement
FID	Free induction decay
FT	Fourier transform
HD	Hydration Difference
IDP	Intrinsically disordered proteins
IR	Infrared (spectroscopy)
LJ	Lennard-Jones (potential)
MD	Molecular dynamics
NDIS	Neutron Diffraction Enhanced by Isotopic Substitution
NIMROD	Near and InterMediate Range Order Diffractometer
NMR	Nuclear magnetic resonance
OPLS	Optimized Potentials for Liquid Simulations
(m)RNA	(messenger) Ribonucleic acid
PDB	Protein Data Bank
PR	Preference Ratio
RDF	Radial distribution function
(R)MC	(Reverse) Monte Carlo
SANDALS	Small Angle Neutron Diffractometer for Amorphous and Liquid Samples
SDM	Spatial Density Map
SPC/E	(Extended) Simple point charge
TIP(3P)	Transferable intermolecular potential with 3 points
TMAO	Trimethylamine N-oxide
TOF	Time-of-flight
TSE	Transmissible spongiform encephalopathy
WMA	Whole Molecule Analysis
UniProt	Universal Protein Database

Chapter 1

Introduction

1.1 Protein Folding and Unfolding

1.1.1 Motivation for Understanding Protein Folding and Unfolding

Proteins are the essential working molecules present in all living organisms and viruses. From catalytic proteins (*enzymes*), transport proteins, structural and motor proteins to signalling and regulatory proteins, there are several specific roles that proteins and shorter peptides can adopt. While the genetic code of DNA molecules encodes information on the sequence of amino acids that make up the protein polymer chain (the *primary structure*), the biological function of the protein depends on its specific spatial arrangement. The transition from an unstructured amino acid polymer to the functional native protein structure is termed *protein folding* [1, 2].

Proteins usually adopt their precise native structures rapidly, reversibly and spontaneously in aqueous solution [3]. This folding process depends on the biological, chemical and physical environment as well as on the amino acid sequence itself. On the one hand, small differences in the amino acid sequence can lead to large differences in the three-dimensional structure [4]. On the other hand, can very similar protein structures arise from distinctly different amino acid sequences [5, 6].

Most protein structures have been found through protein crystallography and X-ray diffraction. Other relevant methods to determine protein structures are NMR [7, 8]

and cryo-EM [9, 10]. Today (October 2017), more than 49,000 unique protein structure clusters with greater than 90 % internal sequence identity are deposited in the Protein Data Bank (PDB [11]) [12]. However, in comparison with the number of known protein sequence clusters with greater than 90 % sequence identity ($> 61,000,000$, [13]) in the Universal Protein Database (UniProt [14]), this number is very small.

Protein crystallography and X-ray diffraction have limitations as the crystalline protein might not represent the structure *in vivo* and cannot give insight into the dynamics of folding. In addition, some proteins may never crystallise, where prominent examples are IDPs (intrinsically disordered proteins) [15]. Further, X-ray diffraction requires very low unphysiological temperatures during data collection to avoid damaging the crystal. These radiation damages might, again, lead to faulty structures, for example breakage of disulfide bonds or decarboxylation of glutamate and aspartate residues [16]. Although the water oxygens can be resolved with high resolution X-ray scattering of protein crystals, the complete hydrating water structure outside the crystal cannot easily be assessed.

Due to the many degrees of freedom of a polypeptide, the relationship between amino acid sequence, protein structure and function is complex and difficult to predict *ab initio*. It is yet unknown how the folding from an unstructured amino acid chain to a functional three-dimensional structure occurs on the atomic scale ($\approx 10^{-10} \text{ m} = 1 \text{ \AA}$).

Folding can be influenced by temperature, pH, solvents and co-solute concentration, other biological macromolecules and compartments such as membranes, liposomes or molecular chaperones. Under non-native conditions, such as unphysiological temperatures or changes in pH, proteins usually lose their folded structure (*denaturation*) and subsequently their function. The unfolded or denatured structures are more complex, flexible and heterogeneous and as such their detailed structural properties are difficult to assess.

Unfolding is often reversible [1] but further aggregation under more extreme conditions is usually irreversible [17, 18]. It is also possible, although less likely, that the protein can adopt an alternative structure under non-native conditions with different biological functions. Notable examples of alternative protein folding that lead to sys-

tematic biological malfunctions and diseases are pathological prions that can lead to Transmissible spongiform encephalopathies (TSEs) or amyloid proteins that are linked to Alzheimers's and Huntington's disease [19, 20, 21]. Another potential application of protein folding prediction could be the treatment of a wide range of diseases. Knowledge about the folding of the involved disease-causing or -preventing proteins will facilitate the development of drugs that target these structures specifically. Understanding protein folding could also be useful for the engineering and *de novo* protein synthesis of enzymes for industrial biotechnology or laboratory applications [22].

Protein folding occurs in solution and the local structure in a liquid is difficult to determine by many experimental methods. Although modelling techniques have been used to obtain folded protein structures from unfolded polypeptide chains since 1975 [23] and in recent studies larger protein structures could be predicted using homology [24] or *de novo* modelling [25], immense computational power is needed to simulate protein folding in atomistic detail [26]. The specific role water has in protein folding is difficult to assess with most experimental techniques and current theories on protein folding often lack verification of this phenomenon on the atomic scale. Due to the lack of a formal and universal mechanism, despite empirical models, the *protein folding problem* remains one of the main challenges in molecular biology [27].

1.1.2 Protein Structure

20 standard proteinogenic amino acids (Figure 1.2) are the building units that constitute the *primary structure* (Figure 1.3) of proteins and the order of these amino acids is directly encoded by the universal genetic code in DNA. In general, the *central dogma of molecular biology* applies and DNA is transcribed into mRNA which is translated, using 3-nucleotide-letter codons, into an amino acid sequence. Unusual proteinogenic amino acids like selenocysteine require special biological pathways to be incorporated into the protein and non-proteinogenic amino acids like hydroxyproline are formed by post-translational modifications [28, 29].

Amino acids are made of a central α -carbon (C_α) which is covalently bonded to a hydrogen (-H), a primary amine group (-NH₂), a carboxylic acid group (-COOH) and an

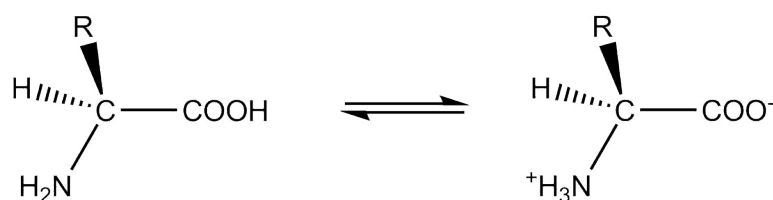
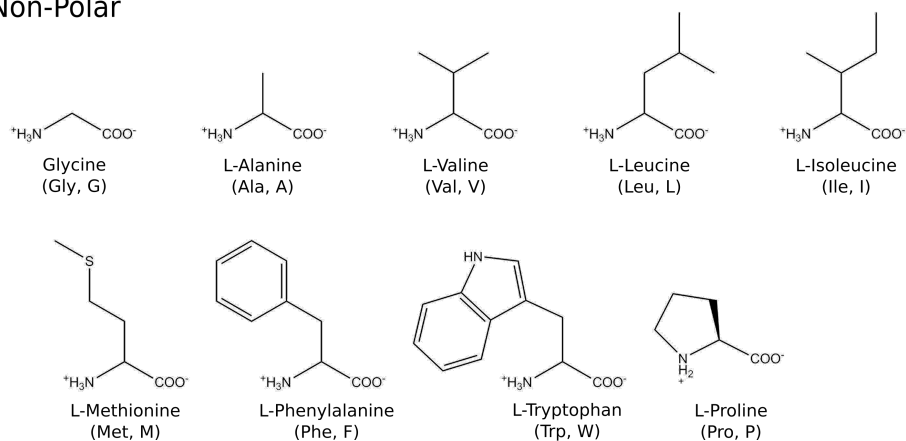


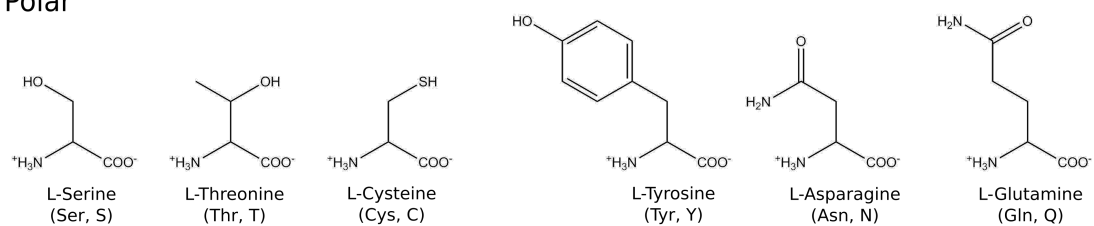
Figure 1.1: Basic structure of an L-amino acid in its neutral form (left) and as a zwitterion (right).

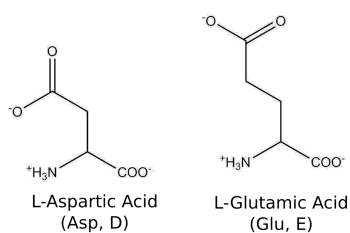
amino acid specific side chain (-R) (Figure 1.1, left). As the central carbon atom of an amino acid has four distinct groups attached, the molecules are chiral and can potentially exist in two stereoisomer forms. In naturally occurring proteins, amino acids are almost always L-stereoisomers. In aqueous solution and under physiological conditions the dominant species of amino acids is the *zwitterion* where the carboxylic acid and the amine group are both ionised, forming an ammonium ($-\text{NH}_3^+$) and a carboxylate group ($-\text{COO}^-$) (Figure 1.1, right): the pH at which the charge balances is called the *isoelectric point*. The amino acids can be classified into four groups with respect to their side chain: charged (forming weak acids or bases), polar and hydrophilic or apolar and hydrophobic (Figure 1.2).

Polymerisation of amino acids occurs *in vitro* and *in vivo* by a condensation reaction between a carboxylic acid group and an amine group of two different amino acids (Figure 1.3, top). Repetition of this process leads to the formation of shorter peptides, with up to 50 amino acid residues, and longer polypeptide chains that can fold into proteins.

The *protein backbone* are the amino acid atoms in the peptide bond chain of the protein ($\dots\text{C}\alpha\text{-C}(=\text{O})\text{-N}(\text{H})\text{-C}\alpha\dots$). Three torsion angles per residue can be identified: ϕ between $\text{C-N-C}\alpha\text{-C}$, ψ between $\text{N-C}\alpha\text{-C-N}$ and ω between $\text{C}\alpha\text{-C-N-C}\alpha$ (Figure 1.4). As the peptide bond between $\text{C}(=\text{O})\text{-NH}$ has a partial double bond character because of electron resonance with the carbonyl group, the dihedral angle of this bond (ω) is usually 180° [32]. However, recent analysis of high resolution protein crystal structures have shown that the root-mean-square deviation from peptide bond planarity is about $\pm 6^\circ$ and can even exceed 20° [33]. A special case of peptide bond is between any amino acid (X) and proline (P) which can not only have a dihedral angle of $\approx 180^\circ$ (*trans*) but also $\approx 0^\circ$ (*cis*). This is because the cyclic structure of the proline side chain and the lack of an amide hydrogen equalise the steric hinderance of *cis* and *trans* isomers. In an unfolded

Non-Polar

Polar

Acidic

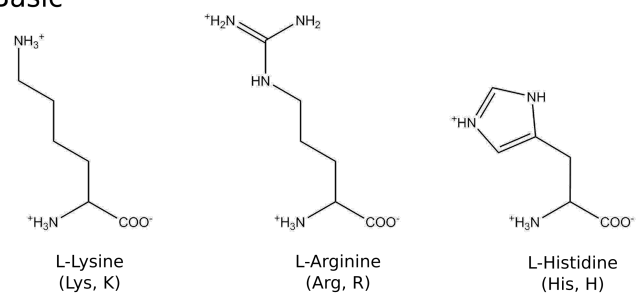
Basic

Figure 1.2: Overview of the 20 standard proteinogenic amino acids as zwitterions with their three-letter and one-letter notation.

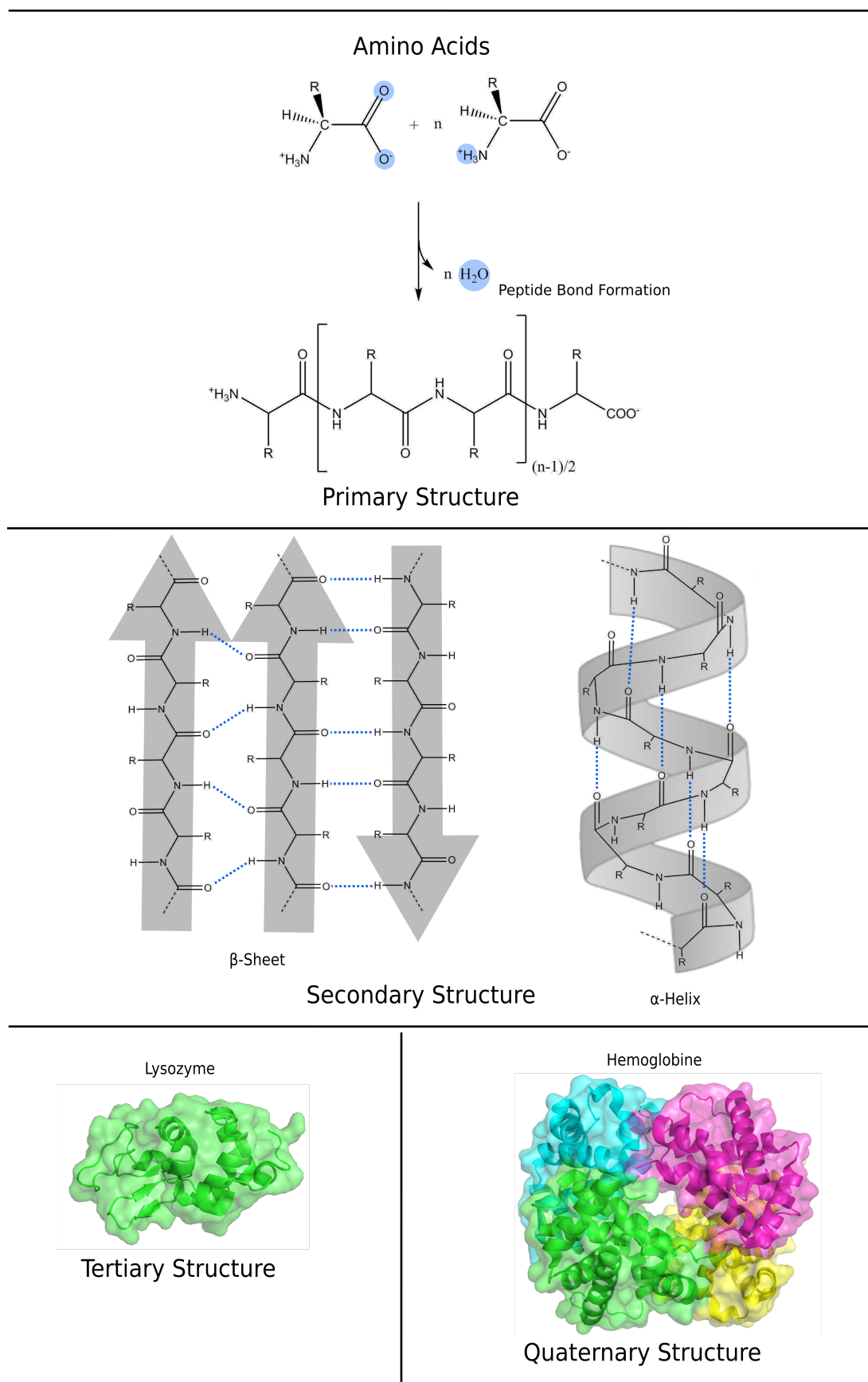


Figure 1.3: Peptide bond formation and overview of levels of protein structure. Protein structures were taken from [30, 31].

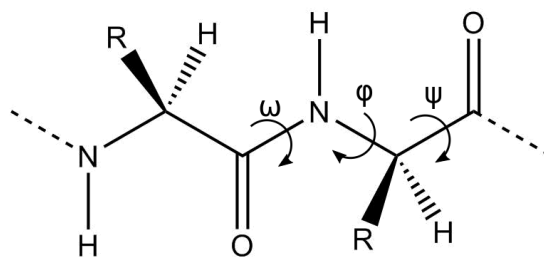


Figure 1.4: Peptide and protein backbone angles.

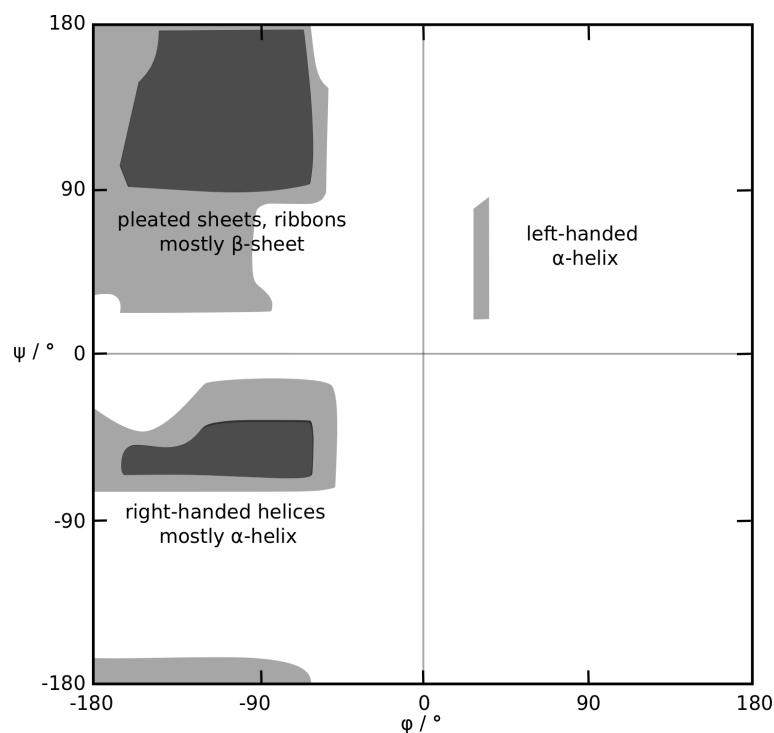


Figure 1.5: Schematic Ramachandran Plot with light-grey areas allowed (ca. 22 %) and dark grey areas favoured (ca. 7 %) in known structures, adapted from [36].

protein the *cis-trans* ratio of an X-Pro bond is at a sequence specific equilibrium. In a folded native protein a specific X-Pro bond is usually exclusively at a *cis* or *trans* conformation [34]. Usually, rotation along the backbone is only possible between the two other dihedral angles ϕ and ψ . Steric hindrances from both the backbone and side-chain atoms limit allowed angles to a very small range. The conformation of the backbone can be described with the pairs of ϕ and ψ angles which are characteristic for specific *secondary structure* elements (Figure 1.3, middle), smaller local structure patterns, such as α -helices and β -sheets, and are visualised in a two dimensional *Ramachandran plot* [35] (Figure 1.5).

The *tertiary structure* is the overall atomic three-dimensional arrangement of sec-

ondary structures and amino acid side chains of the polypeptide. An arrangement of multiple polypeptide chains to a multimeric protein complex is called *quaternary structure*, but this level of structure is not found in all proteins (Figure 1.3, bottom).

1.1.3 Fundamental Thermodynamics and Kinetics of Protein Folding

Gibb's free energy is the work energy associated with any reaction or interaction, including protein folding. The change in Gibb's free energy (ΔG) is given by

$$\Delta G = \Delta H - T\Delta S \quad (1.1)$$

where ΔH is the change in *enthalpy*, T the temperature and ΔS the change in *entropy* of this reaction. The enthalpy is the sum of the internal energy (U) and the product of pressure (p) and volume (V):

$$H = U + pV \quad (1.2)$$

At a constant pressure, the change in enthalpy is a measure of the change in internal energy. A heat-releasing, exothermic, reaction has a decreasing enthalpy ($\Delta H < 0$); a heat absorbing, endothermic reaction has an increasing enthalpy ($\Delta H > 0$). The entropy is a measure of disorder in the system and energy dispersal. It is proportional to the Boltzmann constant (k_B) and the number of possible micro-states (W) within a macro-state:

$$S = k_B \ln W \quad (1.3)$$

A decrease in entropy ($\Delta S < 0$) signifies a reaction that leads to more ordered system while an increase in entropy means ($\Delta S > 0$) that the system becomes more disordered. The *second law of thermodynamics* states that the entropy of a completely isolated system can only increase over time [37]. A spontaneous reaction must have a decreasing

free energy ($\Delta G < 0$). However, it should be noted that this does not take into account the possible kinetic barriers that might lead to a reaction with a negative change in free energy that does not occur spontaneously.

Protein folding is a highly *cooperative* mechanism of non-covalent interactions. Consequently, protein folding from an unfolded, extended state (U) to a folded, native state (F), can, in this context, be simplified as a first-order and two-state mechanism without intermediate states [38]



where k_f and k_u are the kinetic rate constants for folding and unfolding, respectively. Kinetic studies can usually only observe k_{obs} which is the sum of k_f and k_u :

$$k_{obs} = k_f + k_u \quad (1.5)$$

The change in Gibb's free energy during the folding determines the protein stability in the folded state (Figure 1.6):

$$\begin{aligned} \Delta G_{U \rightarrow F} &= -RT \ln K \\ &= -RT \ln \frac{k_f}{k_u} \end{aligned} \quad (1.6)$$

As protein folding is usually a spontaneous reaction, a negative associated ΔG can be assumed. However, the free-energy difference (ΔG_{U-F}) between the biologically active native and the denatured states of the proteins is, between -10 to -45 kJ/mol, very small [39]. Because protein folding leads from an extended unfolded structure to a three-dimensional arrangement with less degrees of freedom, this decrease in entropy is thought to be unfavourable. However, the whole system, including multiple protein states, the solvents and other solutes must be taken into account; and also the energetically favourable non-covalent interactions that lead to protein folding and native structure stability.

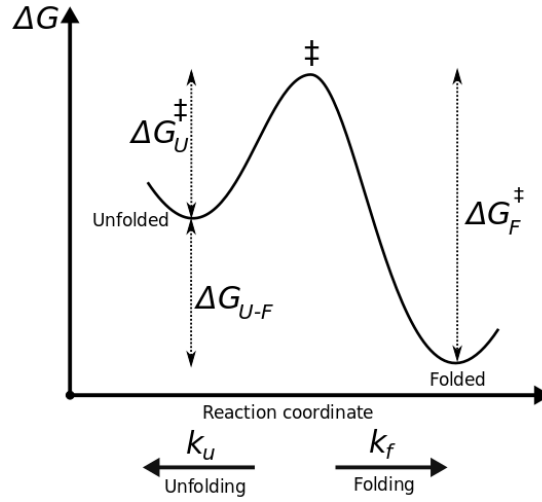


Figure 1.6: Schematic free energy model of a protein in water that exhibits two-state folding (unfolded and folded) with a transition state ($\ddagger$).

1.1.4 Non-Covalent Interactions in Biological Molecules

1.1.4.1 Coulomb and Lennard-Jones Potential

Non-covalent interactions determine the intermolecular interactions and the intramolecular structuring of biological macromolecules including protein folding. A summary of interactions compared to covalent bonds is given in Table 1.1. Apart from the covalent peptide bonds and the bonds in the amino acid units themselves, disulfide bonds between cysteine residues are an important factor in protein folding. In fact, unfolded proteins with incorrect disulfide bonds cannot fold into the correct native structure [40]. While a *covalent bond* is formed by sharing of electron pairs between atoms, a *non-covalent interaction* is determined by more disperse attractive and repulsive factors. Non-covalent interactions are characterised by electrostatics described in the *Coulomb potential* and dispersion and steric repulsion described in the *Lennard-Jones potential*.

The Coulomb potential describes the *electrostatic energy* between two point charges (q_i and q_j) at a distance r_{ij}

$$U_C(r_{ij}) = \frac{1}{4\pi\epsilon\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (1.7)$$

where ϵ_0 is a physical constant, the permittivity in vacuum, and ϵ the medium-dependent

Table 1.1: Summary of covalent and non-covalent interactions in biological molecules. Adapted from [37, 41, 42, 43].

Interaction		Example	Typical Distance / Å	Typical Free Energy / kJ/mol
Covalent Bond		-C(=O)⋯N(H)-	0.6 - 3.0	-(100 - 1000)
Electrostatic	Monopole	-COO ⁻ ⋯ ⁺ HN ₃ -	2.8	-30
	Dipole	H-Cl ⋯ H-Cl	3	-10
Dispersion		C-H ⋯ H-C	3.5	-0.1
Steric Repulsion				very strong +
Hydrogen Bond		N-H ⋯ O=C	2	-25

permittivity or dielectric constant. For water this dielectric constant is ≈ 80 and within a protein it is ≈ 4 [37]. This means that in water electrostatic interactions are reduced and shielded about 20 times more compared to the inside of a protein. The dielectric ‘constant’ inside a folded protein is variable and dependent on the exact position within the protein. It has been estimated theoretically [44, 45] and using modelling techniques [46, 47].

The Coulomb potential applies to integral charges (*salt-bridges*), for example between negatively charged glutamic acid and positively charged lysine residues. For non-integral charges (*dipoles*), for example between water molecules, the electrostatic interactions over all atom pairs, that are not in a covalent bond, are summed:

$$U_C(r_{ij}) = \sum_{\text{all pairs}} \frac{1}{4\pi\epsilon\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (1.8)$$

Any fully or partially charged atom can induce a dipole by distortion of the electron density in an uncharged atom. This will lead to a weak electrostatic attraction between charged and uncharged atoms. Even between completely uncharged atoms and molecules random fluctuations in electron density can lead to complementary temporary induced transient dipoles. These attractive induced dipole interactions are called *dispersion forces*. At a very close distance strong *steric repulsion* due to overlapping electron clouds (*Pauli exclusion principle*) dominates the interaction of atoms. The Lennard-Jones (LJ) potential (Figure 1.7) describes the energy of these attractive and repulsive

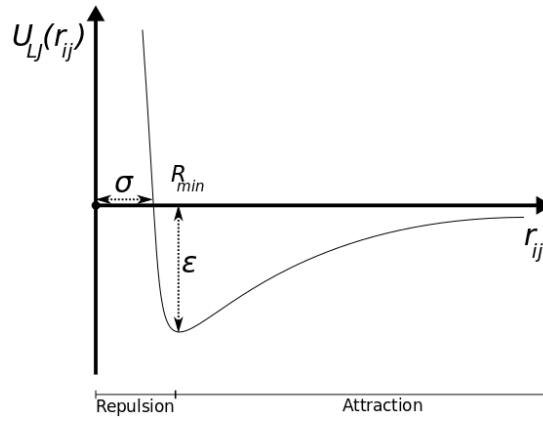


Figure 1.7: Schematic Lennard-Jones Potential.

forces:

$$\begin{aligned}
 U_{LJ}(r_{ij}) &= \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \\
 &= 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
 &= \epsilon_{ij} \left[\left(\frac{R_{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min}}{r_{ij}} \right)^6 \right]
 \end{aligned} \tag{1.9}$$

In these equations the repulsive forces are represented in the r^{-12} -term and the attractive forces in the r^{-6} -term. A_{ij} , B_{ij} , ϵ_{ij} and σ_{ij} are specific to the atom pair. ϵ_{ij} is the energy value at the potential minimum and σ_{ij} is the sum of the *hard-sphere radii* at which the intermolecular potential equals zero. The equilibrium distance where the potential is minimal (R_{min}) is the sum of *van-der-Waals radii*.

The LJ potential applies to all non-covalent atom pairs, not only uncharged atoms. The attractive force between permanent dipoles is called *Keesom force* [48], the force between a permanent dipole and an induced dipole *Debye force* [49] and the aforementioned force between two induced dipoles *London dispersion force* [50]. However, between polar and charged groups a combination of LJ and Coulomb potential will be applied.

Hydrogen bonds are a special case of dipole-dipole interactions. They occur between a hydrogen atom that is covalently bonded to an electronegative atom (in biological molecules usually nitrogen or oxygen), called a *donor*, and another electronegative atom with free electron pairs, called an *acceptor*. A hydrogen bond is shorter and stronger

than electrostatics and van-der-Waals radii would suggest. When only two atoms are involved, there is also a directionality observed where acceptor or the free electron pair of the acceptor, hydrogen and donor atom are often found to be on one axis. This is due to the partial covalent character of a hydrogen bond which has also been confirmed by J -coupling connectivity in NMR spectroscopy between hydrogen bonded atoms [51]. This special case of dipole interactions can be described by

$$U_{HB}(r_{ij}) = \frac{1}{4\pi\epsilon\epsilon_0} \frac{q_i q_j}{r_{ij}} + \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \quad (1.10)$$

as a combination of electrostatic dipole interaction and a partial covalent character where C_{ij} and D_{ij} are atom pair specific constants [37]. The most common example of hydrogen bonds occurs in water, but proteins secondary structure elements like helices and sheets are also stabilised *via* hydrogen bonds between protein backbone amide hydrogens and carbonyl oxygens (Figure 1.3).

1.1.4.2 The Hydrophobic Effect

Non-polar molecules or parts of molecules tend to aggregate in water. These non-polar molecules are therefore also called hydrophobic and the phenomenon of molecular aggregation of these compounds the *hydrophobic effect*. The origin of the hydrophobic effect, on the atomic scale, is not fully understood [52, 53].

This effect cannot be explained by the previously described non-covalent interactions as the Debye forces between non-polar molecules and permanent dipoles like water are stronger than the London dispersion forces of non-polar molecules with each other. A system of aggregated non-polar molecules in water is also more ordered than a mixture of randomly distributed, unaggregated non-polar molecules in water. Therefore the aggregation of non-polar molecules in water is unfavourable both in terms of enthalpy and, in principle, entropy for the non-polar molecules and it is argued that the reason for the hydrophobic effect must arise from water-water interactions.

In solution, water molecules are strongly hydrogen bonded to one another and any

interaction with non-polar molecules requires some of these hydrogen bonds to be broken. Non-polar solute molecules interact only comparably weakly with the water molecules so that the solvation of these molecules would be, overall, enthalpically unfavourable. However, the main reason for the hydrophobic effect is thought to be a favourable change in entropy upon hydrophobic aggregation [54, 55]. In order to maximise the remaining water-water hydrogen bonds, water molecules on the surface of non-polar molecules are assumed to adopt highly ordered cage-like structures (clathrates) which is entropically unfavourable [56]. The aggregation of hydrophobes is thought to lessen this effect by lowering the solvent accessible hydrophobic surface area and thus increasing the entropy of the whole system [37, 57].

1.1.5 The Protein Folding Problem

The protein folding problem is essentially the question how the amino acid sequence and a given environment determine the three-dimensional structure of a protein. There are three aspects associated with this: Firstly, the thermodynamic question as to what forces determine and stabilise the folded structure; secondly how proteins fold rapidly *via* folding pathways to form their native structure; thirdly how to predict the native structure from the amino acid sequence [58].

The *thermodynamic hypothesis* states that the protein adopts the conformation at which Gibbs free energy (G) is the lowest under the given conditions, the native state. Therefore, the sum of interatomic interactions of the amino acid sequence, in a physiological environment, determines the native structure [1, 59].

What small local interactions like hydrogen bonds, ion pairs, van-der-Waals forces and hydrophobic interactions contribute to, and the degree to which they contribute to, protein stability is still debated [60]. It is important to clearly distinguish this side of the protein folding problem, the question of thermodynamic native state stability, from the kinetic folding pathway question. A non-covalent interaction that primarily stabilises the native structure might not be the initial interaction that leads to protein folding.

It has been suggested that electrostatic ion-pairing, for instance, cannot be the driv-

ing force in protein folding as ion pairs are not highly conserved in evolution and are comparatively rare in fully folded proteins [61]. However, genome-wide and structural comparisons of thermophilic and mesophilic proteins indicate that salt bridges may significantly contribute to the enhanced thermal stability of these proteins from thermophilic organisms [62]. Further, recent mutational studies suggest that hydrogen bonds do contribute to protein stability, and that this contribution is as important as the hydrophobic effect, although hydrogen bonds with the solvent tend to be more energetically favourable than intramolecular hydrogen bonds [63].

The fact that a folding protein cannot randomly sample all possible configurations in the time it actually needs to reach the native state (usually within ‘ms’ to ‘s’) is called the *Levinthal paradox* [2, 64]. In fact, the speed of the folding processes is one of the challenges in solving the protein folding problem because folding intermediates, which have been identified [65, 66], will also only exist in that short period [67]. A *Molten globule* often refers to these intermediate species in protein folding. However, this term is usually not clearly defined as it can refer to an intermediate structure without secondary structure elements that is mostly stabilised *via* hydrophobic interactions [68] or a structure with most secondary structure elements intact and a flexible tertiary structure [69].

A high cooperativity of non-covalent interactions is assumed as proteins undergo rapid transitions between unfolded and folded states. The concept of folding funnels states that not only the lowest energy structure must be taken into account but also the folding pathway of the protein on the energy landscape. This conformational landscape is a rugged surface sloped down into a global minimum, representing the native state (Figure 1.8) [70]. Protein folding must be speeded and guided by local interactions. Analogous to the concept of codons, a theory is that small cooperative folding and unfolding units, so called foldons, define the steps along the folding pathway [71, 72]. However, the classical view of one single folding pathway has been disproven as multiple different folding intermediates of the same protein have been found [73]. The current assumption is a multipath funnelled energy landscape with the native structure as the energy minimum [74]. Importantly, what the dominant driving forces of protein folding are and what the sequence of events is that leads to protein folding and unfolding is still

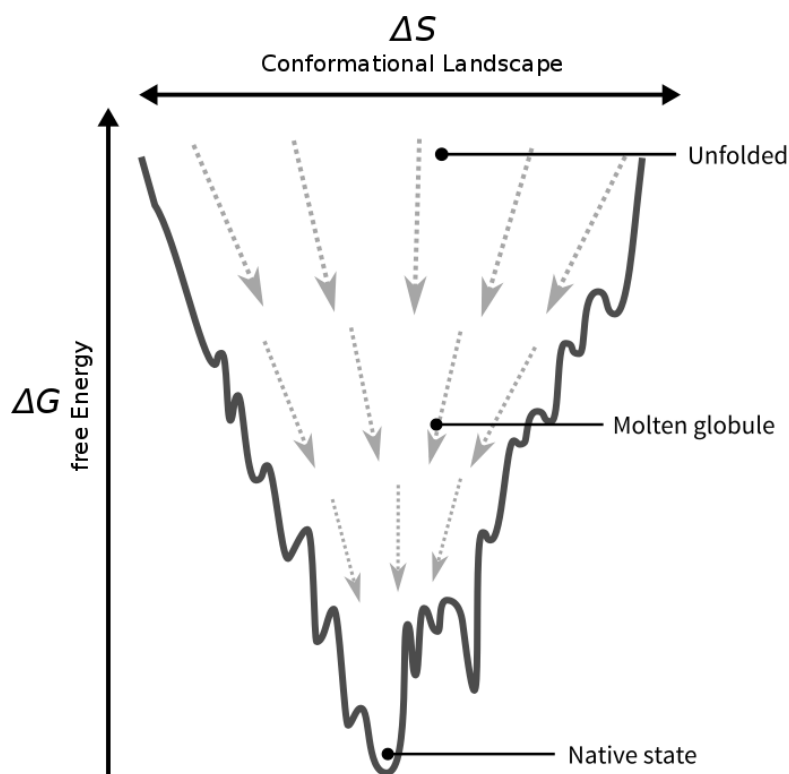


Figure 1.8: Schematic folding funnel. The height of the folding funnel represents the free energy, the width represents the entropy (adapted from [75]).

not understood.

Three main theories that seek to explain the significance of the driving forces [76] behind protein folding can be distinguished:

The *hydrophobic collapse model* emphasises the role of hydrophobic interactions that are said to initiate the tertiary structure by the rapid collapse of hydrophobic clusters to form a compact molten globule structure before the secondary structures develop through hydrogen bonds [77, 78, 79, 80, 81]. However, these hydrophobic interactions are also a subject of debate as it is unknown how unpolar residues would form a specific tertiary structure within the average protein folding time with short ranging hydrophobic interactions [82].

The idea that hydrophobic interactions are the dominant and initial step to protein folding is challenged by the idea that short range interactions such as hydrogen bonds work as initiation sites to enable long-range interactions [83]. This *framework model* supposes that local interactions of intramolecular hydrogen bonds form the secondary

structure of the protein before the long range tertiary structure is formed [84].

The *nucleation condensation model* combines both theories. It proposes a hydrophobic collapse and the forming of intramolecular hydrogen bonds simultaneously without an intermediate structure and that secondary and tertiary structures are mutually dependent [85].

Another related aspect is the question as to what general dynamics determine the folding process. The *Nucleation/growth model* assumes that the rate-limiting step to initiate protein folding is the formation of small folding nuclei [86, 66], whereas the *Diffusion-collision-adhesion model* argues that the diffusion and collision of folded microdomains, is the factor that determines the rate of protein folding [87]. Complexity is further increased by the theory that proteins might have multiple different and intersecting folding pathways depending on intermediates (*Jigsaw puzzle model*) [76].

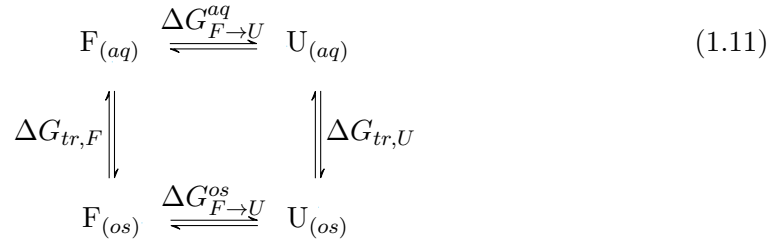
1.1.6 Protein Stability and Protein Solute-Induced Unfolding

Several factors that influence protein stability can be altered to unfold proteins [88, 89]; such as temperature, pH, ionic strength, the concentrations of co-solutes or the solvent itself. The effect of different salts on the solubility and structural stability of proteins is empirically described by the ‘Hofmeister series’ [90, 91]. Protein stability and unfolding has frequently been examined by protein denaturation studies using urea [1, 92, 93]. Another well-known denaturant is the salt guanidinium chloride [94]. Conversely, other molecules, such as trimethylamine N-oxide (TMAO), have a large stabilising effect on proteins [95, 96]. Similar to protein folding, the processes that lead to protein unfolding and denaturation, or conversely increased protein stability, caused by chemical compounds in solution, is unclear.

The equilibrium between folded (F) and unfolded (U) protein states ($U \rightleftharpoons F$) can be shifted both by the presence of stabilising or denaturing osmolytes (*os*). Analogous to protein folding, this two-state transition can be oversimplified. A multi state equilibrium with intermediates states ($U \rightleftharpoons I_1 \rightleftharpoons I_2 \cdots \rightleftharpoons F$) is possible and there are cases of transitions, where stable intermediates exist at a low denaturant concentrations [97, 98].

The addition of urea to a solution stabilises the unfolded protein and decreases the free energy of the unfolded protein state (ΔG_U). Thereby the unfolding free energy difference $\Delta G_{F \rightarrow U}^{os}$ is decreased in urea solutions compared to the unfolding free energy difference in osmolyte (*os*)-free systems ($\Delta G_{F \rightarrow U}^{aq}$). In contrast to urea, the presence of a protecting osmolytes, such as TMAO, destabilises the unfolded state (increases its free energy ΔG_U) and $\Delta G_{F \rightarrow U}$ is increased [99]. These changes in thermodynamic stability are thought to be a result of energetically favourable interactions between urea and the protein surface or, conversely, energetically unfavourable surface exclusion of stabilising osmolyte molecules. As the unfolded state has a larger solvent exposed area, its stability is more affected than the folded state [100, 101].

The energetic changes of the protein upon transfer between in pure water and an osmolyte-solution, leading to osmolyte-induced protein folding and unfolding, is described by the thermodynamic cycle of the Tanford transfer model [99, 102]



where F is the folded protein, U is the unfolded protein in water (*aq*) or osmolyte solution (*os*), $\Delta G_{tr,U}$ and $\Delta G_{tr,F}$ are the free energies of the transfer between water and osmolyte solution for the unfolded or folded protein, respectively. The impact that the osmolyte has on the free energy difference between folded and unfolded state is typically a linear relationship [103]

$$\Delta G_{F \rightarrow U}^{os} = \Delta G_{F \rightarrow U}^{aq} + m[os] \tag{1.12}$$

where m is a factor that indicates the sensitivity of the folding/unfolding transition to the osmolyte concentration. For a denaturant like urea as the osmolyte the m value is negative, whereas for a protecting osmolyte m will be positive. The transfer model postulates that the energetic effects of osmolyte concentration on protein stability is the

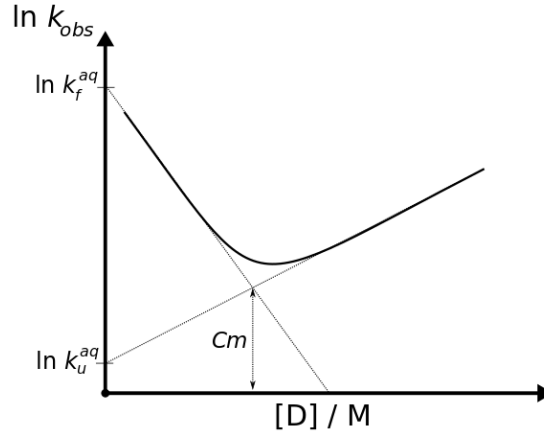


Figure 1.9: Schematic chevron plot (adapted from [108]).

sum of the free energy contributions of different solvent exposed protein groups such as backbone and side chains.

With the aid of the Tanford transfer model previous studies have shown that the impact urea has on protein stability is most notably with the protein backbone [104], although more recent computational studies have shown an equal contribution to $\Delta G_{F \rightarrow U}$ of protein backbone and side chains [105].

As protein stability is related to the folding and unfolding rates (1.1.3) the kinetics of the $U \xrightleftharpoons[k_u]{k_f} F$ reaction are also influenced by the concentration of denaturants and stabilising osmolytes. A logarithmic plot of k_{obs} vs. the denaturant concentration ($[D]$) is called a Chevron-plot which has a characteristic V-shape (Fig 1.9). Higher concentrated denaturants increase the logarithm of the unfolding rate k_u

$$\ln k_u^{os} = \ln k_u^{aq} + m_u[D] \quad (1.13)$$

while lower concentrated denaturants decrease the logarithm of folding rate k_f linearly [106, 107]. The concentration where 50 % of the protein is folded and 50 % unfolded is called the denaturation midpoint Cm .

$$\ln k_f^{os} = \ln k_f^{aq} - m_f[D] \quad (1.14)$$

1.2 Aims and Objectives

The aim of this thesis is to obtain structural insight into protein folding and unfolding and specifically to elucidate the role of water and co-solutes in these processes. For this, neutron and X-ray diffraction and spectroscopic methods are combined with simulations techniques, such as Empirical Potential Structure Refinement (EPSR) and Molecular Dynamics (MD) simulations, in order to analyse model peptide systems. In an inductive approach, the presented results will be used to increase understanding of protein folding on the atomic scale. This thesis will present different aspects of protein folding and unfolding in the results chapters. However, some overarching aims will be relevant throughout:

First and foremost is the aspect of **protein hydration**. One of the main advantages in the use of neutron scattering is the large contribution water has to the scattering signal. By using isotopic substitution a reliable model of the hydrating water around the peptides can be obtained. This information will then be used to elucidate how hydrating water contributes to **protein stability** and how **sequence specific hydration patterns** might influence initial folding and unfolding dynamics.

Another important aspect, in this context of protein solvation, is the question as to how other solutes interact with the model peptides. How **co-solutes** like salts and the common chaotropic denaturant urea influence peptide hydration and how in an **amphiphilic environment** hydrophobic and hydrophilic solvents interact with the peptide.

These findings in peptide solvation will be conjoined with the ensembles of three-dimensional peptide structures, obtained by the computer simulations and experimental techniques, to inductively **describe protein folding and unfolding mechanisms** in the liquid phase on the atomic scale.

Apart from that, the development of **new computational methods** to analyse and characterise molecules on the atomic scale is an important ancillary aim. The employed computer programmes are relatively new and provide an innovative insight into detailed atomistic models based on experimental scattering data.

1.3 Structure and Scope of the Thesis

Chapter 2 will give an overview of the employed experimental methods and their underlying principles and theories. The fundamentals of neutron and X-ray scattering will be described as well as the experimental setup of the scattering data collection on the diffractometers and the subsequent data correction. The employed computer simulation techniques, namely Empirical Potential Structure Refinement (EPSR) will be described along with a brief description of Molecular Dynamics (MD). The principles behind the two employed spectroscopic techniques nuclear magnetic resonance (NMR) and circular dichroism (CD) will briefly be explained. Finally, the principles of the program ANGULA, used for the atomistic analysis of the simulated data, will also be outlined. However, the main focus of this chapter is the description of the principles, techniques and computational analysis involved in the key experimental technique of this thesis, neutron scattering.

Chapters 3 and 4 will focus on the question of how urea is able to unfold proteins. The small model peptides GPG and KGPGK in an aqueous urea solution will be examined and their hydration and solute-interactions compared with and without urea. Chapter 3 will mostly focus on the influence urea has on the GPG peptide hydration and Chapter 4 will look into the interplay of co-solutes influenced by urea.

Further, Chapter 4 analyses the sequence specific hydration of 4 different KGXGK peptides and the probabilities of bridging water molecules between residues. It will discuss if these hydration pattern might indicate a nucleation site to protein folding.

Chapter 5 investigates a group of amphiphilic helix-forming peptides (KLL₁₅ and KLL₂₂) in water, water/methanol mixture and aqueous urea-solution. It should be noted that these peptides are the only ones in this thesis that are long enough to form secondary structures. However, this chapter also presents the most challenging and complicated subject to analyse using the given methods.

The final chapter of this thesis (Chapter 6) summarises the presented findings and describes overarching conclusions and implications of these results regarding protein stability and protein folding and unfolding mechanisms. Open questions will be discussed,

how these findings relate to current theories, as well as suggestions for further work on the subject.

Chapter 2

Methodical Background

Throughout the text, bold letters are used to indicate vectors and non-bold versions of these letters to indicate their magnitude.

2.1 The Structure of Biomolecules in the Liquid Phase

Experimental techniques to study the structure of biomolecules on the atomic to intermediate scale ($\text{\AA}$ to nm) include most notably cryo-electron microscopy (cryo-EM) [10], X-ray [109] and neutron diffraction [110, 111], UV-VIS [112, 113], fluorescence [114], infra red (IR) [112], circular dichroism (CD) [115, 116] and nuclear magnetic resonance (NMR) [117, 118] spectroscopy. Aside from diffraction, these are primarily spectroscopic methods which provide an indirect measurement of certain properties of molecule ensembles and large-scale (nm to μm) structures. The impact water has on biomolecular structure in solution is often not assessed in detail; as biomolecules *in vivo* rely on an aqueous environment to function, studies of liquid phase systems can provide a more realistic view on their structure and structure dynamics.

2.1.1 Local Order in the Liquid Phase

Currently, most atomic-scale ($\approx 10^{-10} \text{ m} = 1 \text{ \AA}$) resolution biological structures have been obtained through protein crystallography and diffraction using X-rays as a probe.

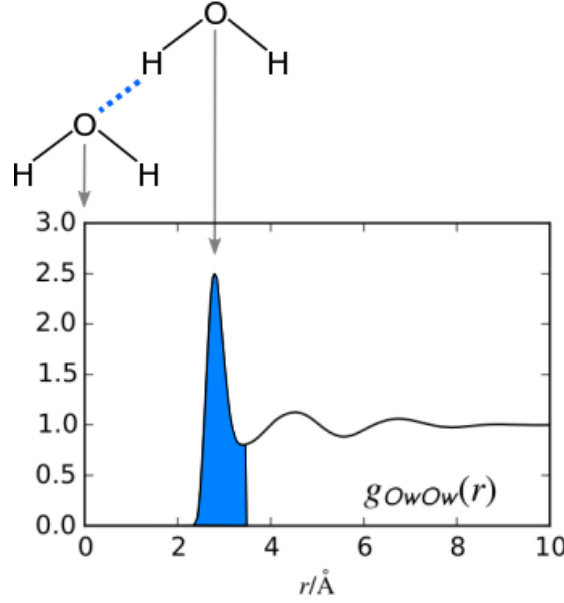


Figure 2.1: Intermolecular water oxygen (OwOw) radial distribution function (RDF) for pure water (data from [121]).

The ordered three-dimensional lattice of a crystal contains repeating unit cells which magnify the diffraction signal and, with additional computational techniques, a structure with atomic-scale resolution can be obtained.

In liquid non-crystalline systems, only local structures, usually up to ≈ 10 Å, are present [119]. This local structural information is often displayed as radial distribution functions (RDFs or $g(r)$ s) [120]. These are functions which give the neighbouring atom (K) at a distance (r) from the central atom (J), normalised to the neighbouring atom concentration (c_K) and the overall atomic density (ρ). RDFs can be obtained from X-ray or neutron diffraction data *via* Fourier transformation of the structure factors (Section 2.2). The average number of neighbouring K atoms around J atoms $n_J^K(r)$ at a distance range (coordination number, CN) between r_{max} and r_{min} is conversely found by integration of the $g_{JK}(r)$:

$$g_{JK}(r) = \frac{n_J^K(r)}{4\pi\rho c_K r^2 dr} \quad \Leftrightarrow \quad n_J^K(r) = 4\pi\rho c_K \int_{r_{min}}^{r_{max}} r^2 g_{JK}(r) dr \quad (2.1)$$

Figure 2.1 shows the intermolecular radial distribution function of water oxygens ($g_{OwOw}(r)$) in pure water [121]. Each peak corresponds to a coordination shell and the

area underneath the curve to the number of neighbouring water oxygens found at this distance range. For example, the first peak at 2.8 Å indicates the most probable first neighbour distance and the blue area underneath the peak relates to a coordination number of 5.2 Ow neighbours between 0 and 3.5 Å.

2.1.2 Techniques to Determine the Structure of Biomolecules in the Liquid Phase

There are a number of experimental and computational techniques which can be used to study the thermodynamic and kinetic properties of biomolecules in solution. However, experimental methods that specifically seek to determine the atomic-scale structure of liquids are, in general, confined to spectroscopic and scattering techniques [122]. Spectroscopic techniques such as NMR, UV-VIS, IR, Raman and CD rely on the absorption and emission of radiation by atoms and subatomic particles which can indirectly be related to structural properties [122, 123]. Scattering techniques directly determine the structure of molecules by measuring the change in direction of incident particles on the sample [119]. Another emerging technique in structural biology is cryo-EM, which determines the three-dimensional structure of macromolecules by combining electron microscopy images of individual molecules [9].

Computer simulations can provide a tool to understand the disordered system of liquids *in silico* [124]. Most notably there are two types of simulation techniques, Molecular Dynamics (MD) [125, 126] and Monte Carlo (MC) [127, 128, 129].

The following sections will introduce the concepts of the complementary experimental and computational methods that have been used in this thesis to study peptides in solution, namely nuclear magnetic resonance (NMR), Circular Dichroism (CD) and Molecular Dynamics (MD). Detailed explanations on neutron (and X-ray) scattering combined with computational techniques is found in 2.2 - 2.4. Additional details on method applications will be described in Chapters 3 - 5.

2.1.2.1 Nuclear Magnetic Resonance

The following sources were used in this section: [130, 131, 132, 133, 8, 7].

Nuclear Magnetic Resonance (NMR) spectroscopy is a standard technique in biochemistry and organic chemistry to determine structures indirectly by monitoring the influence of a magnetic field on nuclei. These nuclei with a non-zero spin ($I \neq 0$) (e.g. ^1H , ^{13}C and ^{15}N) have $2I + 1$ possible orientations that are usually of equal energy.

In an external magnetic field (B_0), each orientation has a distinct energy level with an associated magnetic quantum number m , between $-I \leq m \leq I$. For example, hydrogen has a spin of $I = 1/2$ and consequently there are 2 possible orientations (with $m = 1/2$ and $m = -1/2$) in an applied magnetic field, parallel (energetically preferred) or antiparallel. Within the applied field, the nuclear spin generates a small magnetic field with a magnetic moment (μ). The external magnetic field (B_0) is usually applied along the z-axis. As the electrons surrounding the nuclei are also charged and possess a spin, they produce a magnetic field that generally reduces the external magnetic field (B_0), resulting in a magnetic field (B) at the nucleus which can be related to a chemical environment. The energy of the magnetic moment at the nucleus is therefore given by

$$E = -\mu_z B = -\gamma m \hbar B \quad (2.2)$$

where μ_z is the z-component of the magnetic moment vector μ , B is the magnetic field strength at the nucleus, γ is the gyromagnetic ratio, an atom specific constant and $\hbar$ is Planck's reduced constant ($\hbar = \frac{h}{2\pi}$). This energy contains the *Larmor precession frequency* ($\omega = -\gamma B$) at which the axis of nucleus rotation precesses around in the magnetic field. In a magnetic field, the initial population of the energy levels follows a Boltzmann distribution ($\propto e^{\frac{E}{k_B T}}$).

Electromagnetic radiation pulses are used to excite the nuclei and this resonant energy absorption of the nuclei occurs when the applied radiation frequency (ν) equals the Larmor precession frequency difference between the energy states ($\nu = \frac{\gamma B}{2\pi}$). Consequently, the nuclei have different resonance frequencies depending on their chemical

environment. After the excitation by the electromagnetic pulse, the nuclei relax back into their original Boltzmann distribution of energy levels within the applied magnetic field. This relaxation can be detected as an electric signal called *free induction decay* (FID) and converted into a spectrum by Fourier transformation.

Peak position (chemical shift), multiplicity (splitting), intensity and shape are sensitive to the chemical environment of atoms and allow structural properties of the sample to be determined. Multidimensional spectra can also be obtained by using different pulse sequences to obtain more information on molecular conformations, through space correlations and associations. Quantitative information about the structure of analysed molecules can be obtained, for example, by peak integration. Coordination numbers are usually not obtained, although there are recent studies on the hydration shells of biomolecules using NMR spectroscopy [134].

2.1.2.2 Circular Dichroism

The following sources were used in this section: [135, 136, 137, 138, 139, 38, 140].

Circular dichroism (CD) spectroscopy is a technique based on the difference in absorption of left- and right-handed circularly polarised light. Natural light is generally unpolarised, which means that all orientations of the oscillating electric and magnetic fields of the wave are equally probable. Planarly polarised light consists only of waves that have their electric and corresponding magnetic fields oscillating in the same direction. As the magnetic and electric fields are always perpendicular to each other, only the electric fields and the corresponding amplitude vector $\mathbf{E}$ will be considered from here on. Other polarised light states, such as circular polarised light, can be described as a sum of two linearly polarised states of equal intensity with their $\mathbf{E}$ s horizontal and vertical at a right angles to each other ($\mathbf{E}_h$ and $\mathbf{E}_v$). Depending on the phase difference of the vertical and the horizontal wave, different polarisation states can be generated. If the phase difference is a quarter wave ($\frac{\pi}{2}$) the resultant sum of $\mathbf{E}_h$ and $\mathbf{E}_v$ is rotating and a beam of left-(L) or right-handed (R) polarised light is created. Conversely, linear polarised light can be viewed as a combination of right- and left-handed circular light of equal intensity.

The absorbance of left- and right-handed circularly polarised light (A_L and A_R) is different for chiral molecules ($\Delta A = A_L - A_R \neq 0$). This phenomenon is called circular dichroism. It is described by Lambert-Beer's law

$$\Delta A(\lambda) = \frac{[\epsilon_L(\lambda) - \epsilon_R(\lambda)]LC}{M} = \frac{\Delta\epsilon(\lambda)LC}{M} \quad (2.3)$$

where ϵ is the molar extinction coefficient, L is the path length of the sample, C is the mass concentration and M is the molecular weight.

After monochromatic linear polarised light has passed through a chiral and absorbing sample, the transmitted light will be elliptically polarised, because this light is a combination of left- and right-handed light of different intensities. However, the actual CD measurement in a spectropolarimeter is commonly performed by switching the incident light between left- and right-handed circular light. CD is measured in ellipticity (θ°)

$$\begin{aligned} \theta &= \arctan\left(\frac{b}{a}\right) \\ &= 32.98\Delta A(\lambda) \end{aligned} \quad (2.4)$$

where b and a are the axes of the resulting ellipse. The molar ellipticity $[\theta]$ is commonly given by:

$$\begin{aligned} [\theta] &= 3298 \frac{\Delta A(\lambda)M}{LC} \\ &= 100 \frac{\theta M}{LC} \end{aligned} \quad (2.5)$$

Specific structures in proteins and peptides give rise to different CD spectra. Apart from the aromatic residues which give a tertiary structure fingerprint in near-UV (260 - 320 nm) and organic co-factors which can also create a CD signal, specifically, the secondary structure of a peptide or protein is visible by peptide bond absorption in far-UV, below 240 nm. The associated transitions of electrons in the peptide bond are $n \rightarrow \pi^*$ at 215-222 nm and $\pi \rightarrow \pi^*$ at 190 nm. The intensity and energy of these

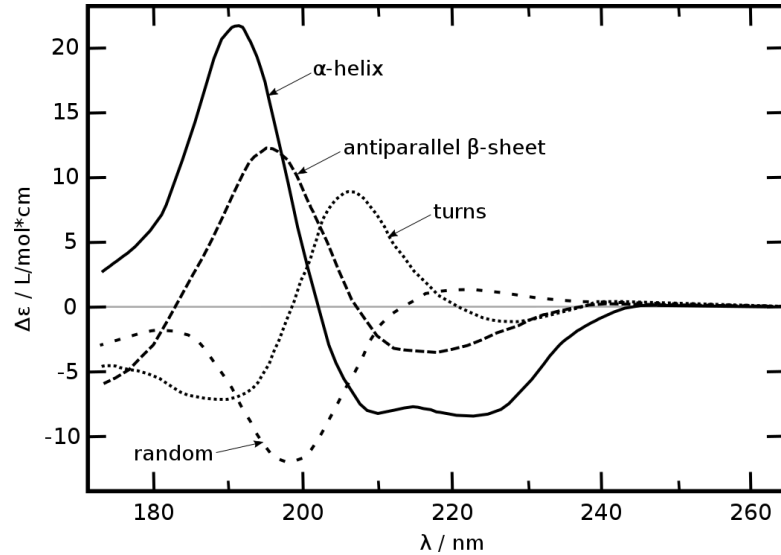


Figure 2.2: Schematic protein secondary structure CD spectra (adapted from [141]).

transitions depends on the backbone dihedral angles ϕ and ψ , the secondary structure elements. Common secondary structure features in protein CD spectra are illustrated in Figure 2.2. Lower wavelength CD data (< 190 nm) is generally only accessible using the intense light of a synchrotron beam as a light source. This synchrotron radiation circular dichroism (SRCD) also significantly improves the signal-to-noise-ratio.

As the CD signal is dependent on the number of peptide bonds, and therefore on number of residues N , an ellipticity per mean residue weight ($\text{mrw} = \frac{M}{N-1}$) $[\theta]_{\text{mrw}}$ is commonly used:

$$[\theta]_{\text{mrw}} = \frac{\theta_{\text{mrw}}}{10LC} \quad (2.6)$$

2.1.2.3 Molecular Dynamics

The following sources were used in this section: [126, 124, 125, 142, 143, 144, 145].

Molecular Dynamics (MD) is a computer simulation technique that describes the movement of molecules on the atomic scale. MD is widely used to study different multi-atomic systems in physics, chemistry, material science and biology. In order to determine the time-dependent behaviour of molecules in the system MD numerically solves Newton's equations for the motion of particles. This classical approach relies on

approximations which lower the computational costs and allows for the study of larger systems on a longer time scale. Specifically, the electron and nuclear degrees of freedom are separated (*Born-Oppenheimer approximation*), the potentials are described by a force field and the dynamics are described by solving Newton's equations of motion. However, these approximations commonly limit MD to studies without chemical reactions and the accuracy of simulations depends on how well quantum mechanics of the system can be simplified by these approximations.

The interatomic and intermolecular interactions between the particles are described by forcefields (e.g. CHARMM [146, 147, 148], AMBER [149], OPLS [150]). These forcefields are an empirical approximation to calculate the potential energy of the system as a sum of pair-potentials by describing the energetics between covalently bonded atoms (harmonic bond-stretching U_{bond} , bond-angle vibrations U_{angle} and proper and improper dihedral rotations $U_{dihedrals} + U_{impropers}$) and interactions between all pairs of atoms by the Coulomb potential $U_{Coulomb}$ and Lennard-Jones potential U_{LJ} (Section 1.1.4.1):

$$\begin{aligned} U &= U_{bonded} + U_{non-bonded} \\ &= U_{bond} + U_{angle} + U_{dihedrals} + U_{impropers} + U_{LJ} + U_{Coulomb} \end{aligned} \quad (2.7)$$

For a system consisting of N particles ($i = 1, 2, 3 \dots N$) the equations of motion are

$$m_i \frac{d^2 r_i(t)}{dt^2} = F_i = -\nabla_i U(r_i) \quad (2.8)$$

where m_i is the particle mass, t is the time, r_i is the position, F_i is the force and U_i is the potential.

As a basic algorithm, MD generates new structures from a starting ensemble of atoms and molecules by calculating the potential and force functions, moving the atoms accordingly over a short time step (Δt , usually few fs) and recalculating potentials and forces.

How to simulate water molecules is an important factor in biological MD simulations. As a solvent it can be represented as a continuous medium (implicit solvation)

or as explicit solvent molecules which is used throughout for this work. Standard water models exist, including SPC, SPC/E (Extended Simple Point Charge [151]) and TIP (Transferable Intermolecular Potential [152]) models.

2.2 Diffraction

The following sources were used in this section: [119, 153, 154, 155].^a

2.2.1 Fundamentals

The *wave-particle duality* is the fundamental quantum physics concept that objects can have both wave-like and particle-like properties. The de Broglie equation relates wavelength λ to the particle momentum p

$$\lambda = \frac{h}{p} \quad (2.9)$$

where h is Planck's constant and the momentum (p) is defined as the product of mass and velocity of the particle ($\mathbf{p} = m\mathbf{v}$).

Scattering occurs when particles experience a change in momentum due to obstructions, like other particles, in their pathway. If the kinetic energy of the incident particle is unchanged and only the direction is altered, the scattering is considered *elastic* whereas if the incident particle also loses or gains energy to or from the sample, this is called *inelastic* scattering (e.g. Raman scattering). *Coherently* scattered wavelengths have a phase relationship with one another while *incoherently* scattered particles have no phase relationship. *Diffraction* is elastic, coherent scattering that can be used to probe the structure of a sample system.

Diffracted waves can produce *interference* scattering when two diffracted waves superpose within the same spatial region resulting in a new wave. This interference is

^aThe mathematical formalism used to describe diffraction is based on *Elementary Scattering Theory* by D.S. Sivia [119].

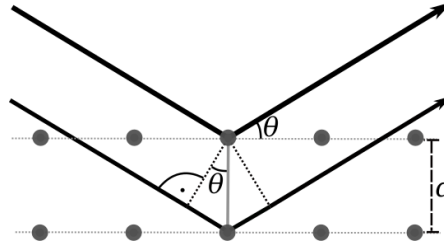


Figure 2.3: Bragg diffraction: Two waves with identical wavelengths scattered by a crystal with an interplanar distance d and a scattering angle θ .

constructive when the two waves are in phase, enhance each other and the resultant wave has a higher amplitude; or destructive when the waves are out of phase and cancel each other to produce a wave with a smaller amplitude or no wave at all.

Constructive interference can be used to obtain spatial information of a target structure. This basic principle for lattices is described by Bragg's law

$$n\lambda = 2d \sin \theta \quad (2.10)$$

where d is the distance between lattice planes, λ is the wavelength, θ is the scattering angle^b between lattice plane and the scattered wave and n is a positive integer (Figure 2.3). The scattering angle 2θ is between incident and scattered wave. θ can be at a maximum of 90° when the particle is scattered in the opposite direction as the incident wave. Therefore it is clear from Equation 2.10 that larger structural distances d can be measured at smaller (half-)scattering angles θ (small-angle scattering) and, conversely, smaller distances are measured at larger scattering angles (wide-angle scattering). It should be noted that this interference is only observed if the wavelength λ and the structural distance d are comparable in size.

In principle, a diffraction experiment aims to measure constructive inference of particles with known incident wavelengths to obtain information on the structure of the sample.

^bThese descriptions will, to begin with, be focused on two-dimensional examples. Three-dimensional aspects will be considered from Section 2.2.4 onwards.

2.2.2 Scattering Probes

Commonly used diffraction probes are neutrons, electrons and X-ray. While electron scattering is often employed to examine surface layers of materials [156] and sub-atomic properties [157], neutron and X-ray scattering are used to determine micro to atomic scale ($\approx 10^{-7} - 10^{-11}\text{m}$) structures. While all of the scattering techniques are based on the same fundamentals, the neutrons are scattered by the target nuclei or through electromagnetic interaction of the magnetic moments of the neutron and the target magnetic fields^c whereas X-ray photons are scattered through an interaction with the electron clouds surrounding the nuclei.

Neutrons are subatomic electrically uncharged particles with a spin (I) of $\frac{1}{2}$. Protons (p) and neutrons (n) constitute the nucleus of every isotope, with the exception of hydrogen nuclei that are only one proton. Between these nucleons acts a strong nuclear force that holds them together. After the existence of an uncharged nucleon was predicted by Rutherford in 1920 [158] neutrons were discovered by Chadwick in 1932 [159, 160]. By reacting α -rays with beryllium neutrons were produced. Due to their lack of charge, neutrons are highly penetrating, yet their interaction is comparably weak and non-damaging.

Although probably not the first to produce X-rays with a Crookes discharge tube, Röntgen was the first who described them in 1895 [161]. X-ray photons are electromagnetic radiation ($\lambda = 10^{-11} - 10^{-8} \text{ m}$, $I = 1$) that interact with the orbital electrons of the sample. This makes them an effective probe for heavy atoms, yet lighter elements, that are more common in biological samples, are often not well resolved in an X-ray diffraction experiment. Although X-ray is an ionising radiation that can disrupt molecular bonds, X-ray is the most commonly used probe for biological (often crystalline) samples.

2.2.3 Scattering by a Single Fixed Atom

Neutrons and X-ray can be described by a wave function ψ :

^cThis magnetic scattering will not be discussed in this work as the presented neutron scattering experiments were performed with non-magnetic samples and an unpolarised neutron beam.

$$\psi = \psi_0 e^{i(\mathbf{k}\mathbf{r} - \omega t)} \quad (2.11)$$

where the wave vector $\mathbf{k}$ points along the wave trajectory, $\mathbf{r}$ is the position vector, ω is the angular frequency, t is time and $i = \sqrt{-1}$. ψ_0 is a complex constant related to the flux of particles, $\Phi = |\psi_0|^2$. The wave vector modulus is related to the wavelength λ and therefore to the kinetic energy and the momentum of the wave:

$$k = \frac{2\pi}{\lambda} \quad (2.12)$$

Scattering results in a change in momentum ($\mathbf{P}$) and, in case of inelastic scattering, energy (E). The momentum transfer ($\mathbf{P}$), is determined by the difference between the wave vectors of incident $\mathbf{k}_i$ and outgoing $\mathbf{k}_f$ wave and the scattering vector $\mathbf{Q}$:

$$\begin{aligned} \mathbf{P} &= \hbar \mathbf{k}_i - \hbar \mathbf{k}_f = \hbar \mathbf{Q} \\ \Leftrightarrow \quad \mathbf{Q} &= \mathbf{k}_i - \mathbf{k}_f \end{aligned} \quad (2.13)$$

The energy transfer, E , is determined by the difference in radial frequencies ω :

$$\begin{aligned} E &= \hbar \omega \\ \Leftrightarrow \quad \omega &= \omega_i - \omega_f \end{aligned} \quad (2.14)$$

For elastic scattering the initial (ω_i) and scattered (ω_f) radial frequency are equal and their difference and the energy transfer equals zero. Additionally, the wave vectors ($\mathbf{k}_i$ and $\mathbf{k}_f$) are equal in magnitude and from this triangle (Figure 2.4) and Bragg's law the following relationships can be derived:

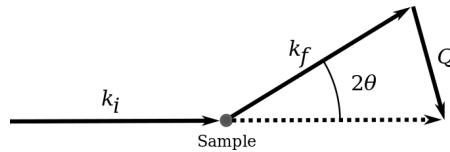


Figure 2.4: Geometric relationship between incident ($\mathbf{k}_i$) and outgoing wave vector ($\mathbf{k}_f$) and scattering vector ($\mathbf{Q}$) when a particle is scattered by a sample.

$$\begin{aligned}
k_i &= k_f = \frac{Q}{2 \sin \theta} = \frac{2\pi}{\lambda} \\
\Leftrightarrow \quad Q &= \frac{4\pi \sin \theta}{\lambda} \\
\Leftrightarrow \quad Q &= \frac{2\pi}{d}
\end{aligned} \tag{2.15}$$

The way a target atom scatters an incident wave is described by the scattering factor $f(\lambda, \theta)$ which is a probability weighting of a particle with a specific wavelength λ to be reflected at an angle θ . For elastic scattering ($k = k_i = k_f$) the incident linear wave (in z -direction, ψ_i) and the isotropic scattered wave (in r -direction, ψ_f) are related (Figure 2.5):

$$\begin{aligned}
\psi_i &= \psi_0 e^{ikz} \\
\psi_f &= \psi_0 f(\lambda, \theta) \frac{e^{ikr}}{r}
\end{aligned} \tag{2.16}$$

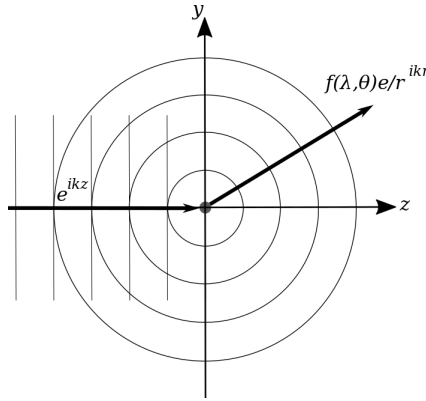


Figure 2.5: Incident (e^{ikz}) plane wave travelling in z direction and outgoing isotropic scattered wave ($f(\lambda, \theta) \frac{e^{ikr}}{r}$).

ψ_0 , in this case, subsumes the temporal variation and phase offset ($\psi_0 = |\psi_0|e^{i(-\omega t)}$) and $\mathbf{r}$ is parallel to $\mathbf{k}_f$ when the target atom is at the origin. For neutron scattering by individual nuclei (and not magnetic interaction), the form factor is an isotope-specific constant^d, $-b$, the scattering length, which is invariant with respect to wavelength (λ) and scattering angle (θ):

^dThe imaginary part of the complex scattering length is negligibly small.

Table 2.1: Bound coherent and incoherent neutron scattering lengths for some common elements and nuclides [164]. Z denotes the atomic number, I the spin.

Isotope	Z	I	b_{coh}	b_{incoh}
^1H	1	1/2	-3.7406	25.2479
^2H	1	1	6.6710	4.0430
^{12}C	6	0	6.6511	0.0000
^{13}C	6	1/2	6.1990	-0.5290
^{14}N	7	1	9.3720	2.0200
^{16}O	8	0	5.8030	0.0000

$$f_{neutron}(\lambda, \theta) = -b \quad (2.17)$$

The interaction of neutron and nuclei depends on the spin combination of the target nucleus and the neutron. The coherent scattering length corresponds to the weighted average of scattering lengths of different nuclear spin states ($\langle b \rangle = b_{coh}$). The corresponding root-mean-square deviation (standard deviation) of scattering lengths of different nuclear spin states is the incoherent scattering length ($\Delta b = b_{incoh}$). Consequently, isotopes with a spin of $I = 0$ have an incoherent scattering length of 0. The magnitude of b indicates the strength of neutron scattering by the nucleus whereas the sign indicates, for thermal neutrons (Section 2.2.7.1), if the scattered wave is in phase ($b < 0$) or out of phase ($b > 0$) of the incident wave. Table 2.1 shows the isotope specificity of the scattering lengths for some common elements in biological compounds. Most notably is the coherent scattering length of hydrogen negative and for deuterium positive. This difference in sign can be exploited by isotopic substitution in order to single out or mask specific interatomic interactions [162, 163].

For X-rays the form factor is dependent on the atomic number Z , but not on the isotope, and decreasing with increasing Q and therefore a function of θ and λ . These X-ray form factors have the same sign for all elements:

$$f_{X-ray}(\lambda, \theta) = Zg(Q)r_e \quad (2.18)$$

where $g(Q)$ is a decay function and r_e is the classical radius of the electron, the Thomson

scattering length ($r_e = 2.818 \cdot 10^{-15}$ m).

2.2.4 Scattering Cross Section

In diffraction experiments the measured quantity is, principally, the number of scattered particles per time (R) and solid angle ($\Delta\Omega$) relative to the incident flux of particles (Φ) and the number of scattering centres in the sample (N). This differential cross section is defined by

$$\frac{d\sigma}{d\Omega} = \frac{R(2\theta, \phi)}{N\Phi\Delta\Omega} \quad (2.19)$$

The solid angle (Ω) gives the area (A) of a spherical shell defined by the two spherical coordinate angles (the scattering angle θ and ϕ the rotation about $\mathbf{k}_i$) per squared sphere radius ($\Omega = A/r^2$). If the scattering experiment is made as a function of the energies of the scattered particles, E_f , then this quantity is defined by the partial differential cross section, $\frac{d^2\sigma}{d\Omega dE_f}$, which can be related to the differential cross section by

$$\frac{d\sigma}{d\Omega} = \int_0^\infty \left(\frac{d^2\sigma}{d\Omega dE_f} \right) dE_f. \quad (2.20)$$

However, as inelastic scattering can be minimised by the experimental setup and subsequent data correction (Section 2.2.7.3) the total scattering can usually be approximated as elastic ($R \approx R_{el}$), ignoring inelastic scattering.

$$\frac{d\sigma}{d\Omega} \approx \left(\frac{d\sigma}{d\Omega} \right)_{el} = \frac{R_{el}}{N\Delta\Omega\Phi}. \quad (2.21)$$

The cross section (σ , in units of barns = 10^{-28} m²) is the hypothetical scattering rate per incident flux ($\sigma = \frac{R}{\Phi}$), the effective nucleus scattering target area and the probability of a specific nucleus to interact with a neutron.

This interaction can, in general, be scattering or the capturing and absorption of the neutron into the nucleus. The total cross section is therefore the sum of the scattering

cross section and absorption cross section ($\sigma_{total} = \sigma_{scatt} + \sigma_{abs}$) and the scattering cross section the sum of incoherent and coherent cross section ($\sigma_{scatt} = \sigma_{incoh} + \sigma_{coh}$). The cross section can be related to the scattering length b (coherent or incoherent) by

$$\sigma = 4\pi b^2 \quad (2.22)$$

Figure 2.6 illustrates the scattering cross sections for hydrogen, deuterium, carbon, nitrogen and oxygen. The total neutron scattering cross section does not increase with the atomic number, whereas the X-ray scattering cross section does.

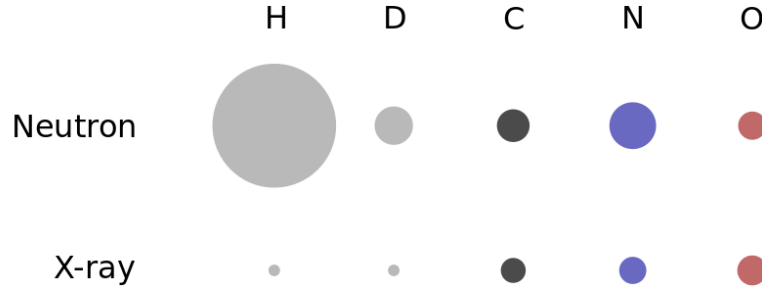


Figure 2.6: X-ray (at 100 keV) [165] and neutron scattering [164] cross sections for hydrogen, deuterium, carbon, nitrogen, oxygen, adapted from [166].

2.2.5 Extracting Structural Information from Scattering Data

For a multiple atom sample the scattered wave function ψ_f will be influenced by other atoms, labelled by the index j . The new wave function ψ_f with the summed contribution from all N atoms is given by this mathematical representation of the superposition principle (Figure 2.7):

$$\psi_f = \psi_0 e^{i\mathbf{k}_f \mathbf{r}} \sum_{j=1}^N f_j(\lambda, \theta) \frac{e^{i\mathbf{Q} \mathbf{r}_j}}{|\mathbf{r} - \mathbf{r}_j|} \quad (2.23)$$

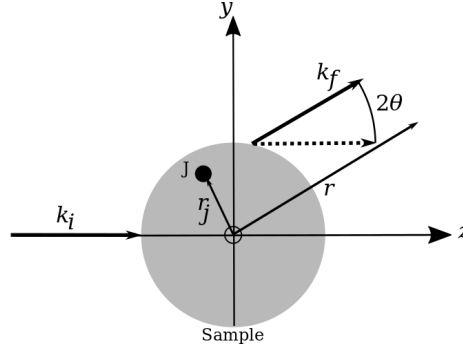


Figure 2.7: Geometric relationship between atom J in the sample, located at r_j relative to the origin at the centre of the sample and r as the distance to the detector.

where $f_j(\lambda, \theta)$ is the interaction characteristics of the j^{th} atom, located at $\mathbf{r}_j$. In contrast, the influence of the scattered waves on the incident neutron beam and multiple scattering effects are ignored because the scattering is comparably weak. This simplification is known as the *Born approximation*. Given that the distance to the detector is much larger than the distances within the sample ($|\mathbf{r} - \mathbf{r}_j| \approx r$) and combining these approximations with Eq. 2.19 and the definition of the incident flux ($\Phi = |\psi_0|^2$) and scattering rate ($R = \sigma\Phi$) this leads to the following relationship of the differential cross section for elastic scattering of an assembly of N atoms:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{el}} \propto \left| \sum_{j=1}^N f_j(\lambda, \theta) e^{i\mathbf{Q}\cdot\mathbf{r}_j} \right|^2 \quad (2.24)$$

The proportionality results from the right-hand side not being normalised by N .

This elastic scattering can itself be divided, for a multi-atom system, into two components of self and distinct scattering contributions:

$$\begin{aligned} \left(\frac{d\sigma}{d\Omega}\right)_{\text{el}} &\propto \left(\frac{d\sigma}{d\Omega}\right)_{\text{self}} + \left(\frac{d\sigma}{d\Omega}\right)_{\text{dist}} \\ \left(\frac{d\sigma}{d\Omega}\right)_{\text{self}} &= \sum_{j=1}^N |f_j(Q)|^2 \\ \left(\frac{d\sigma}{d\Omega}\right)_{\text{dist}} &= F(Q) = 2 \sum_{j>k} \text{Re}\{f_j(\mathbf{Q})f_k(\mathbf{Q})^* e^{i\mathbf{Q}\cdot(\mathbf{r}_j - \mathbf{r}_k)}\} \end{aligned} \quad (2.25)$$

where Re signifies the real term of the complex number and $f_j(Q)$ the form factor of

atom j (i.e. the scattering length b for neutron scattering), located at r_j , as a function of Q . The self scattering accounts for scattering of individual nuclei which does not contain any structural information. The distinct scattering represents the structural factors $F(Q)$, the correlation between the j^{th} and k^{th} atoms and contains structural information on their atomic positions.

Converting and normalising the distinct scattering term for neutrons to the atom concentrations results in

$$F(Q) = \sum_{J \leq K} (2 - \delta_{JK}) c_J c_K b_J b_K (S_{JK}(Q) - 1) \quad (2.26)$$

where $F(Q)$ summates all partial structure factors $S(Q)$. For X-rays, the neutron scattering lengths can be replaced for X-ray form factors. The Kronecker delta function, δ_{JK} , equals 1 if $J=K$ and 0 otherwise, c_J and c_K are the concentrations of atomic species J and K , and b_J and b_K the coherent scattering lengths respectively. A partial structure factor for the atomic components K and J is described by

$$S_{JK}(Q) = 1 + \frac{4\pi\rho}{Q} \int_0^\infty r [g_{JK}(r) - 1] \sin(Qr) dr \quad (2.27)$$

where ρ is the atomic density and r the radial distance between K and J . A Fourier transformation of this equation will extract the radial distribution function $g_{JK}(r)$ and subsequently the coordination number n_J^K (Equation 2.1).

It is possible to obtain individual atomic $S(Q)$ and thereby $g(r)$ s directly from the (corrected) scattering data of systems with a small number of distinct atom types, often in combination with isotopic substitution. This isotopic substitution is the measuring of chemically identical, yet isotopically unique samples, taking advantage of the isotope specificity of neutron scattering lengths. However, this approach is usually not applicable for large biological samples because it is not possible to isotopically substitute all molecule sites and the data collection would take a very long time.

The predominant species in aqueous biomolecular solutions, like the peptide samples

in this thesis, is water. It is therefore not only necessary to correct the scattering signal for inelasticity, attenuation and self- and multiple scattering (Section 2.2.7.3), all of the structural features can only be extracted from a computational model of the system (Section 2.3).

For a system with M as the number of atom types the number of $S(Q)$ and $g(r)$ correlations N will be:

$$N = \frac{M(M+1)}{2} \quad (2.28)$$

This large number of correlations cannot be extracted from the scattering data alone, even combined with isotopic substitution.

2.2.6 Small-Angle Scattering

The larger KLL peptides (Chapter 5) were also examined by small-angle scattering to complement the atomic-scale structural information with information on the bulk structure as these techniques can determine the shape and size of larger molecules. In general, small-angle scattering uses a Q -range below $0.5 \text{ \AA}^{-1}$ and probes structures between approximately $10 - 1000 \text{ \AA}$. The underlying principles are according to Sections 2.2.1-2.2.5 and it should be noted that the scattering intensity, again, depends on the probe used because of the different scattering lengths of X-ray and neutrons. The information extracted from small angle scattering data from a solution ($I(Q)$) is a combination of structure factor $S(Q)$ and a form factor $f(Q)$ [167].

$$I(Q) = N_p V_p^2 (\rho_p - \rho_s)^2 S(Q) f(Q) \quad (2.29)$$

where N_p is the number of particles, V_p the particle volume and ρ are the scattering length of particle and solvent. By a Fourier transformations this scattering data gives a distance distribution function $p(r)$:

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty Q^2 I(Q) \frac{\sin(Qr)}{Qr} dQ \quad (2.30)$$

2.2.7 Neutron Scattering Experiments

2.2.7.1 The SANDALS and NIMROD Instruments

All neutron diffraction data in this thesis was collected at the ISIS facility (STFC, UK). The neutron diffraction data of liquids containing the shorter GPG and KGPGK peptides (presented in Chapters 3-5) was obtained with the Small-Angle Neutron Diffractometer for Amorphous and Liquid Samples (SANDALS, built 1989, [168], Figure 2.9). The longer amphiphilic KLL-15 and KLL-22 peptides (presented in Chapter 5) were measured using the Near and InterMediate Range Order Diffractometer (NIMROD, built 2009 [169]). Both instruments are specifically built for the study of disordered systems, such as liquids and amorphous materials. SANDALS can measure static structure factors ($S(Q)$) between 0.2 and 50 Å⁻¹. The NIMROD instrument measures over larger Q -range between 0.02–100 Å⁻¹.

Neutron production at ISIS [170] begins with an ion source, that produces negatively charged hydrogen ions (H⁻) from hydrogen (H₂) gas *via* an electric discharge. The H⁻ are accelerated and focused into bunches, which are 4.94 ns apart, by a radio frequency quadrupole accelerator. In a linear accelerator the H⁻ bunches are accelerated to 37 % of the speed of light. The ions pass a 0.3 µm thin aluminium oxide foil which removes the electrons, turning the H⁻ ions into H⁺, *i.e.* protons. The protons enter the synchrotron, a 163 m circumference ring of magnets, and are further accelerated into two bunches travelling at 84 % of the speed of light. After the bunches of protons are extracted from the synchrotron with magnets, they collide with a tungsten target and ‘spallate’ (around 11 neutrons per proton) from the target nuclei. The high energy protons drive neutrons from the tungsten nuclei, producing a high energy neutron pulse. The spallation source for the SANDALS instrument at target station 1 produces neutron pulses at a 40 Hz frequency. Target station 2, which NIMROD uses, produces neutron pulses at a frequency of 10 Hz.

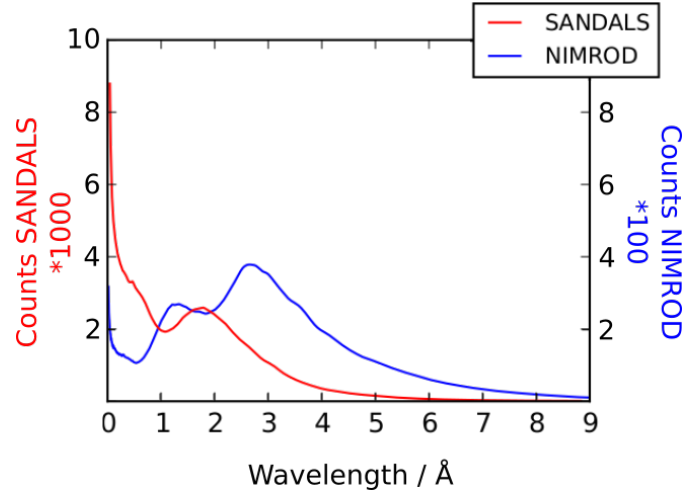


Figure 2.8: The spectrum of the moderated neutrons used by SANDALS and NIMROD.

This method of neutron production via heavy nuclei emitting particles after being hit by high energy particle pulses is called (pulsed) spallation (an example of a continuous spallation source is SINQ in Villingen, Switzerland). Another method to produce a continuous flux of neutrons is used in nuclear reactors, such as located at the Institut Laue-Langevin (ILL) in Grenoble, France; where neutrons are produced by nuclear induced fission.

A moderator containing hydrogenated liquids is used to reduce the neutron energy *via* inelastic scattering to be suitable for condensed matter experiments [154] before they are directed to the different instruments. Hydrogenated moderators are used because of the large absorption cross sections of hydrogen nuclei. Moderation reduces the neutron energy to approximately 0.025 eV. These neutrons used by the SANDALS and NIMROD diffractometers are called thermal neutrons and have a Maxwellian wavelength distribution (Figure 2.8). Higher energy neutrons are called epithermal neutrons. The flux of thermal neutrons (ϕ) is therefore related to the moderator temperature (T) and the energy of the incident neutrons (E) and the Boltzmann constant (k_B):

$$\phi \propto e^{\frac{-E}{k_B T}} \quad (2.31)$$

The SANDALS moderator contains liquid methane at 110 K. The NIMROD moderator contains liquid H₂ at 17 K. The moderated neutron pulses with wavelength between

0.05 and 4.95 Å for SANDALS or 0.04–10 Å for NIMROD are focused in a collimator. Neutrons with shorter wavelengths will be faster and arrive at the instrument detector first. Therefore a time-dependent (time of flight, TOF) experiment can resolve the neutrons by their wavelength for a pulsed neutron beam and monochromation of the beam is unnecessary. Figure 2.8 shows the spectra of moderated neutrons for NIMROD and SANDALS.

SANDALS has 18 groups of detectors over a range of 3.8–41° at a distance of 0.75–4.0 m to the sample. NIMROD also has 18 detector groups covering a range 0.5–40°. For SANDALS this makes a Q -range of 0.1–50 Å⁻¹ accessible which corresponds to structure distances between 0.1–62.8 Å. NIMROD measures a Q -range of 0.02–100 Å⁻¹ corresponding to 0.06–314 Å in real space. The SANDALS and NIMROD instrument uses ZnS scintillation detectors, which are doped with silver and lithium ions. This neutron detection begins with the following reaction



where the Lithium nucleus absorbs a neutron and emits helium (⁴He, an α -particle), tritium (³H) and 4.8 MeV of energy. This energy excites an electron within the ZnS lattice which combines with an electron hole to an exciton, a quasi-particle. Upon this exciton reaching a silver atom a photon is emitted which passes through a photomultiplier tube (PMT) to be detected [171].

The scattering data is collected at a vacuum in the sample tank (1–10⁻² mbar) to minimise signals due to scattering by air molecules and attenuation (the actual sample pressure is 1 atm, due to the vacuum tight sample containers). A set of B₄C jaws can adjust the size of the neutron beam to fit to the sample size.

2.2.7.2 Measurements

The aim of any diffraction experiments is to measure the differential cross section $(\frac{d\sigma}{d\Omega})_{\text{coh}}$ of the sample on an absolute scale. Four types of measurements are necessary to obtain

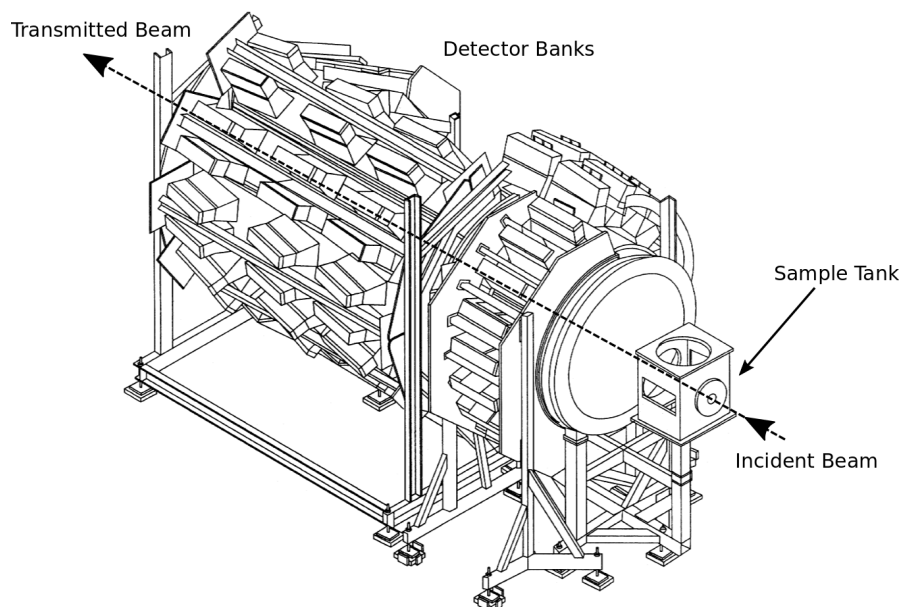


Figure 2.9: Schematic diagram showing SANDALS at the ISIS neutron source (modified from [168]). The principle setup is very similar to NIMROD [169].

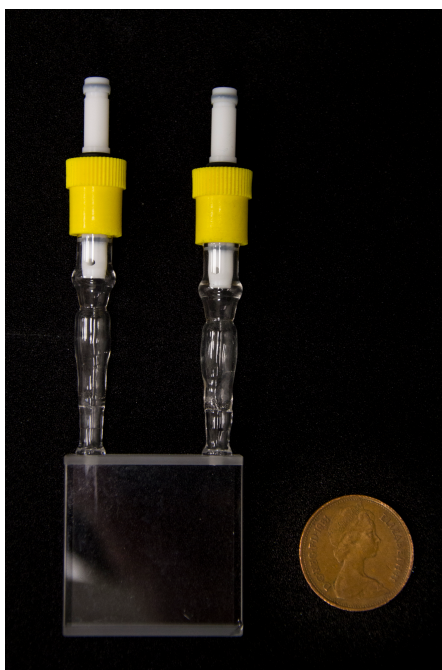


Figure 2.10: Quartz cell used in the neutron diffraction experiments next to a GBP penny for size comparison. Photograph courtesy of Richard Gillams.

this data of a sample in a container: a standard for calibration, the sample background, the empty sample container and the sample in its container. The merging, processing and correcting of the data was carried using the Gudrun program [172], based upon the ATLAS package [173]. The type of container used in this thesis are quartz (SiO_2) cells (Figure 2.10) with vacuum tight Young's taps. The normalisation standard measured is a 3.5 mm vanadium plate, without any container, which is measured at the same conditions as the sample. Vanadium is an almost completely incoherent (Q -independent) scatterer with only small Bragg peaks. This allows the diffraction data to be put on an absolute scale, eliminating calibration constants [174]. The differential cross section of the sample is therefore obtained by:

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Sample}} = \frac{\left(\frac{d\sigma}{d\Omega}\right)_{\text{Sample in Container}} - \left(\frac{d\sigma}{d\Omega}\right)_{\text{Empty Container}}}{\left(\frac{d\sigma}{d\Omega}\right)_{\text{Vanadium Standard}} - \left(\frac{d\sigma}{d\Omega}\right)_{\text{Background}}} \quad (2.33)$$

As the incoherent scattering length of hydrogen is larger than the incoherent scattering length of deuterium, the measured total differential cross section is higher for hydrogen containing samples compared to deuterated samples (Figure 2.11).

2.2.7.3 Data Processing

The scattering datasets are corrected using Gudrun [172]. The aim of this correction is to extract the distinct scattering data in order to obtain a $F(Q)$ function for each sample. As a first stage the data of faulty detectors is removed. This 'purge' routine groups the detectors, compares their standard deviations and removes detectors with an unusually high or low standard deviation.

Following this, the data is corrected for several effects: Fluctuations of the incident beam are corrected for by comparing the data to the incident beam counts of the monitor detector. The detector dead time is accounted for and the fraction of neutrons transmitted by the sample is calculated by dividing the count rate of the transmission monitor, placed after the sample, by the count rate measured by the monitor detector. The transmission rate is also used to identify the sample by calculating the number of neutrons scattered relative to the expected number given by the sample composition (DCS Level).

Multiple scattering is accounted for by an iterative process [175]. Neutron absorption is accounted for by calculating the total cross section ($\sigma_{total} = \sigma_{scatt} + \sigma_{abs}$) from the transmission cross section ($\sigma_{trans} = e^{-\rho\sigma_{total}(\lambda)L}$), where L is the sample thickness and ρ the density). The baseline of scattering corresponds to the self scattering which needs to be subtracted to obtain the distinct scattering cross section for the sample.

This inelastic scattering effects are most difficult to correct for at low Q , contributing mostly to self scattering and affects mostly light atoms, such as hydrogen. This inelasticity is usually corrected for using the Placzek correction [176, 177] which uses a Taylor expansion of the total structure factor $F_{tot}(Q)$ to determine the elastic component. But this correction relies on the sample atoms being heavier than the incident neutrons and is therefore not sufficient in the case of biomolecular samples and was only used for the vanadium standards. For the aqueous biomolecular samples, the Q -dependent ‘pseudo-background’, including inelastic (self) scattering, was subtracted. This pseudo-background ($I'(Q_e)$) is calculated via a convolution of the data ($I(Q_e)$) with a top hat function ($T(Q)$) so that this correction only applies to low Q values.

$$I'(Q_e) = \int_{Q_T} I(Q_e - Q'_e) T(Q'_e) dQ'_e \quad (2.34)$$

where Q_e is the elastic scattering vector and Q_T is the top hat function width. In addition, a series of exponentials ($B(Q)$) is subtracted from the data in an *ad hoc* method

$$B(Q) = \sum_i A_i e^{\frac{-Q}{\gamma_i}} \quad (2.35)$$

where A_i and γ_i are the exponential parameters that need to be adjusted in Gudrun for each sample. There is also an automated iterative correction available in Gudrun. However, this function was found to be less accurate and was not used for the results presented in this thesis.

Figure 2.11 (left) shows the unprocessed differential cross section data of five isotopically distinct samples. Due to the larger incoherent cross section of hydrogen

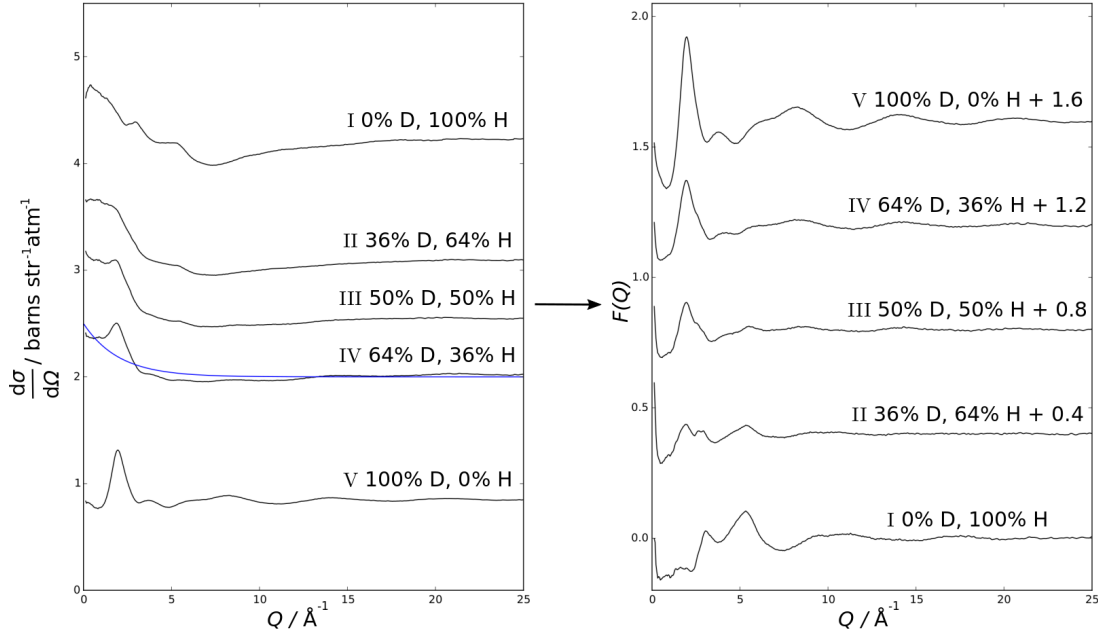


Figure 2.11: The GPG peptide in aqueous urea solution (Chapter 3) data before data processing (left) as differential cross sections ($\frac{d\sigma}{d\Omega}$) and processed data as $F(Q)$ (right) as solid lines. For Sample IV the exponential decay function ($B(Q) = 0.5e^{-Q/2}$) that is used to correct the data at low Q is also displayed as a blue line.

(≈ 80 barns) compared to deuterium (≈ 2 barns [164]), the differential cross sections are larger for samples that contain more hydrogen atoms. The inelastic scattering effects of hydrogen at low Q are also more pronounced for samples with a larger hydrogen content, where after normalisation and data correction these effects are minimised in the resultant $F(Q)$ data (Figure 2.11 (right)).

2.3 Empirical Potential Structure Refinement

After the neutron (and X-ray) scattering data has been collected and processed, the partial structure factors ($S(Q)$) and consequently the atomic-scale information of a system containing a large number of different atoms can only be obtained by a computational model that fits to the neutron or X-ray diffraction data (Section 2.2.5). EPSR (Empirical Potential Structure Refinement) is a software package specifically developed to create this computer model for disordered systems [178, 179]. A summary of the EPSR process is shown in Figure 2.12.

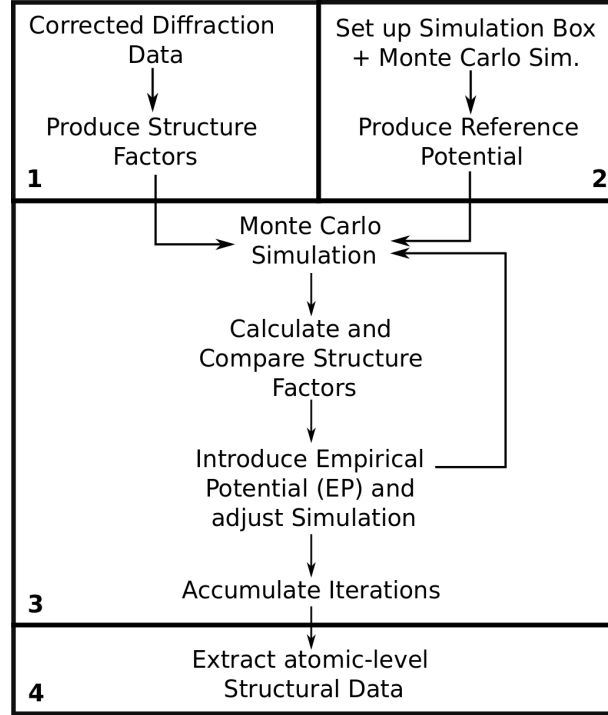


Figure 2.12: General EPSR flowchart: Parallel to completing the diffraction experiments and correcting the data (1) a computational reference model of the system is set up (2). The simulation is compared and adjusted to the experimental data, introducing the empirical potential (3). Finally, atomic-level information is extracted from the simulation (4).

The following description is based on the EPSR manual [180]:

2.3.1 The Reference Potential

As a first step EPSR requires the creation of a molecular model of the system with the correct molar ratios, density and atomic potentials. This model needs to be energetically equilibrated and gives the energy term of the simulation box, the reference potential U_{ref} . This potential consists of an intra- and inter-molecular part.

$$U_{Ref} = U_{Inter} + U_{Intra} \quad (2.36)$$

The intramolecular potential between atoms α and β is defined by:

$$U_{\text{Intra}}(r_{\alpha_i\beta_i}) = C \sum_i \sum_{\alpha\beta \neq \alpha} \frac{(r_{\alpha_i\beta_i} - d_{\alpha\beta})^2}{2w_{\alpha\beta}^2} \quad (2.37)$$

where r is the distance between the atoms compared to the average distance d , C is a constant, w is a broadening function to account for thermal fluctuations in atom positions ($w = \frac{d_{\alpha\beta}}{\sqrt{\mu_{\alpha\beta}}}$) and $\mu_{\alpha\beta}$ is the reduced mass of the atom pair ($\mu_{\alpha\beta} = \frac{M_\alpha M_\beta}{M_\alpha + M_\beta}$)

The intermolecular potential is a combination of Lennard-Jones and Coulomb potential (Section 1.1.4.1):

$$U_{\text{Inter}}(r_{\alpha\beta}) = 4\epsilon_{\alpha\beta} \left[\left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}} \right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}} \right)^6 \right] + \frac{q_\alpha q_\beta}{4\pi\epsilon_0 r_{\alpha\beta}} \quad (2.38)$$

The Lennard-Jones potential parameters for each atom pair $\sigma_{\alpha\beta}$ and $\epsilon_{\alpha\beta}$ are combined from the atomic parameters by the Lorentz-Berthelot mixing rules:

$$\sigma_{\alpha\beta} = \frac{1}{2}(\sigma_\alpha + \sigma_\beta) \quad \epsilon_{\alpha\beta} = \sqrt{\epsilon_\alpha \epsilon_\beta} \quad (2.39)$$

The reference potential parameters are often obtained from MD forcefields. Specifically, in this work, CHARMM36 [146, 147, 148] and the water potentials from the (extended) Simple Point Charge (SPC/E) [151] or the TIP3P model [152] were used as starting potentials for the EPSR simulations.

Long-range effects are eliminated by truncating the reference potentials. This is necessary to save computational time and does not have a significant impact on shorter-range interactions. The Lennard-Jones term of the reference potential is truncated using a smooth decaying function, $T(r)$:

$$T_{LJ}(r) = \begin{cases} 1, & \text{if } r < r_{\text{minpt}} \\ 0.5 \left(1 + \cos \pi \left(\frac{r - r_{\text{minpt}}}{r_{\text{maxpt}} - r_{\text{minpt}}} \right) \right), & \text{if } r_{\text{minpt}} < r < r_{\text{maxpt}} \\ 0, & \text{if } r > r_{\text{maxpt}} \end{cases} \quad (2.40)$$

The distances at which truncating starts r_{minpt} and ends r_{maxpt} can be modified according to size of the molecules in the system. The truncation factor for the Coulomb potential $T_C(r)$ is

$$T_C(r) = \left(1 - \frac{r}{r_{\text{maxpt}}}\right)^4 \left(1 + \frac{8r}{5r_{\text{maxpt}}} + \frac{2r^2}{5r_{\text{maxpt}}^2}\right) \Theta(r_{\text{maxpt}} - r) \quad (2.41)$$

where $\Theta(r_{\text{maxpt}} - r)$ is the Heaviside function; which equals 0 when $r_{\text{maxpt}} < r$ and 1 when $r_{\text{maxpt}} > r$. For the peptide systems studied in this thesis r_{minpt} and r_{maxpt} have been set to 9 and 12 Å respectively.

Due to computational limits the size of the simulation box can only be smaller than the actual scattering sample. However, the movement of the molecules within the box is not constrained by the box-size, as when a particle passes through one side of the (usually cubic) simulation box it re-enters on the opposite side. This is called the *periodic boundary condition*.

2.3.2 The Monte Carlo Simulation

The Monte Carlo method employed by EPSR moves atoms, molecules and rotational groups randomly and accepts or rejects these movements based on the Metropolis condition in order to obtain a lower energy configuration [181]. Specifically there are four types of movements that EPSR performs: atomic and molecular translations and molecular and group rotation. The Metropolis condition states that a move is accepted if the it results in a decrease in total energy ($\Delta U < 0$) of the simulation box; if the move increases the total energy ($\Delta U > 0$), the move can be accepted with the probability P . This probability depends on the movement type. For an atomic translation only the intramolecular potential determines the acceptance of movements:

$$P_{\text{atm}} = e^{-\Delta U_{\text{Intra}}} \quad (2.42)$$

Other molecular movements are influenced by other inter or intramolecular groups

and the probability of a movement acceptance is

$$P_{\text{mol}} = e^{-\left(\Delta U_{\text{Intra}} + \frac{\Delta U_{\text{Ref}} + \Delta U_{\text{EP}}}{kT}\right)} \quad (2.43)$$

It should be noted that the empirical potential U_{EP} , which is related to the experimental diffraction data and will be defined in the next section, is optional in this step of energy minimisation. Usually, the simulation is equilibrated only using the reference potential before the empirical potential is introduced.

2.3.3 The Empirical Potential

The empirical potential accounts for differences between the experimental structure factors $F(Q)$ and the simulated structure factors. It has been found that the most successful way to represent $U_{\text{EP}}(r)$ is as a series of Poisson functions:

$$U_{\text{EP}}(r) = k_B T \sum_{\mathbf{k}} C_{\mathbf{k}} p_{\mathbf{n}_{\mathbf{k}}}(r, \sigma_{\mathbf{r}}) \quad (2.44)$$

where

$$p_{\mathbf{n}}(r, \sigma_{\mathbf{r}}) = \frac{1}{4\pi\rho\sigma_{\mathbf{r}}^3(n+2)!} \left(\frac{r}{\sigma_{\mathbf{r}}}\right)^n e^{-\frac{r}{\sigma_{\mathbf{r}}}}. \quad (2.45)$$

$C_{\mathbf{k}}$ is a real positive or negative number, k_B is the Boltzmann constant, T is temperature, ρ is the atomic density of the system, n is a natural number and σ is a width function. A Fourier Transform converts $p_{\mathbf{n}}(r, \sigma_{\mathbf{r}})$ into reciprocal Q -space *via*

$$\begin{aligned} p_{\mathbf{n}}(Q, \sigma) &= 4\pi\rho \int p_{\mathbf{n}}(r, \sigma) e^{i\mathbf{Q}\mathbf{r}} d\mathbf{r} \\ &= \frac{1}{(n+2)(\sqrt{1+Q^2\sigma^2})^{(n+4)}} \left[2\cos(n\alpha) + \frac{(1-Q^2\sigma^2)}{Q\sigma} \sin(n\alpha) \right] \end{aligned} \quad (2.46)$$

where $\alpha = \arctan(Q\sigma)$. This allows the coefficient, $C_{\mathbf{k}}$ to be estimated directly from the

diffraction data and the reciprocal space equivalent of $U_{\text{EP}}(r)$ is given by

$$U_{\text{EP}}(Q) = \sum_{\mathbf{k}} C_{\mathbf{k}} p_{\mathbf{n}_{\mathbf{k}}}(r, \sigma_Q) \quad (2.47)$$

The coefficient $C_{\mathbf{k}}$ and therefore the empirical potential $U_{\text{EP}}(Q)$ is recalculated for each EPSR iteration and the total potential is used in the refinement to fit the simulation to the data by minimising the total energy U_{Tot} :

$$U_{\text{Tot}} = U_{\text{Ref}} + U_{\text{EP}} \quad (2.48)$$

2.3.4 Refining the Empirical Potential

To indicate the link between experimental structure factor $F(Q)$ (corrected for self- and inelastic scattering) and the empirical potential $U_{\text{EP}}(Q)$ the structure factor of the i^{th} experimental data set can set be written as

$$F_i(Q) = \sum_{j=1}^N \gamma_{ij} S_j(Q) \quad (2.49)$$

where j refers to an atomic pair $\alpha\beta$ and γ_{ij} is a weights factor related to the atomic concentrations c and the scattering lengths b (with $\delta_{\alpha\beta}$ as the Kronecker delta function):

$$\gamma_{ij} = (2 - \delta_{\alpha\beta}) c_{\alpha} c_{\beta} b_{\alpha} b_{\beta} \quad (2.50)$$

All weights factors γ_{ij} can be put into a $M \times N$ matrix with M as the number of data sets and N as the number of pair correlations. To obtain the partial structure factors, $S(Q)$, this weights matrix has to be inverted. Usually there are not enough isotopically labelled samples to directly invert the $M \times N$ matrix and a confidence factor, f , is assigned to each experimental dataset and a modified matrix with γ'_{ij} , is then formed, where input from the experimental data is accepted with the confidence f and input from the model is accepted with a confidence $(1-f)$:

$$\gamma'_{ij} = \begin{cases} f\gamma_{ij}, & \text{if } 1 \leq i \leq M \\ (1-f)\delta_{(i-M)j}, & \text{if } M \leq i \leq (M+N) \end{cases} \quad (2.51)$$

This leads to an overdetermined $(M+N) \times N$ matrix. By pseudo-inversion of this matrix it is then possible to obtain the partial structure factors $S(Q)$ and the coefficients $C_k^{(j)}$

$$C_{k,m+1}^{(j)} = C_{k,m}^{(j)} + \sum_{i=1}^M \gamma_{ji}^{-1} C_k^{(i)}. \quad (2.52)$$

which can be used to determine the empirical potential $U_{EP}(Q)$ as shown in Equation 2.44. This process is repeated until the difference coefficients and therefore U_{EP} reach a lower limit. The cycle of Monte Carlo simulation, structure factor and empirical potential calculation (Figure 2.12) is then continued in order to accumulate ensemble information of the simulation box.

It should be noted that the structure obtained by an EPSR simulation is not unique and represents one physically plausible result that is consistent with the experimental data.

2.4 ANGULA

ANGULA is a program which can be used to analyse the three-dimensional atomic-scale information from computational models of liquids and amorphous systems provided by EPSR or MD [182, 183].^e It allows for the relative positions and orientations of molecules and atoms to each other to be obtained. ANGULA can also be used to analyse and plot frequency distributions of molecular properties and their possible correlation (e.g. inter- or intramolecular distances, angles, dihedrals) and to select molecules with any of these criteria. Further, these analysis options can be combined with each other or other scripts to provide a comprehensive range of analytical techniques.

^eANGULA was written by Luis Carlos Pardo of the Universitat Politècnica de Catalunya

2.4.1 Assigning Coordinate Systems

As the first step in using ANGULA, a large number of coordinate files (> 1000) from EPSR (ATO files accumulated) or MD (PDB or XYZ files from MD trajectories) are used where orthonormal Cartesian coordinate systems are assigned to molecules or certain fragments of larger molecules chosen for analysis. As an example, the assignment of coordinate systems for urea and water (Figure 2.13) has been performed as follows:

Each coordinate system is defined *via* the addition of four pseudo-atoms: the first in the position of the origin ($\odot$) and the others ($\otimes$, $\odot$ and $\odot$) located on the intended x-, y- and z-axes. The origin of the coordinate system will usually be placed in the centre of a small molecule or at a specific site, such as a carbonyl group, of a larger molecule. In the case of urea, the origin is placed at the central carbon atom and the pseudo-atom $\odot$ is superimposed onto the same position as the urea carbon and the z-axis pseudo-atom $\odot$ is superimposed on one of the nitrogen atoms. The definition of the x- and y-axes is not easily accessible by using existing atoms. The positions of the related pseudo-atoms ($\otimes$ and $\odot$) need to be defined via the vector product. For the y-axis, the z-axis from origin to nitrogen and the vector from origin to the oxygen map out a plane to which the cross product resulting vector is perpendicular, according to the right-hand-rule. It should be noted that the vector product is not commutative, meaning that the order of vectors determines the direction of the resulting vector. As the vector product only defines a direction and not a position, the distance to the origin (usually 1 Å) needs to be defined as well for the positioning of $\odot$. Subsequently, the x-axis is defined by placing $\otimes$ at a distance of 1 Å to the origin on the vector product of the y- and z-axis. Similarly, an example coordinate system can be assigned to water with the origin at the water oxygen, the z-axis along one of the OH-bonds and the x- and y-axes defined *via* cross products.

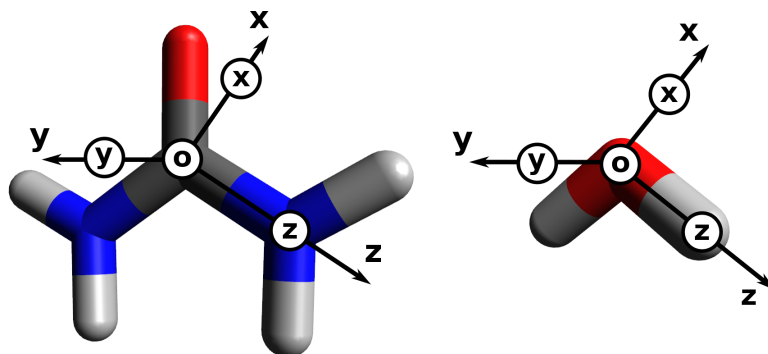


Figure 2.13: Example coordinate systems assigned to urea and water.

It should be noted that multiple coordinate systems can be added to the same molecular species. This can be useful if different parts of a larger molecule or multiple geometric properties are analysed.

2.4.2 Site Specific Analysis

The relative position of molecules to each other can be obtained by calculating a spatial distribution [183]. A central and a surrounding molecular species with coordinate systems can be chosen, e.g. water around urea. Typically, the distribution of origin-to-origin positions will be calculated. However, any atom or pseudo-atom close to the origin could be chosen. In the example, not only could the distribution of water oxygens around the urea carbon be calculated, but also the distribution of water oxygens around the urea nitrogen (i.e. the z-axis pseudo-atom). Figure 2.14 shows this urea nitrogen to water oxygen distribution as RDFs and as SDMs (spatial density maps [184]). This calculation can be used for nearest (or second/third...nearest) neighbour analysis (Figure 2.14 bottom), where for each frame the nearest (or second/third...nearest) surrounding molecule atoms to the central molecule atom will be used; or for a distance range analysis where all surrounding molecule atoms within a distance to the central molecule atom will be selected (Figure 2.14, top).

Site-specific distributions are displayed as spatial density maps (SDMs) in this work [183].^f Each surrounding atom within the defined distance range or as the nearest

^fThe script to make SDMs has been developed by Sebastian Busch. Small modifications, like different colour-schemes for multiple molecule neighbours, have been implemented by myself.

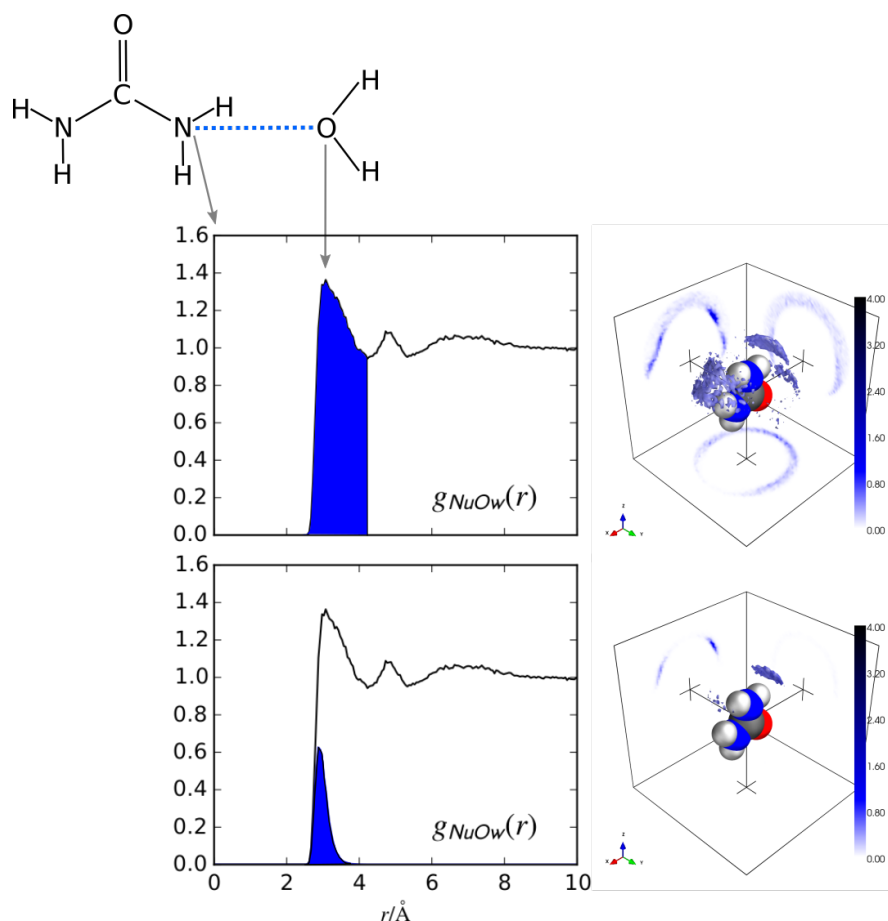


Figure 2.14: Site specific analysis of water oxygens (Ow) around one of the urea nitrogens (Nu) at a distance range from 0–4.2 Å (top) or nearest neighbour Ow (bottom). Corresponding radial distribution functions (left) with the blue areas indicating the atoms included in the spatial density map (right, each showing the top 10 % atoms as a blue cloud). The data has been taken from the EPSR simulation of the GPG peptide in aqueous urea solution (Chapter 3).

neighbour is represented by a density point in the SDM. A nearest neighbour analysis will therefore always contain as many data points as the number of frames times the number of central molecules per frame. For the distance range calculation all points within the defined distance will be included and any SDM is normalised to the number of frames and the number of central molecules within the simulation box. Therefore, a comparison between different simulation boxes may be valid if the relative molecular ratios are the same.

The density of points is indicated by a scale bar, which relates the colour-scheme to the absolute density of atoms per $\text{\AA}^3$. In most cases it is not practical to display 100 % of the data points as it will represent all conformations, not only the most likely, and only the top 50 %, or less, within the density distribution will be shown. Consequently, the displayed data point clouds make up a surface of equal density (i.e. isopycnic). The SDMs also include displaced projections of centre cut-throughs in all three planes which are independent from the isopycnic level of displayed points.

ANGULA calculates the position of surrounding atoms as spherical polar coordinates $(r, \theta_{CM}, \phi_{CM})$ from the central molecule origin. The calculation of SDMs relies on the conversion of spherical polar-coordinates calculated by ANGULA to Cartesian xyz-coordinates:

$$x = r \sin(\theta_{CM}) \cos(\phi_{CM}) \quad (2.53)$$

$$y = r \sin(\theta_{CM}) \sin(\phi_{CM}) \quad (2.54)$$

$$z = r \cos(\theta_{CM}) \quad (2.55)$$

where r is the distance between central and surrounding molecule origin and $_{CM}$ distinguishes these angles from the relative orientation Euler angles $_{OR}$ (Section 2.4.4).

2.4.3 Whole Molecule Analysis

It is also possible to calculate the density of surrounding molecules not only for one specific central atom but for a whole molecule [185]. This Whole Molecule Analysis (WMA) displays the distribution of neighbouring molecule atoms (usually the origin pseudo atoms) within a specific distance range from any of the central molecule atoms. Similar to the site specific analysis, this calculation relies on the spherical polar coordinates of the neighbouring atoms to the central molecule coordinate system. For small central molecules this method can easily be applied without further steps. Larger and more flexible molecules, like peptides, need to be analysed for their conformational properties as well. As the position of the neighbour atom is only determined by one coordinate system, the further this neighbour atom is away from the central origin, the more the position will be influenced by the central molecule conformation. There are multiple criteria to determine and select for a specific conformation, such as intramolecular distances, bond angles or dihedrals (2.4.5). As a WMA can only be performed for a specific central molecule conformation, a WMA for larger and more flexible central molecules, like proteins or lipids would be very challenging. Instead, a site specific analysis of certain parts of the central molecule would be more informative.

Figure 2.15 (right) shows a simple example of the WMA of water oxygens around the whole urea molecule. The corresponding RDF (Figure 2.15 left) shows two peaks. The first at 2.0 Å and the second at 2.8 Å. These peaks have been separated in the WMA where the density clouds represent the molecules between 0 and 2.3 Å (blue) and from 2.3 to 4.0 Å (teal).

2.4.4 Relative Molecule Orientations

The orientation of molecules relative to each other can be determined from an ANGULA calculation over a distance range or for a nearest neighbour analysis. This technique needs a coordinate system assigned to the central and the surrounding molecule and uses Euler angles (θ_{OR} , ϕ_{OR} and ψ_{OR}) to quantify their relative orientation. This relative orientation depends on the position of the molecules to each other and can therefore only be calculated for a specific neighbour molecule position.

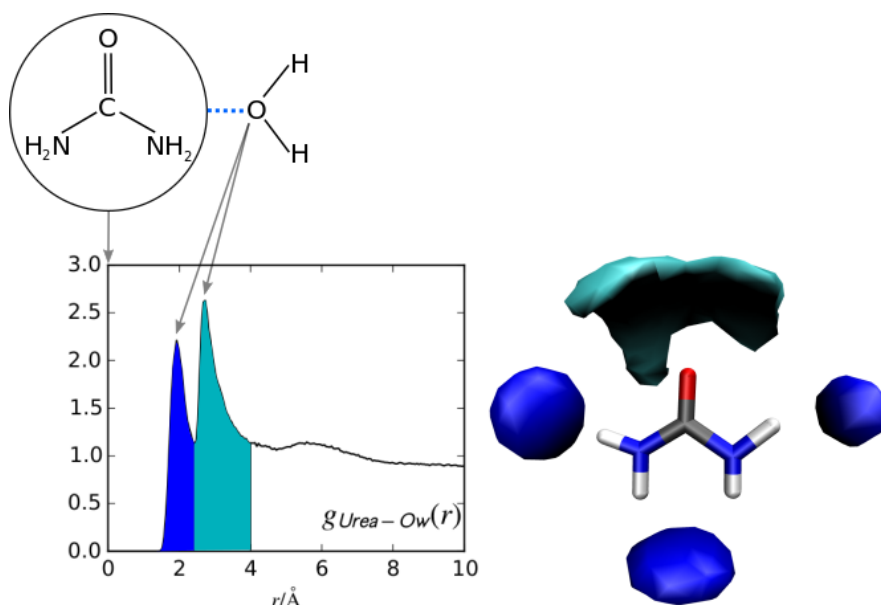


Figure 2.15: Whole molecule analysis of water oxygens (Ow) around urea. The radial distribution function (left) shows two hydration peaks. The area under the first and second peak correspond, according to their colour, to the density clouds of the WMA (right). These blue and teal clouds represent the top 50 % and top 30 % of Ow density respectively.

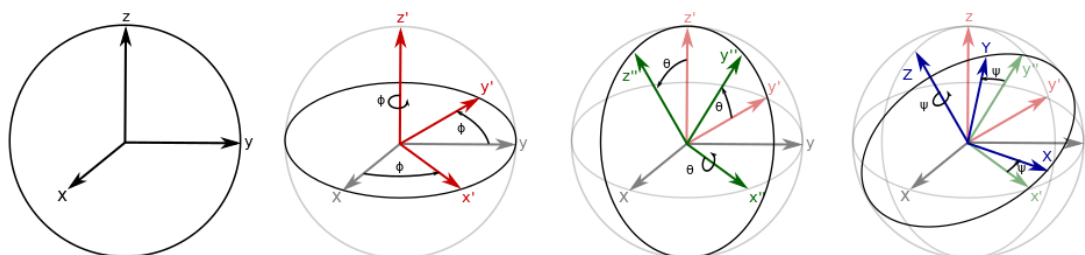


Figure 2.16: Definition of Euler angles (θ_{OR} , ψ_{OR} and ϕ_{OR}) for ANGULA

Figure 2.16 shows exemplarily how Euler angles are defined and used in ANGULA to describe the relative orientation of two coordinate systems (xyz, and XYZ): A rotation around the z-axis gives ϕ_{OR} , a rotation around the new x'-axis θ_{OR} and a rotation around the new (and final) Z-axis defines ψ_{OR} . For clarity, both coordinate systems have the same point of origin. A neighbouring molecule origin would, naturally, be at a distance from the central molecule origin.

Figure 2.17 shows how the information on position and orientation can be combined for the nearest neighbour water molecule to a urea nitrogen. The water position is determined by the distance r which can be extracted from the RDF ($r \approx 3.0$ Å) and the spherical polar angles θ_{CM} and ϕ_{CM} which can be depicted as 2D density plot. There

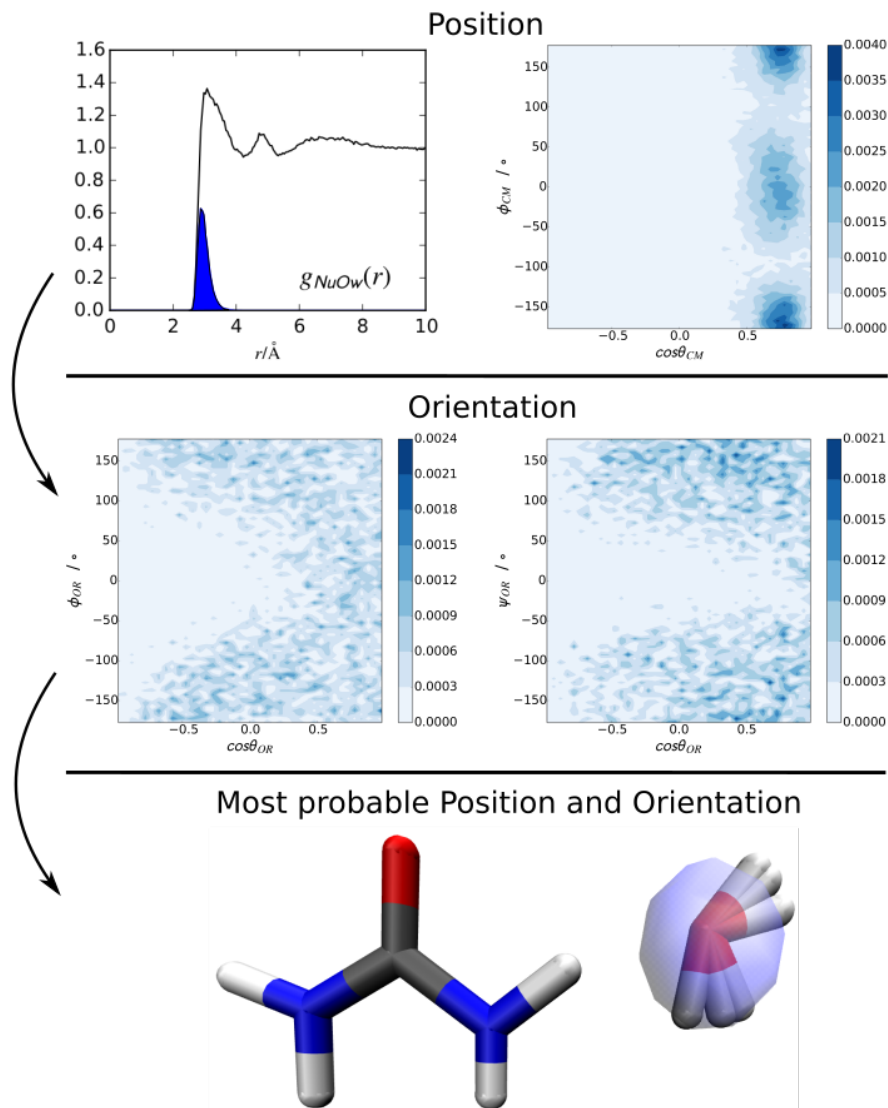


Figure 2.17: Combining the information on water position and orientation towards a urea nitrogen (Nu): The top panel shows the distance of nearest neighbour Ow to Nu in an RDF and the spherical polar angles θ_{CM} and ϕ_{CM} in a 2D density plot determining the position in space. The middle panel shows the relative orientation of water and urea molecule at the most probable position in 2D plots of the Euler angles θ_{OR} , ϕ_{OR} and ψ_{OR} . The bottom panel shows urea next to water molecules at the right orientation within the transparent water density cloud.

is a clear maximum for $\cos(\theta_{CM})$ to be at ≈ 0.75 and ϕ_{CM} to be at $\approx 180^\circ$. The water orientation at this position is determined by the Euler angles θ_{OR} , ϕ_{OR} and ψ_{OR} which are also depicted as two 2D density plots. These plots show a preference for ϕ_{OR} and ψ_{OR} to be $\approx 180^\circ$, whereas a clear preference for θ_{OR} cannot be extracted. This is also reflected in the representative figure at the bottom that shows urea next to water molecules at the preferred orientations within the transparent water density cloud which represents the preferred water position.

2.4.5 Further Applications

Combining the different ANGULA applications with each other is often necessary. Certain variables like dihedral angles or atomic distances can be found to be correlated with one another. The information about molecule position and orientation can be combined to determine the most probable position and orientation of molecules to each other. For WMA of large molecules, their conformational properties and the most probable conformation must be determined. For the peptides in this work the backbone dihedral angle distributions were usually determined. Overlaying 2D bivariate plots of the dihedrals ϕ and ψ for each residue can produce a Ramachandran plot (Section 1.1.2 and Chapter 5).

To determine interactions between all atoms of the same or different molecular species over a certain distance range, a contact map (CM) can be calculated, which is a matrix with the atom numbers of each molecule corresponding to the row or column number. This matrix can also be displayed as a heat map showing the relative density of the neighbouring molecule atoms to each other as a colour scheme.

Chapter 3

Influence of Urea on the GPG-NH₂ Peptide

3.1 Introduction

Similar to protein folding, the process that leads to protein unfolding and denaturation caused by chemical compounds, such as urea, in solution, is unclear. There are multiple theories as to how the presence of urea denatures proteins, yet there is no consensus on the mechanism by which this occurs on the atomic level. There are two broad categories of current theories concerning this process: The ‘indirect’ model suggests that urea disrupts or interferes with the water structure surrounding the protein which indirectly leads to a loss of its native state stability. One proposed mechanism for this is that urea indirectly ‘dries’ the protein, thus lowering the protein hydration and weakening the hydrophobic effect [186, 187, 188].

There is a much larger number of ‘direct’ theories, that propose a direct urea-protein interaction as the main cause of protein unfolding. However, the way in which urea specifically interacts with the protein in these ‘direct theories’ is still debated. Reversible folding and unfolding replica exchange simulations have shown preferential urea contacts with the side chains [189]. Other MD simulations have found that urea does not cause, but stabilises initial unfolding conformations by favourable interactions with less polar

residues [190]. It has also been suggested, from MD simulations, that urea unfolds proteins by altering electrostatic interactions either by solvating the charged residues or by engaging in hydrogen bonds with the protein backbone [191]. Urea has also been shown to destabilise the folded structure by breaking hydrogen bonds within the protein [192, 193]. Another recent study has shown that urea replaces hydrating water in lysozyme crystals when these are soaked in urea solution [194]. Stepwise denaturation processes of proteins by urea have also been proposed, where an all-atom MD simulation of lysozyme suggested that once urea has expelled the first hydration layer around the protein, the hydrophobic core is penetrated by urea as opposed to water [195, 196]. This is in opposition to other investigations which found that urea, on the one hand, altered the local water structure and thus encouraged the water solvation of the hydrophobic core and, on the other hand, also interacted directly with polar residues [197].

In this chapter, the structure of the short glycine-proline-glycine peptide (GPG-NH₂; Figure 3.1) in the presence of urea in water has been determined using neutron scattering combined with EPSR and MD. GPG-NH₂ contains a motif sequence (PG) which is commonly found in turns in many proteins [198, 199, 200, 201, 202, 203] and allows for the observation of a peptide in a variety of conformations, so that the role of water and urea can be probed directly. This peptide is also small and soluble enough to be investigated experimentally using neutron diffraction techniques *in solution*. The atomic-scale hydration and conformation of this short chain peptide have been previously measured and was found not to aggregate at the same peptide concentration [198, 183],^a thus allowing for a direct understanding of how urea and water interact with this molecule in solution and comparing the systems with and without the presence of urea.

^aThese data has been used for this chapter as a comparison regarding the RDFs. These RDFs are throughout printed in blue to distinguish them from the GPG-NH₂:urea:water data (red RDFs) which are the main focus of this chapter.

3.2 Methods

3.2.1 NMR Spectroscopy Sample Preparation and Measurement

Glycyl-L-prolyl-glycinamide·HCl (GPG-NH₂ ·HCl, C₉H₁₇N₄O₃Cl; CAS No 141497-12-3) was purchased from Bachem (Bubendorf, Switzerland), urea (CH₄N₂O; CAS No 57-13-6) from Sigma Aldrich and both were used without further purification. For the samples which required deuterium labelling, GPG-NH₂ ·HCl was dissolved in 99.8% D₂O (CAS No 7789-20-0 from Sigma Aldrich) and subsequently freeze-dried to remove the heavy water solvent. This process was repeated three times to ensure an adequate level of deuteration of the exchangeable hydrogens on GPG-NH₂. The same procedure was followed to deuterate urea. The level of deuteration was confirmed by NMR spectroscopy (Section 3.3.1).

NMR samples were prepared with O₂ degassed water and under N₂ flow using a Schlenk line [204] and NMR tubes fitted with Young's taps to prevent exposure to the atmosphere and contamination. NMR samples were prepared by mass, using a Satorius scale accurate to 4 d.p. NMR spectra were collected on a 500 MHz and a 600 MHz spectrometer. The 1D ¹H spectra were measured at 8k points for 128 scans with a recycle delay of 1 s.

Apart from peak assignment and to ensure peptide sample purity, NMR spectra were taken to ensure that the high peptide concentration in the neutron sample or the urea concentration had no influence on the *cis/trans* ratio of the peptide or other structural features visible in NMR spectroscopy.

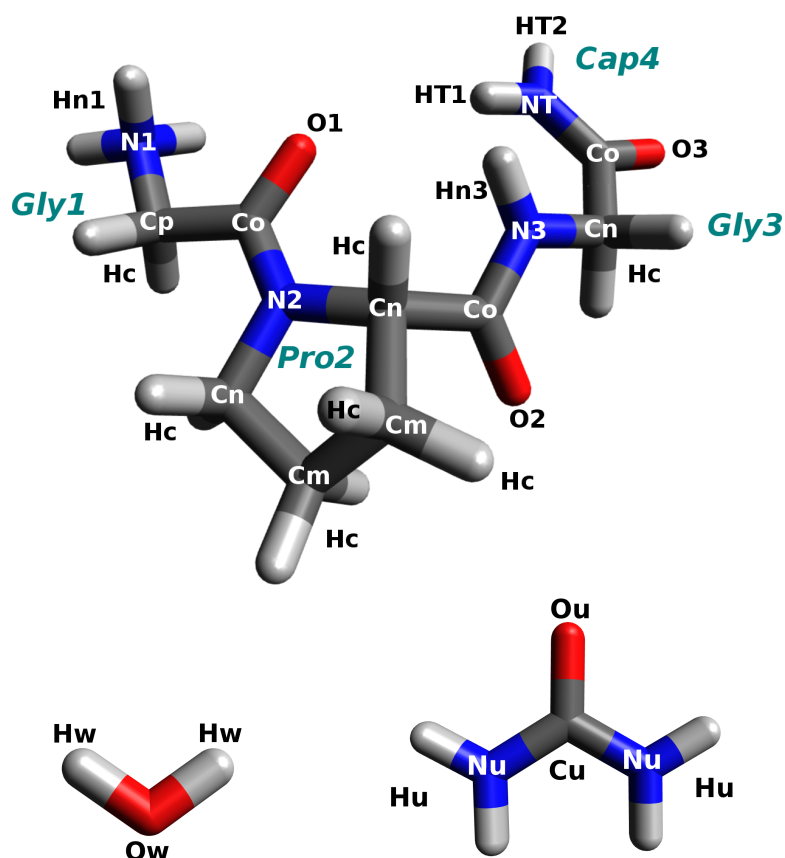


Figure 3.1: Molecular structure and labelling scheme of the GPG-NH₂ peptide in its *trans* conformation and the water and urea labelling scheme.

3.2.2 Neutron Diffraction and Sample Preparation

Isotopomeric NDIS (neutron diffraction enhanced by isotopic substitution) samples were prepared by weight under a N₂ atmosphere with O₂-degassed H₂O and/or D₂O (99.994%, from Sigma Aldrich) to prevent the exchange of D₂O with H₂O in atmosphere and to ensure sample purity. The final molecular ratio of GPG-NH₂:urea:water was 1:4:58 for each sample (Table 3.1) which corresponds to concentrations of ≈ 0.8 M GPG-NH₂ and ≈ 3.3 M urea. The pH of the neutron diffraction samples was 6.0. In comparison, the neutron samples of aqueous GPG-NH₂ solution without urea had a pH of 4.9 [183]. This change in pH is expected due to urea being a weak base ($pK_B = 13.80$ [205])

Neutron diffraction measurements were performed at 298 K on the SANDALS instrument (Section 2.2.7.1) using five isotopically substituted samples (Table 3.1). The level of deuteration is with respect to the exchangeable hydrogens in the sample (Figure 3.1;

Hn1, Hn3, HT1, HT2, Hu and Hw). The samples were contained in SiO_2 cells (Section 2.2.7.2). Diffraction data were collected for between 8 and 9.5 hours per sample (Table 3.1). Data were also collected for the empty cells, the empty instrument and a vanadium standard for background subtraction and normalisation. The data were corrected and converted to $F(Q)$ using the GUDRUN programme (Section 2.2.7.3). The measured scattering levels were within 10% of the expected theoretical values (differential cross section, DCS levels in Table 3.1).

Table 3.1: List of samples used for the collection of neutron diffraction data. The molecular ratio of $\text{GPG-NH}_2 \cdot \text{HCl}:\text{urea}:\text{water}$ was 1:4:58 The level of deuteration refers to the exchangeable hydrogens on GPG-NH_2 , urea and water (Hn1, Hn3, HT1, HT2, Hu and Hw).

Sample Label	GPG-NH ₂ + urea + water	μA	DCS level
I	100% H	1432	95.0%
II	64% H, 36% D	1507	98.9%
III	50% H, 50% D	1437	95.6%
IV	36% H, 64% D	1169	95.7%
V	100% D	1174	101.9%

3.2.3 Empirical Potential Structure Refinement

For the EPSR simulation, the modeling box contained 20 GPG-NH_2 molecules of the same starting conformation (Figure 3.1), 20 Cl^- ions, 80 urea molecules (the urea starting conformation was planar) and 1160 water molecules at the measured (non-deuterated) density ($\rho = 1.091 \text{ g/ml} = 0.102 \text{ atoms} \cdot \text{\AA}^{-3}$) at 298 K. The density was measured using a 100 μl Gilson pipette and a Satorius scale (see above). Starting parameters for the water molecules were taken from the TIP3P water model [152] and starting parameters for the GPG-NH_2 molecules, urea molecules and Cl^- ion were from the CHARMM36 forcefield [148, 206, 207, 208, 209] and modified in order to adjust for different atomic labelling in EPSR compared with MD while ensuring electroneutrality of the simulation box. The full list of the parameters used as starting potentials and parameters can be found in Table 3.2. Other parameters like starting conformation, dihedral angles and rotational groups are in Appendix A.

Similar to previous studies on GPG-NH_2 in water without urea [183, 198], the EPSR simulation contained a mixture of *cis* and *trans* GPG-NH_2 molecules (with respect to

Table 3.2: The parameters for GPG-NH₂, Cl⁻ and urea used in the EPSR simulations from CHARMM36 and TIP3P water. The corresponding labelling scheme is shown in Figure 3.1

Atom Label	ϵ / kJ mol ⁻¹	σ / Å	q / e
N1	0.83680	3.2963	-0.300
Hn1	0.19246	0.4000	0.330
Cp	0.23012	3.8754	0.130
Hc	0.92050	2.3520	0.090
Co	0.37656	3.5636	0.530
O1	0.50208	3.0291	-0.530
N2	0.83680	3.2963	-0.270
Cn	0.15481	3.9645	0.000
Cm	0.23012	3.8754	-0.180
O2	0.50208	3.0291	-0.530
N3	0.83680	3.2963	-0.490
Hn3	0.19246	0.4000	0.310
O3	0.50208	3.0291	-0.530
NT	0.83680	3.2963	-0.620
HT1	0.19246	0.4000	0.320
HT2	0.19246	0.4000	0.300
Cl ⁻	0.62760	4.0447	-1.000
Cu	0.29288	3.5636	0.600
Nu	0.83680	3.2963	-0.690
Hu	0.19246	0.4000	0.340
Ou	0.50208	3.0291	-0.580
Ow	0.63639	3.1507	-0.834
Hw	0.19246	0.4000	0.417

the Gly1-Pro2 peptide bond) in ratios which correspond with that measured by ¹H 1D NMR spectroscopy (10% *cis*; 90% *trans*). This *cis/trans*-ratio was unchanged by the addition of urea up to 6 M (Section 3.3.1). Although peptide bonds can *cis-trans* isomerise [210], this was not incorporated into the non-dynamic EPSR simulation and 2 *cis* and 18 *trans* GPG-NH₂ molecules with non-rotating peptide bonds were used as starting conformations.

3.2.4 Molecular Dynamics Simulations

Two MD simulations at the same molecular ratios as the neutron samples, were also performed;^b one with all of the GPG-NH₂ molecules in the *trans* conformation (regarding the Gly1-Pro2 peptide bond) and one with all of the peptide molecules in *cis* conforma-

^bThese simulations were performed by Chris Lorenz. They were analysed and converted to theoretical $F(Q)$ s/ $G(r)$ s by myself using EPSR.

tion. For the peptide hydration and urea interaction analysis, only the *trans* simulation results are discussed in the results as this is the dominant species in solution and there are no appreciable differences between *cis* and *trans* simulations regarding these aspects; Figure 3.14 and 3.16 (below) include the *cis* simulations to exemplify this. Nevertheless, there is a clear difference in *cis* and *trans* peptides regarding the conformation which will be discussed in Section 3.3.5.

Each MD simulation contained 64 GPG-NH₂ molecules, 64 Cl⁻ ions, 256 urea molecules and 3712 water molecules to give the same molar ratio as the experimental neutron data. Similar to EPSR, the GPG-NH₂ molecules and the Cl⁻ ions were modelled using the CHARMM36 force field [207, 148], and the water molecules were modeled using TIP3P [152] modified for the CHARMM force field [211]. All of the bonds and angles for the water molecules were constrained using the SHAKE algorithm [212] and both simulations were conducted using GROMACS 4 [213]. The initial configuration for the simulation was constructed using the Packmol software [214]. For both simulations, an energy minimisation simulation was used initially to eliminate any atomic overlaps that resulted from the construction of the initial configurations, then a 2 ns simulation using the NVT ensemble with a target temperature of 300 K was performed in order to equilibrate the temperature of the system. After these two initial steps, a 2 ns NPT simulation was performed with a target temperature of 300 K and a target pressure of 1 atm, in order to equilibrate the pressure and volume of the simulated systems. Finally, a NPT production simulation was performed at 300 K and 1 atm for 50 ns with a timestep of 2 fs. The Nose-Hoover thermostat [215, 216] was used in all simulations to control the temperature, while the Martyna-Tuckerman-Tobias-Klein (MTTK) barostat [217] was used in the NPT simulations to control the pressure. A cut-off of 14 Å was used for the van der Waals interactions, and the long range Coulomb interactions were calculated using the particle mesh Ewald (PME) algorithm [218, 219].

3.2.5 ANGULA

Orthonormal coordinate systems were assigned to different fragments of the GPG-NH₂ molecule, to the water and urea molecules as shown in Figure 3.2. Using these coordinate

systems, the distribution of the neighbouring water or urea molecules can be plotted relative to specific sites (SDMs) or the whole molecule (WMAs) [220, 184] as described in Section 2.4. For EPSR 4000 snapshots and for MD 1250 snapshots of the simulation boxes were used resulting in the same number of GPG-NH₂ molecules (80000) analysed.

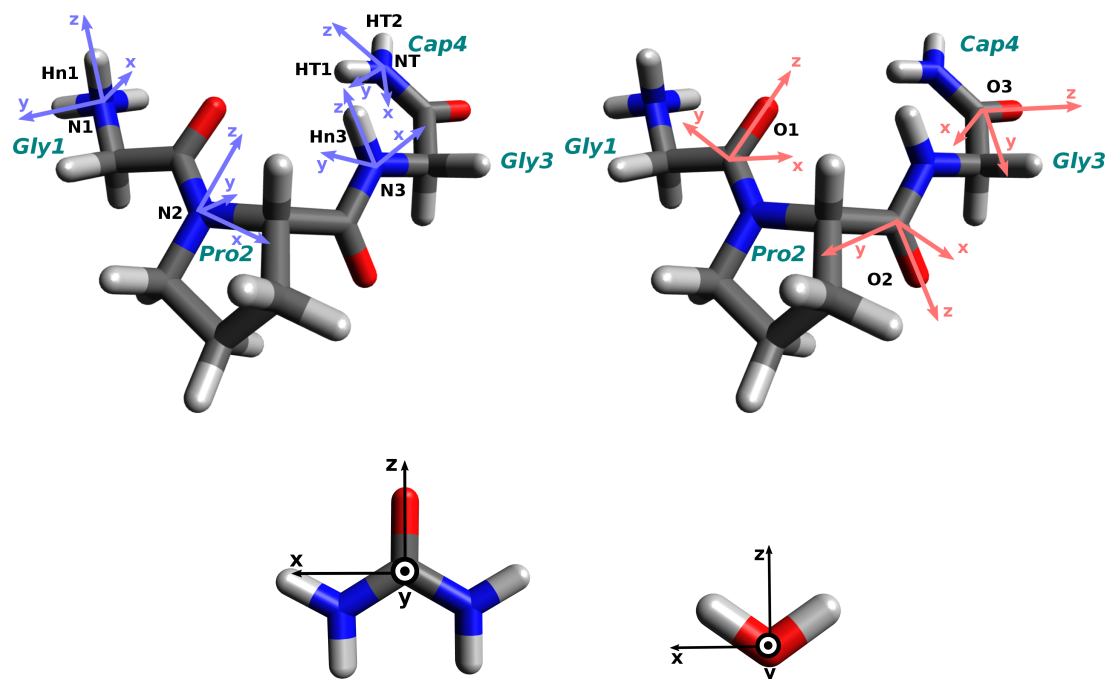


Figure 3.2: The coordinate systems assigned to GPG-NH₂, urea and water.

3.2.6 Preference ratios

In order to quantify the preference of water or urea around specific sites on GPG-NH₂, a contact preference ratio (PR) analysis was performed. Previously, a similar analysis was used for MD simulations for small peptides in aqueous urea solutions [221], where an interaction was defined as any atom on a water or urea molecule being within 3.5 Å of a particular atom on the amino acid. Given the variety of first peak positions in the RDFs in Figs. 3.14, 3.16 and 3.20, this approach was modified so that the relative preference of urea or water around a specific site was determined using the distance of the first minimum in the RDF as a criteria. This relative preference ($P_{UW_X}^r$) or preference ratio (PR) is defined by:

$$P_{UW_X}^r = \frac{n_X(gpg)^U}{n_X(gpg)^W} \cdot \frac{N_W}{N_U} \quad (3.1)$$

where $n_X(gpg)^U$ is the coordination number of urea around a specific GPG-NH₂ site (X) and $n_X(gpg)^W$ is the corresponding coordination number of water around the same site at the same distance r . The ratio of the coordination numbers of urea and water atoms around a GPG-NH₂ atom X were weighted by the ratio of water and urea atoms in the system $\frac{N_W}{N_U}$. This corrects for the fact that there are more water than urea molecules in the solutions. The PRs have been determined for all of the GPG-NH₂ atoms and are tabulated in Tables 3.9, 3.10, 3.11 and 3.13. Similar to previous investigations [221], a contact coefficient greater than 1 indicates that urea is preferred over water and less than 1 indicates that water is preferred over urea.

3.3 Results

3.3.1 NMR Spectroscopy

Figure 3.3 shows the assignment of a 1D ¹H NMR spectrum of 20 mM GPG-NH₂ in water. The *cis/trans* ratio, of 10% *cis*; 90% *trans* (regarding the Gly1-Pro2 peptide bond), was determined by integration of the HT1/HT2 and proline H β *cis* and *trans* peaks. This ratio was unchanged by the addition of urea (up to 6 M). This is shown in Figure 3.4 where urea (1-6 M) was titrated into a 20 mM solution of GPG-NH₂. The level of deuteration of the lyophilysed samples (20 mM) was checked by NMR spectroscopy and is shown in Figure 3.5. It is clearly visible that, within the sensitivity of NMR, the exchangeable hydrogens Hn3 and HT1/HT2 (including *cis* peaks) are absent in this spectrum. The 100% H neutron sample with a higher peptide concentration is also shown in Figure 3.5 to ensure that an increased peptide concentration does not change structural features visible in NMR spectroscopy.

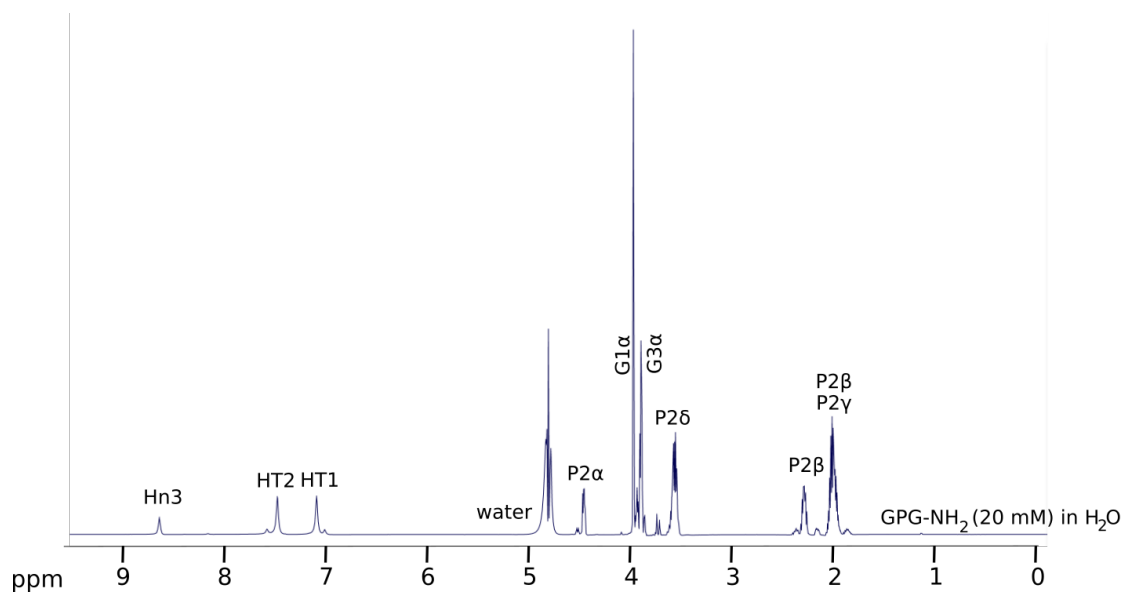


Figure 3.3: ¹H 1D NMR spectrum of 20 mM GPG-NH₂ in H₂O with peak assignment.

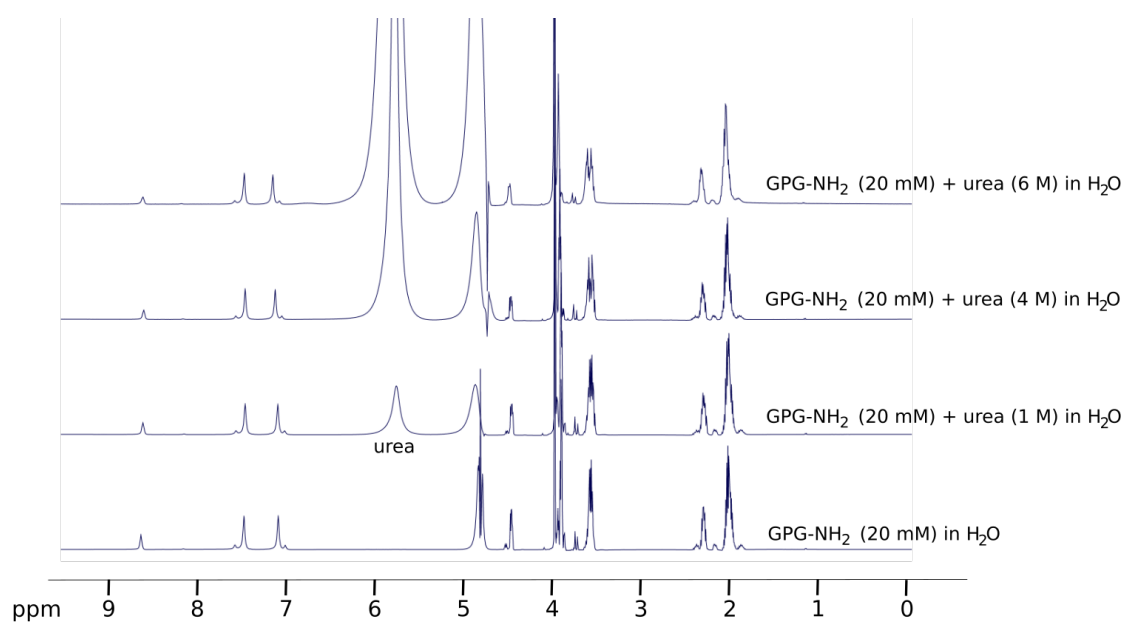


Figure 3.4: ¹H 1D NMR spectra of 20 mM GPG-NH₂ in H₂O with varied urea concentrations (1-6 M).

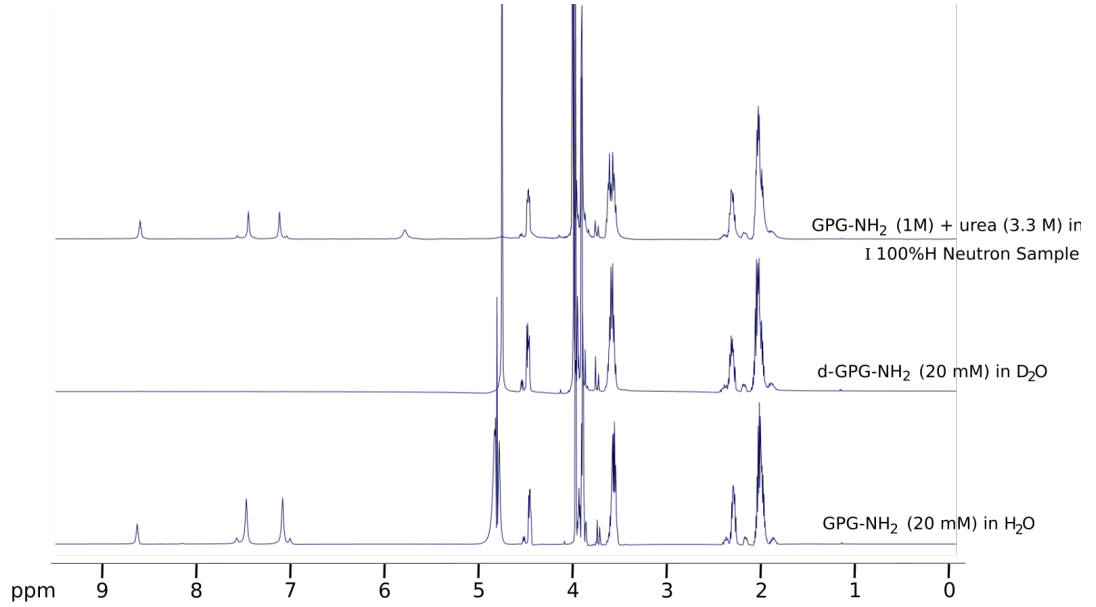


Figure 3.5: ^1H 1D NMR spectrum of 20 mM D-GPG-NH₂ in D₂O and the I (100% H) neutron sample in comparison with 20 mM GPG-NH₂ in H₂O

3.3.2 Diffraction Data and EPSR Fits

Figure 3.6 (left) shows the measured $F(Q)$ data along with the EPSR fits to these data and the difference between the two. Additionally, Figure 3.7 (left) shows comparisons between the measured data and $F(Q)$ functions which have been derived from the MD trajectories (*cis* and *trans*). The sample labels I-V correspond to the level of deuteration in listed in Table 3.1. From these figures, the EPSR fits are good, with only small differences occurring at low Q values where in the measured data the inelasticity effects are the most difficult to correct for. The MD derived data does not provide as good of comparison to the measured data compared with the EPSR fits. The scattering data at low Q , especially at higher levels of deuteration, is usually not in good agreement with the MD simulation (Figure 3.7), but the general shape of the MD simulated scattering data is similar to the experimental data. The differences between *cis* and *trans* MD simulation are not visible in the $F(Q)$ and $G(r)$ data. The corresponding Fourier transformations which show the data in real space $G(r)$ are also shown in Figures 3.6 and 3.7 (right) along with the EPSR fits and MD fits in comparison. Again, EPSR shows a good agreement with the neutron scattering data whereas the MD simulations tend to have a over-pronounced peak at 1 Å resulting from the more rigid intramolecular O-H

distance in MD water molecules. Improving this MD model using data extracted from EPSR is principally possible. However, it is not possible to extract the refined ϵ and σ parameters from EPSR. Only the observable properties like atom distances and bond angles could be used as new starting configuration for the MD simulation.

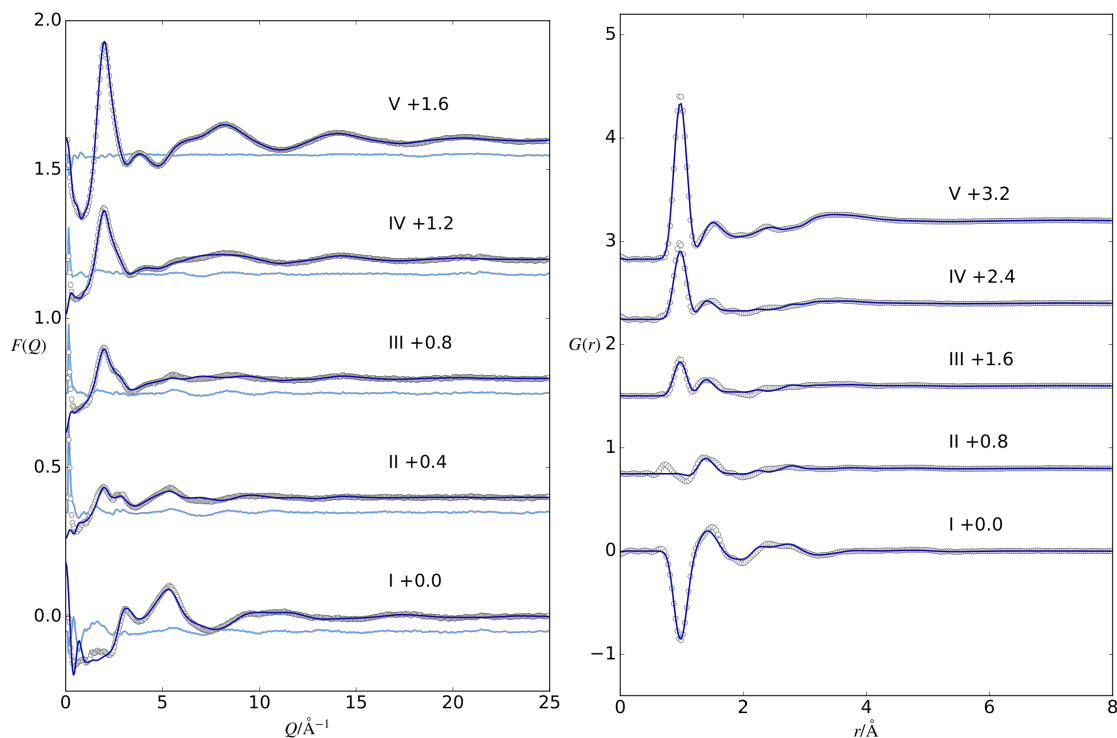


Figure 3.6: Measured diffraction data (grey dots) for GPG-NH₂ and urea in aqueous solution compared with the **EPSR** fits (dark blue lines) in reciprocal space ($F(Q)$, left) and in real space ($G(r)$, right). The difference between simulation and data are shown as a light blue lines (left). The data and corresponding fits have been shifted for clarity. The sample labels I-V correspond to the level of deuteration in listed in Table

3.1.

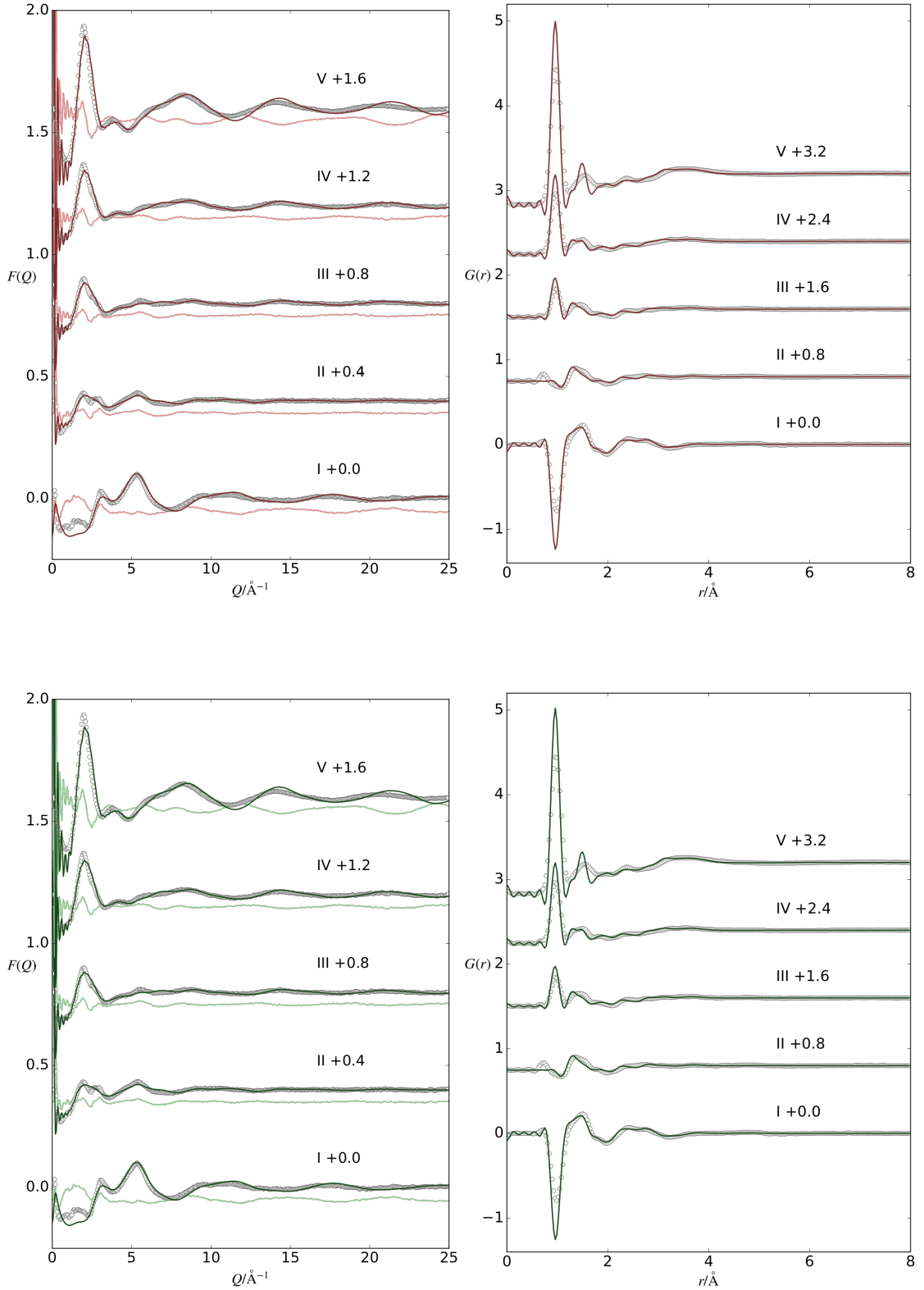


Figure 3.7: Measured diffraction data (grey dots) for GPG-NH₂ and urea in aqueous solution compared with the MD (*trans*-GPG-NH₂: top, *cis*-GPG-NH₂: bottom) fits dark red lines in reciprocal space ($F(Q)$, left) and in real space ($G(r)$, right). The difference between simulation and data are shown as a light red lines (left). The data and corresponding fits have been shifted for clarity. The sample labels I-V correspond to the level of deuteration in listed in Table

3.3.3 Solvent Structure

Figure 3.8 shows the intermolecular water RDFs ($g_{OwOw}(r)$, $g_{HwOw}(r)$ and $g_{HwHw}(r)$) for the GPG-NH₂:urea:water system compared with the same functions for pure water [121], GPG-NH₂:water in the absence of urea [198], and for urea:water solutions at a ratio of 1:16 [222], approximately the same concentration in the present GPG-NH₂:urea:water solution of urea:water 1:14.5. Table 3.3 shows the corresponding coordination numbers of the water-water RDFs shown in Figure 3.8. The addition of GPG-NH₂ lowers all CNs of the first neighbouring water and the addition of urea lowers this coordination even further.

The $g_{OwOw}(r)$ function, is often considered a ‘signature’ for the tetrahedral coordination of pure water [223, 224], as it is sensitive to changes in this bulk water structure. For these RDFs, the first neighbour water distances in $g_{OwOw}(r)$ are similar with all of solutions showing first peaks at around 2.8 Å, while the second nearest neighbour water shells are different in each function. The MD simulations have, overall, higher n_{Ow}^{Ow} coordination numbers than EPSR. For GPG-NH₂ containing solutions in EPSR, the water-water shells show two second neighbour distances, between 3 and 5.5 Å, compared to pure water where only one second neighbour peak at 4.3 Å is visible. For GPG-NH₂ in pure water, the closer of these two second shells most likely occurs from both the Cl[−] ion-water interactions and GPG-NH₂-water interactions (GPG-NH₂ is cationic), as ions have been observed to have a constricting effect on the second nearest neighbour distance in the $g_{OwOw}(r)$ in solutions [223]. Interestingly, in the present GPG-NH₂:urea:water solutions, the second neighbour at 4.5 Å peak has shifted to slightly higher r values relative to GPG-NH₂ in the absence of urea. This shift to higher values has also been observed for higher concentrations of urea in solution [225], while in contrast mixtures of trimethylamine N-oxide (TMAO) and urea in solution showed a more limited perturbation to this second hydration shell in solution [226]. It should be noted that for both MD simulations the second peak in $g_{OwOw}(r)$ is not visible due to the TIP3P water modelling which usually does not accurately predict the second neighbour Ow-Ow shell, as previously shown [227].

The shape of the first neighbour peak at 1.9 Å for the $g_{OwHw}(r)$ is unchanged by the

addition of urea or GPG-NH₂. The second neighbour peak on the other hand becomes less clearly defined towards a larger distance when urea is added and it also shifts slightly from 3.5 Å in pure water to 3.7 Å in GPG-NH₂ and urea. GPG-NH₂ on its own in water does not significantly change the $g_{OwHw}(r)$.

The $g_{HwHw}(r)$ shows that the more solutes are added to pure water, the more narrow does the first neighbour peak at 2.4 Å become. A small peak shoulder at 2.0 Å in the EPSR simulation of GPG-NH₂ is lost when urea is added. The second neighbouring peak at 3.8 Å becomes more shallow, less defined and broader when urea is added to water. This effect is not visible, to this degree, in GPG-NH₂ containing simulations.

These findings of a slightly altered water-water interaction is not enough to support the theory of indirect urea denaturation *via* disruption of the local water structure. The bulk water structure is not substantially perturbed by the addition of urea and a similar effect can be observed for other solutes that form hydrogen bonds with water [228]. A unique effect that explains the urea-caused denaturation *via* the bulk water structure is not observed. However, any effect at this relatively low urea concentration (3.3 M), where GPG-NH₂ is not yet denaturing, might be stronger at commonly used denaturing conditions (≥ 8 M) where 0.8 M GPG-NH₂ was found to be insoluble.

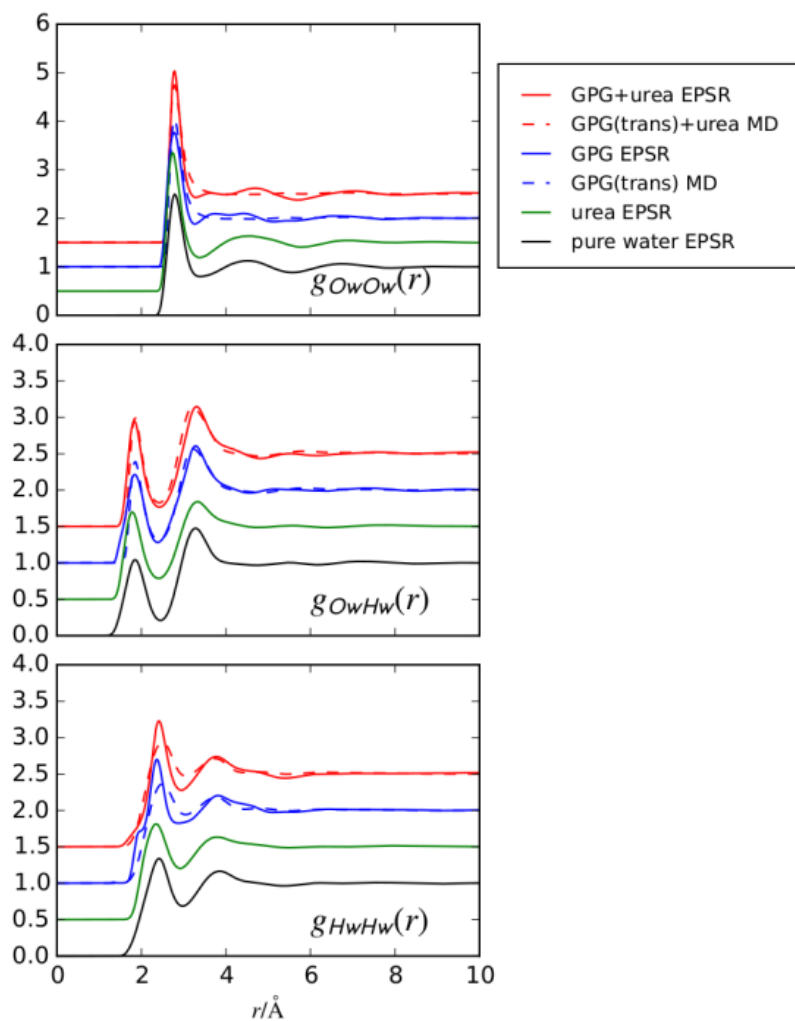


Figure 3.8: RDFs for intermolecular water-water interactions from the GPG-NH₂:urea:water EPSR fits to the neutron data and MD (only *trans* GPG-NH₂) compared with the same functions in pure water [121], GPG-NH₂:water [183], and urea:water [222] solutions. The different RDFs have been shifted for clarity.

Table 3.3: Coordination numbers (n_{α}^{β}) for the water-water RDFs in Figure 3.8, including pure water [121], GPG-NH₂:water [183], and urea:water [222].

	$r_{max} / \text{\AA}$	pure water EPSR	urea EPSR	GPG-NH ₂		GPG-NH ₂ + urea	
				EPSR	MD	EPSR	MD
Ow-Ow	3.5	5.2	4.4	4.8	5.0	4.2	4.6
Ow-Hw	2.4	1.9	1.7	1.8	1.7	1.6	1.6
Hw-Hw	3.0	5.4	4.7	5.2	5.1	4.3	4.7

3.3.4 Co-Solute Structure

3.3.4.1 Urea-Water Interactions

Figure 3.9 shows the urea-water RDFs, comparing the urea:water and the GPG-NH₂:urea:water systems. The hydrogen bonding between water and urea seems mostly unchanged by the addition of GPG-NH₂. Although the first neighbour peaks of Ou-Ow and Ou-Hw appear higher and more narrow with GPG-NH₂ compared to the pure urea mixture, the corresponding coordination numbers in Table 3.4 show only a slight decrease in urea hydration with the addition of GPG-NH₂ to the solution.

A small decrease in hydration of the urea-NH₂ groups when GPG-NH₂ is added is visible in the $g_{HuOw}(r)$, as the first hydration peaks shift away from urea (from 1.8 Å to 2.0 Å) and become more shallow and less clearly defined. This could also be an effect of increased urea-chloride coordination reducing the urea-water contacts.

The decreases in urea hydration are slight and it should be noted that the urea:water ratio in the peptide containing simulation (1:14.5 urea:water) is not exactly the same as in the pure urea:water mixture (1:16 urea:water). Nevertheless, this decrease in urea-hydration upon the addition of a peptide suggests the presence of direct urea-peptide interactions.

Table 3.4: Coordination numbers (n_α^β) for the urea-water RDFs in Figure 3.9, including for urea:water [222].

	$r_{max} / \text{\AA}$	urea EPSR	GPG-NH ₂ + urea EPSR	MD
Cu-Ow	5.3	17.0	15.1	14.4
Ou-Ow	3.3	2.7	2.4	2.4
Nu-Ow	4.3	8.2	7.1	6.8
Hu-Ow	2.4	0.8	0.6	0.6
Cu-Hw	3.1	2.9	2.5	2.3
Ou-Hw	2.5	2.2	2.1	2.1
Nu-Hw	4.5	19.7	17.6	16.7
Hu-Hw	3.0	3.9	3.2	2.9

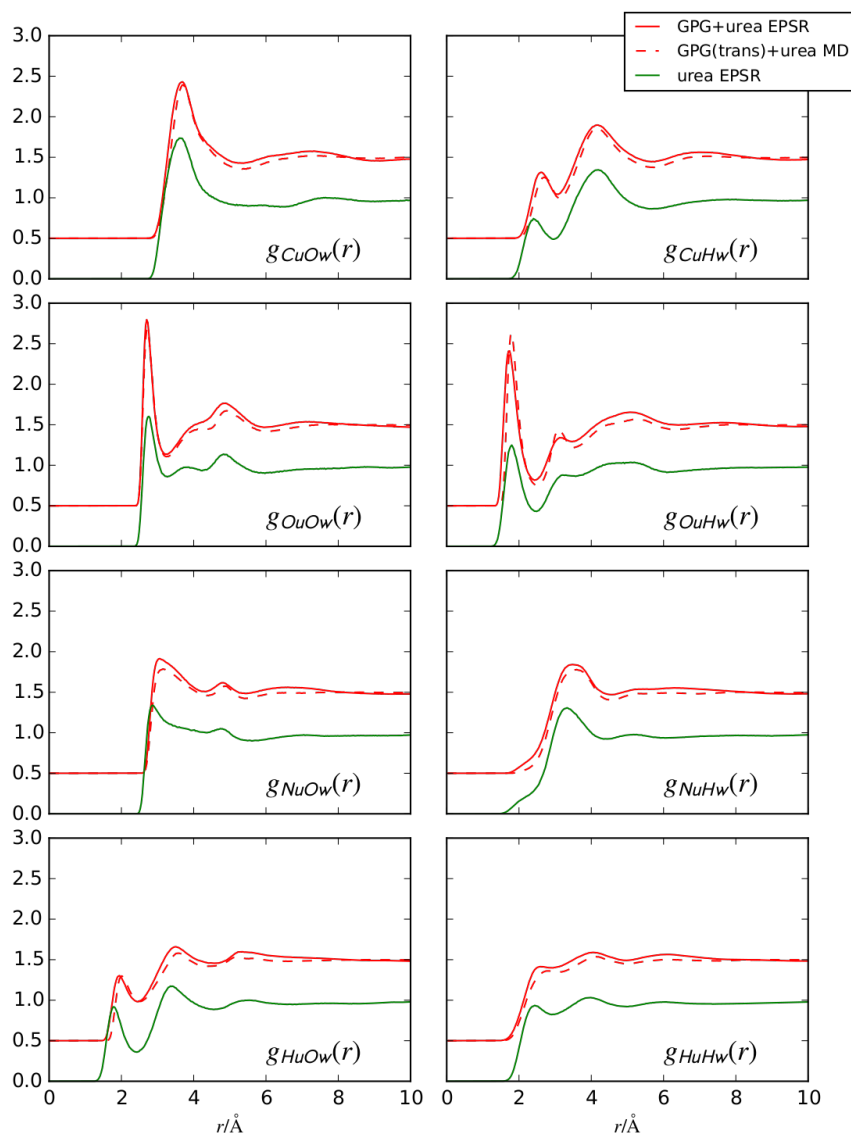


Figure 3.9: RDFs for urea-water interactions from the GPG-NH₂:urea:water EPSR fits to the neutron data and MD compared with the same functions in urea:water [222] solutions. The different RDFs have been shifted for clarity.

3.3.4.2 Chloride-Water Interactions

Figure 3.10 shows the chloride-water and chloride-chloride RDFs, comparing the GPG-NH₂:water ([183]) and the GPG-NH₂:urea:water systems. The chloride-chloride interaction shows the most striking difference between EPSR and MD simulations. The addition of urea to the GPG-NH₂:water mixture increases the first neighbour chloride-chloride interaction from a coordination number (Table 3.5) of 0.2 to 0.5 for the EPSR simulations. MD does not show this clear difference between urea containing mixtures and the peptide:water mixture. On the contrary, MD even shows a slight decrease in the

nearest chloride-chloride interactions upon urea addition. However, it should be noted that because of the relatively small chloride concentration in the system, the contribution of Cl-Cl interactions to the neutron scattering signal is small (less than 0.8% for the 100D sample) and as such will not be highly weighted in the subsequent EPSR model. A follow-up experiment with X-ray scattering, where the heavier chlorine has more impact on the scattering data, might help to disprove or confirm this result (Chapter 4).

The hydration of chloride is decreased by the addition of urea to the system across both simulation techniques indicating that urea draws hydrating water away from ions. This decrease is more pronounced in EPSR than MD. For example decreases the $g_{ClOw}(r)$ hydration peak at 3.1 Å from a coordination number of 6.9 to 5.2 for EPSR and 7.1 to 6.6 for MD.

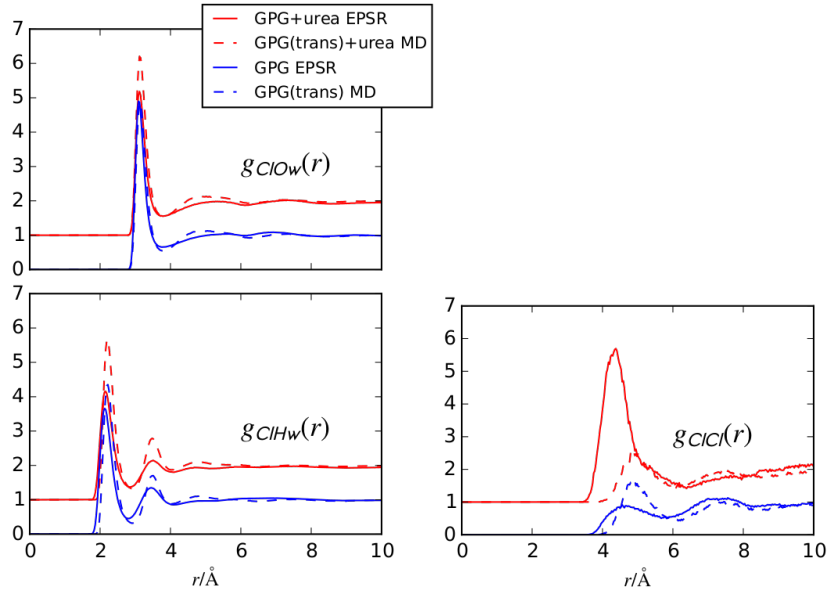


Figure 3.10: RDFs for chloride-water and chloride-chloride interactions from the GPG-NH₂:urea:water EPSR fits to the neutron data and MD compared with the same functions in GPG-NH₂:water [183] solutions. The different RDFs have been shifted for clarity.

Table 3.5: Coordination numbers (n_{α}^{β}) for the chloride-water RDFs in Figure 3.10, including for GPG-NH₂:water [183].

	$r_{max} / \text{\AA}$	GPG-NH ₂		GPG-NH ₂ +urea	
		EPSR	MD	EPSR	MD
Cl-Cl	5.8	0.2	0.3	0.5	0.2
Cl-Ow	3.8	6.9	7.1	5.2	6.6
Cl-Hw	2.8	6.0	6.5	4.6	6.0

3.3.4.3 Urea-Chloride Interactions

Figure 3.11 shows the urea-chloride RDFs. This interaction is largely through the urea hydrogens with the first neighbour peak being located at 2.2 Å. Although the corresponding CNs in Table 3.6 are small, because of the low chloride concentration, these interactions are very similar between EPSR and MD, also with respect to the shape of the RDF-curves.

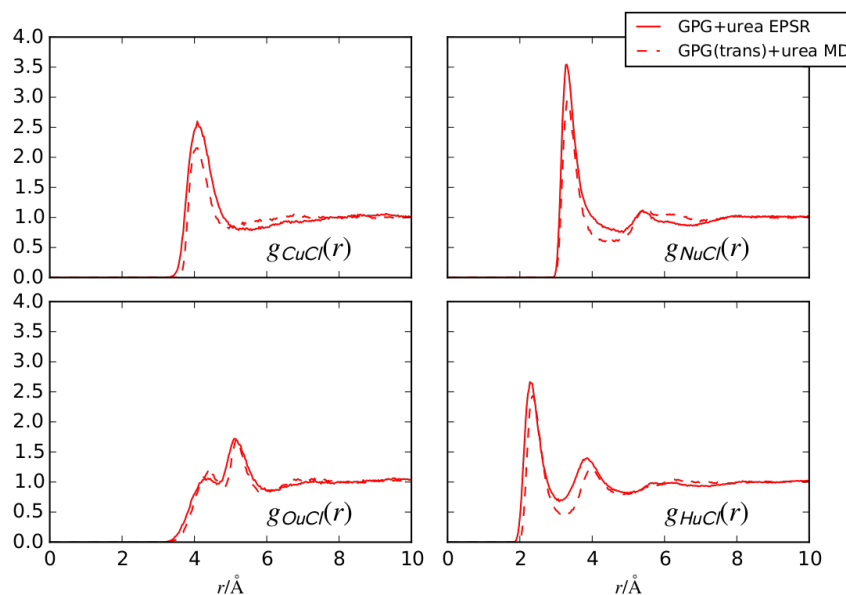


Figure 3.11: RDFs for Urea and Chloride interactions from the GPG-NH₂:urea:water EPSR fits to the neutron data and MD.

Table 3.6: Coordination numbers (n_{α}^{β}) for the urea-chloride RDFs in Figure 3.9.

	$r_{max} / \text{\AA}$	GPG-NH ₂ +urea	
		EPSR	MD
Cu-Cl	5.5	0.3	0.2
Ou-Cl	4.6	0.1	0.1
Nu-Cl	4.9	0.2	0.2
Hu-Cl	3.2	0.1	0.1

3.3.4.4 Urea-Urea Interactions

Figure 3.12 shows the intermolecular urea-urea RDFs for urea in water and the GPG-NH₂:urea:water mixture. The addition of GPG-NH₂ increases all the intermolecular urea-urea interactions and especially the hydrogen bonding between Hu and Ou at 2.0 Å.

Although the corresponding coordination number (Table 3.7) at 2.6 Å only increases from 0.2 without GPG-NH₂ to 0.3 with the addition of GPG-NH₂, the increased urea-urea association upon GPG-NH₂ addition is clearly visible in all RDFs in Figure 3.12 for both MD and EPSR. Almost all urea-urea RDFs have higher and clearer defined peaks with GPG-NH₂ in the mixture with both simulation techniques. It is not clear if this effect is due to the peptide concentration or a general increased ionic strength in the GPG-NH₂ containing mixture.

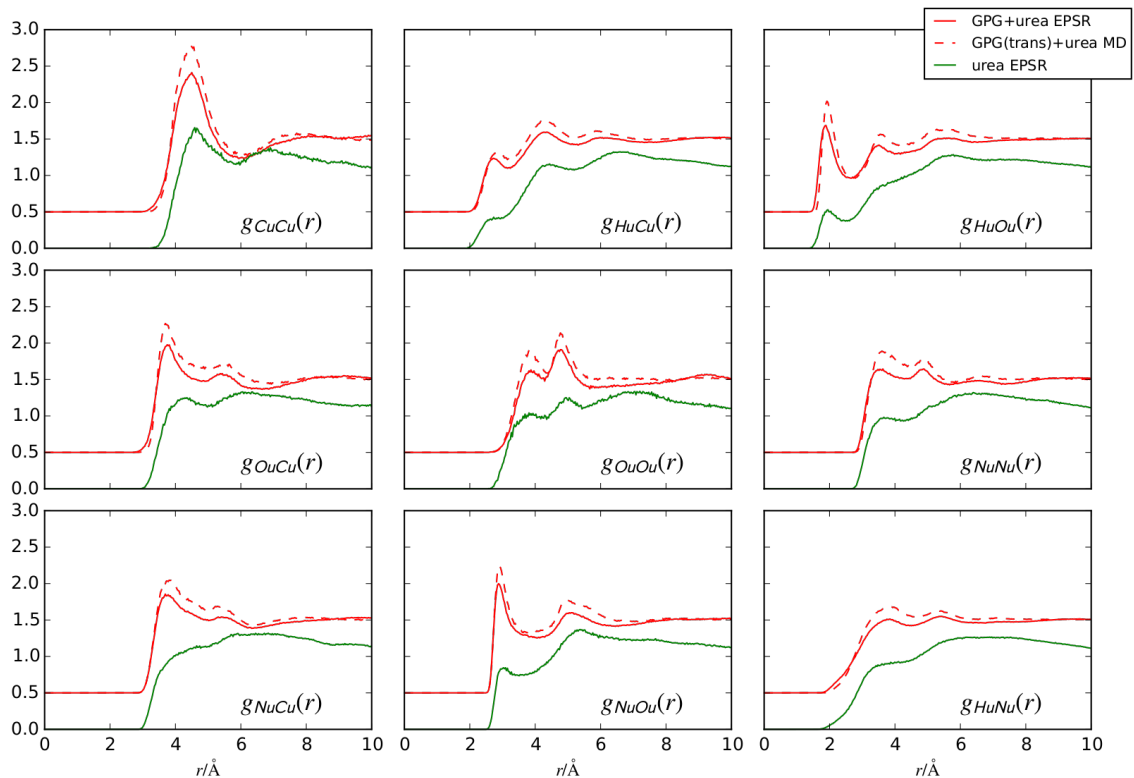


Figure 3.12: RDFs for intermolecular urea-urea interactions from the GPG-NH₂:urea:water EPSR fits to the neutron data and MD compared with the same functions in urea water:urea mixtures [222] solutions. The different RDFs have been shifted for clarity.

3.3.5 Peptide Structure and intramolecular distance

The intramolecular distance between the first carbonyl oxygen of GPG-NH₂ (O1) and the C-terminal amide cap hydrogen (HT1) (Figure 3.13) has previously been used as a measure for peptide folding [183]. Here a clear difference between *cis* and *trans* peptides with or without urea in MD can be observed. The *trans* peptides in MD adopt a larger

Table 3.7: Coordination numbers (n_{α}^{β}) for the urea-urea RDFs in Figure 3.12, including for urea:water [222].

	$r_{max} / \text{\AA}$	urea EPSR	GPG-NH ₂ EPSR	+urea MD
Cu-Cu	6.0	1.6	1.4	1.6
Cu-Ou	4.9	0.7	0.7	0.8
Cu-Nu	5.0	1.5	1.5	1.8
Cu-Hu	3.2	0.3	0.4	0.4
Ou-Ou	4.3	0.4	0.3	0.4
Ou-Nu	4.1	0.6	0.6	0.7
Ou-Hu	2.6	0.2	0.3	0.3
Nu-Nu	4.3	0.7	0.8	0.9
Nu-Hu	4.5	1.9	1.9	2.3

range of different conformations and, on average, shorter intramolecular distances than the *cis* peptides which have a clear preference to be relatively open. For the EPSR simulation, which contains a mixture of *cis* and *trans* peptides, two peaks at 3.5 and 6 Å and a clear minimum at 4 Å is visible. The conformational analysis in Section 3.3.7 will show that for EPSR the peptides adopt a range of different conformations. The peaks in the intramolecular distribution do not represent unique structures but a range of different conformations as this intramolecular distance is only one possible characteristic.

The addition of urea has a different effect on this intramolecular distance depending on the simulation technique. On the one hand, the *trans* peptides in MD adopt closer intramolecular distances upon the addition of urea. This is visible also in the CNs in Table 3.8 where the proportion of peptides with a close intramolecular distance between 0-4 Å doubles for MD(*trans*) from 12% to 24%. On the other hand, the peptides in EPSR adopt, on average, conformations with a larger intramolecular distance and the proportion of more compact peptides between 0-4 Å lowers, after the addition of urea from 11% to 4%. This slight change in peptide conformation was not visible in NMR.

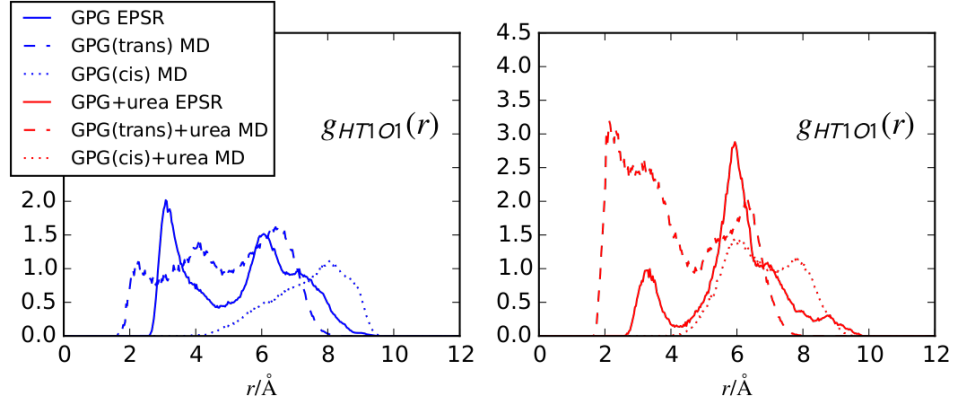


Figure 3.13: Intramolecular RDFs of GPG-NH₂ between HT1 and O1 with (right) and without urea (left) [183].

Table 3.8: Coordination numbers (n_{α}^{β}) for the intramolecular distances in Figure 3.13, including GPG-NH₂:water [183].

$r/\text{\AA}$	GPG-NH ₂			GPG-NH ₂ +urea		
	EPSR	MD		EPSR	MD	
		<i>trans</i>	<i>cis</i>		<i>trans</i>	<i>cis</i>
0-4.0	0.11	0.12	0	0.04	0.24	0
4.0-5.5	0.12	0.24	0.03	0.11	0.24	0.07
5.5-7.0	0.41	0.50	0.20	0.52	0.46	0.37
7.0-10.0	0.36	0.14	0.77	0.32	0.05	0.56

3.3.6 Peptide Solvation

3.3.6.1 Peptide Backbone Oxygen Solvation

Figure 3.14 shows the RDFs for water and urea around the peptide bond carbonyl oxygens in GPG-NH₂ (Figure 3.1) from both EPSR fits to the neutron data and the MD simulations. It is immediately evident that all of the hydrogen bonds from either water or urea hydrogens (H_w and H_u) to the GPG-NH₂ oxygens occur at the same distances ($\approx 1.9 \text{ \AA}$). The coordination numbers for the RDFs in Figure 3.14 are listed in Table 3.9.

What is also clear from Figure 3.14 and Table 3.9 is that the GPG-NH₂ oxygens interact with water to varying degrees, specifically the hydration increases from O1 to O3 in both simulations. This same trend is also observed for urea-GPG-NH₂ interactions, with O3 showing the highest number of urea-hydrogen bonds and the O1 oxygen the least. In general, the hydration numbers from the EPSR fits to the neutron data

are larger than from the MD trajectories, and MD shows comparatively more peptide carbonyl oxygen-urea contacts. For instance has O1 on average 0.1 neighbouring urea hydrogens (Hu) for EPSR and 0.2 for MD.

Comparing the GPG-NH₂ backbone oxygen hydration with and without urea shows that the addition of urea leads to a loss of hydrating water for all oxygens. The loss of hydrating water upon urea addition follows for EPSR a similar trend to the direct urea interaction. Where O3 has relatively the most hydrogen bonding urea with a CN of 0.3, it also loses the most hydrating water (-17%) compared to the simulation without urea. This might suggest that urea molecules replace hydrating water. However, this trend is not observed in MD where O3 exhibits the lowest loss of water contacts and yet has the most urea contacts. Interestingly, although the hydration and urea contacts around the oxygen atoms showed the same trend in coordination from both MD and EPSR, the preference ratios reveal that in MD all of the oxygens preferentially interact with urea over water (PR > 1, Table 3.9), while in EPSR only the Gly-amide-Cap (O3) shows this preference. These two methods also show a large difference in the relative preference ratio for O1 with MD showing a large preference for urea (PR $\approx$ 1.8) while in EPSR this atom shows a marked preference to be hydrated (PR $\approx$ 0.7).

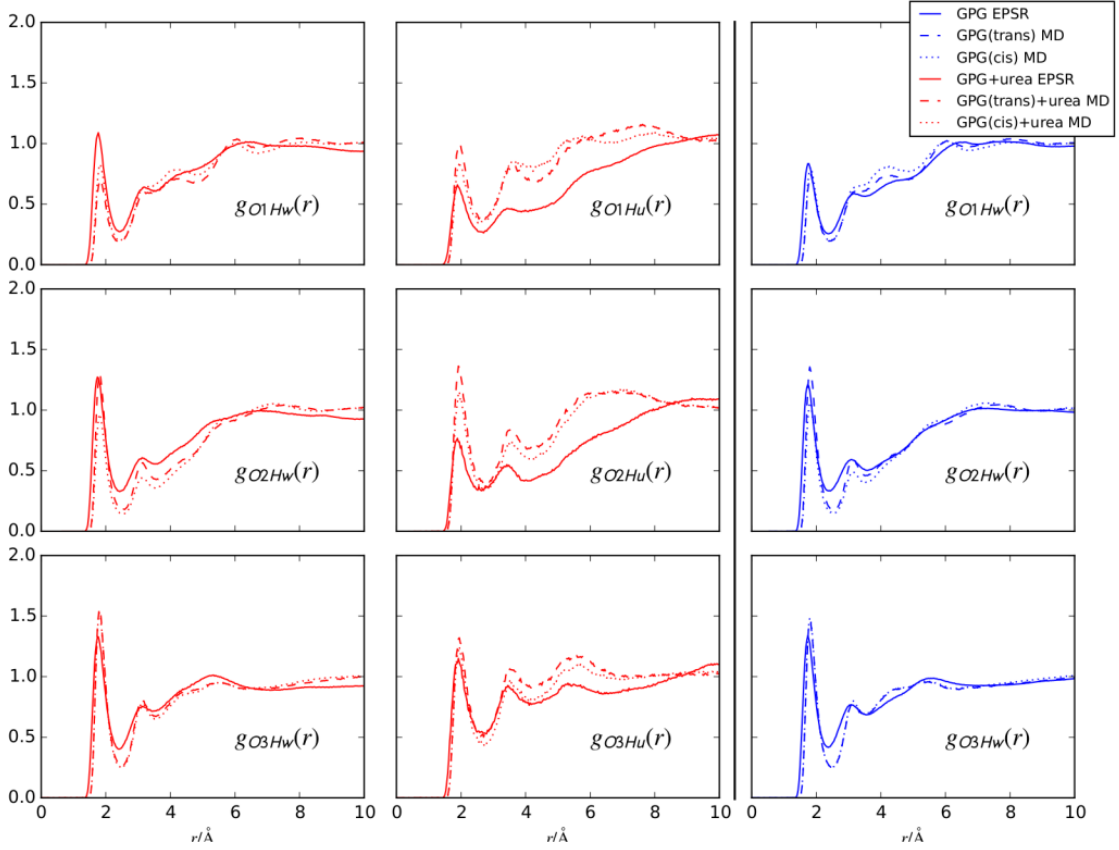


Figure 3.14: RDFs for GPG-NH₂ peptide bond carbonyl oxygens and Hw/Hu atoms from water and urea, including for GPG-NH₂:water [183].

Table 3.9: Coordination numbers (n_{α}^{β}) for the water hydrogens (Hw) and urea hydrogens (Hu) around GPG-NH₂ oxygens from EPSR and MD simulations in Figure 3.14, including for GPG-NH₂:water [183]. Also shown are the hydration difference (HD) and the urea/water preference ratio (PR).

	$r_{max} / \text{\AA}$	GPG-NH ₂		GPG-NH ₂ + urea		HD / %		PR	
		EPSR	MD	EPSR	MD	EPSR	MD	EPSR	MD
O1-Hw	2.5	1.3	1.1	1.3	0.9	-2	-16	-	-
O2-Hw	2.5	1.8	1.6	1.6	1.4	-13	-15	-	-
O3-Hw	2.5	2.1	1.8	1.8	1.8	-17	-4	-	-
O1-Hu	2.5	-	-	0.1	0.2	-	-	0.7	1.8
O2-Hu	2.5	-	-	0.2	0.3	-	-	0.9	1.4
O3-Hu	2.5	-	-	0.3	0.3	-	-	1.1	1.2

This may be a result of how compact (‘folded’) or less compact (‘open’) the peptides are in solution (Section 3.3.5). The MD simulation has more compact ‘folded’ peptides, as a comparison between the O1-HT1 intra-peptide contacts, which is indicative of β -

turn formation [198]. Figure 3.13 shows a larger number of ‘folded’ GPG-NH₂ molecules in the MD simulation ($\approx 24\%$) compared with EPSR ($\approx 4\%$) and oddly here there are slightly more ‘folded’ GPG-NH₂ peptides in the presence of urea than without in the MD simulations [183]. More unfolded peptides would be more solvent accessible and therefore somewhat higher hydration might be expected around the (Gly-Pro) O1 atom. To test this, the O1-Hw coordination numbers were taken for ‘folded’ peptides and compared to unfolded peptides in the MD simulation, where GPG-NH₂ was considered to be ‘folded’ when the HT1-O1 distance was less than 4.5 Å. The coordination numbers do show that the O1 atom is slightly less hydrated when ‘folded’ ($n_{O1}^{Hw}(r) = 0.8$; $r = 2.5$ Å) compared with unfolded (HT1-O1 > 4.5 Å; $n_{O1}^{Hw}(r) = 0.9$; $r = 2.5$ Å), however O1 also showed more urea interactions when GPG-NH₂ was unfolded ($n_{O1}^{Hu}(r) = 0.2$; $r = 2.5$ Å) and as such the O1 atom still shows a preference for urea when the GPG-NH₂ molecules are unfolded (PR = 1.6 for unfolded molecules in MD).

Figure 3.15 shows SDMs which depict the most probable location of water and urea molecules in 3-dimensions around each of the GPG-NH₂ oxygens within a specific distance range. In both simulations, the SDMs are similar, where for each oxygen there is a broader distribution of urea molecules compared with water molecules in the surrounding shells. This can, in part, be explained by the SDM setup; both urea and water spatial distributions are indicating the positions of the central atoms, water oxygen (Ow) or urea carbon (Cu). As hydrogen bonds towards the peptide carbonyl oxygens will be *via* the water or urea hydrogens (Hw and Hu) and the latter are two bonds distant from the central atom (Cu) positions depicted a broader distribution of urea in the SDMs is the result. However, the broader urea distribution can be observed for the Gly1-Pro2 peptide bond oxygen (O1; Figure 3.1) where the urea molecules have a high probability of not only being located behind the oxygen atom (in the $-x$ direction towards the proline ring) but also on either side of the O1 atom. Water molecules, on the other hand, show a more narrow distribution of locations with a high probability of being located behind the O1 atom. To varying degrees this is true for all oxygen sites; urea also preferentially occupying spaces where the water molecules have lower probability of being located. Despite the differences in coordination numbers between the two simulation techniques, the SDMs for both MD and EPSR are similar, with both

showing urea in different locations to the surrounding water shell.

3.3.6.2 Peptide Backbone Nitrogen Solvation

Similar to Figure 3.14, Figure 3.16 shows RDFs between urea and water oxygens (Ou and Ow, respectively) and the GPG-NH₂ nitrogens (Figure 3.1) and the coordination numbers for these functions are listed in Table 3.10. In both simulations, EPSR fits to the neutron data and MD, the water and urea interactions with the GPG-NH₂ nitrogens are broadly similar although MD shows significantly more urea interactions compared with EPSR (Table 3.10) for N2 at a larger distance where the coordination numbers are 1.8 for EPSR and 2.3 for MD.

The -NH₃⁺ (N1) nitrogen clearly shows the most prominent peaks for the water-nitrogen and urea-nitrogen interactions, which are a result of hydrogen bonding from the N-terminal hydrogens to the Ow and Ou oxygens. The Pro2-Gly3 peptide bond nitrogen (N3) also shows sharp peaks (at 2.8 Å) for both water and urea molecules at relatively low r values, which again are a result of hydrogen bonding between the Hn3 hydrogen (Figure 3.1) and the Ow and Ou atoms. Interestingly, while the N1 and N3 first peaks are sharp and well defined (at 2.8 Å), both the proline ring nitrogen (N2) and Gly3-Ncap (NT) nitrogen-water RDFs show broader and less well defined first peaks (at 3.0 Å), indicating a more disordered distribution of water molecules around the Pro2 ring and Ncap portions of GPG-NH₂. For N2, this is unsurprising given the lack of a hydrogen on the proline ring to which water or urea molecules could potentially form hydrogen bonds. For the NT, the relatively broad peak in the $g_{NTOW}(r)$ function is rather unexpected given that direct hydrogen bonds are formed between the both HT1 and HT2 atoms and water oxygens. However, this hydration pattern is characteristic for amide-caps in general, as will be shown in the following chapters. A likely explanation might be that at the C-terminus of the peptide, water molecules are more delocalised because they are not spatially constrained by a continuous peptide chain and can hydrogen bond to two possible hydrogens (HT1 and HT2).

Similar to the peptide backbone oxygens, the backbone nitrogens also lose water contacts upon urea addition. This can be observed for all N-Ow interactions in both

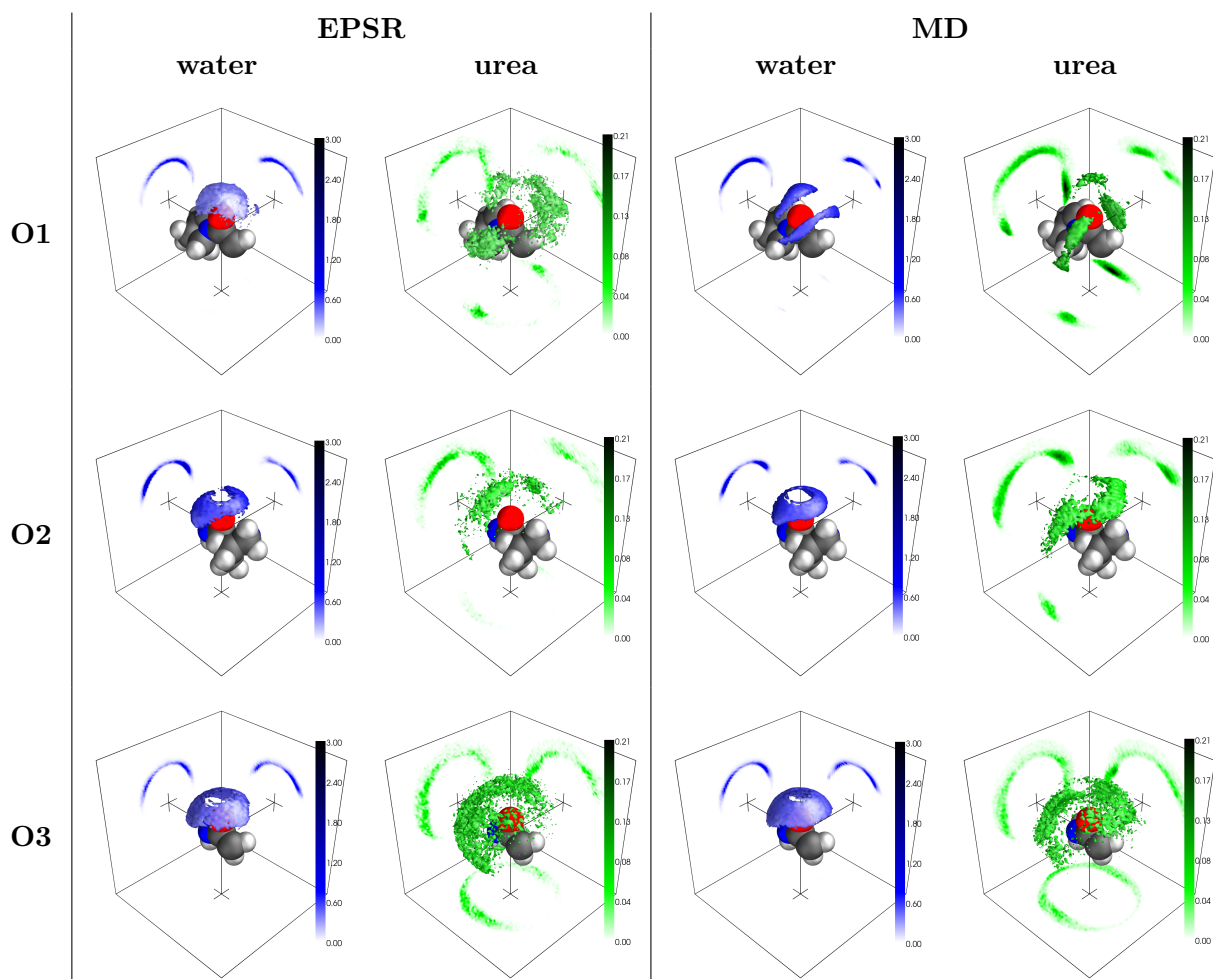


Figure 3.15: Spatial Density Maps (SDMs) of Ow (blue clouds) and Cu (green clouds) around O1, O2 and O3 from EPSR (left columns) and MD (right columns). The 3D shells show the top 40% of all molecules in a distance range from 0–3.3 Å for water and 0–5 Å (20%) for urea around GPG-NH₂ oxygens on the central axis. The scale bar show the local density of neighbours around the central nitrogen in units of atoms/Å³.

EPSR and MD. For example, the addition of urea decreases the water coordination of NT from 7.0 to 6.2 ($\approx -11\%$) for EPSR and 6.9 to 5.8 ($\approx -16\%$) for MD. In MD especially the proline nitrogen N2 and the amide cap nitrogen NT lose relatively more water contacts ($\approx -15\%$) than the more polar groups N1 or N3. In EPSR, N3 loses the most water contacts when urea is added ($\approx -15\%$).

In this peptide:urea:water system all GPG-NH₂ nitrogens preferentially interact with urea (*via* the carbonyl oxygen Ou) over water (*via* oxygen Ow; Table 3.10), this is especially clear for the polar N1 and N3 where urea shows especially in MD a relatively large preference for these groups. The Gly1-Pro2 (N2) and Ncap (NT) nitrogen atoms on the other hand, show a less marked preference for urea. For the EPSR fits to the neutron data these backbone nitrogens (N2 and NT) show neither a preference for urea nor water.

Figure 3.17 shows the SDMs for water and urea around the nitrogen atoms in GPG-NH₂. As opposed to the SDMs in Figure 3.15 where water and urea are in different locations around the oxygen atoms, both water and urea occupy similar locations around each nitrogen for both EPSR fits to the neutron data and the MD simulations. This indicates that urea is likely to directly replace the waters around the peptide-bond nitrogen groups, rather than coordinating to the peptide in a different location. The only slight exception to this is the EPSR for the proline (N2) nitrogen which shows a more diffuse shell of urea molecules around the proline nitrogen compared with the MD. This is, in part, likely a result of these closest urea molecules being spread over a much larger distance range as evidenced by the $g_{N2Ou}(r)$ function.

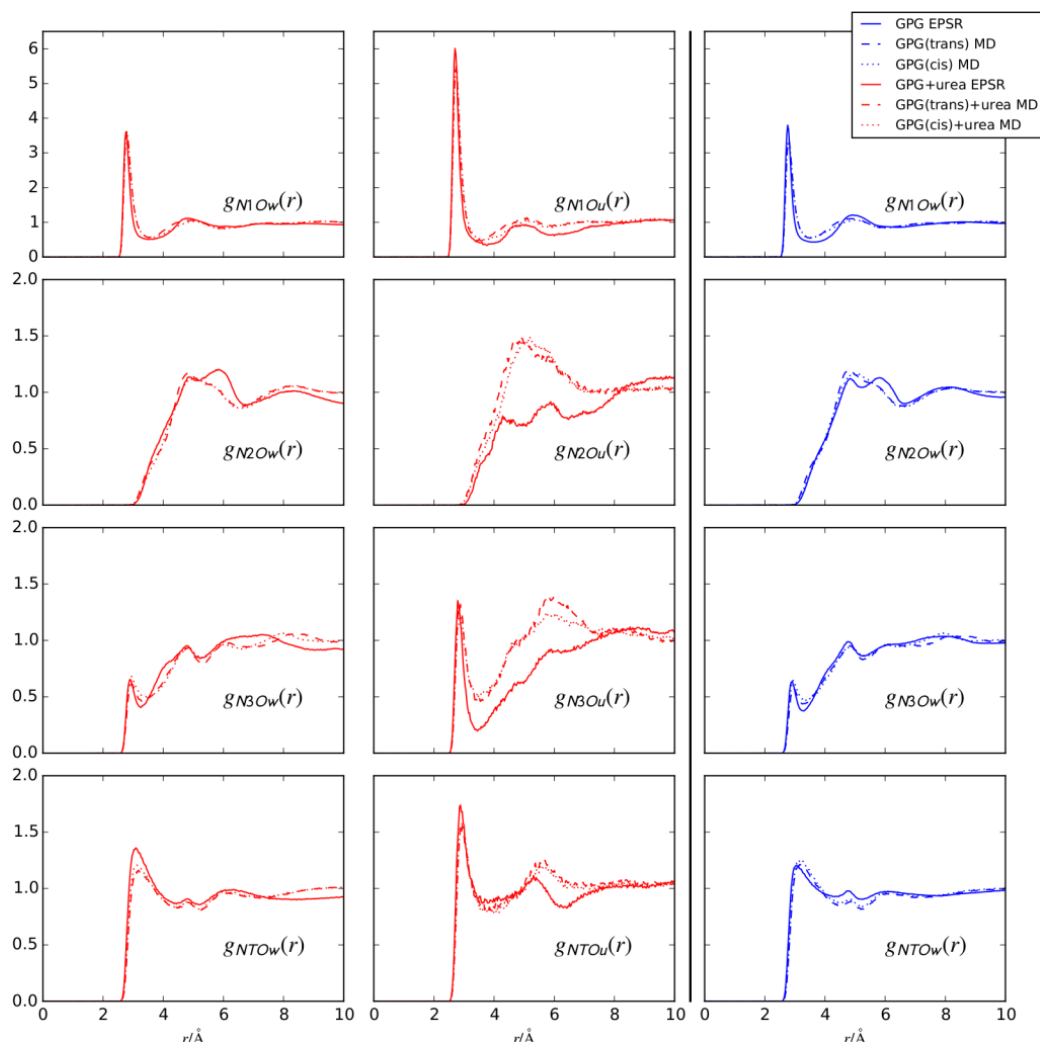


Figure 3.16: RDFs for GPG-NH₂ peptide bond nitrogens and N terminal nitrogen and Ow/Ou atoms from water and urea, respectively, from MD and EPSR, including for GPG-NH₂:water [183].

Table 3.10: Coordination numbers (n_{α}^{β}) for the water oxygens (Ow) and urea oxygens (Ou) around GPG-NH₂ nitrogens from EPSR and MD simulations in Figure 3.16, including for GPG-NH₂:water [183]. The minima where the n_{α}^{β} s were taken are also listed. Also shown are the hydration difference (HD) and the urea/water preference ratio (PR).

	$r_{max} / \text{\AA}$	GPG-NH ₂		GPG-NH ₂ +urea		HD /%		PR	
		EPSR	MD	EPSR	MD	EPSR	MD	EPSR	MD
N1-Ow	3.6	3.9	4.3	3.4	4.0	-12	-7	-	-
N2-Ow	6.8	31.0	30.7	27.1	26.4	-12	-14	-	-
N3-Ow	3.4	1.1	1.1	1.0	1.1	-15	-6	-	-
NT-Ow	4.3	7.0	6.9	6.2	5.8	-11	-16	-	-
N1-Ou	3.6	-	-	0.3	0.4	-	-	1.3	1.3
N2-Ou	6.8	-	-	1.8	2.3	-	-	1.0	1.3
N3-Ou	3.4	-	-	0.1	0.1	-	-	1.7	1.8
NT-Ou	4.3	-	-	0.4	0.4	-	-	1.0	1.0

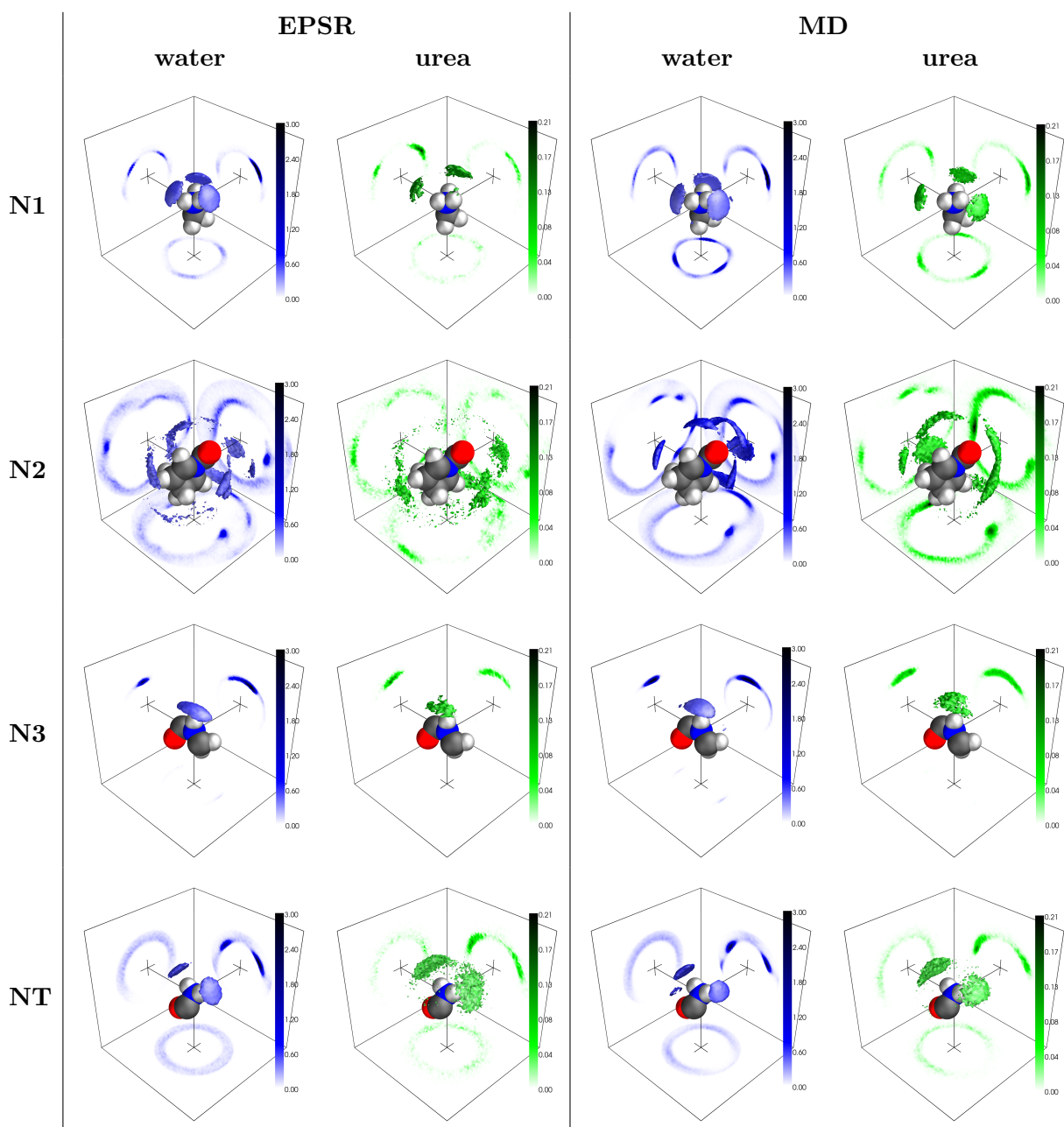


Figure 3.17: Spatial density maps (SDMs) of Ow (blue) and Ou (green) around N1, N2, N3, NT from EPSR (left column) and MD (right column). In each case, the 3D shells show the top 40% of neighbour water molecules and 20% of urea molecules in a distance range from 0 Å to the r_{max} distance shown in Table 3.10 for each atom. The scale bar show the local density of neighbours around the central nitrogen in units of $\text{atms}/\text{\AA}^3$.

Figure 3.18 shows the RDFs of the backbone hydrogens bound to the GPG-NH₂ nitrogens with urea and water oxygens. Similar to the backbone nitrogen hydration and urea interaction, it is clear that there is a direct hydrogen bonding between these peptide sites and the urea molecules, as indicated by hydrogen bond peaks in the RDFs. Comparing the CNs in Table 3.11 there is, again, a slight loss in hydrating water upon the addition of urea to the GPG-NH₂:water mixture visible in both simulation techniques. For example, the CN of the $g_{Hn1Ow}(r)$ decreases from 0.9 to 0.8 in both simulations.

While in EPSR all of the hydrogens lose hydrating water upon urea addition, where this is most pronounced for the hydrogen Hn1 (-13%), in MD only the N- and *cis* (to O3) C-terminal hydrogens lose water contacts. MD shows a clear site preference for HT2 (*cis* to O3) to be less hydrated upon urea addition and for HT1 (*trans* to O3) to have more water contacts in the presence of urea. In EPSR, all of these backbone hydrogens have a clear preference to form hydrogen bonds with urea rather than water. The same is true for the MD simulation, however in MD the polar Hn3 has very strong preference for urea while HT1 only has a very small preference for urea.

The backbone nitrogen-chloride interaction RDFs are shown in Figure 3.19. For EPSR the addition of urea leads to a clear increased interaction between the chloride anions and the polar amide nitrogens and even more with the positively charged amine N-terminus (N1) where the CN almost doubles from 0.3 to 0.6 for EPSR. This effect does not occur in MD where the coordination of the backbone nitrogens and chloride is almost unchanged by urea.

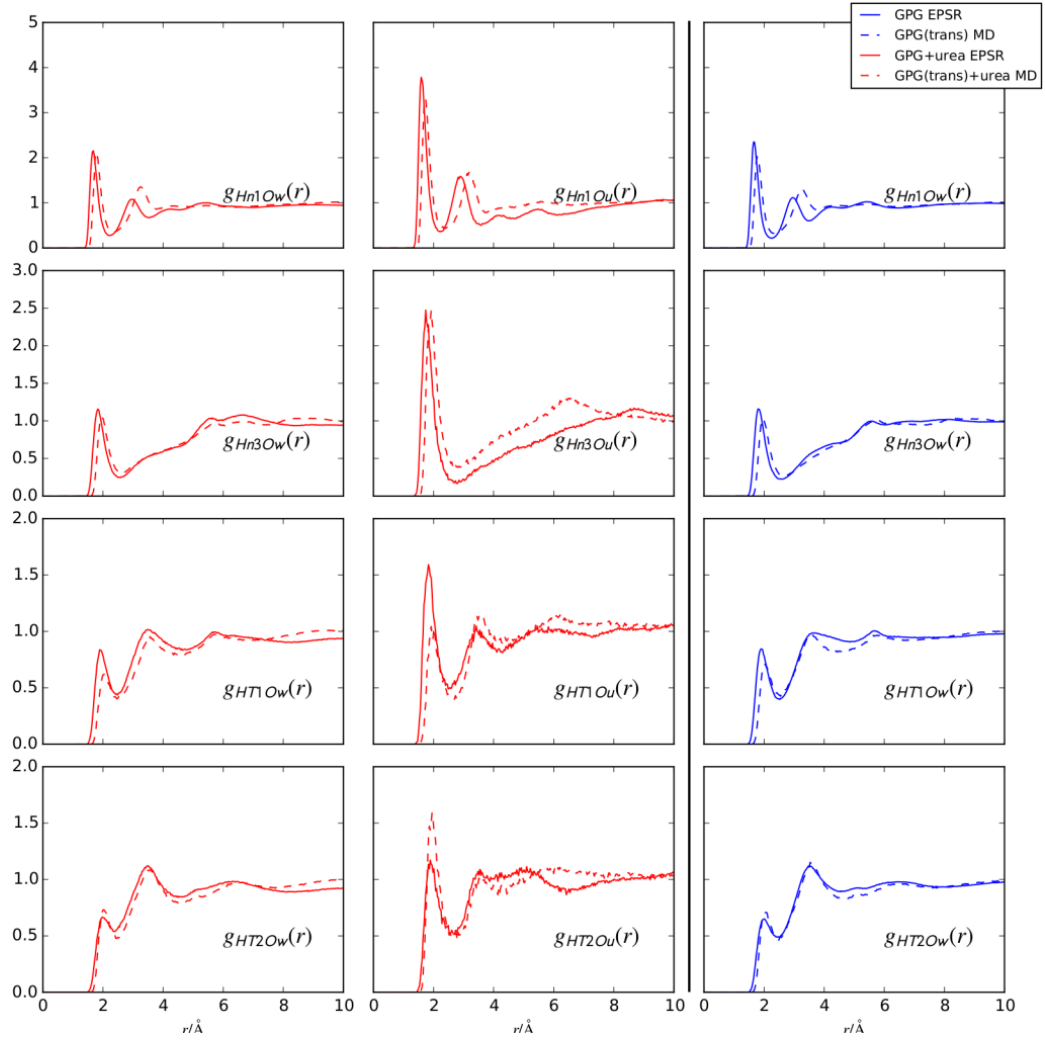


Figure 3.18: RDFs for GPG-NH₂ peptide bond hydrogens and Ow/Ou atoms from water and urea, respectively, from MD and EPSR, including for GPG-NH₂:water [183].

Table 3.11: Coordination numbers (n_{α}^{β}) for the peptide backbone hydrogens and water oxygens (Ow) and urea oxygens (Ou) around GPG-NH₂ NH hydrogens from EPSR and MD in Figure 3.18, including for GPG-NH₂:water [183]. The minima where the n_{α}^{β} were taken are also listed. Also shown are the hydration difference (HD) and the urea/water preference ratio (PR).

	$r_{max} / \text{\AA}$	GPG-NH ₂		GPG-NH ₂ +urea		HD / %		PR	
		EPSR	MD	EPSR	MD	EPSR	MD	EPSR	MD
Hn1-Ow	2.3	0.9	0.9	0.8	0.8	-13	-7	-	-
Hn3-Ow	2.5	0.8	0.8	0.8	0.8	-4	+1	-	-
HT1-Ow	2.5	0.8	0.7	0.7	0.7	-10	+10	-	-
HT2-Ow	2.5	0.7	0.7	0.7	0.6	-8	-13	-	-
Hn1-Ou	2.3	-	-	0.1	0.1	-	-	1.5	1.6
Hn3-Ou	2.5	-	-	0.1	0.2	-	-	1.1	2.8
HT1-Ou	2.5	-	-	0.1	0.1	-	-	1.8	1.0
HT2-Ou	2.5	-	-	0.1	0.1	-	-	1.8	1.7

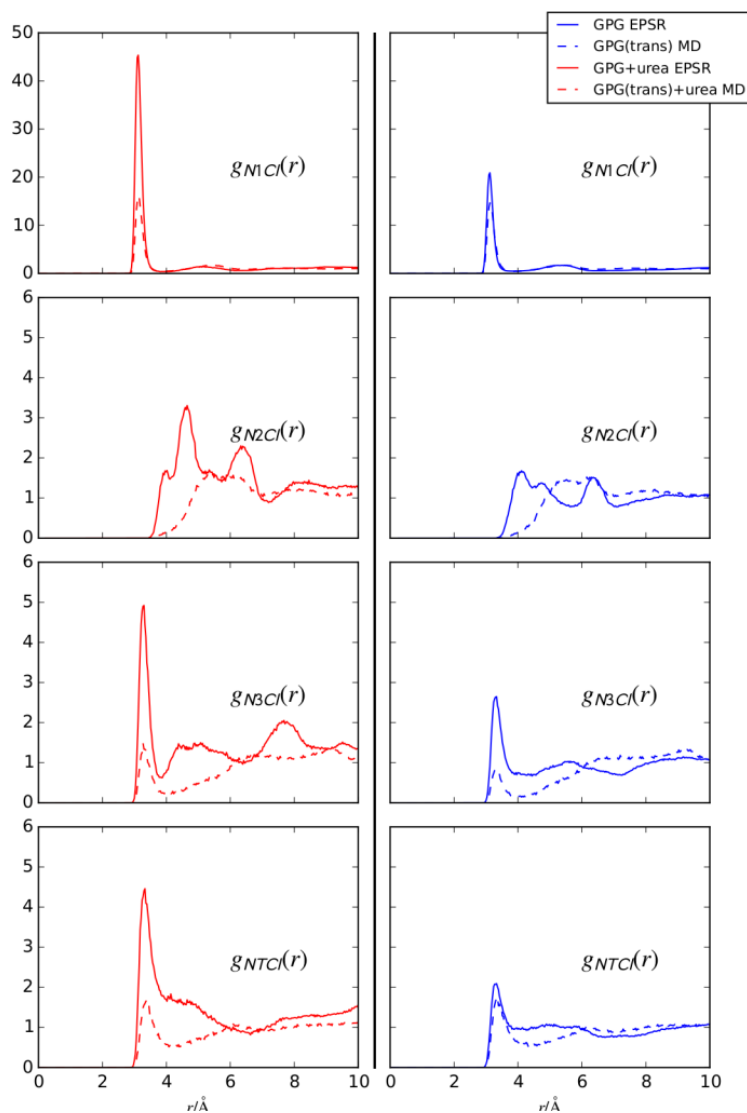


Figure 3.19: RDFs for GPG-NH₂ peptide bond nitrogens and N terminal nitrogen and Cl atoms from MD and EPSR, including for GPG-NH₂:water [183].

Table 3.12: Coordination numbers (n_{α}^{β}) for chloride and GPG-NH₂ nitrogens from EPSR and MD simulations in Figure 3.19, including for GPG-NH₂:water [183]. The minima where the n_{α}^{β} s were taken are also listed.

	$r_{max} / \text{\AA}$	GPG-NH ₂		GPG-NH ₂ +urea	
		EPSR	MD	EPSR	MD
N1-Cl	3.9	0.3	0.3	0.6	0.3
N2-Cl	5.7	0.3	0.3	0.5	0.2
N3-Cl	4.2	0.1	<0.1	0.1	0.1
NT-Cl	4.1	0.1	0.1	0.2	0.1

3.3.6.3 Apolar Solvation

The hydration and urea coordination of remaining peptide atom types (C α s and H α s, carbonyl Cs and proline side chain carbons and hydrogens) is shown in Figure 3.20. The hydration of the mostly apolar groups decreases slightly when urea is added, for example decreases the CN (Table 3.13) of the Co atoms from 4.9 to 4.4 in EPSR and from 5.0 to 4.2 in MD. The loss of water coordination is on average higher ($\approx$ -13%) in the MD simulations than in the EPSR fit to the neutron data ($\approx$ -9%). This might be due to a slight preference for urea in MD; whereas in EPSR the hydrophobic sites have a preference for water interaction. Water is relatively disordered around these groups and the addition of urea leads to a loss of water at a larger distance ($< 5 \text{ \AA}$) but no clear dehydration. In fact the magnitude of loss of water contacts is comparable with the polar backbone sites in Section 3.3.6.1-3.3.6.2.

Table 3.13: Coordination numbers (n_{α}^{β}) at $2.52 \text{ \AA}$ for the GPG-NH₂ apolar atoms and water and urea oxygens (Ow) and (Ou) from EPSR and MD simulations in Figure 3.20, including for GPG-NH₂:water [183]. The minima where the n_{α}^{β} s were taken are also listed. Also shown are the hydration difference (HD) and the urea/water preference ratio (PR).

	$r_{max} / \text{\AA}$	GPG-NH ₂		GPG-NH ₂ +urea		HD /%		PR	
		EPSR	MD	EPSR	MD	EPSR	MD	EPSR	MD
Cp-Ow	4.7	9.0	9.6	8.3	8.5	-8	-12	-	-
Co-Ow	4.2	4.9	5.0	4.4	4.2	-10	-16	-	-
Cm-Ow	4.2	4.5	4.5	4.2	3.9	-8	-15	-	-
Cn-Ow	4.5	6.0	6.1	5.5	5.2	-8	-14	-	-
Hc-Ow	3.5	2.6	2.4	2.3	2.1	-9	-11	-	-
Cp-Ou	4.7	-	-	0.6	0.7	-	-	1.0	1.2
Co-Ou	4.2	-	-	0.2	0.3	-	-	0.7	1.0
Cm-Ou	4.2	-	-	0.2	0.3	-	-	0.7	1.1
Cn-Ou	4.5	-	-	0.3	0.4	-	-	0.8	1.2
Hc-Ou	3.5	-	-	0.1	0.2	-	-	0.8	1.2

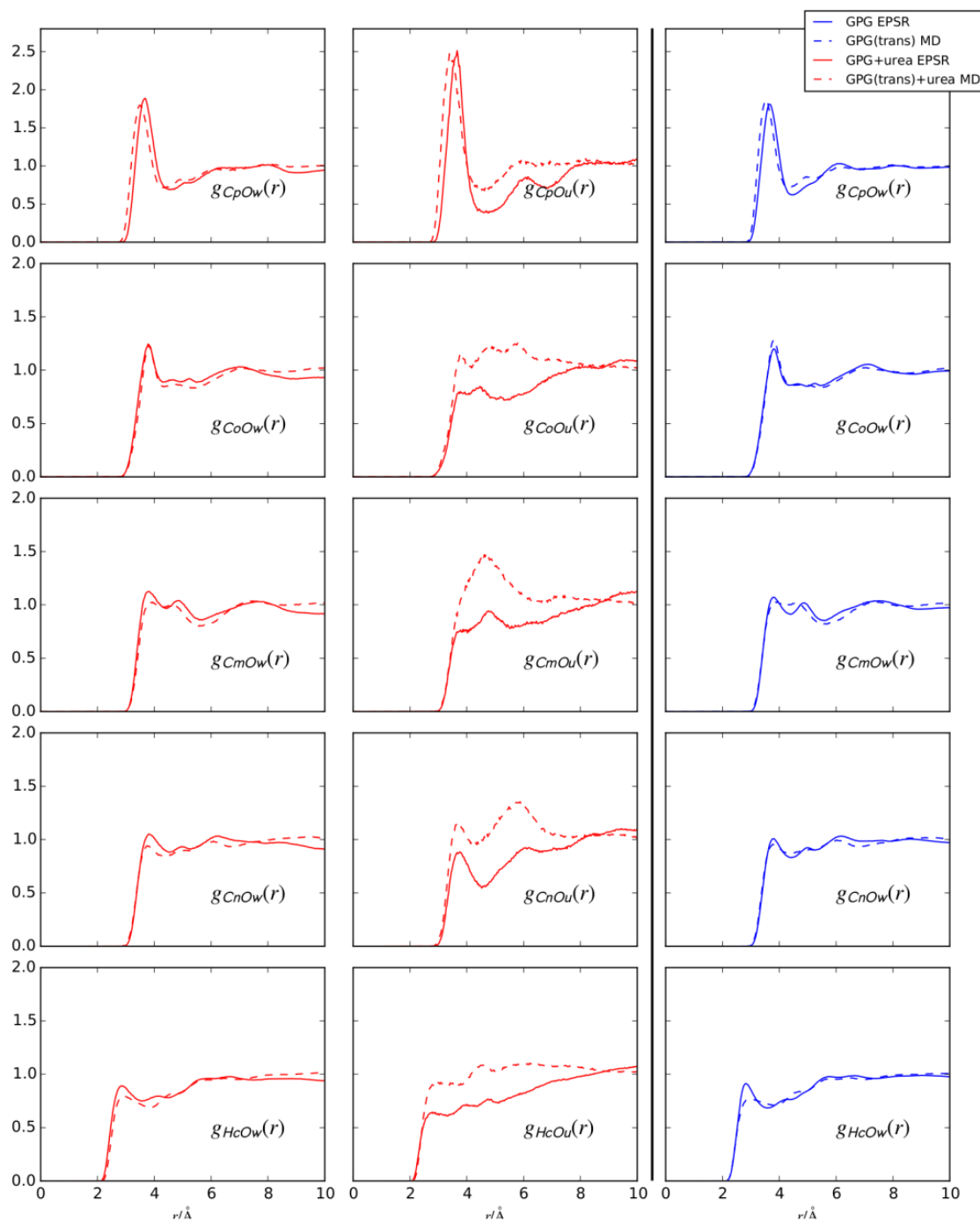


Figure 3.20: RDFs for GPG-NH₂ apolar atoms and Ow/Ou atoms from water and urea, respectively, from MD and EPSR, including for GPG-NH₂:water [183].

3.3.7 Whole molecule analysis

Whole molecule analysis (WMA) has also been performed using ANGULA [229]. In the WMA in Figure 3.24, the GPG-NH₂ molecules at the center of these plots are a representative of the distribution of the GPG-NH₂ conformations in each of the simulations. While it cannot be prevented that the selected peptide conformations influence the observed solvent distribution around them, these conformations were determined by selecting for the most likely conformation by selecting the backbone dihedral angles in Table 3.14. Figure 3.21 shows the dihedral labelling of GPG-NH₂ and Figure 3.22 the dihedral angle distributions. Correlations between these dihedral distributions were ruled out by the bivariate analysis. In general, the GPG-NH₂ peptides show more confined conformations in MD and more flexibility in EPSR. Unlike the intramolecular distance in Figure 3.13 the dihedral angles show that there are more than two unique peptide conformations in EPSR. Apart from the Gly1-Pro2 peptide bond ($\omega_{\text{Gly1-Pro2}}$: $\text{C}\alpha_{\text{G1}}\text{-C(=O)-N2-C}\alpha_{\text{P2}}$) both *cis* and *trans* MD simulations have almost identical dihedral angle preferences. In EPSR the peptide bond angles $\omega_{\text{Gly1-Pro2}}$ and $\omega_{\text{Pro2-Gly3}}$ ($\text{C}\alpha_{\text{P2}}\text{-C(=O)-N3-C}\alpha_{\text{G3}}$) do not have clear maxima at 180° (or 0° for *cis*) but are slightly shifted by $\approx 30^\circ$ and ψ_{Pro2} ($\text{N2-C}\alpha_{\text{P2}}\text{-C(=O)-N3}$) exhibits no clear dihedral preference. As a further work, it might be interesting to compare these dihedral angle preferences to GPG-NH₂ without urea.

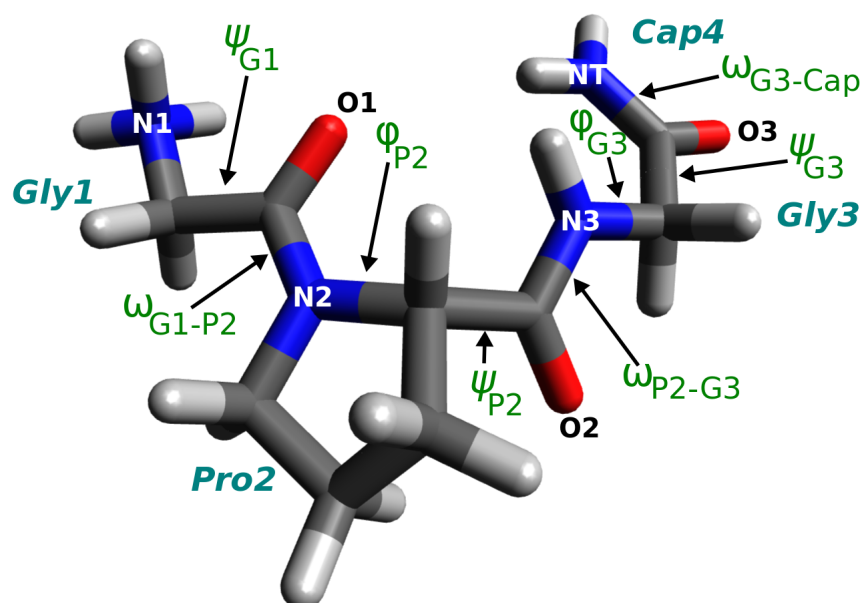


Figure 3.21: Dihedral angles along the GPG-NH₂ backbone. The labelled bond refers to the respective rotational axes.

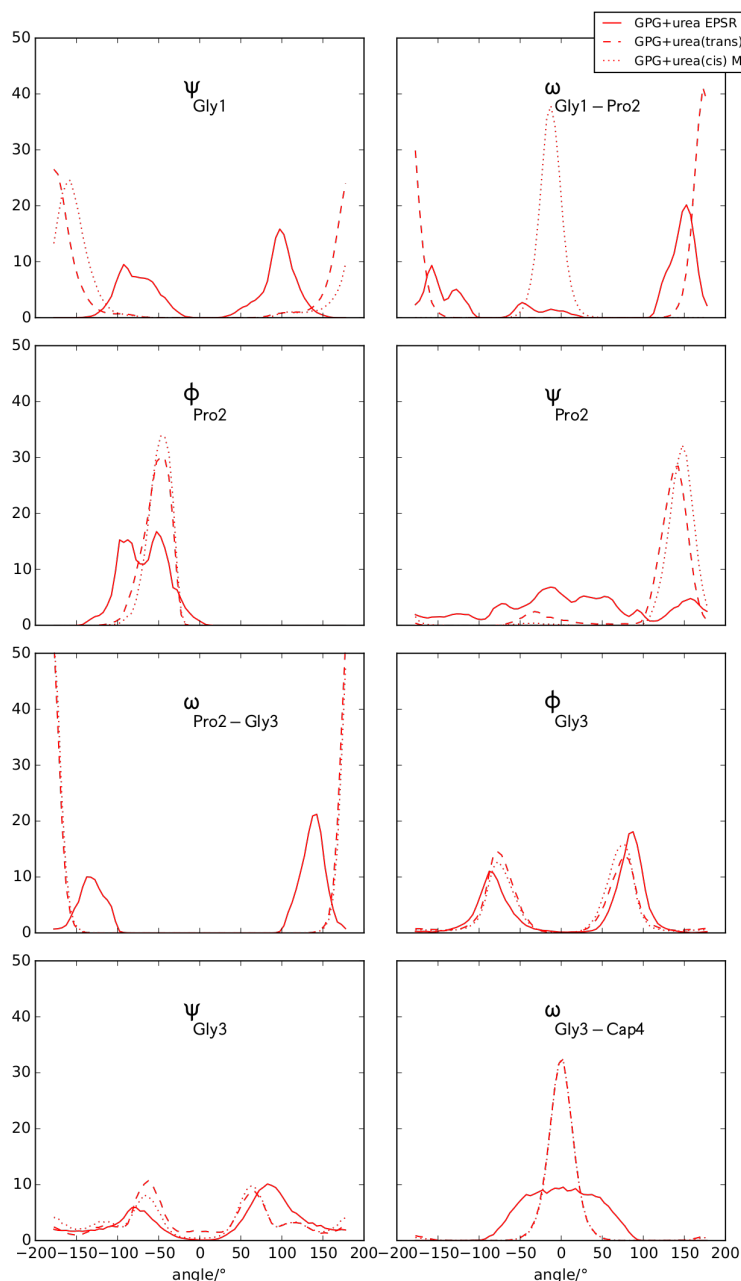


Figure 3.22: Dihedral angle distributions along the GPG-NH₂ backbone (labels in Figure 3.21) in the urea containing simulations.

Figure 3.23 shows the related whole molecule RDFs of urea carbon and water oxygen around GPG-NH₂. The nearest neighbour peak at 1.8 Å of the water RDF represents the close water interactions *via* the positively charged N-terminus and the polar amine groups. In EPSR, this peak is slightly closer to the peptide than in both MD simulations. The second peak at 2.8 Å accounts for other water hydrogen bonds with the peptide and is higher for EPSR than for MD. To account for all urea-peptide interactions, including urea as a hydrogen bond donor and acceptor, the central urea carbon was chosen for

Table 3.14: Dihedral angles selected ($\pm 20^\circ$) for the whole molecule analysis depictions in Figure 3.24. The related dihedral distributions are shown in Figure 3.22 and the labels in Figure 3.21.

Label	Dihedral between	GPG-NH ₂ +urea		
		EPSR	MD <i>trans</i>	MD <i>cis</i>
ψ_{Gly1}	N1-C α_{G1} -C(=O)-N2	100	180	150
$\omega_{\text{Gly1-Pro2}}$	C α_{G1} -C(=O)-N2-C α_{P2}	150	180	0
ϕ_{Pro2}	C(=O)-N2-C α_{P2} -C(=O)	-50	-50	-50
ψ_{Pro2}	N2-C α_{P2} -C(=O)-N3	-20	150	150
$\omega_{\text{Pro2-Gly3}}$	C α_{P2} -C(=O)-N3-C α_{G3}	140	180	180
ϕ_{Gly3}	C(=O)-N3-C α_{G3} -C(=O)	80	80	80
ψ_{Gly3}	N3-C α_{G3} -C(=O)-NT	75	75	75
$\omega_{\text{Gly3-Cap}}$	C α_{G3} -C(=O)-NT-HT1	0	0	0

the WMA distribution. This results in a whole molecule RDF with a broad peak at $\approx 3.0 \text{ \AA}$ in Figure 3.23 for all simulations. However, in EPSR this peak is markedly smaller, indicating less peptide-urea contacts than in MD simulations. This is consistent with the preceding peptide site analysis in Section 3.3.6 where, overall, more peptide-urea contacts were visible in MD than in EPSR.

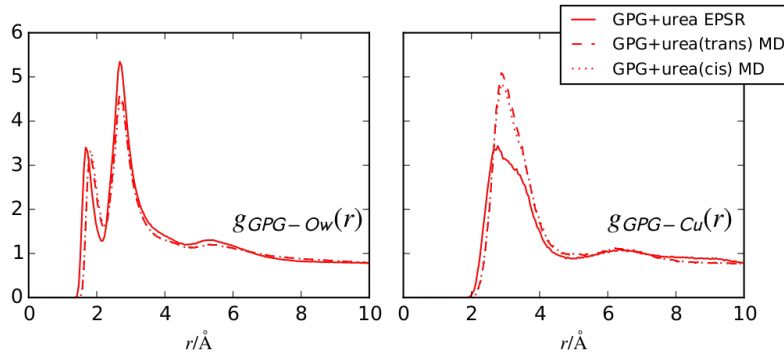


Figure 3.23: RDF of water oxygen (Ow) and urea carbon (Cu) around the whole GPG-NH₂ peptide.

While the first column of Figure 3.24 depicts the chosen peptide conformation, the probability density of water molecules around GPG-NH₂ is shown in the second column of Figure 3.24 within a distance range of 0-4.0 $\text{\AA}$ from any atom on the peptide. In both MD and EPSR there is a marked preference for water to be located around the polar regions of the peptides with MD showing a higher propensity of waters being either around the N-terminus or around the Pro2-Gly3 peptide bond in a fairly broad

distribution, consistent with the RDFs and SDMs in Figures 3.16 and 3.17. EPSR, on the other hand, shows a similar hydration shell but also shows density on the other side of the GPG-NH₂ peptide and overall a broader distribution of hydration compared to MD. In each of these hydration WMA plots, there is a relative absence of water around the proline ring portions of the peptide. This is in accordance with the side chain hydration RDFs (Figure 3.20) where the first hydration peak was usually small.

The third column of in Figure 3.24 shows the most probable distributions of urea molecules around GPG-NH₂, with EPSR and MD showing fairly different density distributions. For EPSR, urea preferentially occupies spaces where water is absent, while MD shows a broader distribution of urea and water and urea have a similar propensity in MD to occupy many the same spaces around the GPG-NH₂ peptide. This is most clear for the band of urea density around the Pro2-Gly3 portion of the molecule in Figure 3.24. As opposed to the hydration density, the GPG-NH₂-urea WMA shows density around the proline ring, with especially MD showing some density near this ring, which is in accordance with the findings in Section 3.3.6.3 where apolar groups on GPG-NH₂ had a preference for urea over water contacts in the MD simulation.

3.4 Conclusions

In this investigation, the bulk water structure in the GPG-NH₂:water:urea system does not appear to be significantly perturbed by the presence of urea in solution apart from the slight perturbation of the second neighbour shell in the $g_{OwOw}(r)$ s in Figure 3.8 [183, 230]. This is consistent with previous investigations on TMAO/urea in aqueous solution where it was found that large increases in the amount of urea in solution did not significantly perturb the bulk water structure [226]. This lack of perturbation suggests that an ‘indirect’ mechanism of protein denaturation by urea altering the bulk water structure is not dominating, in agreement with previous investigations [225, 231]. But the slight change in local water structure might, especially at higher urea concentrations, still be a secondary effect of urea on the solvent.

As opposed to previous simulations which found that water was preferred over urea

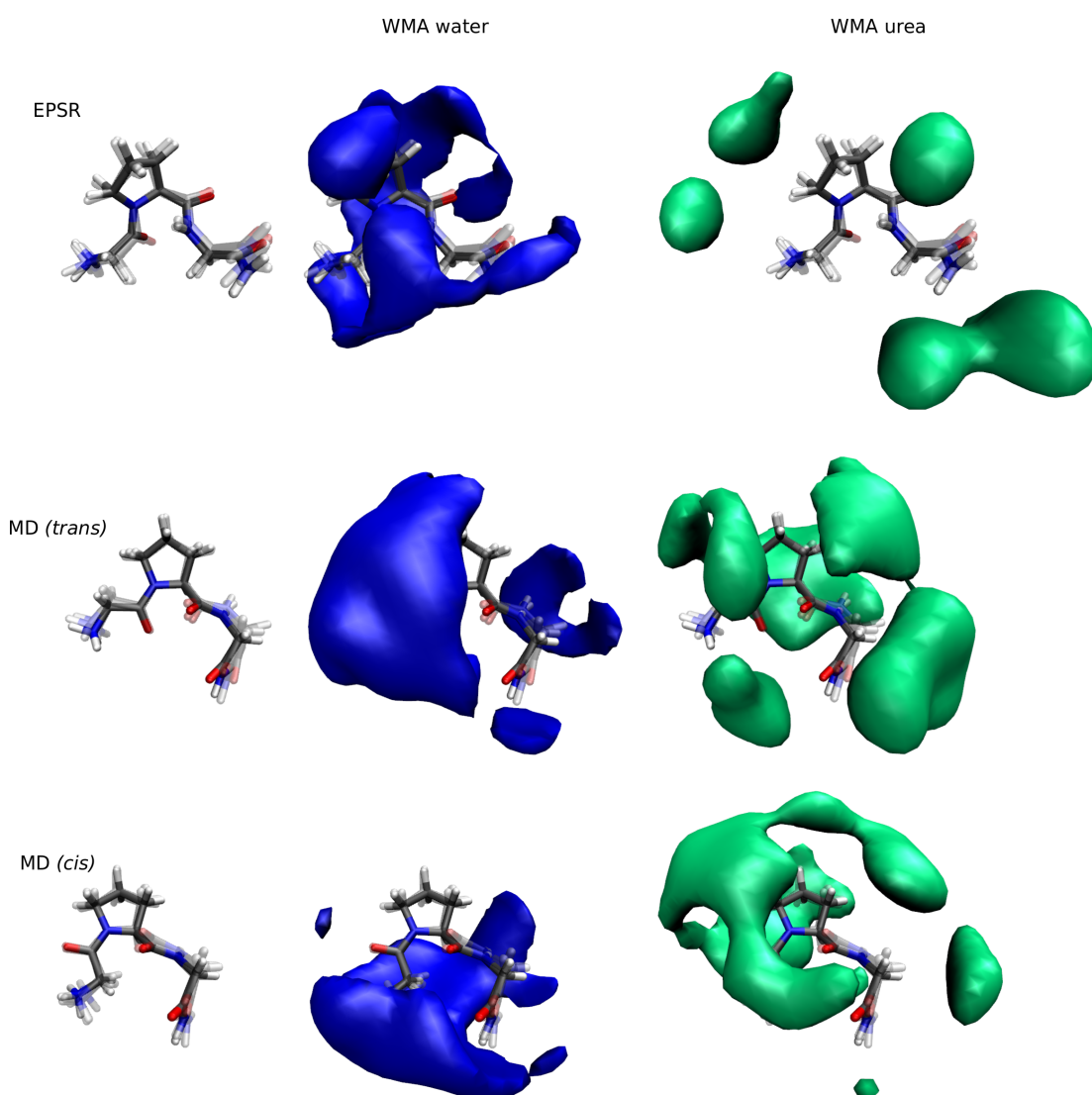


Figure 3.24: Most probable conformations and whole molecule analysis of solvent clouds around the most probable conformation of GPG-NH₂. Water oxygens (Ow, blue, 35%, up to 4.0 Å) and urea carbons (Cu, green, 50%, up to 5.0 Å)

for the charged and polar portions of the peptides [221]; here, the charged groups of the GPG-NH₂ peptide in almost all cases showed a preference to form hydrogen bonds with urea rather than water. The only exception to this preference is the Gly1-Pro2 peptide bond carbonyl oxygen (O1) from the EPSR fits to the neutron data which showed a clear preference to be hydrated rather than to form hydrogen bonds with urea, which may be due to the fact that in MD GPG-NH₂ has more of a ‘folded’ conformation.

Interestingly, urea was preferred over water at a larger distance (> 3.4 Å) from apolar groups of the peptide, in agreement with previous MD studies [221]. However in the EPSR fit to the neutron data, this preference for urea on apolar sites was not observed. Instead, EPSR showed that urea is preferred at mostly polar or charged groups that are in this case found on the peptide backbone.

In general, MD also showed a slightly higher level of urea-GPG-NH₂ contacts compared to EPSR where this preference was most notable for the the Gly1-Pro2 (N2) nitrogen (Figure 3.16). This preference for urea in the MD simulations is accentuated by the WMA shown in Figure 3.24 where there is a much broader distribution of GPG-NH₂-urea interactions compared with the WMA for the EPSR fits to the neutron data. Even though the distribution of urea is quite different for each simulation, they both show a fairly significant probability of urea-proline ring interactions while there is an absence of water-proline contacts in GPG-NH₂ WMAs. This observation indicates that urea preferentially associates with the hydrophobic sites of proteins rather than water as previously proposed [195]. The urea density in the WMAs, for both simulation techniques, is located around different portions of the GPG-NH₂ backbone and the proline ring. This observation supports simulation studies on group transfer free energies which suggest that urea denatures proteins due to largely favourable interactions with the protein backbone and side chains [104, 105].

That urea is preferred over water around most of the peptide bonds indicates that urea can easily replace water along the GPG-NH₂ backbone, thus lowering the overall hydration of the peptide. This is what seems to occur for the amide nitrogens in the backbone, with the SDMs in Figure 3.17 suggesting that water molecules are simply replaced by urea in the same locations around these nitrogen containing groups, this

is especially evident for the Pro2-Gly3 peptide bond. For the peptide bond carbonyl oxygens however, the SDMs in Figure 3.15 indicate that while there is some replacement of water by urea, that urea also has a high probability to bind to the backbone oxygens at different locations in space. The fact that the addition of urea lowers the peptide hydration regardless of a preference for urea or water at that site, suggests that the effect of urea is more than a simple replacement of water. Taken together, these two phenomenon suggest that the amide hydrogens in the protein backbone might be the initial site of water replacement by urea forming strong hydrogen bonds with -NH- groups in a peptide or, by extension, proteins. Although in the current work only a small model peptide and a relatively low urea concentration has been used, it may be that when initiating denaturation of a peptide or protein backbone, urea first attacks the backbone by displacing the waters around the amide groups and then as a secondary step bridges from the -NH- group to the peptide bond oxygens - which eventually will ‘unzip’ the backbone, exposing the hydrophobic core to the surrounding solvent molecules.

A potential urea-effect on co-solute ions would need to be confirmed by additional experiments. This aspect will be discussed in the next chapter.

Chapter 4

Conformation of KGXGK Peptides and Influence of Urea on the KGPGK Peptide

4.1 Introduction

Previous studies have shown an ambiguous influence of salts such as LiCl, CaCl₂ and KCl on the ability of urea to denature proteins. Lower salt concentrations have been shown to stabilise against urea-induced denaturation in CD and other spectroscopic measurements [232, 233]; whereas higher salt concentrations appear to destabilise the protein in the presence of urea [232]. In addition, osmolytes such as trimethylamine N-oxide (TMAO) and proline are known to protect against urea-induced denaturation [234], and certain proteins from halophilic organisms are resistant against urea denaturation in the presence of high salt (NaCl) concentrations [235]. Recently, the protective properties of osmolytes was shown by NMR spectroscopy to be caused by a direct hydrogen bonding interaction between the osmolyte betaine and urea [236]. Under other conditions, a low [237] or high pH [238], has been shown to facilitate urea-induced denaturation of proteins in solution.

Interestingly, although it is well-known that proteins only function under very spe-

cific conditions, many *in vitro* investigations of proteins, such as X-ray crystallography [239] and other biochemical methods use high-salt buffers, organic polymers and unphysiological pH conditions in order to stabilise and solubilise proteins. There is clearly a conundrum which is not easy to overcome in biophysical investigations of proteins; while *in vivo* proteins only function in their native form under heavily prescribed conditions, unphysiological conditions must often be used in order to perform studies on proteins *in vitro*.

The model peptide lysine-glycine-proline-glycine-lysineamide (in the following referred to as KGPGK; Figure 4.2), both with and without urea in solution, has been investigated and the atomic scale details analysed. The lysine residues were chosen to be added to the GPG-NH₂ sequence in order to increase both the solubility and flexibility of the peptide within the solution. However, the addition of lysine residues also lowers the pH significantly compared to the GPG-NH₂ solutions. In both KGPGK solutions, with and without urea, the anion trifluoroacetic acid (TFA) is present, where TFA⁻ is a commonly used counter anion in peptide synthesis.

As is apparent throughout this work, proline is a unique amino acid as it is occasionally found in a *cis* peptide bond (Figure 4.1) conformation within the amino acid chain of certain proteins [240, 241]. But unlike glycine, the cyclic structure of proline makes it one of the conformationally more rigid amino acids within a folded protein structure. However, in an unfolded peptide [242], folding polypeptide chain [34] or a flexible protein part [243] proline can flip between *cis* and *trans* conformations. Its tertiary amine structure, within a peptide bond, introduces a kink into the polypeptide chain often interrupting repetitive secondary structure elements, such as sheets and helices, which explains its prevalence in turns and unique protein structures such as in left-handed collagen helices [244, 245].

The pentapeptide KGPGK contains a sequence which commonly forms turns in many proteins (XPGX) [199, 200, 201, 202, 203], is not folded in solution, but exists in a variety of conformational states. KGPGK allows for the observation of a peptide in the process of folding, so that the role of water can be probed directly. The XGPGX sequence is found 1347 times in the UniProt Protein Database [246], mostly in loop-

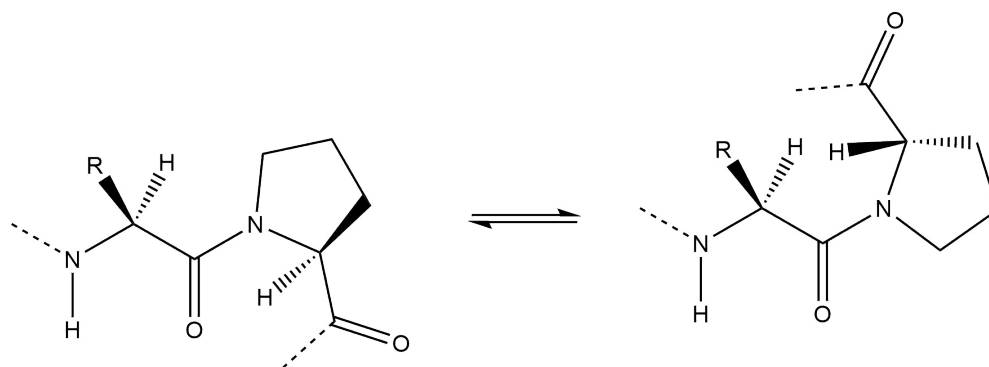


Figure 4.1: A proline peptide bond in *trans* and *cis* conformation.

sequences; 39% in β -turns and 10% in γ -turns. The positively charged lysine residues next to the GPG motif were chosen in order to add more flexibility and solubility to the system in order to observe a transient conformation in the process of folding.

In addition, previous studies have shown a possible water mediating between the β -turn typical $i \rightarrow i+3$ hydrogen bond in GPG-NH₂ which might precede a direct hydrogen bonding in a protein structure; where water perhaps ‘nucleates’ the beginning of protein folding [198].

Firstly, this chapter will focus on the balance between a salt and urea, and how they interact with different parts of a peptide in solution. Secondly, this chapter will probe the conformational and hydration specificities that proline introduces into the peptide chain. Specifically, four model peptides KGPGK, KGGGK, KGSGK and KGLGK will be structurally analysed. These peptides (in the following referred to as KGXGK) were chosen on the basis of the KGPGK peptide to study proline in comparison with a nonpolar (L), a polar (S) and a highly flexible (G) residue. The data has been derived from an EPSR fit to the neutron and X-ray diffraction data and MD in case of KGPGK, only *trans* conformations will be considered in this chapter) and only by MD in the case of KGGGK, KGSGK and KGLGK.

4.2 Methods

4.2.1 NMR Spectroscopy Sample Preparation and Measurement

Lysine-glycine-proline-glycine-lysineamide with trifluoroacetic acid ($\text{KGPGK}^{3+} \cdot 3\text{TFA}^{-}$) was purchased from Proteogenix (Schiltigheim, France) and used without further purification. Peptide composition and sample purity were confirmed by elemental analysis (Medac Ltd., Table 4.3) and NMR spectroscopy (Figure 4.6). The NMR samples and other deuterated samples were prepared and measured as described in Section 3.2.1. The *cis/trans* ratio of the Gly2-Pro3 peptide bond was determined by integration of proline *cis* and *trans* peaks. Apart from spectra that were used to assign the peaks and ensure sample purity, complete deuteration of the 100% D samples was confirmed by NMR spectroscopy and the influence of the higher peptide concentration in the neutron scattering samples and the influence of urea and a neutral pH were also assessed by NMR spectroscopy.

4.2.2 Neutron and X-ray Diffraction and Sample Preparation

The samples for neutron and X-ray diffraction measurements were prepared by weight as described in Section 3.2.2 with the appropriately isotopically labelled peptide, urea, H_2O and/or D_2O (99.98%) to give a molecular ratio of 1:346 (peptide:water) and 1:2.75:346 (peptide:urea:water) which was found to be the upper limit of peptide and urea solubility. This corresponds to a peptide concentration of ≈ 0.15 M, a TFA concentration of ≈ 0.45 M and a urea concentration of ≈ 0.4 M. The level of deuteration is with respect to the exchangeable hydrogens in the sample (H3n, H2, H4, H5, H6a, H6b, Hu and Hw). The measured pH of KGPGK:TFA in water was 2.3 and in the KGPGK:urea:TFA solution the measured pH was 4.4.

Neutron diffraction data for five isotopically substituted samples (Table 4.1) of KGPGK in solution, both with and without urea, were collected on the SANDALS diffractometer at the ISIS Facility (STFC, UK) for ≈ 8 hours per sample at 298 K and 1 atm pressure and as described in Section 3.2.2. For each data set, the measured scat-

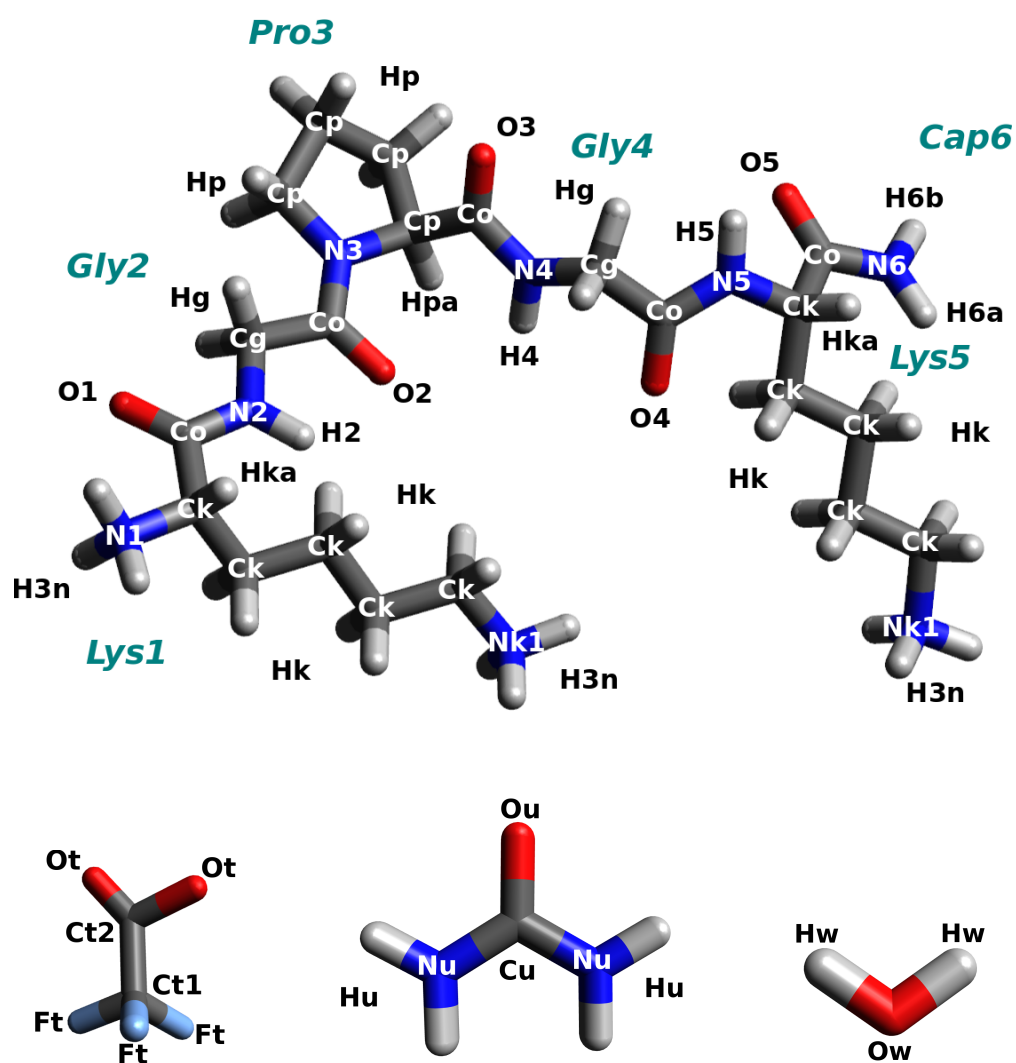


Figure 4.2: Molecular structure of the $\text{H}_3\text{N-KGPGK-NH}_2^{+3}$ (*trans*) pentapeptide, TFA, urea and water including the labelling scheme used in this and the following chapter.

tering levels were within 10% of the expected theoretical values (DCS levels in Table 4.1). The neutron scattering data were corrected using the GudrunN programme (Section 2.2.7.3). X-ray diffraction data were collected for the fully protiated sample contained in 0.1 mm diameter borosilicate glass capillary using a Panalytical X’pert Pro diffractometer, also available at the ISIS facility, which has an incident wavelength of 2 Å. The measured X-ray scattering data were reduced and corrected for effective density, absorption, bremsstrahlung and Compton scattering and normalised to the total form factor and converted to a structure factor using GudrunX [172].

Table 4.1: List of samples used for the collection of neutron diffraction data. For each dataset the 100% H sample was also used for X-ray diffraction. The molecular ratio of KGPGK·3TFA:water was 1:346 and KGPGK·3TFA:urea:water 1:3:346. The level of deuteration refers to the exchangeable hydrogens on KGPGK, urea and water (H3n, H2, H4, H5, H6a, H6b, Hu and Hw).

Sample Label	Deuteration	KGPGK + water		KGPGK +urea + water	
		μA	DCS level	$\mu / \text{\AA}$	DCS level
I	100% H	1572	90.9%	1664	97.0%
II	64% H 36% D	1452	91.3%	1634	94.7%
III	50% H 50% D	1282	93.6%	1620	93.6%
IV	36% H 64% D	1353	92.6%	1579	93.1%
V	100% D	801	92.6%	1568	96.5%

4.2.3 Empirical Potential Structure Refinement

For the work presented here, the modeling box contained 20 KGPGK molecules (atomic labeling for the KGPGK molecules is shown in Figure 4.2), 60 TFA[−] ions and 6920 water molecules at the measured density ($\rho = 1.098 \text{ g/ml} = 0.1006 \text{ atoms}\cdot\text{\AA}^{-3}$) and a temperature of 298 K. For the simulations with urea, 55 molecules of urea were added to this mixture ($\rho = 1.044 \text{ g/ml} = 0.1008 \text{ atoms}\cdot\text{\AA}^{-3}$). Starting potentials for water molecules were taken from the TIP3P water model [152] and parameters for the peptide and urea molecules and TFA ions from the CHARMM36 forcefield [148, 206, 207, 208, 209] and modified in order to adjust for different atomic labeling in EPSR compared with MD. These parameter are shown in Table 4.2. Other parameters like starting conformation, dihedral angles and rotational groups are in Appendix A. The EPSR simulation contained a mixture of *cis* and *trans* KGPGK molecules (with respect to the Gly2-Pro3 bond) in ratios which correspond with that measured by ¹H NMR spectroscopy (15% *cis*;

85% *trans*). 3 *cis* and 17 *trans* KGPGK molecules with non-rotating peptide bonds were used as starting conformations.

4.2.4 Molecular Dynamics Simulations

MD simulations^a were performed at the same molecular ratios as the diffraction measurements and EPSR simulations. Each system contained 20 KGPGK-NH₂ molecules of the same starting conformation with *trans* Gly2-Pro3 peptide bonds which were observed to be the significant majority from the NMR spectroscopy experiments, 60 TFA⁻ counter ions, 55 urea molecules for the urea containing simulation and 6920 water molecules. The KGPGK-NH₂ molecules were modeled using the CHARMM36 forcefield [148, 206, 207] and the TFA⁻ ions and urea molecules were modeled using the CHARMM General force field [208, 209]. The water molecules were modeled with the TIP3P model [152] which has been modified for the CHARMM force field [211]. The MD simulation was performed as described in Section 3.2.4 for 80 ns with a timestep of 2 fs.

Further MD simulations were performed on KGGGK, KGSGK and KGLGK (each with a C-terminal amide cap).^b Each simulation contained 20 peptides, 60 TFA molecules and 6920 water molecules. The MD setup was according to Section 4.2.4 using the TIP3P water model and the CHARMM36 and CHARMM General force field.

4.2.5 ANGULA

Figure 4.3 shows the orthonormal coordinate systems assigned to the KGPGK peptide nitrogens and the water, TFA and urea molecules for the ANGULA analysis presented here. These coordinate systems and 4000 snapshots of both the EPSR and MD simulation boxes were used resulting in the same number of KGPGK molecules (80000) analysed. The position of the first neighbour molecules or molecules within a given distance range of each fragment (SDM) or the whole molecule (WMA) can be depicted as described in Section 2.4.

^aThese simulations were performed by Anna Genina. They were analysed and converted to $F(Q)$ s/ $G(r)$ s by myself.

^bThese simulations were performed by Chris Lorenz. They were analysed by myself.

Table 4.2: The parameters used in the EPSR simulations of the KGPGK/TFA/urea/water samples. The first 28 atom types are on the KGPGK peptide. For TFA, urea and water the second letter of the atom label (t, u, w) indicates the molecule to which the atom belongs, where t is for TFA, u is for urea and w is for water.

Atom Label	ϵ / kJ mol ⁻¹	σ / Å	q / e
Cg	0.23430	3.5814	-0.020
Ck	0.21422	3.5778	-0.038
Co	0.42677	3.5636	0.518
Cp	0.19351	3.9200	-0.085
H2	0.19246	0.4001	0.310
H3n	0.19246	0.4001	0.330
H4	0.19246	0.4001	0.310
H5	0.19246	0.4001	0.310
H6a	0.19246	0.4001	0.320
H6b	0.19246	0.4001	0.300
Hg	0.11715	2.3876	0.090
Hk	0.14226	2.3876	0.080
Hka	0.09205	2.3520	0.095
Hp	0.14226	2.3876	0.090
Hpa	0.09205	2.3520	0.090
N1	0.83680	3.2963	-0.300
N2	0.83680	3.2963	-0.470
N3	0.83680	3.2963	-0.290
N4	0.83680	3.2963	-0.470
N5	0.83680	3.2963	-0.470
N6	0.83680	3.2963	-0.620
Nk1	0.83680	3.2963	-0.300
Nk5	0.83680	3.2963	-0.300
O1	0.50208	3.0291	-0.510
O2	0.50208	3.0291	-0.510
O3	0.50208	3.0291	-0.510
O4	0.50208	3.0291	-0.510
O5	0.50208	3.0291	-0.550
Ct2	0.00837	4.1000	0.812
Ct1	0.71128	3.2000	0.661
Ft	0.40585	2.8500	-0.295
Ot	1.00420	3.4700	-0.794
Cu	0.29288	3.5636	0.600
Nu	0.83680	3.2963	-0.690
Hu	0.19246	0.4000	0.340
Ou	0.50208	3.0291	-0.580
Ow	0.63639	3.1507	-0.834
Hw	0.19246	0.4000	0.417

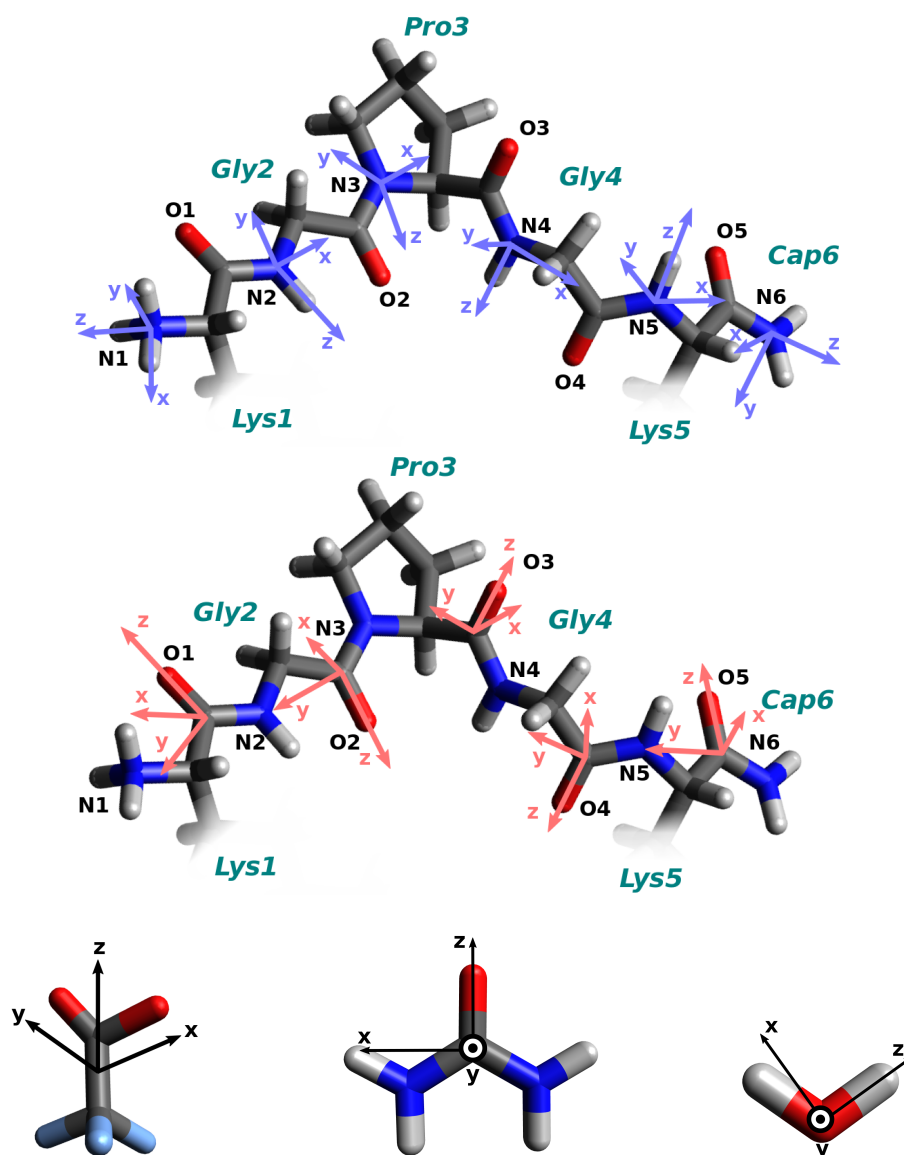


Figure 4.3: Coordinate systems used in ANGULA for the KGPBK and KGPBK+urea systems. The centre of the TFA molecules has been assigned the pseudoatom Ctm.

Figure 4.4 shows the basic KGXGK peptide labelling of residues and backbone dihedral angles. Orthonormal coordinate systems were assigned to the peptide backbone oxygens and nitrogens and the water molecules according to Figure 4.3. 4000 snapshots of both the EPSR and MD simulation boxes were used resulting in the same number of KGXGK molecules (80000) analysed.

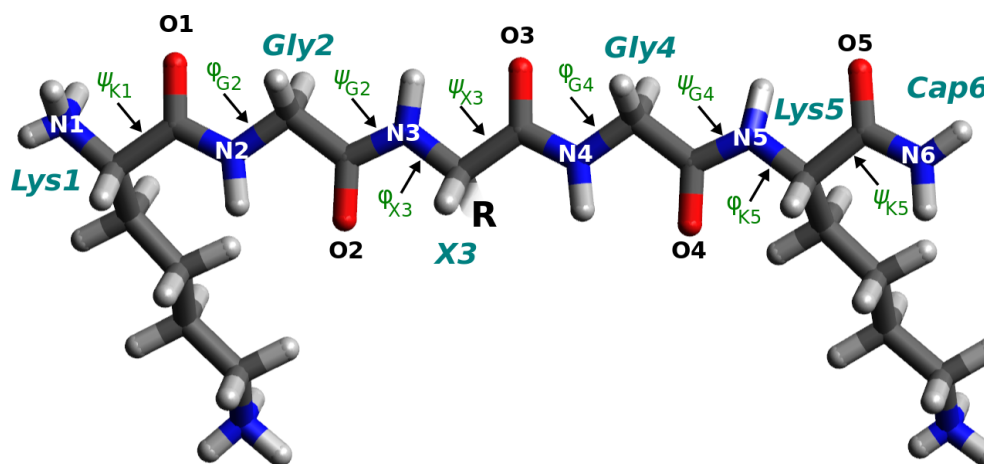


Figure 4.4: The basic structure and labelling of the KGPGK, KGGGK, KGSGK, KGLGK peptides. The labelled bond refers to the respective rotational axes of the dihedral angles. The varied residue of the third amino acid is $R = H$ (for Gly3), $R = C_3H_6$ (for Pro3, -H on N3), $R = CH_2OH$ (for Ser3), $R = CH_2CH(CH_3)_2$ (for Leu3).

4.2.5.1 Bridging Water Analysis

In order to analyse if a water molecule bridges between a transient hydrogen bond (for example between i and $i + 3$; O2 and N5) in the KGXGK peptides two nearest neighbour analyses of water molecules next to two sites of the peptide (e.g. the hydrogen bonding donor O2 and acceptor site N5) were performed and the resulting data sets of water molecules compared. If water molecules were the nearest neighbour for both sites they were included in the set of bridging water molecules or shared nearest neighbour molecules (Figure 4.5).

This analysis was not only performed for nearest neighbour water molecules, but also for second neighbour and first or second neighbour water molecules.

Table 4.3: Elemental analysis of KGPGK: %_m (Analysis) is the result of the elemental analysis performed by MEDAC Ltd., %_m (theor.) corresponds to the expected mass percentages of pure [H₃N-KGPGK-NH₂] $\cdot$ 3TFA.

Element	% _m (Analysis)	% _m (theor.)
C	37.09	39.23
H	3.77	5.24
N	12.14	13.55
O	24.88	21.29
Cl	< 0.10	-
F	22.87	20.69

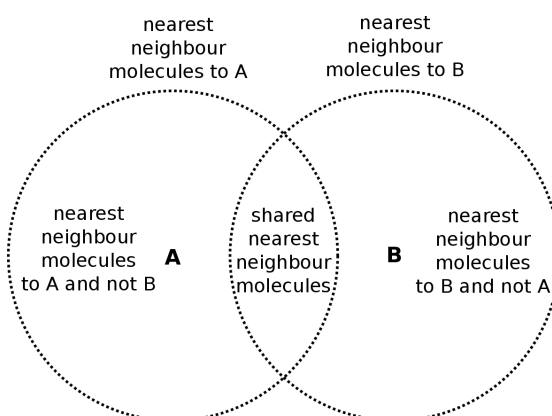


Figure 4.5: Schematic bridging/shared neighbour analysis between atoms A and B, which can be on different molecules or the same.

4.3 Results

4.3.1 NMR Spectroscopy and Elemental Analysis

Table 4.3 shows the result of the elemental analysis compared with the theoretical elemental mass percentages of pure [H₃N-KGPGK-NH₂] $\cdot$ 3TFA. Although the fluorine content is slightly higher than expected, a peptide:TFA ratio of 1:3 is still the best estimate; also considering that the neutron diffraction samples were within the expected differential cross section (Table 4.1).

The 1D ¹H NMR spectrum of 15 mM KGPGK in H₂O is shown in Figure 4.6. Smaller peaks corresponding to peptides in the *cis* conformation of the Gly2-Pro3 peptide bond are visible in the amide proton peaks for Gly4 (H4) and Lys5 (H5). The *cis*/*trans* ratio

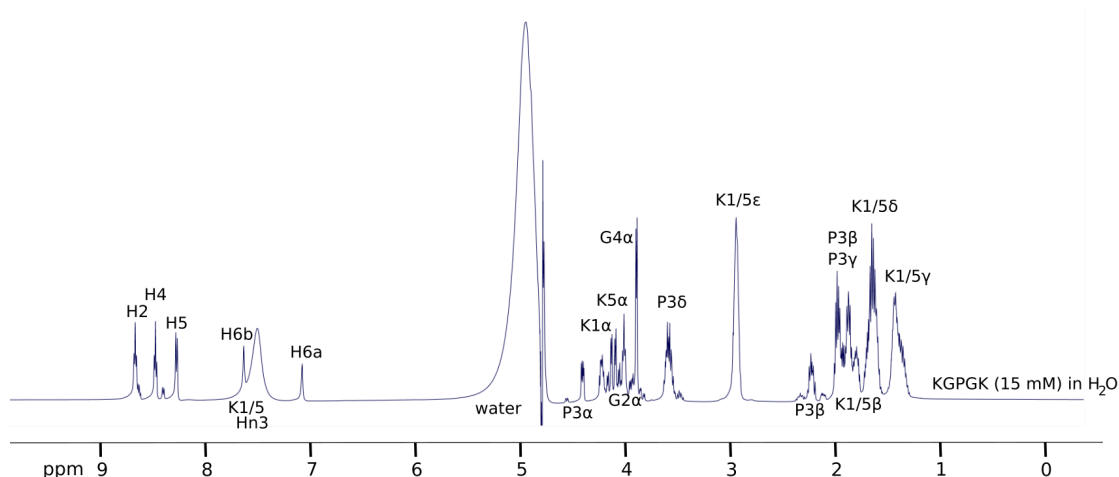


Figure 4.6: ^1H 1D NMR spectrum of 15 mM KGPGK in H_2O with peak assignment.

of 15% *cis*; 85% *trans* was determined by integration of the proline $\text{H}\beta$ *cis* and *trans* peaks.

Figure 4.7 shows the 1D ^1H NMR spectra of KGPGK under varying conditions. The *cis/trans* ratio is unchanged by an increased peptide concentration, as used for the neutron scattering experiments, the addition of 2 M urea or sodium hydroxide to achieve a neutral pH. The peptide backbone amide hydrogens and the lysine residue amine hydrogens are absent in the deuterated water sample so a complete hydrogen/deuterium exchange of these -NH hydrogens in the deuterated neutron scattering samples can be verified. The only other visible difference is that these hydrogen peaks are pH sensitive is that they are significantly decreased, due to a higher exchange rate with the solvent, at a neutral pH.

4.3.2 Diffraction Data and EPSR Fits

Figure 4.8 shows the collected neutron and X-ray diffraction data for KGPGK-3TFA solution samples with (Figure 4.8, bottom left) and without urea (Figure 4.8, top left) compared to the simulated $F(Q)$ fits from EPSR. The Fourier transformations for these functions into real space are also shown (Figure 4.8, right). Similarly, the comparison of the diffraction data with the $F(Q)$ and $G(r)$ data of the MD trajectories are shown in Figure 4.9.

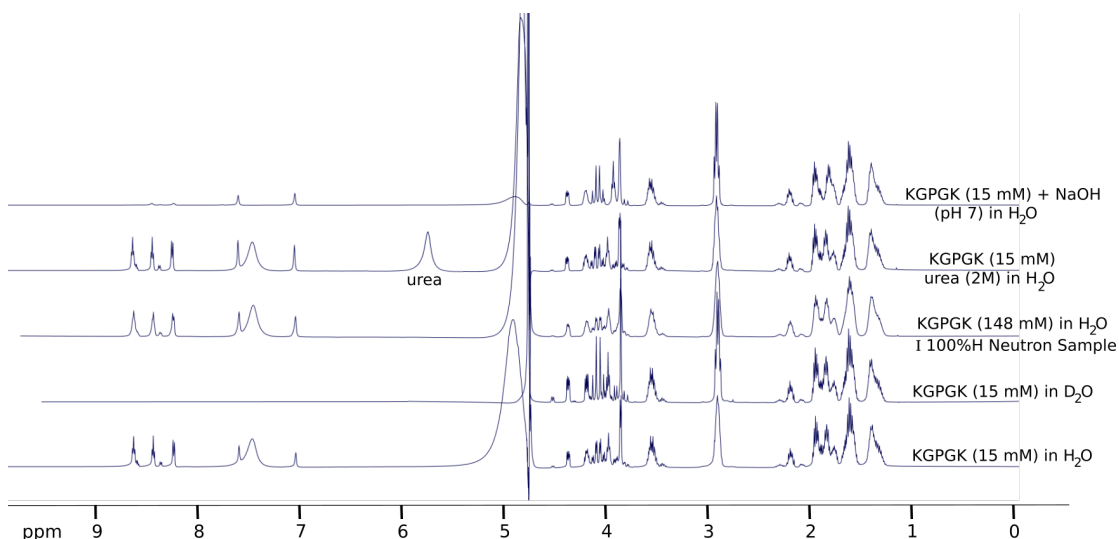


Figure 4.7: ^1H 1D NMR spectra of 15 mM D-KGPGK in D_2O , the I (100% H) neutron sample, 15 mM KGPGK in water with 2 M urea and 15 mM KGPGK in water at a neutral pH in comparison with 15 mM KGPGK in H_2O

EPSR fits the neutron and X-ray data well except at low Q values in the neutron diffraction data, where discrepancies in this region are likely due to under-estimated inelastic scattering corrections [247]. In MD, $F(Q)$ oscillations ranging from 6-10 $\text{\AA}^{-1}$ are often different from the experimental data which predominantly arise from the TIP3P water model as it has a fixed intramolecular O-H bond distance of 0.95 $\text{\AA}$, which is shorter than the measured OH position in the diffraction patterns and in the EPSR simulation. The X-ray scattering data is significantly different from the theoretical MD trajectories $F(Q)$. This might indicate a difference in possibly fluorine or oxygen correlations between experimental and MD simulated data, as they will have a slightly larger contribution to the X-ray scattering signal.

4.3.3 Peptide Backbone Interactions

4.3.3.1 Peptide Backbone Nitrogen Interactions

Figure 4.10 shows the RDFs for the peptide backbone nitrogen interactions with the water oxygens (Ow), TFA (Ot) and urea (Ou). The corresponding coordination numbers of the first peaks in the RDFs can be found in Table 4.4. From this figure, the two simulation techniques (EPSR and MD) show very similar results for the hydration of

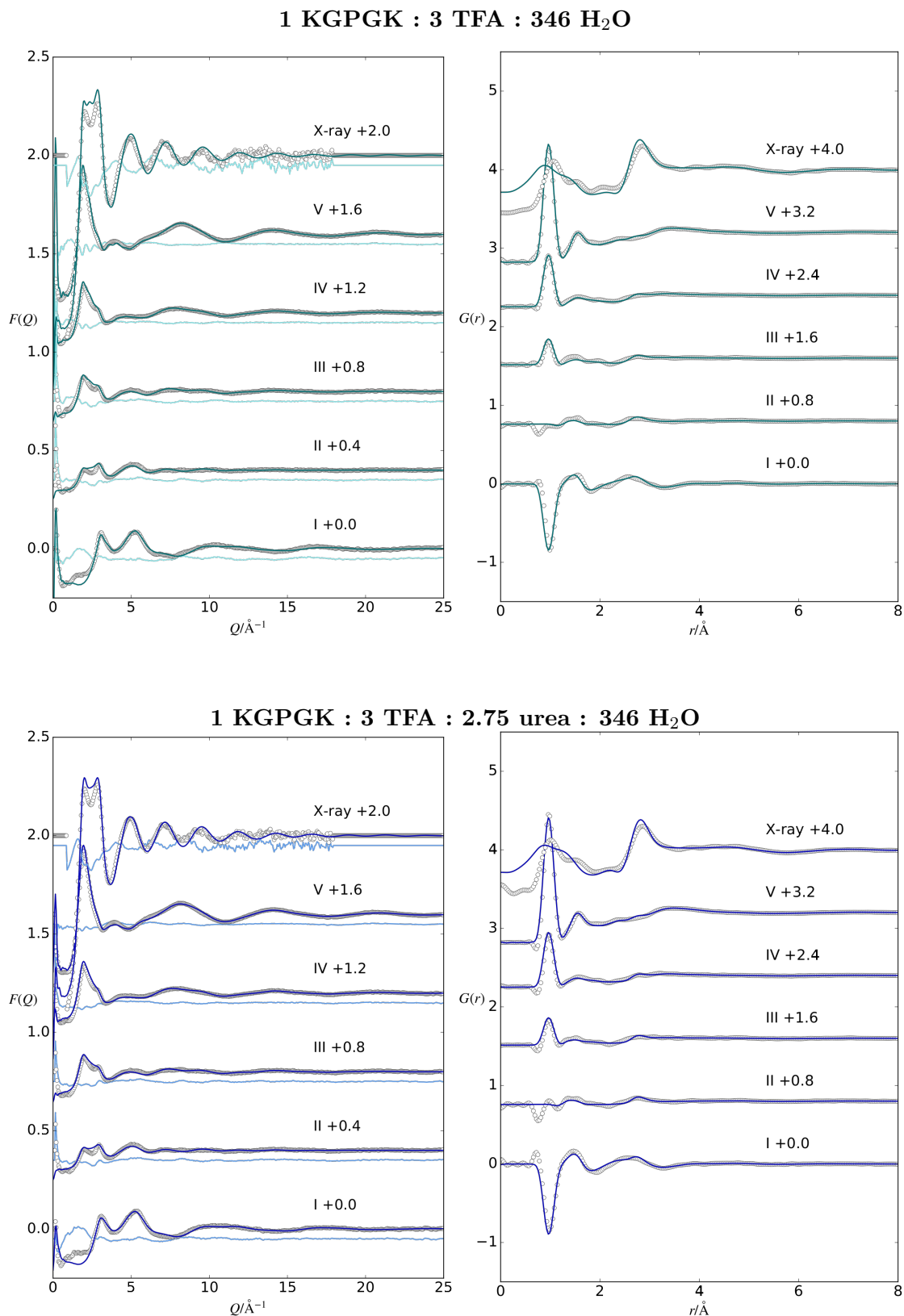


Figure 4.8: Measured diffraction data (grey dots) for KGPGK in aqueous solution with (bottom) and without urea (top) compared with the **EPSR** fits (dark cyan and blue lines) in reciprocal space ($F(Q)$, left) and in real space ($G(r)$, right). The difference between simulation and data are shown as a light cyan and blue lines. The data and corresponding fits have been shifted for clarity. The sample labels correspond to the level of deuteration in listed in Table

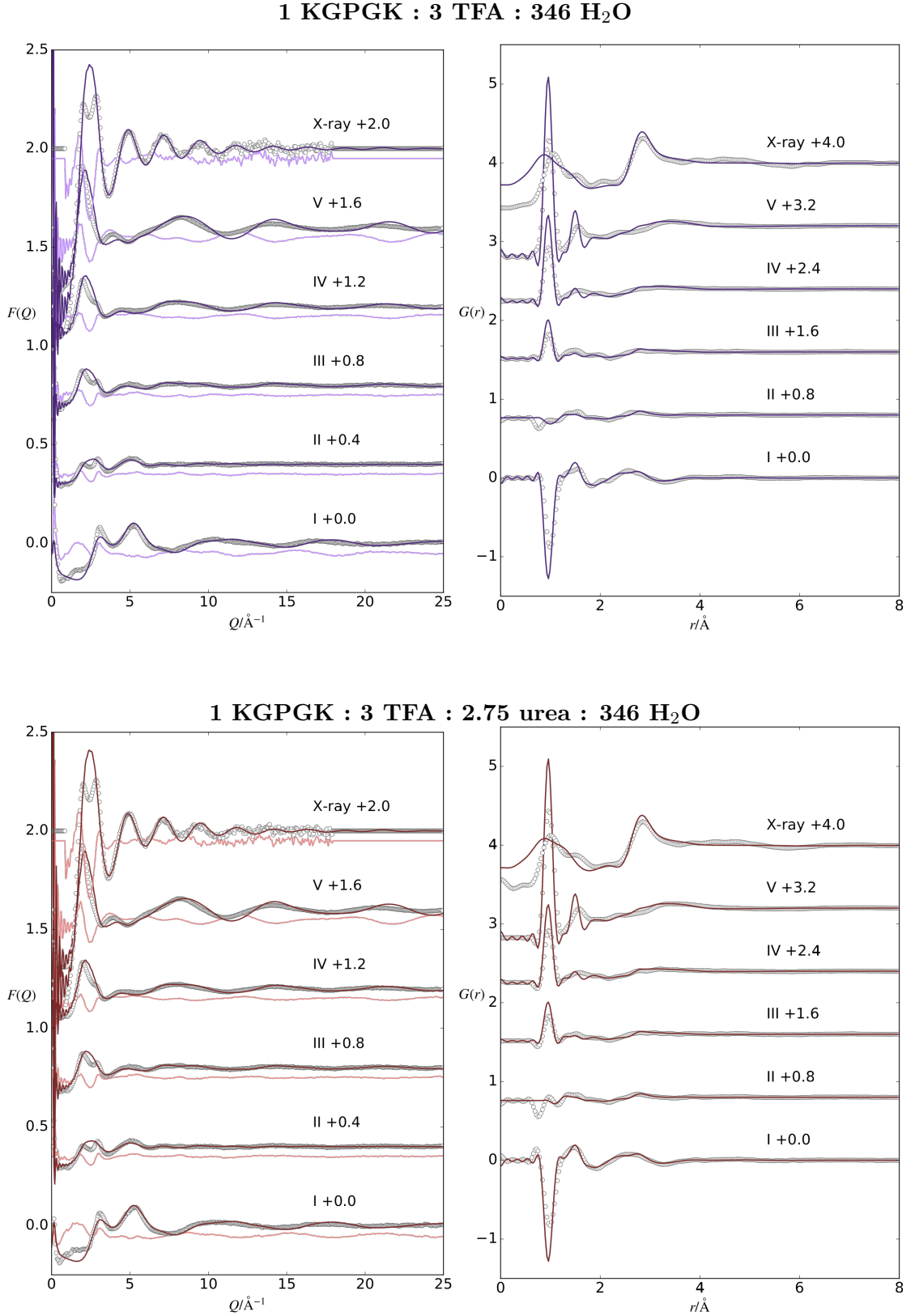


Figure 4.9: Measured diffraction data (grey dots) for KGPGK in aqueous solution with (bottom) and without urea (top) compared with the MD fits (dark purple and red lines) in reciprocal space ($F(Q)$, left) and in real space ($G(r)$, right). The difference between simulation and data are shown as a light purple and red lines. The data and corresponding fits have been shifted for clarity. The sample labels correspond to the level of deuteration in listed in Table

all peptide nitrogens with the water oxygens (Ow, left column), with and without urea. The positively charged N1 is, relatively, the most hydrated with a distinct peak at 2.7 Å (CN $\approx$ 4.0). The NH backbone nitrogens N2, N4 and N5 have a smaller first hydration peaks at 2.8 Å (CN $\approx$ 1.2). In EPSR, this hydration increases or decreases, depending on the residue, with the addition of urea without a clear pattern. In MD, the peptide backbone nitrogen hydration is almost unchanged and only slightly lowered for all residues upon the addition of urea (hydration difference, HD $\approx$ -3%) to the solution.

Unlike the peptide hydration, the peptide nitrogen – TFA oxygen (Ot) interactions (Figure 4.10, middle column) vary significantly between the simulation techniques, with and without urea and are residue specific. In the MD simulations, the N-terminus (N1) is preferentially coordinated by TFA compared with the polar -NH groups on the peptide backbone which show overall fewer interactions with this anion. There is no influence of urea on these -NH-TFA interactions visible in MD as the coordination numbers are unchanged by the presence of urea (Table 4.4). In EPSR, on the other hand, the TFA-oxygen to peptide nitrogen interactions are significantly lowered (- 30-80%) upon the addition of urea to the solution. Additionally, the first peak in the N1-Ot RDF at $\approx$ 3 Å is further away from the nitrogen compared with the same RDF in MD ($\approx$ 2.7 Å). In general, the interactions of TFA with the charged N1 N-terminus are lower for EPSR compared to MD while TFA interactions with the -NH backbone nitrogens are significantly higher in the EPSR simulations and, similar to N1, shifted ($\approx$ 3 Å) further away from the backbone nitrogens. In EPSR, the addition of urea has a clear, overall negative effect on the nitrogen-TFA interactions. Especially, the strong TFA interactions with the charged N-terminus (N1) and the backbone nitrogens N2 and N5 are significantly lowered in the urea-containing simulation which might explain the increase in hydration at these residues.

The $g(r)$ s between backbone nitrogen and urea oxygen (Ou) (Figure 4.10, right column) again differ between EPSR and MD simulations. For EPSR, the first shell of urea interaction *via* Ou varies widely between the backbone nitrogen sites with N1 showing the highest preference for interacting with the urea, N2, N3 and N6 showing an absence of hydrogen bonding to urea and N5 showing strong hydrogen bonding interactions with Ou, visible in a sharp peak in the N5-Ou RDF at 2.7 Å. However the

Table 4.4: Coordination numbers (n_{α}^{β}) for the water oxygens (Ow), TFA oxygens (Ot) and urea oxygens (Ou) around KGPGK nitrogens from EPSR and MD simulations. The corresponding RDFs are shown in Figure 4.10. Also shown are the hydration difference (HD) and the urea/water preference ratio (PR).

	r_{max} / Å	KGPGK		KGPGK+urea		HD /%		PR	
		EPSR	MD	EPSR	MD	EPSR	MD	EPSR	MD
N1-Ow	3.6	3.9	4.1	4.1	4.0	+6	-2	-	-
N2-Ow	3.4	1.1	1.4	1.3	1.3	+16	-3	-	-
N3-Ow	6.8	27.2	30.6	28.7	29.5	+6	-4	-	-
N4-Ow	3.4	1.1	1.2	1.1	1.2	-2	-4	-	-
N5-Ow	3.4	1.2	1.2	1.2	1.2	+1	-2	-	-
N6-Ow	4.3	7.6	6.8	7.0	6.6	-8	-2	-	-
N1-Ot	3.6	0.5	0.5	0.2	0.5	-	-	-	-
N2-Ot	3.4	0.2	0.1	0.1	0.1	-	-	-	-
N3-Ot	6.8	2.3	0.7	1.1	0.6	-	-	-	-
N4-Ot	3.4	0.1	<0.1	0.1	<0.1	-	-	-	-
N5-Ot	3.4	0.2	<0.1	<0.1	<0.1	-	-	-	-
N6-Ot	4.3	0.3	0.1	0.2	0.1	-	-	-	-
N1-Ou	3.6	-	-	0.1	0.1	-	-	3.0	1.6
N2-Ou	3.4	-	-	<0.1	<0.1	-	-	1.0	1.9
N3-Ou	6.8	-	-	0.4	0.4	-	-	1.6	1.7
N4-Ou	3.4	-	-	<0.1	<0.1	-	-	2.3	2.1
N5-Ou	3.4	-	-	0.1	<0.1	-	-	6.4	2.1
N6-Ou	4.3	-	-	<0.1	0.1	-	-	0.7	1.3

associated coordination number is still very small ($CN \approx 0.1$) for this RDF. In the MD simulation, urea interacts much less with the peptide nitrogens in general compared with EPSR, showing a preference to bind to the charged N-terminus over any of the backbone -NH groups. Although the preference ratios (see Section 3.2.6) for all of the peptide nitrogens show that urea interactions are preferred over water, the urea-backbone nitrogen coordination numbers are small (Table 4.4), due to the low urea concentration compared to TFA or water coordination to the peptide.

The SDMs of solvent and solute molecules in Figure 4.11 (for EPSR) and Figure 4.12 (for MD) depict the three-dimensional arrangement of nearest neighbour water (light blue), TFA (magenta) and urea (green) molecules around the peptide backbone nitrogens. These are different to the SDMs in Chapter 3, which were plotted over a distance range. The advantage of the nearest neighbour SDMs is that they represent the same number of neighbour molecules regardless of neighbour molecule concentration and localisation.

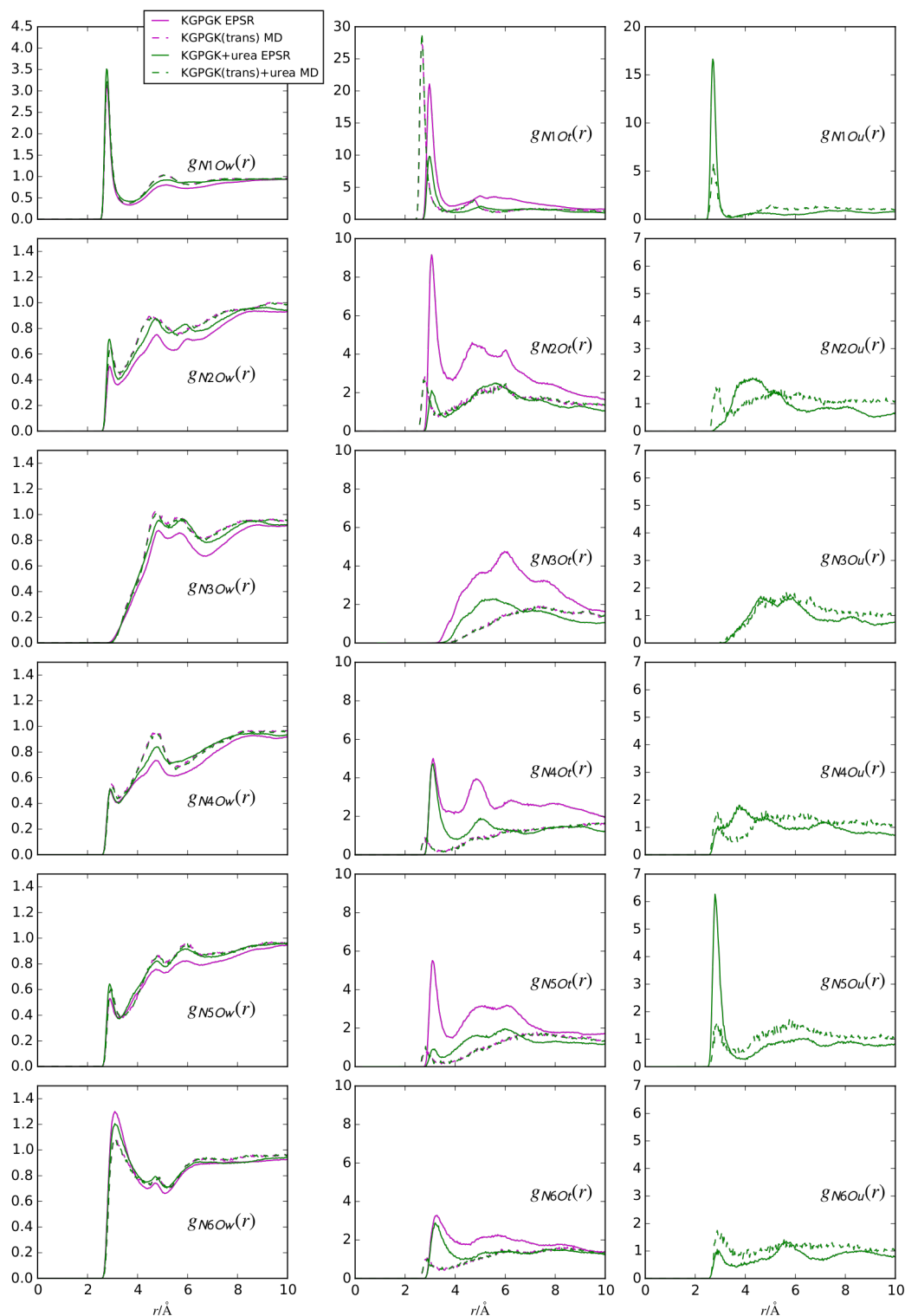


Figure 4.10: Radial distribution functions between the KGPGK backbone nitrogens and the solvent and solute molecules (column 1: water oxygen Ow, column 2: TFA oxygen Ot, column 3: urea oxygen Ou). Urea-containing simulations are shown in green, non-urea containing simulations are shown in magenta. The EPSR $g(r)$ s are shown as solid lines, the MD $g(r)$ s as dashed lines. The corresponding coordination numbers for these functions are in Table 4.4.

In general, the position of the nearest neighbour water molecules is similar in both simulation techniques. Only the water density around N2 and N5 is visibly lower in the urea-free EPSR simulation compared to the EPSR simulation containing urea and both MD simulations with and without urea.

This can possibly be explained by the close TFA association with N2 and N5 in EPSR which is visible as highly dense and localised TFA clouds around these residues, where this strong interaction might prevent an increased hydration. Thus, as these highly localised and directed TFA-nitrogen interactions are disrupted when urea is added to the EPSR simulation, denser hydration clouds are observed at these residues and the TFA cloud becomes scattered around the nitrogens similar to the accordant SDM in MD (Figure 4.12). This is consistent with the comparably large interaction with urea molecules at N5 as shown in the RDFs in Figure 4.10. However, as there is no such direct interaction of urea visible for N2, urea likely only partially replaces TFA but also acts to increase the hydration around the N2 and N5 site. The SDMs for the other nitrogens in EPSR, show that both N1 and N2 have a decrease in interactions with TFA upon urea addition. In MD, the TFA density clouds are more disperse compared to EPSR with the exception of a localised interaction with charged N-terminus and a, possibly related, interaction with the adjacent Gly2 nitrogen (N2). Similar to the RDFs in Figure 4.10, the lower TFA density in MD is unchanged by the addition of urea. The urea density clouds are localised only in the EPSR simulation, where this is especially clear around N1, N2 and N5. In MD, the urea density is very low and unlocalised around the nitrogen sites.

Overall, these results show a comparable hydration of the peptide backbone nitrogens in MD and EPSR both in the presence and absence of urea. On the other hand, the MD simulations only show limited interactions between the solutes TFA and urea with the peptide backbone, whereas by contrast in the EPSR, urea disrupts the strong TFA-peptide interaction at certain residues and increases peptide hydration.

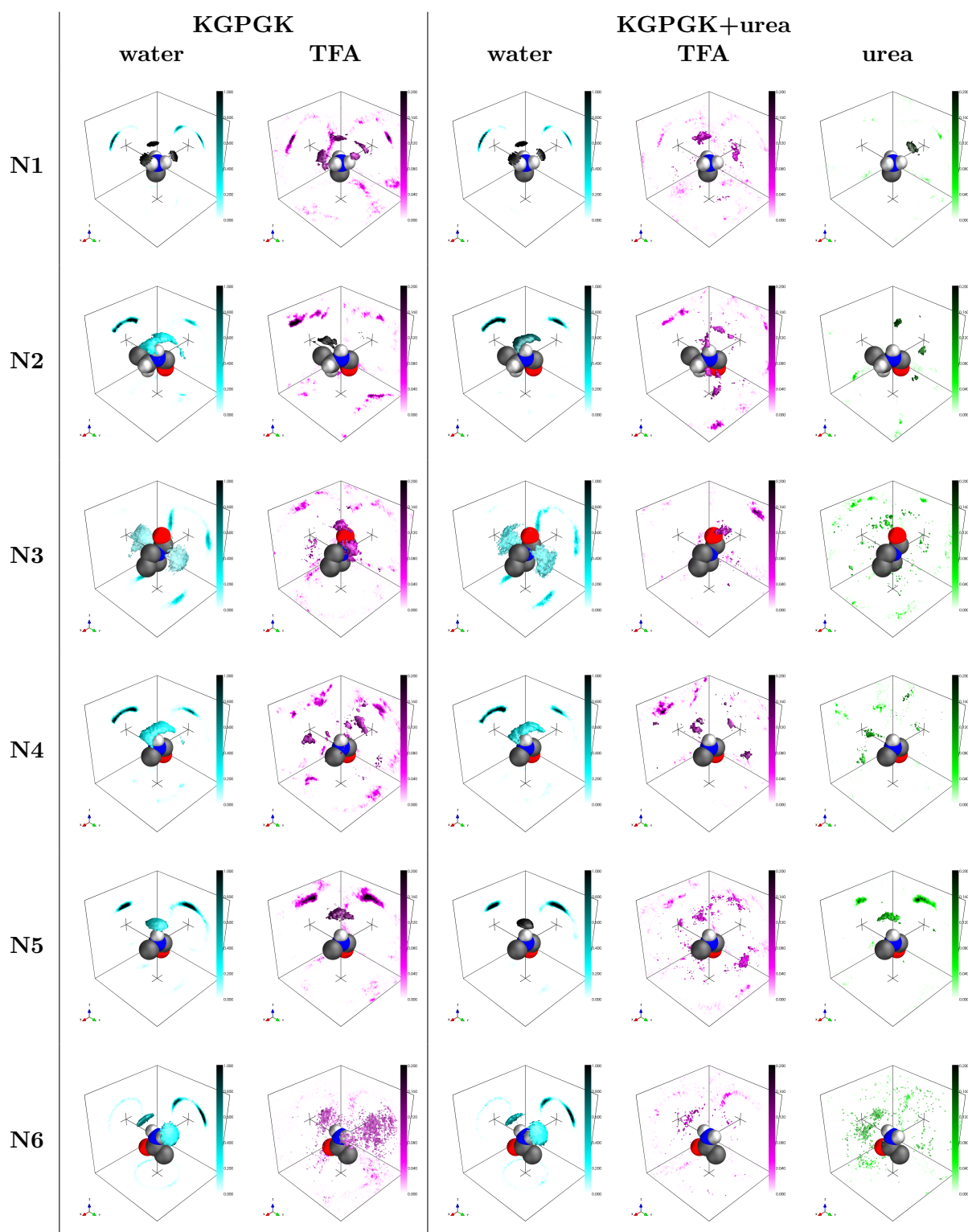


Figure 4.11: Spatial density maps (SDMs) show the location of the first neighbour water (light blue), TFA (magenta) and urea (green) molecules around the peptide backbone nitrogens for the **EPSR** simulations. The isocontour surfaces enclose the most dense 50% for water and 10% for TFA and urea. The scale bar show the local density of neighbours around the central nitrogen in units of $\text{atoms}/\text{\AA}^3$.

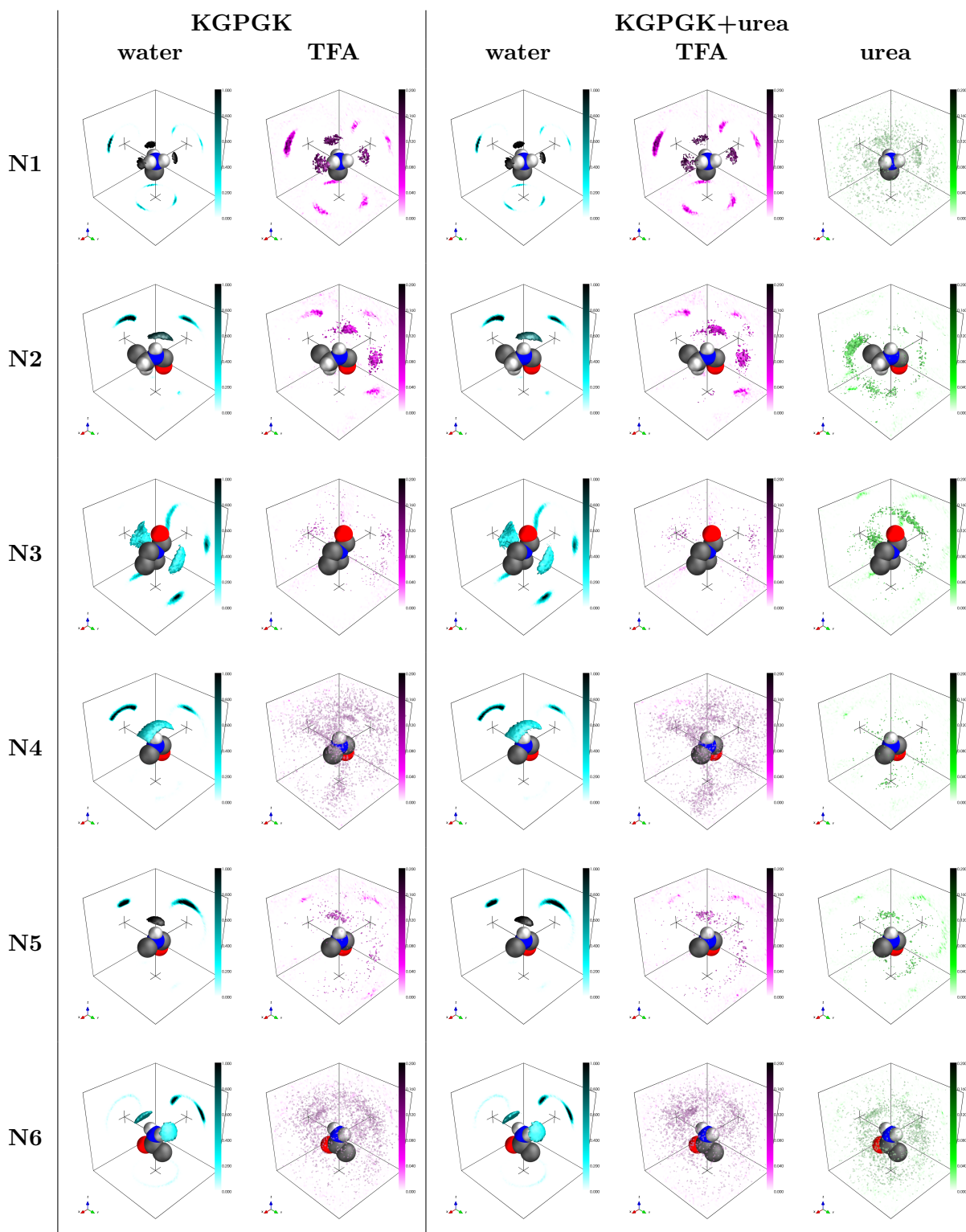


Figure 4.12: Spatial density maps (SDMs) show the location of the first neighbour water (light blue), TFA (magenta) and urea (green) molecules around the peptide backbone nitrogens for the **MD** simulations. The isocontour surfaces enclose the most dense 50% for water and 10% for TFA and urea. The scale bar show the local density of neighbours around the central nitrogen in units of atoms/ $\text{\AA}^3$.

4.3.3.2 Peptide Backbone Oxygens Interactions

Figure 4.13 shows the RDFs of the KGPGK backbone oxygens with water hydrogens (Hw, left column) and urea hydrogens (Hu, right column). The corresponding coordination numbers are in Table 4.5. For the negatively charged TFA molecules, no clear interaction with the peptide backbone oxygens is visible.

The hydration of the backbone oxygens is, in general, similar between EPSR and MD and only slightly effected by the addition urea. All backbone oxygens have a first hydration peak at ≈ 1.8 Å and a second hydration peak at ≈ 3.2 Å. The level of hydration increases towards the C-terminus, where the O1-Hw interaction has a coordination number of ≈ 1.7 and the CN for O5-Hw is ≈ 2.4 for EPSR. The oxygens in MD show a similar trend, however, the peptide hydration is generally lower in MD than in EPSR. Although there is no significant increase or decrease in oxygen hydration visible upon urea addition and the urea coordination numbers are very small, most peptide oxygens markedly prefer urea over water within the first hydration shell distance up to 2.5 Å. While there is no clear residue specificity for urea in MD, the oxygens in the middle of the peptide chain (O2-O4) show a diminished first urea hydrogen bond peak in EPSR. This might be related to steric hinderance due to differences in peptide conformation (Section 4.3.4) or localised TFA molecules around the adjacent N4 and N5 nitrogens (Figure 4.10).

Figure 4.14 and 4.15 depict the SDMs of nearest neighbour water (light blue) and urea (green) molecules around the peptide backbone oxygens for EPSR and MD respectively. Although most water molecules in both simulations are located similarly in two spots in front ($-x$ direction) or behind ($+x$ direction, Figure 4.3) the peptide chain there are a few exceptions. For O5, which is not within but at the end of the peptide chain, the nearest neighbour water molecules at this site are less localised and arranged in a halo around the C-terminal carbonyl oxygen in both EPSR and MD. For O1 and O2 in MD, the nearest neighbour water molecules are more localised towards the carbonyl carbon than the oxygen. This might be due to the close positive charge of N1 or due to a specific peptide conformation (Section 4.3.4).

The localisation of urea density around the backbone oxygens shows no clear hydro-

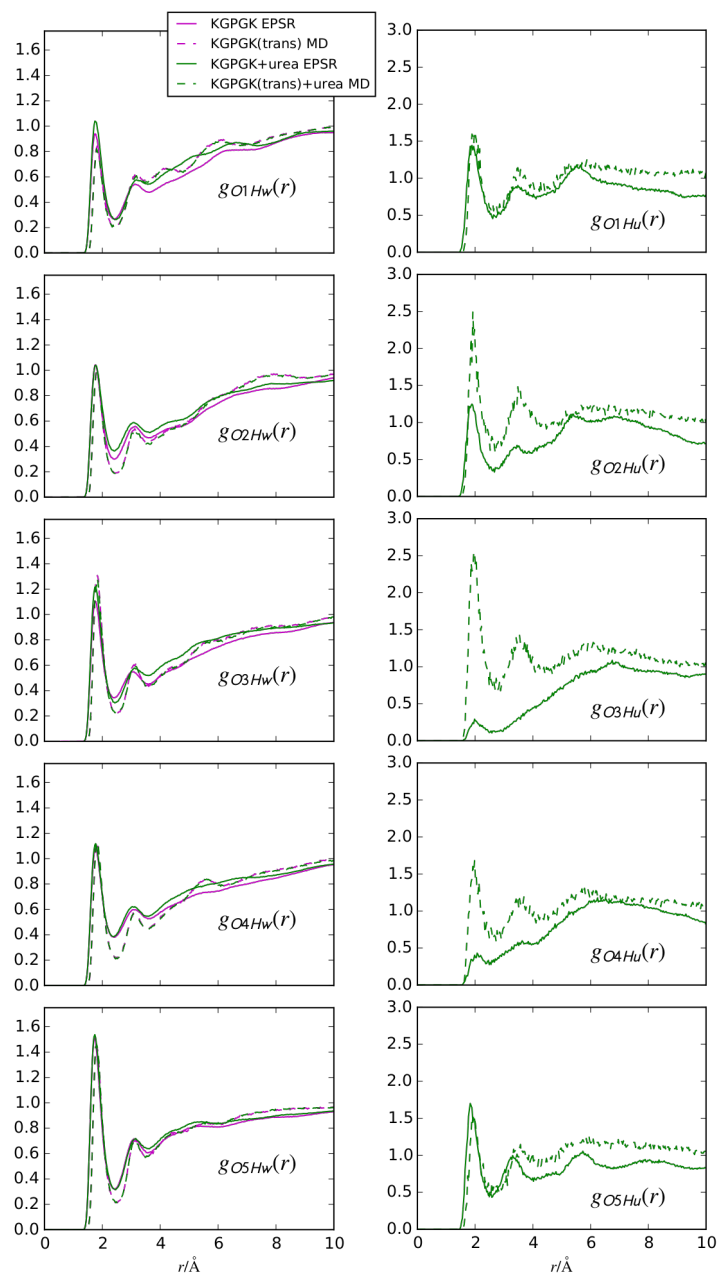


Figure 4.13: Radial distribution functions between the KGPBK backbone oxygens and the solvent and solute molecules (column 1: water hydrogens Hw, column 2: urea hydrogens Hu). Urea-containing simulations are shown in green, non-urea containing simulations are shown in magenta. The EPSR $g(r)$ s are shown as solid lines, the MD $g(r)$ s as dashed lines. The corresponding coordination numbers for these functions are in Table 4.5.

Table 4.5: Coordination numbers (n_{α}^{β}) for the water hydrogens (Hw) and urea hydrogens (Hu) around KGPGK oxygens from EPSR and MD simulations. The corresponding RDFs are shown in Figure 4.10 Also shown are the hydration difference (HD) and the urea/water preference ratio (PR).

	$r_{max} / \text{\AA}$	KGPGK		KGPGK+urea		HD /%		PR	
		EPSR	MD	EPSR	MD	EPSR	MD	EPSR	MD
O1-Hw	2.5	1.7	1.3	1.7	1.2	+1	-4	-	-
O2-Hw	2.5	1.8	1.4	1.9	1.4	+2	-2	-	-
O3-Hw	2.5	2.0	1.8	2.0	1.7	-3	-4	-	-
O4-Hw	2.5	2.1	1.6	2.0	1.6	-5	-1	-	-
O5-Hw	2.5	2.4	1.9	2.3	1.9	-4	-1	-	-
O1-Hu	2.5	-	-	0.1	0.1	-	-	3.8	5.2
O2-Hu	2.5	-	-	<0.1	0.1	-	-	2.7	5.5
O3-Hu	2.5	-	-	<0.1	0.1	-	-	0.7	5.1
O4-Hu	2.5	-	-	<0.1	0.1	-	-	1.3	4.0
O5-Hu	2.5	-	-	0.1	<0.1	-	-	2.8	2.7

gen bonding positions, similar to the SDMs in Figures 4.11 - 4.12. However, in EPSR the urea molecules are slightly more localised in areas of higher urea density compared to MD where the urea molecules are more disperse around all of the backbone carbonyl oxygens.

4.3.4 Whole Molecule Analysis

To determine the most likely peptide conformations and the solvent and solute densities around them, the dihedral angles along the peptide backbone (labels in Figure 4.16) were calculated using ANGULA (angle distributions in Figure 4.17). Similar to GPG-NH₂ (Section 3.3.7), the peptide conformation is more flexible in EPSR than in MD and has peptide bond dihedral angles (ω) often slightly shifted from 180°. The MD peptide conformations are also unaffected by the addition of urea.

The conformation independent solvent interaction RDFs for the whole peptide molecule are shown in Figure 4.18. Again there is no difference in MD for the hydration or TFA interaction with and without urea. The EPSR simulations, on the other hand, show a slight decrease in hydration and TFA-peptide interactions in the urea-containing simulation. Overall, the TFA-peptide interactions are slightly closer to the peptide in MD than in EPSR, whereas the urea-peptide interactions are marginally closer in EPSR

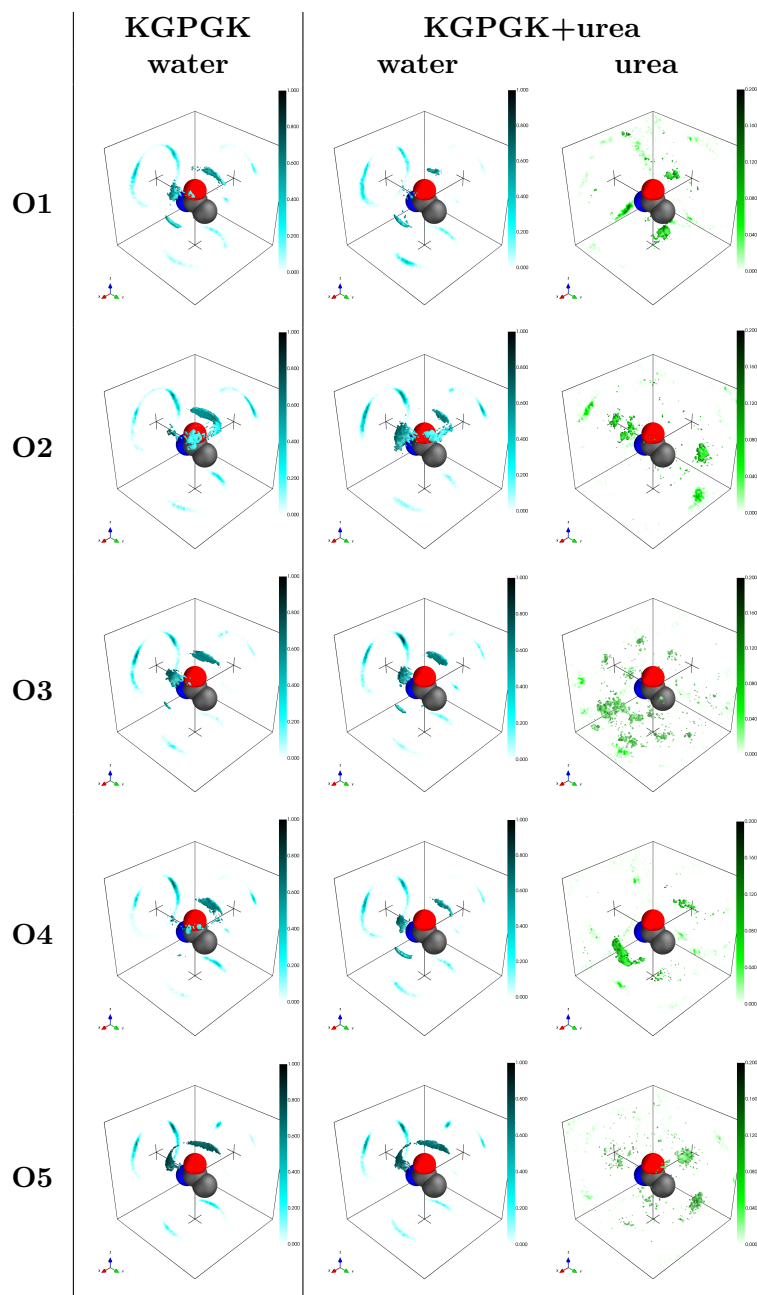


Figure 4.14: Spatial density maps (SDMs) show the location of the first neighbour water (light blue) and urea (green) molecules around the peptide backbone oxygens for the **EPSR** simulations. The isocontour surfaces enclose the most dense 20% for water and 10% for urea. The scale bar show the local density of neighbours around the central nitrogen in units of $\text{atoms}/\text{\AA}^3$.

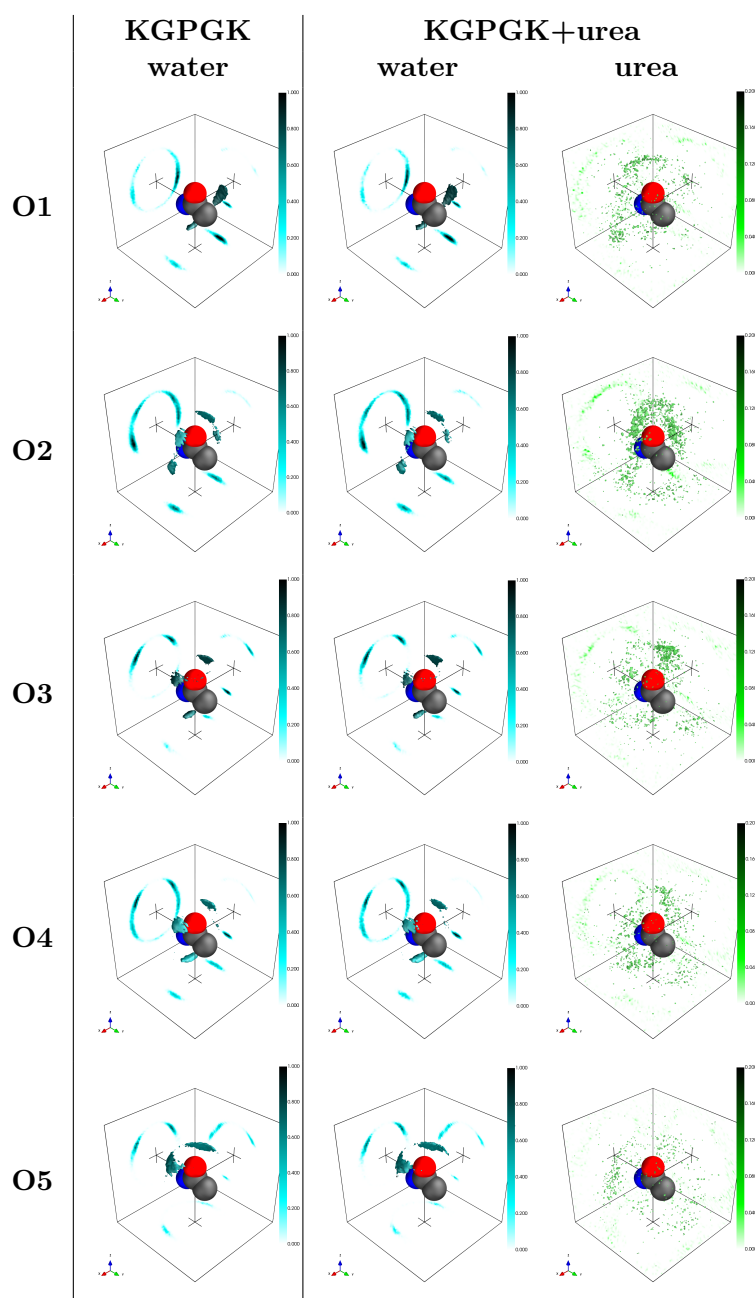


Figure 4.15: Spatial density maps (SDMs) show the location of the first neighbour water (light blue) and urea (green) molecules around the peptide backbone oxygens for the MD simulations. The isocontour surfaces enclose the most dense 20% for water and 10% for urea. The scale bar show the local density of neighbours around the central nitrogen in units of atoms/ $\text{\AA}^3$.

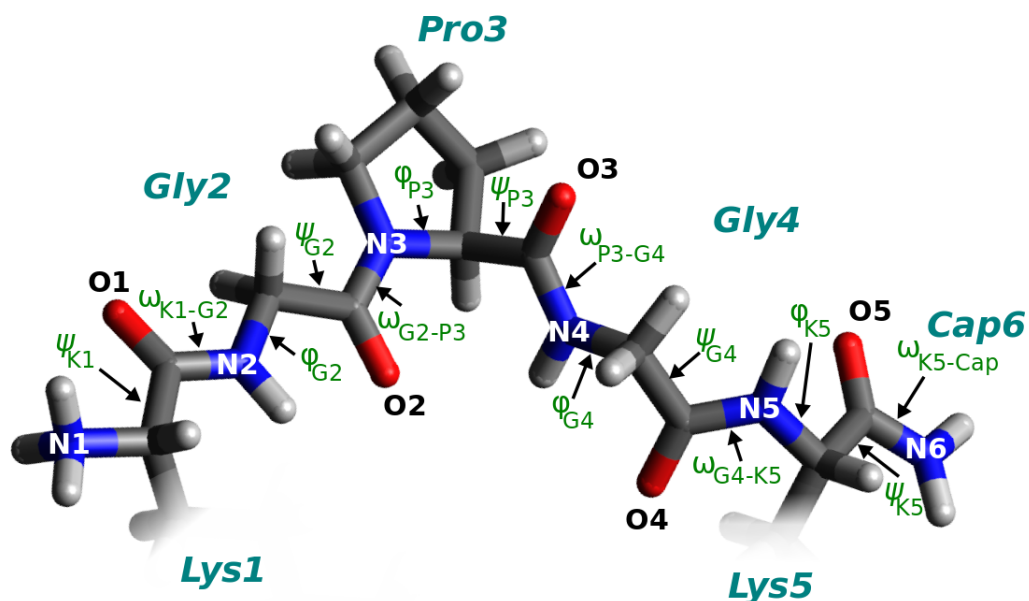


Figure 4.16: Dihedral angles along the KGP GK backbone. The labelled bond refers to the respective rotational axes.

than in MD.

To compare the solvent interaction across the whole molecule, the most probable backbone conformations for each simulation was chosen. Correlations between dihedral angles were analysed using bivariate analysis. The selected peptides exhibit dihedral angles in Table 4.6. These peptide conformations and the solvent clouds are shown in Figure 4.19 for EPSR and Figure 4.20 for MD (top row).

The hydration in EPSR is, unaffected by urea, distributed broadly around the backbone and charged lysine residues whereas in MD water molecules are located mostly around the backbone and the proline ring, excluding the charged N-terminus and lysine residues. The broad water distribution around the proline ring is even more visible in the urea-containing MD simulation. Overall, that the hydration is unchanged by the addition of urea is in accordance with the presented results in Section 4.3.3.1-4.3.3.2

In MD, the negatively charged TFA molecules are preferentially located around the positively charged N-terminus. This interaction is almost unchanged, though slightly broadened, by the addition of urea. Another likely TFA location in MD, in the middle between the lysine residues and the backbone chain is shifted towards the C-terminus upon urea addition and a small cloud of TFA density next to the proline ring disappears,

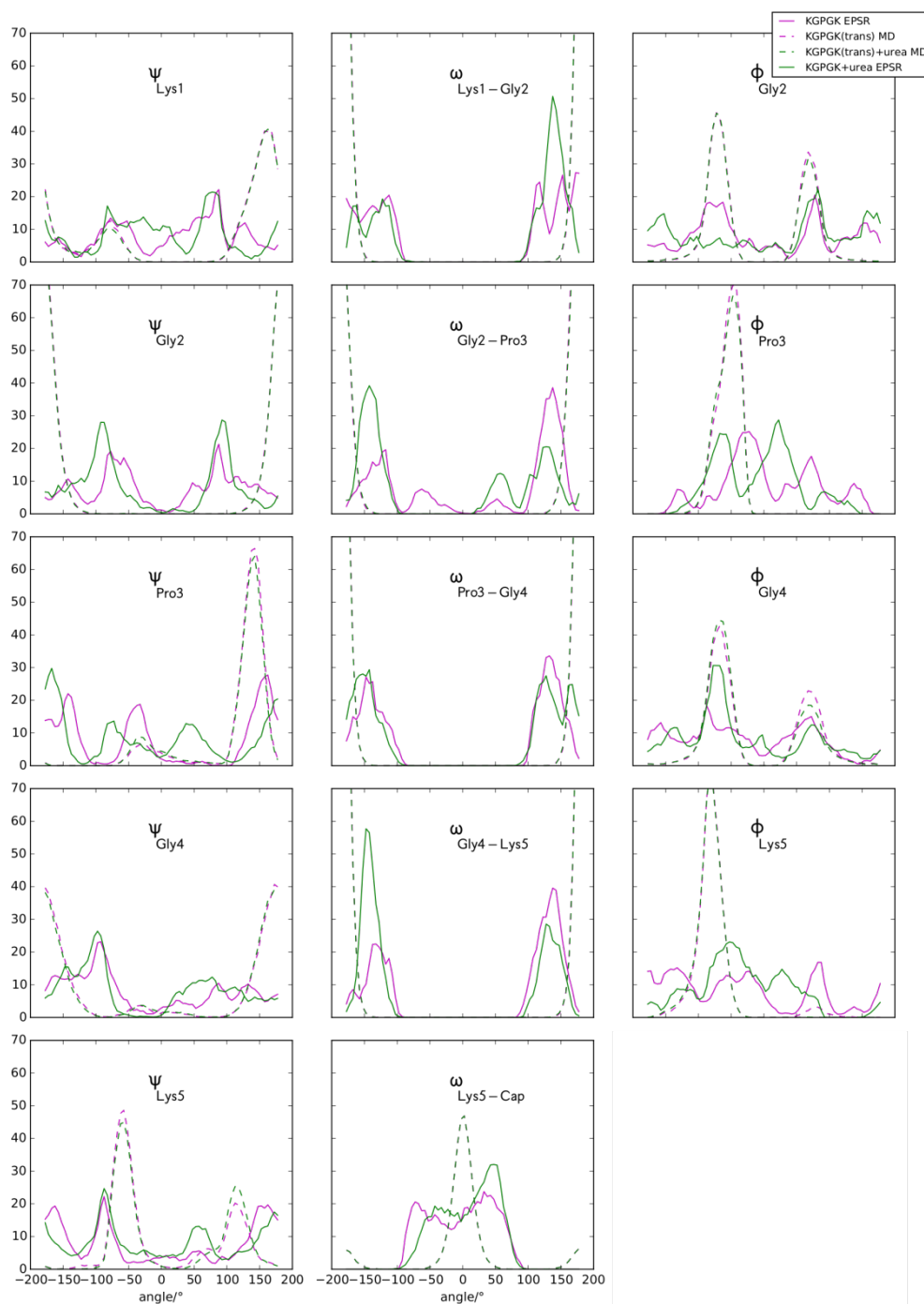


Figure 4.17: Dihedral angle distributions along the KGPGK backbone (labels in Figure 4.16).

Table 4.6: Dihedral angles selected ($\pm 20^\circ$) for the whole molecule analysis depictions in Figures 4.19 and 4.20. The related dihedral distributions are shown in Figure 4.17 and the labels in Figure 4.16.

Label	Dihedral between	KGPGK		KGPGK+urea	
		EPSR	MD	EPSR	MD
ψ_{Lys1}	N1-C α_{K1} -C(=O)-N2	75	150	80	150
$\omega_{\text{Lys1-Gly2}}$	C α_{K1} -C(=O)-N2-C α_{G2}	180	180	150	180
ϕ_{Gly2}	C(=O)-N2-C α_{G2} -C(=O)	-75	-75	80	-75
ψ_{Gly2}	N2-C α_{G2} -C(=O)-N3	90	180	100	180
$\omega_{\text{Gly2-Pro3}}$	C α_{G2} -C(=O)-N3-C α_{P3}	-125	180	-140	180
ϕ_{Pro3}	C(=O)-N3-C α_{P3} -C(=O)	-50	-50	-50	-50
ψ_{Pro3}	N3-C α_{P3} -C(=O)-N4	140	140	140	140
$\omega_{\text{Pro3-Gly4}}$	C α_{P3} -C(=O)-N4-C α_{G4}	-140	180	180	180
ϕ_{Gly4}	C(=O)-N4-C α_{G4} -C(=O)	-75	-75	-70	-75
ψ_{Gly4}	N4-C α_{G4} -C(=O)-N5	-75	180	-140	180
$\omega_{\text{Gly4-Lys5}}$	C α_{G4} -C(=O)-N5-C α_{K5}	125	180	125	180
ϕ_{Lys5}	C(=O)-N5-C α_{K5} -C(=O)	-170	-75	-25	-75
ψ_{Lys5}	N5-C α_{K5} -C(=O)-N6	180	-50	180	-50
$\omega_{\text{Lys5-Cap}}$	C α_{K5} -C(=O)-N6-H6a	0	0	0	0

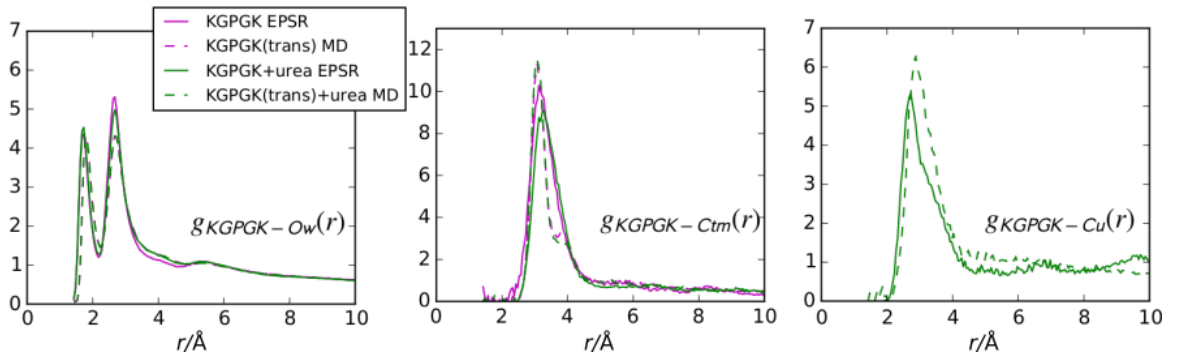


Figure 4.18: Whole molecule analysis RDF of water (Ow), TFA middle (Ctm) and urea (Cu) around the whole KGPGK peptide.

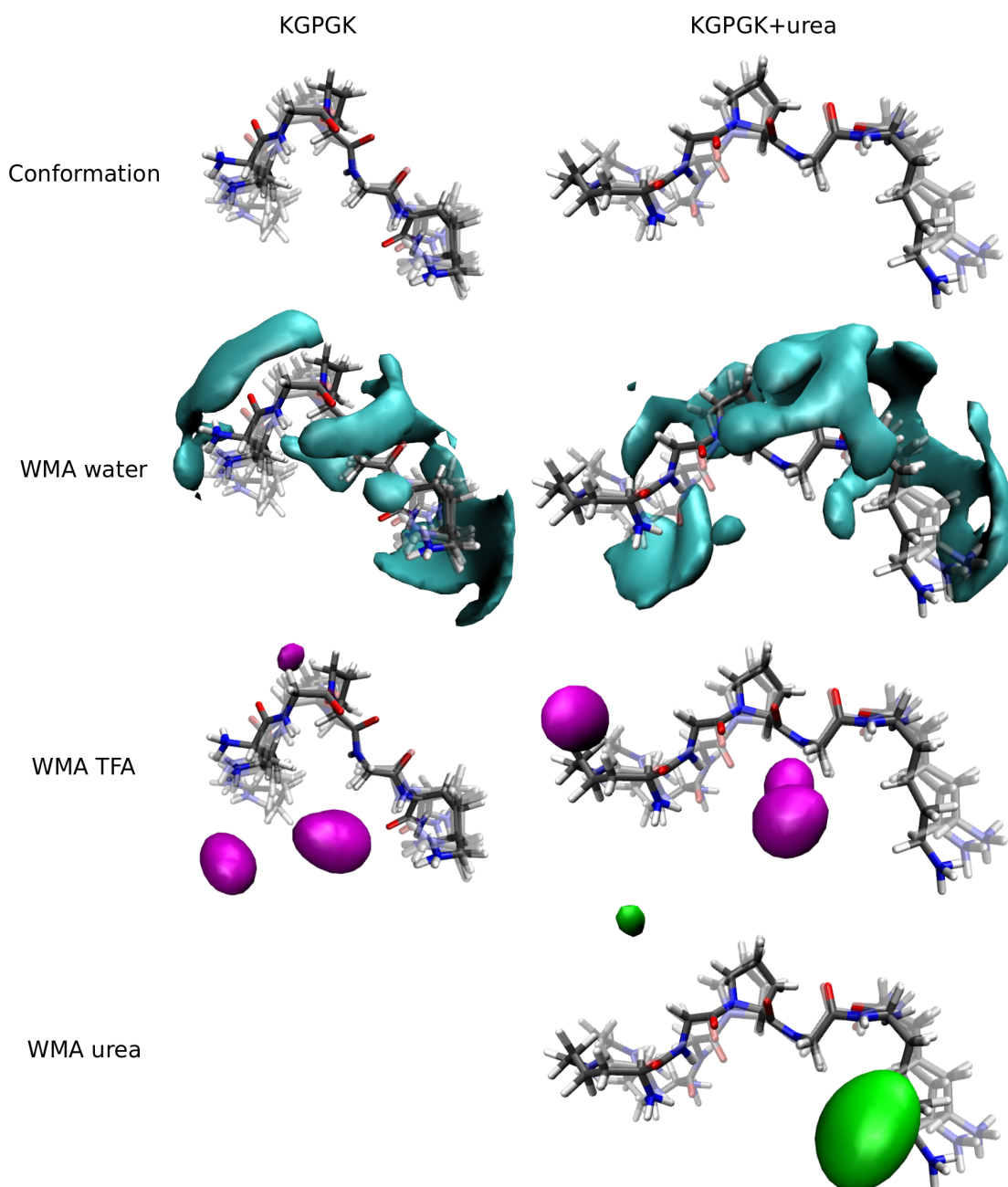


Figure 4.19: Most probable conformations and whole molecule analysis of solvent clouds around the most probable conformation of KGPGK in **EPSR**. Water oxygens (Ow, blue, 30%, up to 4.0 Å), TFA central pseudoatom (Ctm, magenta, 50% up to 5.0 Å) and urea carbons (Cu, green, 50% up to 5.0 Å)

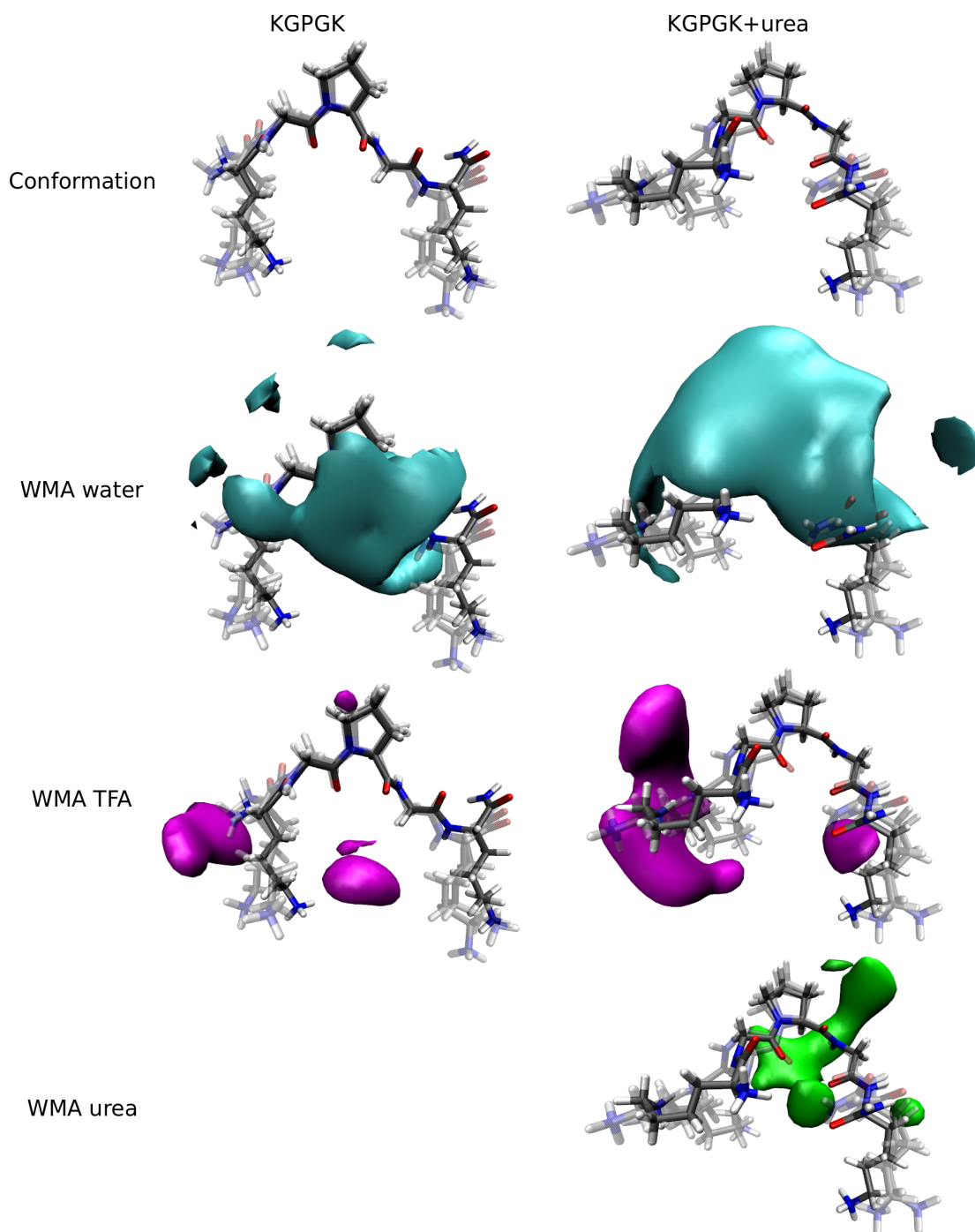


Figure 4.20: Most probable conformations and whole molecule analysis of solvent clouds around the most probable conformation of *trans*-KGPGK in MD. Water oxygens (Ow, blue, 30%, up to 4.0 Å), TFA central pseudoatom (Ctm, magenta, 50% up to 5.0 Å) and urea carbons (Cu, green, 50% up to 5.0 Å)

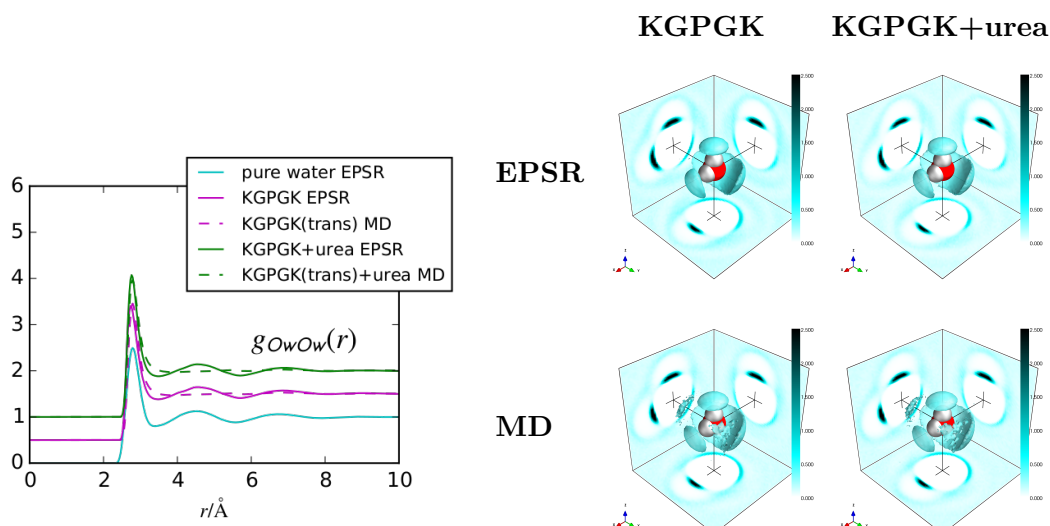


Figure 4.21: Left: Radial distribution function ($g(r)$ s) between the water oxygens (Ow) with urea (green) and without urea (magenta) compared with those from pure water (cyan) [121]. The EPSR $g(r)$ s are shown as solid lines and the MD $g(r)$ s as dashed lines. The corresponding coordination numbers are in Table 4.7. Right: Spatial density maps (SDMs) show the location of the water (cyan) molecules up to a 10 Å distance around a central representative water molecule for the EPSR (top row) and MD (bottom row) simulations. The isocontour surfaces enclose the most dense 5% for water. The scale bars show the local density of neighbours around the central oxygen in units of atoms/Å³.

at the depicted density level. This proline-TFA association is also visible in the urea-free EPSR simulation. Other TFA localisations in EPSR are closer to the positively charged lysine residues and the backbone. These associations are shifted away from the charged sites towards the middle of the backbone when urea is added.

In EPSR urea is mostly located close to the Lys5 residue and the proline ring, approximately in a similar location as TFA in the urea-free simulation. On the other hand, urea is located around the middle of the peptide backbone and the proline ring in the MD simulation, in similar places as water. In general, this effect of urea disrupting TFA associations in EPSR but not MD is in accordance with the TFA and urea association for N2 and N5 in Section 4.3.3.1.

Table 4.7: Coordination numbers for the RDFs in Figure 4.21, 4.22 and 4.23

	$r_{max} / \text{\AA}$	pure water [121] EPSR	KGPGK		KGPGK+urea	
			EPSR	MD	EPSR	MD
Ow-Ow	3.5	5.2	5.3	5.3	4.9	5.2
Ot-Ow	4.0		4.8	5.3	4.8	5.3
Ft-Ft	3.9		0.5	0.1	0.2	0.1

4.3.5 Solvent and Co-Solute Interactions

4.3.5.1 Water-Water Interactions

To determine whether urea changes not only the peptide interactions with both water and TFA but also changes the bulk water-water structure, as is proposed in the theory of indirect urea-induced protein denaturation [188], Figure 4.21 (left) shows the water (Ow-Ow) RDF of both simulation techniques compared to the $g_{OwOw}(r)$ from EPSR simulations of pure water [121]. From this figure, it is clear that neither the nearest neighbour water-water nor the second nearest neighbour water-water structure is significantly changed from that of pure water, with the exception of the second nearest neighbour peak in MD, as expected from the TIP3P water model from MD simulations [227]. This result is different to Chapter 3 where urea had an effect on the second neighbour Ow-Ow distance. From the coordination numbers in Table 4.7, EPSR shows an increase in water-water interaction with KGPGK addition and a small decrease with urea addition, while MD shows a slight increase in water-water interaction upon the addition of urea. The water-water SDMs in Figure 4.21 (right) show no difference to the tetrahedral water-water structure upon urea addition to the solution in either simulation, however, again as expected the MD simulations show a slightly more diffuse second nearest neighbour water coordination due to the TIP3P water model.

4.3.5.2 TFA-Water Interactions

One of the largest differences between the simulation methods are the resulting TFA-peptide interactions, which is, perhaps, due to the change in TFA-TFA and TFA-water association upon the addition of urea between the simulations. Figure 4.22 (left) shows

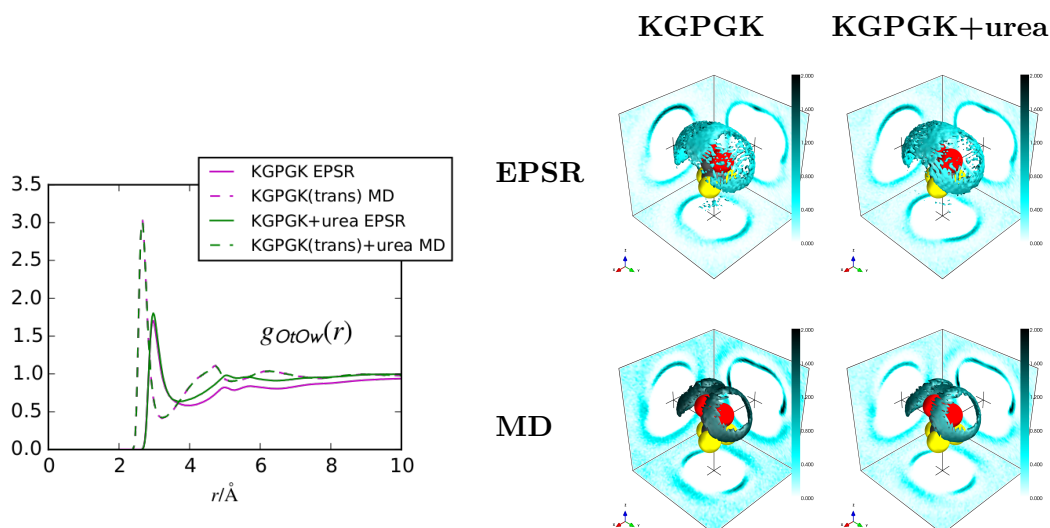


Figure 4.22: Left: Radial distribution function ($g(r)$ s) between the TFA oxygen (Ot) and water oxygens (Ow) with (green) and without urea (magenta) for EPSR (solid lines) and MD (dashed lines) simulations. The corresponding coordination numbers are in Table 4.7. Right: Spatial density maps (SDMs) show the location of the water (cyan) molecules up to a 10 Å distance around the middle of a central representative TFA molecule for the EPSR (top row) and MD (bottom row) simulations. The isocontour surfaces enclose the most dense 5% for water. The scale bar shows the local density of neighbours around the central oxygen in units of atoms/Å³.

the $g(r)$ between TFA oxygens (Ot) and water oxygens (Ow) and the corresponding spatial density maps in Figure 4.22 (right) show the location of neighbour water molecules up to 4 Å around the whole TFA molecule. In the MD simulations, the negatively charged TFA molecules interact more with water compared with the EPSR fits to the neutron diffraction data, where the most localised hydration is visible around the negatively charged TFA oxygens. The hydration around TFA is denser and more localised for MD compared to EPSR, both in terms of the three-dimensional arrangement and the radial distance from the TFA oxygens (Figure 4.22, left). It should be noted that the first MD $g(r)$ peak is at ≈ 2.7 Å, closer to the TFA oxygen than in EPSR (≈ 3.0 Å). Similar to peptide N-terminus N1 hydration, both simulations show no significant difference in TFA hydration for the KGPGK mixture with urea (see Table 4.7). However, there is also a lack of increase in hydration upon the addition of urea which might be expected if urea were to replace interactions of TFA either with itself (see below) or with the peptide backbone. This is likely a result of the low urea concentration and urea forming direct hydrogen bonds to TFA, explaining the lack of additional hydration around TFA in solution.

4.3.5.3 TFA-TFA Interactions

A possible explanation for the lack of TFA hydration in EPSR compared with MD is the higher degree of TFA-TFA association in EPSR. In order to probe this, Figure 4.23 shows the $g(r)$ s for the TFA fluorine atoms with other TFA molecules in solution with the coordination numbers for these functions shown in Table 4.7. Not only does EPSR show a greater number of TFA-TFA interactions, these interactions are markedly lowered upon the addition of urea, unlike in MD where the TFA-TFA $g(r)$ is unchanged by urea addition. From the SDMs in Figure 4.23, TFA-TFA association in EPSR mostly occurs *via* the fluorine and carbon atoms as visible in the high TFA density next to the TFA carbon atoms. It should be noted that TFA-TFA interactions are heavily weighted in the X-ray data and EPSR is in better agreement with this data set compared with the MD $F(Q)$ data. Therefore the TFA-TFA association in EPSR is probably more realistic than the decreased, and urea-unaffected, TFA-TFA interaction in the MD simulations.

Figure 4.23 also shows the results of a clustering analysis [248] between TFA molecules in solution for both MD and EPSR simulations. A cluster was defined as two TFA molecules within a 0-4 Å intermolecular Ft-Ft distance. Most TFA molecules are not associated with each other but there is a small proportion of larger TFA-TFA clusters mostly present in the EPSR simulations in the absence of urea. The addition of urea significantly decreases the relative cluster size and the proportion of TFA molecules in clusters.

Overall these results show a significant difference between EPSR and MD regarding the TFA interactions. MD underestimates TFA-TFA association (Figure 4.23) and TFA-peptide interaction (Figure 4.10) and overestimates TFA-hydration (Figure 4.22).

4.3.6 Conformation Independent Peptide Hydration

Figure 4.24 shows the hydration of the KGXGK peptides backbone oxygens and nitrogens over all peptide conformations within each simulation. The respective coordination numbers are listed in Table 4.8. Overall, the hydration of backbone sites is very similar between the peptides in MD simulations. The hydration of KGPGK oxygens in

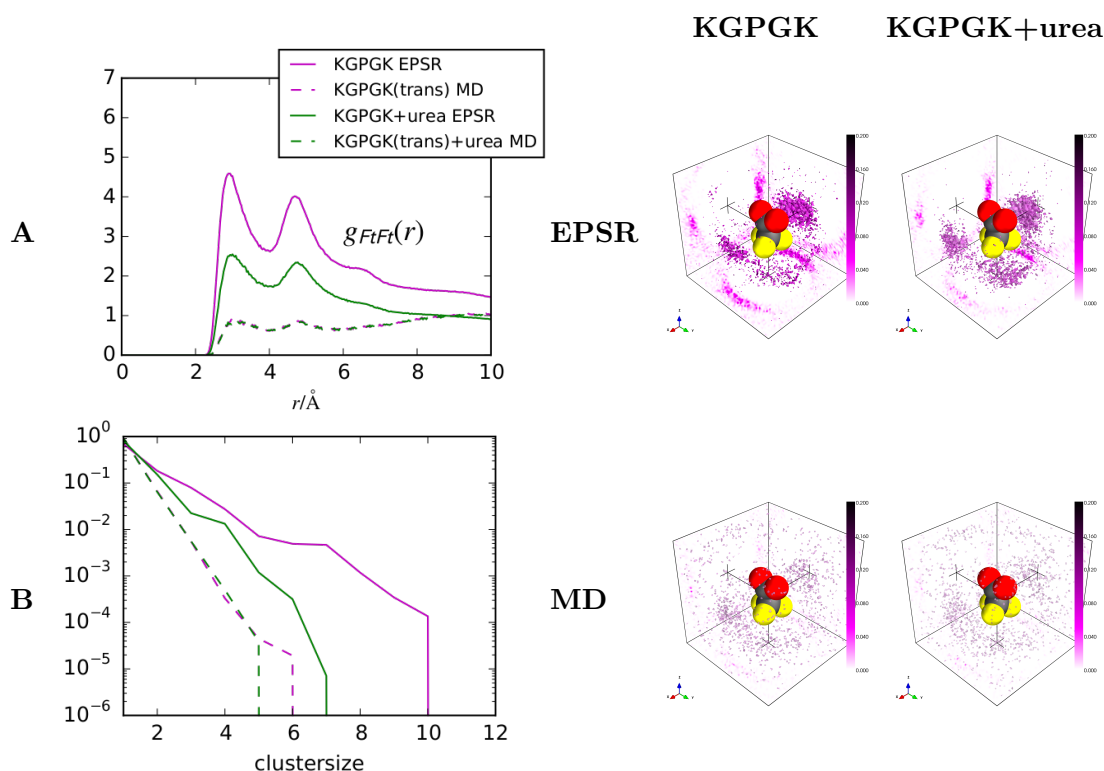


Figure 4.23: Left: A) Intermolecular $g(r)$ s (EPSR solid lines, MD dashed lines) between the TFA fluorines (Ft). B) Proportion of TFA-TFA association within a 0-4 $\text{\AA}$ intermolecular Ft-Ft distance. Urea-containing simulations are shown in green, non-urea-containing simulations are shown in magenta. The corresponding coordination numbers are in Table 4.7. Right: Spatial density maps (SDMs) show the location of the TFA (magenta) molecules up to a 10 $\text{\AA}$ distance around a central representative TFA molecule for the EPSR (top row) and MD (bottom row) simulations. The isocontour surfaces enclose the TFA molecules within the top 10% density. The scale bar shows the local density of neighbours around the central TFA molecules in units of $\text{\AA}^3$.

EPSR is generally higher compared with MD, except for the carbonyl oxygen O1 between Lys1 and Gly2. Unsurprisingly, the largest differences in hydration occur at O2, O3 (and O4) close to the varied residue (X); where KGGGK is the most hydrated and KGPGK and KGLGK are the least hydrated. Therefore smaller amino acid residues seem to facilitate backbone oxygen hydration, whereas larger hydrophobic amino acid residues obstruct hydration of the peptide backbone oxygens. The unusual shape of the hydration RDFs around O1 in MD is, in part, a result of being adjacent to the positively charged N-terminus. This influence is not visible in EPSR where the O1-Ow RDF shape is similar to other residues. In general, the oxygens at the peptide termini (especially the C-termini) are often more hydrated than oxygens in the middle of the peptide. This might, partially, be explained by the better accessibility for water molecules at the end of the peptide chain.

Similar to the backbone oxygen hydration, the nitrogens showing the largest differences in hydration are the ones adjacent to the third varied residue (N3 and N4). The significantly less hydrated N3 on KGPGK, for both EPSR and MD simulations, can be explained by the less polar structure of proline. An apolar leucine residue lowers the N3 hydration only slightly ($CN \approx 1.2$) compared to a proline residue ($CN \approx 0.3$). The polar serine residue has a more hydrated backbone nitrogen N3 compared to proline or leucine ($CN \approx 1.4$) but is less hydrated than a small glycine residue ($CN \approx 1.6$). Again, smaller amino acid residues seem to promote backbone hydration more than polar hydrophilic residues. The two different KGPGK simulations, MD and EPSR, show similar backbone nitrogen hydration. In EPSR, the N-terminal N1 and N2 are slightly less hydrated in EPSR compared with MD and the C-terminal N6 is significantly more hydrated than in MD. As stated above, all of the peptides show a higher backbone hydration at the termini than in the middle of the peptide chain. Especially the positively charged N1 is always the most hydrated nitrogen for all simulations.

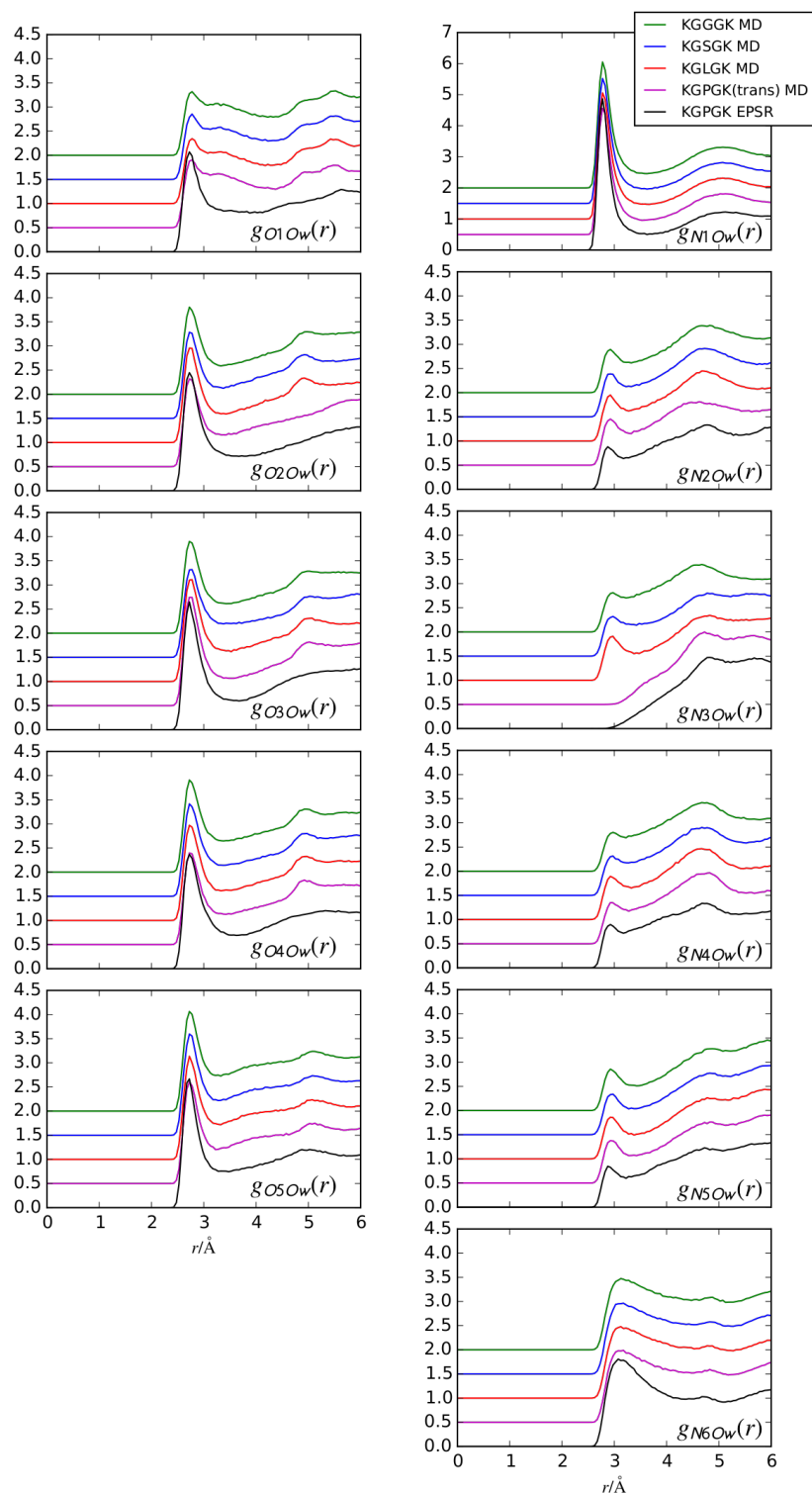


Figure 4.24: Radial distribution functions between backbone oxygens and nitrogens and water oxygens for the KGXGK peptides. The corresponding coordination numbers are listed in Table 4.8.

Table 4.8: Coordination numbers for the water oxygens around the KGXGK peptide backbone nitrogens and oxygens (Figure 4.24).

	$r_{max} / \text{\AA}$	KGPGK EPSR	KGPGK MD	KGLGK MD	KGSGK MD	KGGGK MD
O1-Ow	3.5	2.6	2.7	2.6	2.7	2.7
O2-Ow	3.5	2.5	2.0	2.0	2.2	2.2
O3-Ow	3.5	2.5	2.3	2.3	2.4	2.4
O4-Ow	3.5	2.8	2.2	2.2	2.3	2.4
O5-Ow	3.5	3.3	2.7	2.8	2.8	2.8
N1-Ow	3.5	3.7	3.9	3.9	3.8	3.9
N2-Ow	3.5	1.2	1.5	1.5	1.5	1.5
N3-Ow	3.5	0.3	0.2	1.2	1.4	1.6
N4-Ow	3.5	1.3	1.4	1.3	1.4	1.6
N5-Ow	3.5	1.3	1.3	1.2	1.3	1.3
N6-Ow	3.5	3.4	2.7	2.7	2.7	2.8

4.3.7 Peptide Conformation Analysis

4.3.7.1 Time Dependent Conformation Analysis

To determine if the O2 $\cdots$ N5 distance changes over the course of the MD simulation forming a β -turn structure this intramolecular distance is plotted for one KGPGK molecule in Figure 4.25. The analysis of the other 19 KGPGK peptides is in Appendix B. This intra-peptide distance is mostly at $\approx 6.5 \text{\AA}$ and rarely changes to a water-mediated hydrogen bonding distance of $\approx 3.5 \text{\AA}$ during the course of the simulation. However, this transition from an open to a closed peptide conformation and *vice versa* seems to be correlated with a change in the intramolecular N4 $\cdots$ N5 distance (black) which also shortens from $\approx 3.8 \text{\AA}$ to $\approx 2.8 \text{\AA}$, at the same time or just before a transient O2 $\cdots$ N5 hydrogen bond is formed.

4.3.7.2 Dihedral Angles

There are several approaches to determine the amino sequence dependent peptide conformations. The backbone dihedral (labels in Figure 4.4) angle analysis (Figure 4.26) using ANGULA (as described in Section 4.3.4) show that for the MD simulated peptides only the ϕ_{X3} (C(=O)-N3-C α_{X3} -C(=O)) and ψ_{X3} (N3-C α_{X3} -C(=O)-N4) dihedral angles of the varied middle residue are different between the peptides. Figure 4.26

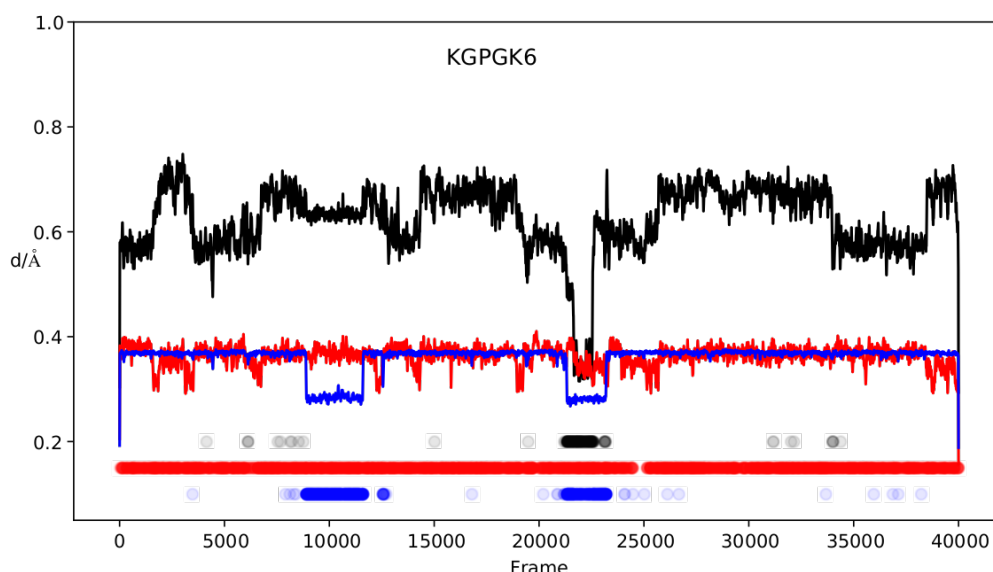


Figure 4.25: Analysis of a KGPGK peptide in the MD trajectory. The top graph shows the intramolecular distances $N4 \cdots N5$ (blue), $O2 \cdots N5$ (black) and $O2 \cdots N4$ (red). The bottom graph illustrates the water molecule mediating this interaction.

depicts these dihedral angle distributions for $X3$ in comparison with the, for MD, unchanging distributions of Gly2 and Lys5. The other dihedral angle distributions of MD simulated peptides are almost identical to the ones of KGPGK(*trans*) in Figure 4.17. The EPSR simulation of KGPGK, on the other hand, shows more variation in peptide conformation across all residues (Figure 4.17) even compared to the same peptide in MD.

While in Figure 4.26 ψ_{X3} is only markedly different for glycine as the third residue, where an angle of 180° is due to the lack of a large amino acid residue preferred over an angle of 140° which is preferred in the other peptides, the most variation is seen in ϕ_{X3} . In general, all peptides, except KGPGK, have two ϕ_{X3} peaks at $\approx 80^\circ$ and, usually, a larger one at $\approx -80^\circ$. For KGPGK in MD this peak is slightly shifted to $\approx -50^\circ$ and no second peak is visible. In EPSR, on the other hand, two peaks are visible at $\approx -30^\circ$ and $\approx 80^\circ$.

To determine how this is reflected in the peptide conformation the correlation of ϕ_{X3} and ψ_{X3} angles is shown in Figure 4.27 in bivariate heat maps next to example peptides in the respective conformations (without lysine residues). These two-dimensional representations are, effectively, Ramachandran plots [35] for the $X3$ residue.

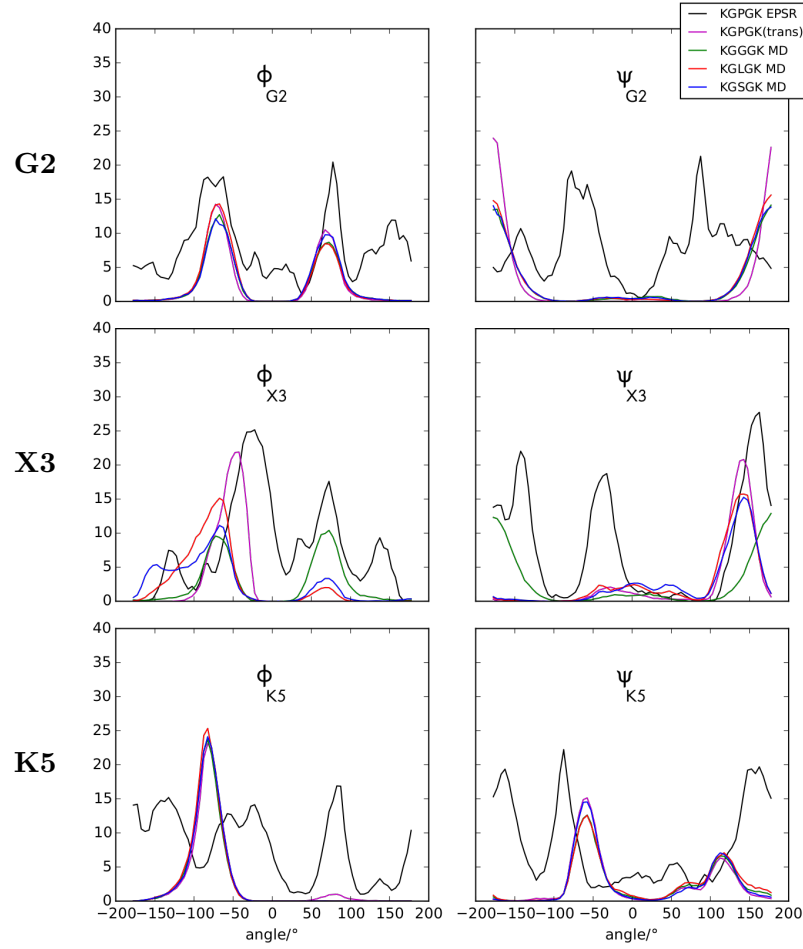


Figure 4.26: Selected dihedral angles along the KGXGK peptide backbones: Only the X3 residue dihedrals are significantly different between the peptides in MD. Other dihedral angles for the peptides simulated with MD are identical to those in Figure 4.17 and Table 4.6 for KGPGK (MD). The labels correspond to the dihedral angles in Figure 4.4.

KGPGK, KGLGK and KGSGK in the MD simulations all have one preferred conformation ($\phi_{X3} \approx -80^\circ$, $\psi_{X3} \approx 140^\circ$). KGGGK has two equally preferred conformations that are both more ‘open’ ($\phi_{G3} \approx -70^\circ / +70^\circ$, $\psi_{G3} \approx 180^\circ$) than any conformations preferred in the other peptides. Comparing the preferred conformations of peptides in MD, it is clear that KGPGK has a slightly different, less ‘open’, backbone conformation than the other peptides, due to the shift in ϕ_{P3} . In EPSR another equally preferred KGPGK conformation ($\phi_{P3} \approx -40^\circ$, $\psi_{P3} \approx -30^\circ$) is even more compact with the residues before and after proline in closer proximity to one another.

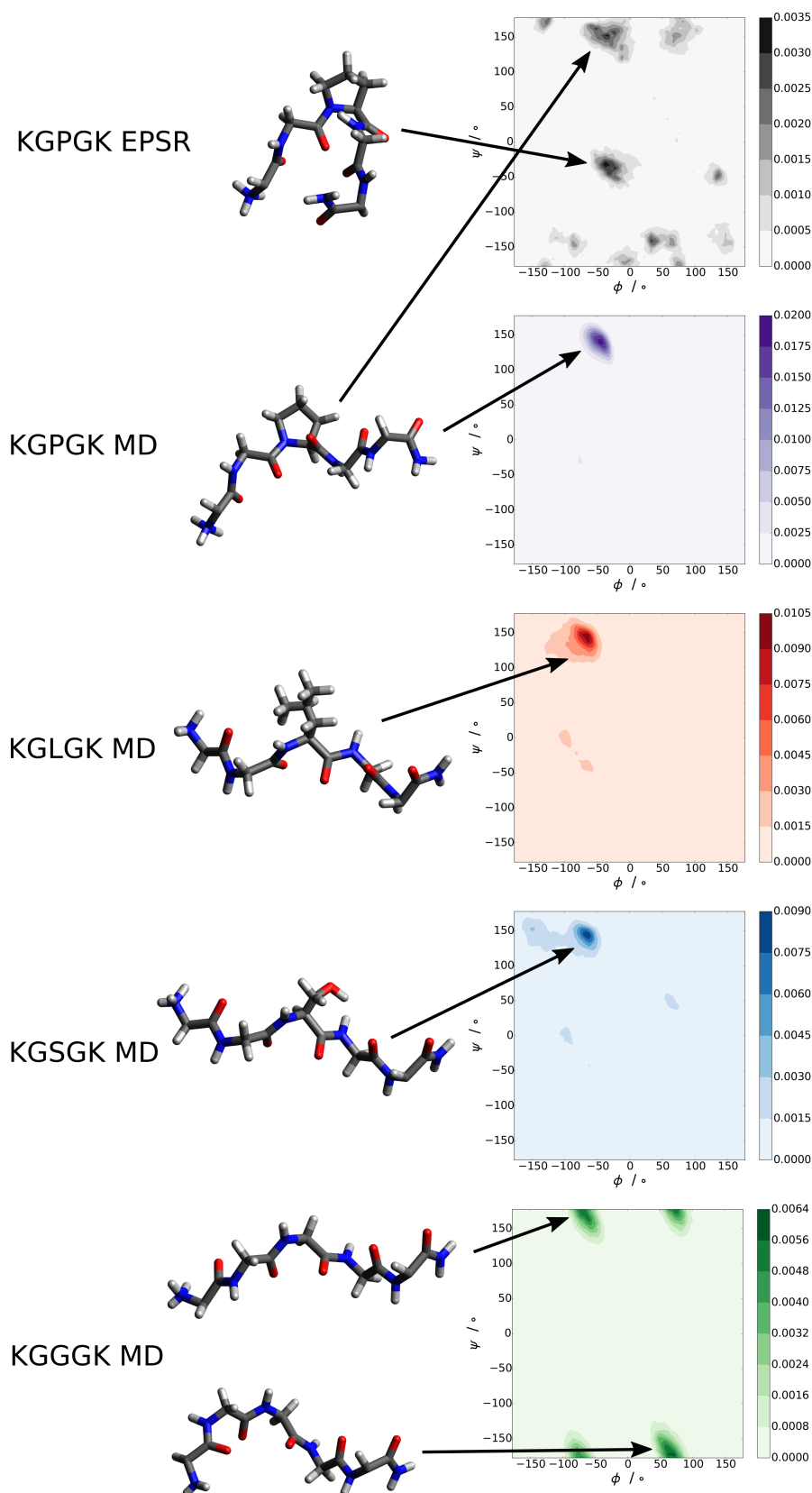


Figure 4.27: Ramachandran plots of the dihedral angle pair ϕ_{X3} / ψ_{X3} for the KGXGK peptides. The corresponding conformations (without lysine residues to enhance clarity) are also displayed. Other dihedral angles for the peptides simulated with MD are identical to those in Figure 4.17 and Table 4.6 for KGPGK (MD).

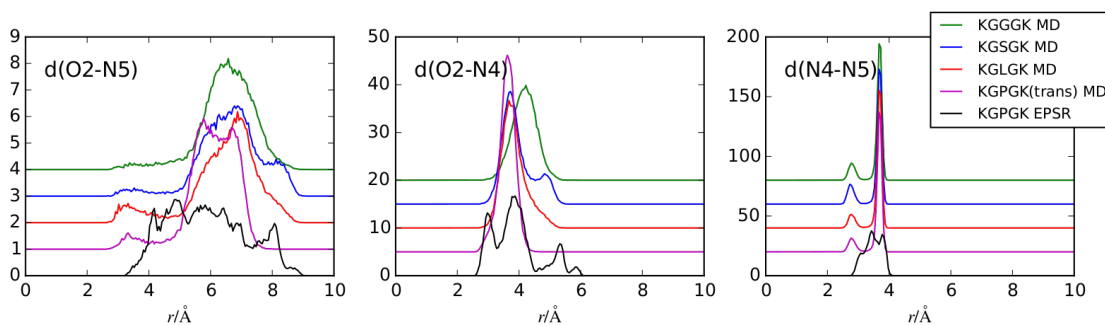


Figure 4.28: Intramolecular distance distributions of selected residues for the KGXGK peptides.

4.3.7.3 Intramolecular distances

Another way to analyse the residue specific peptide conformations is to assess the intra-peptide distances. Within the KGXGK peptides, the $O2 \cdots N5$ intra-peptide distance can give a first insight into the range of folded states in the peptide, as this is the $i \rightarrow i + 3 / O \cdots H-N$ intra-molecular contact position in β -turns. Figure 4.28 shows the intramolecular distance distributions for $O2 \cdots N5$ next to the $O2 \cdots N4$ and $N4 \cdots N5$ as these distances might be correlated (Section 4.3.7.1).

It is clear that the intramolecular peptide distances show a marked difference between EPSR and MD simulations. All MD simulations show a $O2 \cdots N5$ intramolecular RDF with a clear maximum at $\approx 6-7 \text{ \AA}$ and a small proportion of peptides at a closer ($\approx 3.5 \text{ \AA}$) intrapeptide distance; where these peptides are stabilised by a direct $O2 \cdots H-N5$ hydrogen bond. The peptides in EPSR show a broader range of $O2 \cdots N5$ distances and have a considerable proportion of peptides with a $O2 \cdots N5$ distance at $\approx 4.5 \text{ \AA}$ where all MD simulations have a clear minimum at $4.5 \text{ \AA}$.

In the $O2 \cdots N4$ distance, the difference between peptide conformation flexibility is even more pronounced. All of the peptides (except KGSGK with two maxima at $3.8 \text{ \AA}$ and $5.0 \text{ \AA}$) show a clear monomodal distribution (with one maximum at $\approx 4 \text{ \AA}$) of intrapeptide distances in MD whereas EPSR KGPGK shows a broad distribution of distances.

A relatively close $N4 \cdots N5$ distance has been shown to precede a closer $O2 \cdots N5$ distance in Figure 4.25. A clear bimodal distribution of $N4 \cdots N5$ distances is visible in

Table 4.9: Percentages of peptide conformations based on the O2...N5 distance in Figure 4.28 (left).

	$r_{\text{O2-N5}} / \text{\AA}$	KGPGK EPSR %	KGPGK MD %	KGLGK MD %	KGSGK MD %	KGGGK MD %
open	6.0 – 10	61.2	63.4	83.5	84.4	84.7
medium	4.5 – 6.0	31.2	34.5	13.8	14.4	14.0
closed	0 – 4.5	7.6	2.2	2.7	1.2	1.3

all MD simulations, with a small peak at $\approx 2.8 \text{ \AA}$ and a larger one at ≈ 3.8 , identical to the distances observed in Section 4.3.7.1. This might indicate that the observed correlation between a close N4...N5 distance and O2...N5 is not unique to KGPGK. The EPSR-KGPGK N4...N5 distances are in the same close range (3-4 $\text{\AA}$), but do not show this bimodality.

The O2...N5 RDF from the MD simulations can be used to broadly classify the peptides into three conformational states during ‘folding’: closed ($\text{O2}\cdots\text{N5} < 4.5 \text{ \AA}$), medium ($4.5 \text{ \AA} < \text{O2}\cdots\text{N5} < 6.0 \text{ \AA}$) and open ($\text{O2}\cdots\text{N5} > 6.0 \text{ \AA}$). Table 4.9 lists the proportions of peptides in these groups. The conformation distributions of KGPGK in EPSR are similar to the MD simulated KGPGK, although in EPSR the KGPGK molecules sample a wider range of O2...N5 intra-peptide distances, particularly in the closed distance range where EPSR shows fewer peptides with O2...N5 distances less than $3.0 \text{ \AA}$, but shows a much larger number of molecules with an O2...N5 distance between 3.0 and $4.5 \text{ \AA}$.

From the proportions of peptides at different intramolecular O2...N5 distances it is clear that, with the chosen classification, KGPGK is, in both simulation techniques, markedly different from the other peptides. All of the peptides are predominantly populate the ‘open’ conformation, but KGPGK only has $\approx 62\%$ in this conformational group, compared with all other peptides which have $> 83\%$ of the peptide conformations in this group. The proportion of peptides in the ‘medium’ conformation group is double the amount for KGPGK than for all other peptides. The ‘close’ conformation is much more populated in EPSR-KGPGK than in all of MD simulated peptides. Interestingly, the peptides with an apolar X in KGXGK (KGPGK and KGLGK) show a still higher proportion of close conformation peptides compared with KGSGK and KGGGK.

Table 4.10: Coordination numbers for the water oxygens around the KGXGK peptide backbone nitrogens N4 and N5 depending on their conformational group defined in Table 4.9.

	d(O2-N5)	KGPGK EPSR	KGPGK MD	KGLGK MD	KGSGK MD	KGGGK MD
N4-Ow	total	1.3	1.4	1.3	1.4	1.6
	open	1.3	1.4	1.3	1.4	1.6
	medium	1.5	1.3	1.2	1.3	1.5
	closed	1.2	1.4	1.3	1.5	1.5
N5-Ow	total	1.3	1.3	1.2	1.3	1.3
	open	1.2	1.4	1.2	1.3	1.3
	medium	1.2	1.3	1.2	1.2	1.2
	closed	1.5	0.9	0.6	0.6	0.6

4.3.8 Conformation Dependent Peptide Hydration

The shape of the peptide backbone hydration RDFs in Figure 4.24 do not appear to be heavily influenced by the peptide conformation group in all simulations. Only the hydration of the peptide backbone nitrogens N4 and N5 is affected by the conformation of the peptides, based on the intra peptide O2...N5 distance in Table 4.9. These N4-Ow and N5-Ow RDFs are shown in Figure 4.29.

In EPSR-KGPGK the first N4 hydration peak is smaller in the ‘closed’ conformations ($CN \approx 1.2$ compared to $CN \approx 1.3$ for ‘open’) whereas it becomes more pronounced in the apolar MD-KGPGK and MD-KGLGK peptides, but this is not reflected in a change in N4-Ow coordination numbers ($CN \approx 1.3$ -1.4). There is no obvious conformation influence on the N4 hydration in KGSGK and KGGGK. Conversely, N5 becomes more hydrated in EPSR upon peptide closing ($CN \approx 1.5$ compared to $CN \approx 1.2$ for ‘open’) whereas the first N5 hydration peak shrinks in all MD simulated peptides upon peptide closing ($CN \approx 0.6$ -0.9 compared to $CN \approx 1.3$ for ‘open’). All other peptide backbone nitrogens or oxygens are not largely affected by a change in the intrapeptide O2...N5 distance.

Overall these results indicate that, at a closer O2...N5 distance, the MD simulated peptides are in a direct intra-peptide hydrogen bond without mediating water molecules whereas ‘close’ KGPGK peptides in EPSR are in a transient O2...N5 hydrogen bond with bridging water molecules.

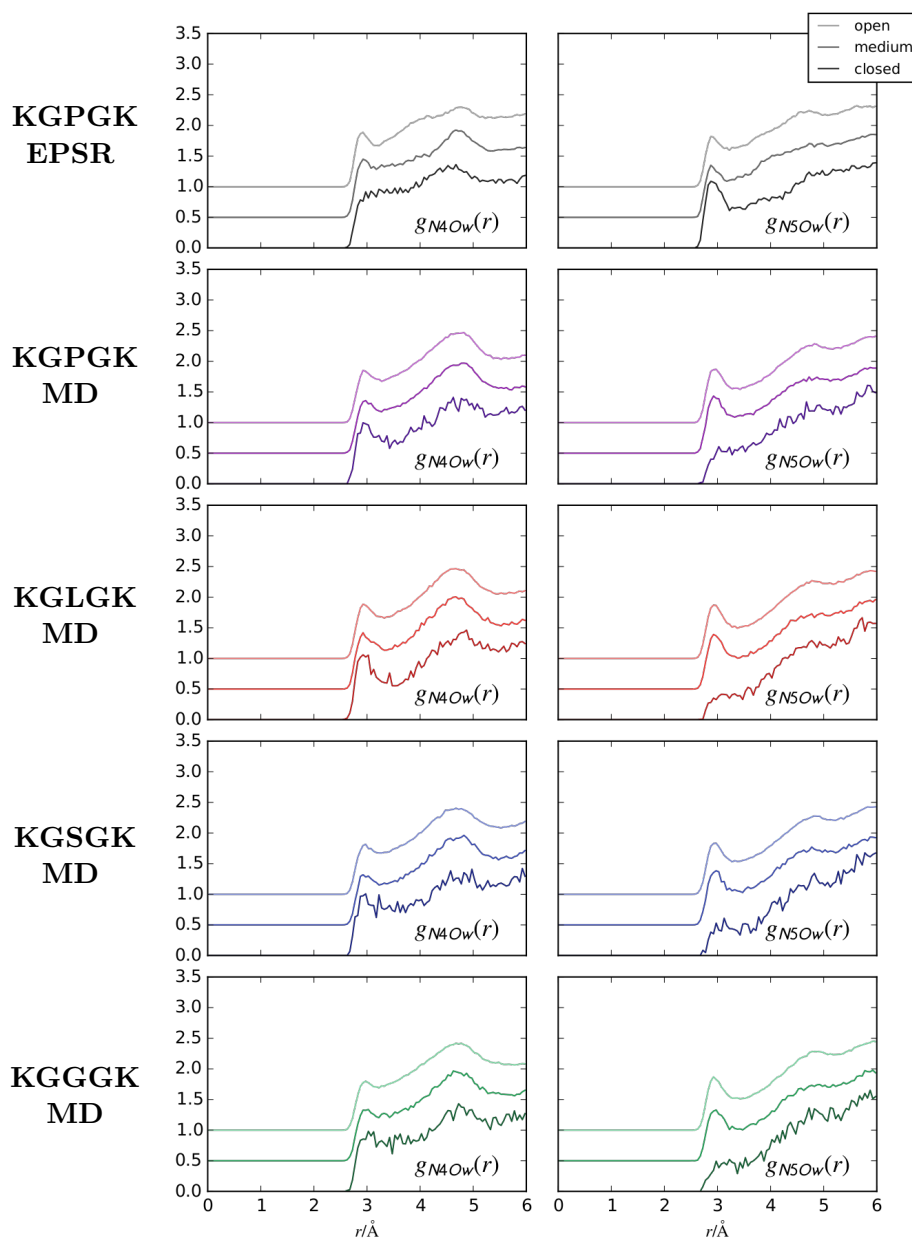


Figure 4.29: Radial distribution functions between backbone nitrogens N4 and N5 and water oxygens for the KGXGK peptides separated into three conformational groups according to Table 4.9. The corresponding coordination numbers are in Table 4.10.

4.3.9 Shared nearest neighbour water molecules

Figure 4.30 shows the same intramolecular distance distributions between $O2 \cdots N5$, $O2 \cdots N4$ and $N4 \cdots N5$ as in Figure 4.28, including the distributions of distances where the nearest neighbour water molecule is shared between the two peptide sites. The corresponding proportions of peptides with shared water molecules within these distances are listed in Table 4.11.

Again, there is a clear difference between EPSR and MD simulations of KGPGK. Between $O2 \cdots N5$ KGPGK-EPSR has, by far, the most shared first nearest neighbour water molecules. Although there are still very few water molecules, there are more shared nearest neighbour water molecules in MD for KGPGK and KGLGK (within the ‘close’ distance range) than for KGGGK and KGSGK. Conversely, the MD simulations of all $KGXGK$ peptides have more bridging water molecules between $O2 \cdots N4$ than in KGPGK-EPSR. There is a clear trend within the MD simulated peptides regarding the number of peptides with water bridging between $O2 \cdots N4$, where peptides with larger and more hydrophobic X -residues have more shared nearest neighbour water molecules than peptides with small or hydrophilic X -residues. While for KGPGK-MD 5.1% of the peptides have a bridging water between $O2 \cdots N4$ only 3.1% of KGSGK and 1.3% of KGGGK have a bridging water in this position. Although there are fewer bridging water molecules between $O2 \cdots N5$ for the MD simulations, a similar trend can be observed where KGPGK-MD and KGLGK peptides show only 0.8% of their conformations with a shared first nearest neighbour water molecule and only 0.4%-0.5% of KGSGK and KGGGK share a nearest neighbour water between $O2 \cdots N5$.

Many of the conformations at a close ($\approx 2.8 \text{ \AA}$) $N4 \cdots N5$ distance in MD have one (nearest neighbour) mediating water between $N4 \cdots N5$ and a shared second nearest neighbour water is found in conformations with a larger $N4 \cdots N5$ distance ($\approx 3.8 \text{ \AA}$). Similar to Section 4.3.8 these results indicate that in MD bridging water is found between $O2 \cdots N4$ and between $O2 \cdots N5$ in EPSR.

Table 4.11: Percentages of conformations that share neighbouring water oxygens.

	shared	KGPGK EPSR	KGPGK MD	KGLGK MD	KGSGK MD	KGGGK MD
O2-N5	1st NN	3.1	0.8	0.8	0.4	0.5
	2nd NN	2.7	1.2	0.7	0.5	0.9
	1st-2nd NN	4.7	1.7	1.5	0.8	1.2
O2-N4	1st NN	3.1	5.1	4.8	3.1	1.3
	2nd NN	3.6	9.1	9.2	6.8	4.2
	1st-2nd NN	7.6	17.1	16.0	11.1	6.3
N4-N5	1st NN	5.9	3.4	3.4	4.0	3.6
	2nd NN	5.1	3.1	3.7	2.9	2.7
	1st-2nd NN	13.9	5.3	5.8	5.8	5.5

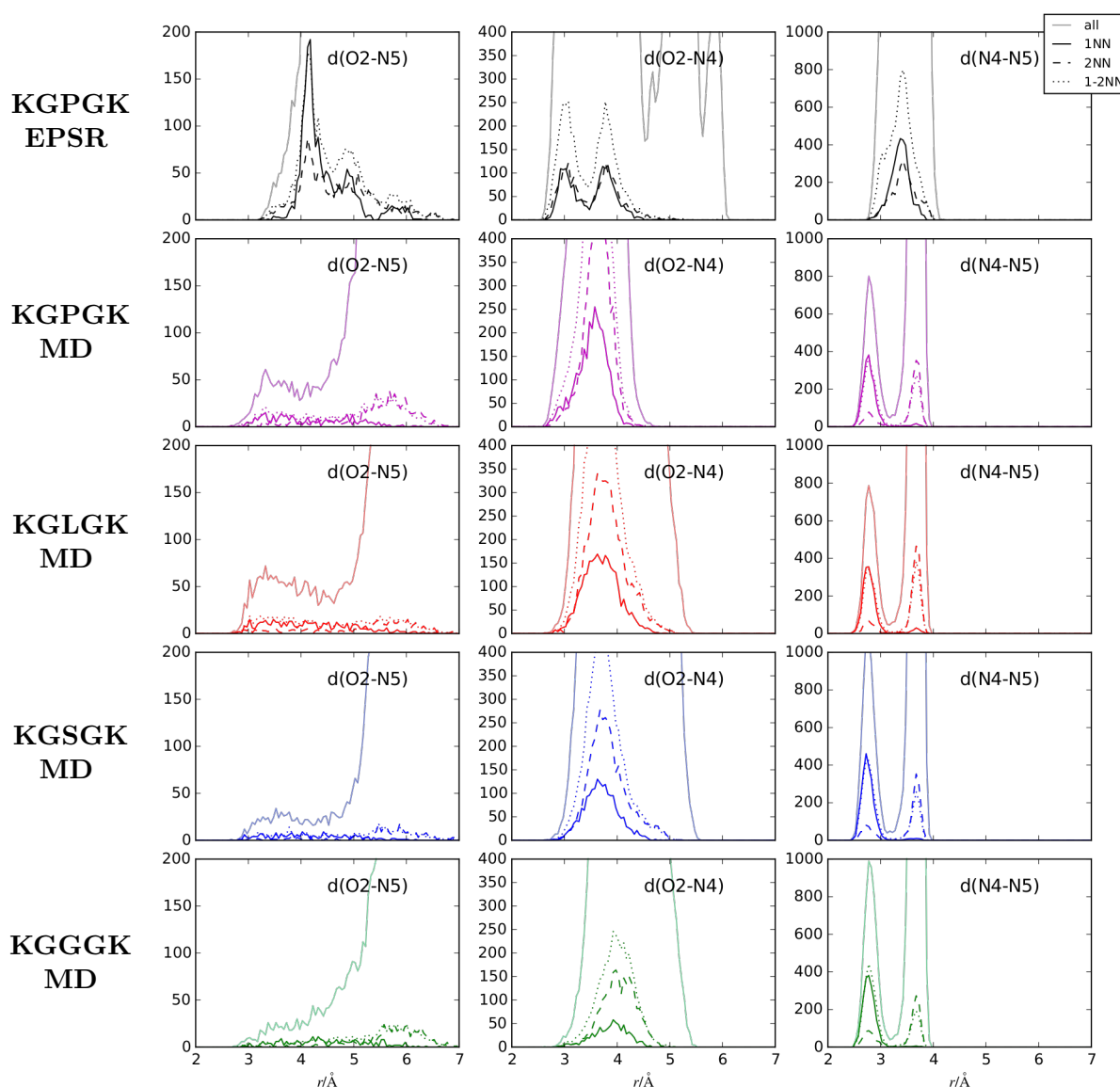


Figure 4.30: Intramolecular distance distributions of selected residues for the KGXGK peptides separated into intramolecular distance of molecules that share their first (1NN), second (2NN) or first/second (1-2NN) nearest neighbour water oxygen and all molecules. The corresponding percentages are shown in Table 4.11.

4.4 Conclusions

This chapter showed that the hydration of the KGPGK peptide is either mostly unchanged (for MD) or slightly increased (for EPSR) upon urea addition to the solution. This is contrary to what occurs for the small peptide, GPG-NH₂, in aqueous solution with Cl⁻ as the counter ion, where the presence of urea (3.3 M) lowers the peptide backbone hydration by up to -20% [249]. A similar replacement of water by urea has also been observed in X-ray crystallography experiments on lysozyme [194]. The relative absence of water replacement around the KGPGK backbone upon the addition of urea in the present work, suggests that the presence of TFA in the solution largely prevents the loss of hydrating water around the peptide, at this low urea concentration, perhaps, effectively protecting the peptide from denaturation. Although direct urea association is preferred over water, no urea impact on the peptide hydration, only on the TFA-interaction, is visible. This is similar to the reduced Cl⁻-peptide interaction that was observed for GPG-NH₂ upon urea addition in Chapter 3. Similarly, the addition of urea to the solution does not largely disrupt either the bulk water structure or the TFA hydration in the solutions, where only minor changes in the coordination numbers are observed in either simulation.

In the absence of urea, MD shows very few peptide-TFA interactions along the peptide backbone. These interactions are more prevalent in the EPSR simulation and vary, along the backbone. Upon the addition of urea to the solution, these peptide-TFA interactions decrease in EPSR and remain roughly the same in the MD simulation, presumably because there were relatively few interactions present in this simulation to begin with. In EPSR, urea disrupts the TFA molecules interacting with the peptide, in some cases by replacing these interactions but only to a minor extent, given the very low coordination numbers for peptide-urea interactions (Tables 4.4 and 4.5). This difference between EPSR and MD simulations appears to be a result of the TFA being less hydrated and associating more with itself and the peptide in EPSR compared to MD.

Interestingly, overall the TFA interactions with either the peptide or between TFA molecules is decreased upon the addition of urea to the solution in the EPSR simulation,

while these interactions remain relatively constant in the presence and absence of urea in MD. The discrepancies between the two simulation methods may be due to the fact MD does not as accurately predict the positions for TFA as EPSR does in this particular solution. This is probably due to the problematic TFA starting parameters from CHARMM36.

The addition of urea leads to a decrease in charge-charge interaction in the solution - both between the polar peptide bonds in KGPGK and TFA and also between TFA molecules with themselves. This, perhaps, suggests that the ability of urea to denature proteins is dependent both upon the pH and ionic strength of the solution. It may be that urea, at low pH and high ionic strength (as in this chapter), first disrupts the peptide-salt interactions by directly interacting with the anions in solution and only at higher urea concentrations directly interacts with the peptide, replacing the hydrating water molecules, as has been previously observed in the absence of high salt concentration at a higher pH (Chapter 3) [249, 250, 221, 104, 105]. Similarly, urea-small molecule interactions have been suggested to contribute to the osmolytic activity of betaine [234], where it was found that betaine effectively removed urea from the surface of a protein preventing its subsequent denaturation [236].

Further, the fact that high salt concentrations can protect against denaturation by urea is in accordance with previous studies showing the protective influence of higher salt concentrations [233] against urea-induced unfolding on human serum albumin protein. MD simulations on amino acid side chain analogues and protein backbone mimic molecules with urea and NaCl have suggested that the high salt protection against unfolding of human serum albumin is due to an opposing salting effect between urea and the salt [251]. This hypothesis would also explain protective properties of high-salt buffers for proteins or the salt resistance of certain organisms [252].

The emerging picture from this chapter is that at high ionic concentrations, the ion-urea interactions can help to protect peptides and proteins from denaturation by intercepting the urea molecules by direct ion-urea contacts or indirect replacement at peptide interaction sites. It may be that at even higher salt concentrations, that salts can also effectively block access to the peptide-charged sites as there is a small propensity

for the salt in the present solution, to interact with the polar portions of the peptide backbone even at the current salt concentrations. Similarly, it may also be that at higher urea concentrations, salt-urea interactions will not be sufficient to prevent denaturation.

The backbone hydration of KGXGK peptides, regardless of conformation, is only marginally affected by the variation in the *X*-residue (Figure 4.24). Large and hydrophobic *X*-residues obstruct the backbone hydration only very slightly, whereas the tertiary amine nitrogens of proline are significantly less hydrated, compared to other residues. Additionally, proline has an effect on the preferred peptide conformation, where, in *trans* KGPGK isomers, hydration sites next to proline are in closer proximity to one another than in any other KGXGK peptide. This is also reflected in the significantly higher percentage of KGPGK peptide conformations within a ‘medium’ O2...N5 distance, compared to the other KGXGK peptides (Table 4.9). The MD simulated peptides have, generally, less variation in their conformation compared to the EPSR simulated KGPGK. This explains why one particular ‘compact’ peptide conformation is only found in EPSR-KGPGK (Figure 4.27). In fact, this unique conformation might explain the conformation and conformation-dependent hydration differences between MD and EPSR simulated KGPGK: While very few KGXGK peptides in MD are either found in a direct O2...N5 hydrogen bond or with a mediating water between O2...N4, which might precede a direct O2...N5 hydrogen bond, many more EPSR-KGPGK peptides are found with a mediating water between O2...N5 (Figure 4.30).

A peptide conformation analysis over a time period is only possible using the MD trajectories, as EPSR does not sample configurations time-dependently. Extracting the peptide conformations of one KGPGK peptide during the conformational changes leading to a direct O2...N5 hydrogen bond from the MD trajectory (Figure 4.31) gives rise to the following proposed mechanism:

As a first step the hydration density around the N4...N5 site initiates the folding process. Once the peptide adopts a conformation which places N5 and its hydrating water close enough to the Gly2-Pro3 peptide bond oxygen, the presence of a hydrophobic residue (e.g. proline) forces the water into a preferred position around O2 in order to complete the folding process. This preferred hydration position orients the O2 water

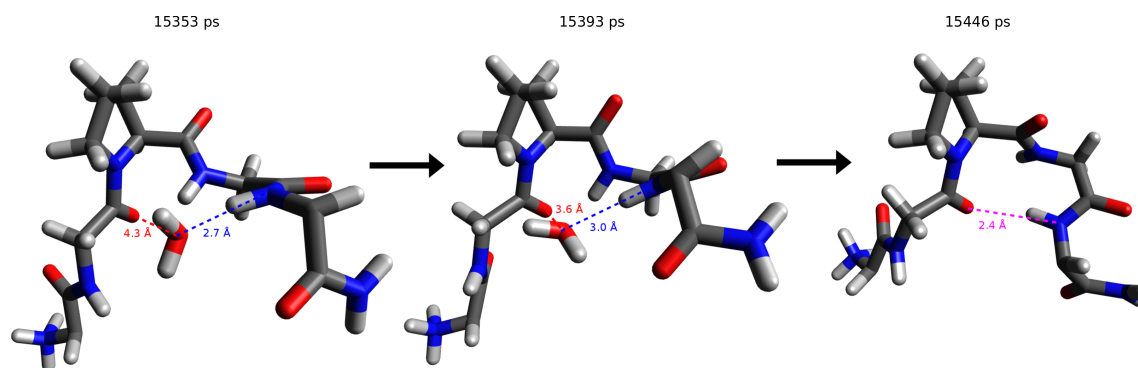


Figure 4.31: Sequence of events that lead to a direct hydrogen bond between O2 and N5 in KGPGK. Taken from the molecule analysed in Figure 4.25. The molecule is displayed without lysine residues to enhance clarity.

molecules in such a way that they can bridge to the $i + 3$ peptide bond nitrogen, thus this site-specific hydration around N5 seems to precede and initiate the folding and the altered orientation of water around O2 drives the peptide folding process to completion.

Overall these results indicate an interplay of proline inducing a specific peptide conformation which facilitates a hydration pattern that leads to a direct peptide-peptide hydrogen bond. The uniqueness of proline is not only due to its backbone dihedral influence but also because of its hydrophobicity and that it is a relatively large residue. A similar behaviour, with slightly more peptides in a direct O2 $\cdots$ N5 hydrogen bond than MD-KGPGK, is observed for KGLGK (Table 4.10), complementing the evidence for the importance of a larger hydrophobic residue to facilitate a $i \rightarrow i + 3$ hydrogen bond, although leucine has not been shown to be preferred in this (second or $i + 1$) position in β -turns [253].

While it is possible that other peptides without proline also exhibit this conformational transition that could lead to a β -turn formation [254], there is a clear conformational preference (Table 4.9) for a more compact conformation and more bridging water molecules within the involved intra-peptides distances (Table 4.11) for KGPGK than for any of the other peptides analysed. This suggests that a proline residue increases the probability of this conformational transition. Additionally, the preference of PG-motif peptides to form a type II β -turn was also found experimentally using directed evolution [254] and NMR spectroscopy [255]. It is possible that this driving force of a turn formation could initiate peptide, and by extension, protein folding in solution [256].

However, as the analysed peptides are very small, do not form a repetitive secondary structure or a hydrophobic core in solution, a comparison to larger proteins remains highly speculative.

Chapter 5

Solvent and Urea Influence on the Amphiphilic KLL-Peptides

5.1 Introduction

Although there are several approaches to successfully predict secondary structures from the amino acid sequence [257, 258, 259], the mechanism of secondary structure folding [260, 261] and unfolding [262, 263] is debated. As the hydrogen bond dependent folding of secondary structures is suggested by the ‘framework model’ (Section 1.1.5) to be the initiating step in protein folding [264], simple secondary structure model peptides have previously been used to probe protein folding and unfolding [265, 266].

Amphiphilic helices are secondary structure elements with hydrophobic and hydrophilic residues on opposing sites of the helix (Figure 5.1). They are found in several biological contexts for example as cytotoxic and hemolytic peptides [267], signalling peptides [268, 269], apolipoproteins [270], transmembrane proteins [271] or antimicrobial peptides [272, 273] that can disrupt bacterial cell membranes and might provide a new treatment for drug resistant infections.

While the peptides analysed in Chapters 3-4 only contained small structural motifs and no secondary or higher protein structure, this chapter will focus on the structure of two larger amphiphilic helix-forming peptides, KLL₁₅ (Ac-KLLKLLKLLKLLK-

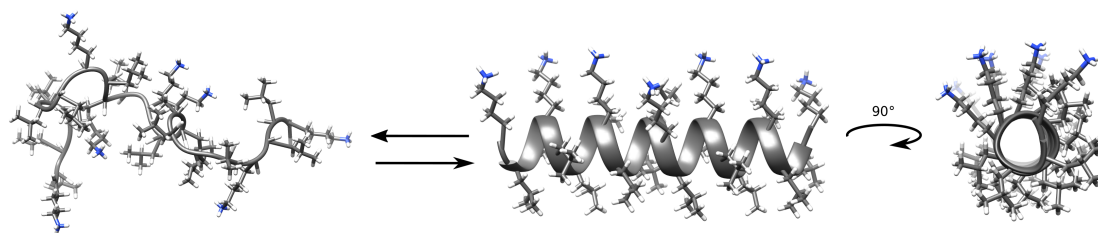


Figure 5.1: KLL₂₂ peptide in a random conformation (left) and as a helix (middle and right). In the helical conformation the hydrophilic lysine residues and the hydrophobic leucine residues are on opposite sites of the helix.

NH₂) and KLL₂₂ (Ac-KLLKLLKLLKLLKLLKLLK-NH₂). These peptides (in the following referred to as KLL) have previously been designed *de novo* and were found to form α -helices in solution, where longer KLL peptides were found to be more helical in aqueous solution than shorter KLL peptides and all were found to have a higher helix content in more apolar solvents [274]. Figure 5.1 shows KLL₂₂ in an example random conformation and fully folded as an α -helix. These peptides were also found to be highly hemolytic, more than natural occurring cytotoxic peptides like melittin in bee venom. However, a mode of action similar to melittin, where the peptide increases the ion permeability of the cell membrane by forming a hydrophilic pore [275], is probable.

The aim of these investigations are, on the one hand, to explore the influence of an amphiphilic solvent environment compared to pure water for KLL₁₅ and, on the other hand, to probe the urea influence in an aqueous solution on the KLL₂₂ peptide; to universally investigate the interactions of solution components with the peptide structure.

These peptides were selected to be investigated experimentally by neutron diffraction because of their high solubility and their redundant sequence of only two different amino acid residues which makes them, despite their length, relatively simple to model in EPSR. As the structure of the KLL peptides contains a larger secondary structure small angle neutron scattering was also performed.

Due to the molecule and simulation box size limitations of EPSR this chapter presents a more challenging and complicated subject to analyse with the given methods in comparison with the peptides presented in the previous chapters.

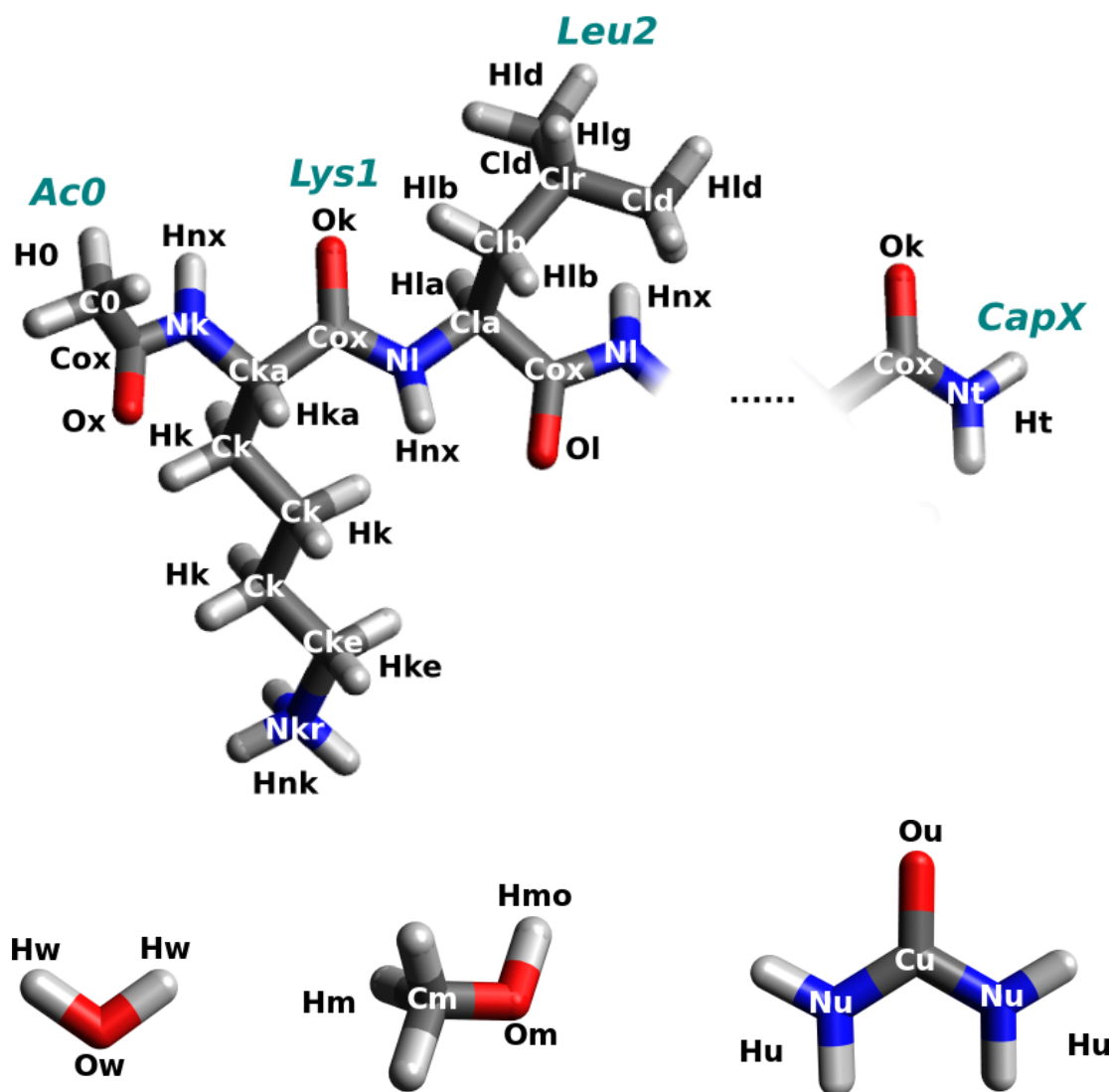


Figure 5.2: Basic KLL-peptides, water, methanol and urea labelling scheme.

5.2 Methods

5.2.1 NMR Spectroscopy Preparation and Measurement

KLL₁₅·5HCl and KLL₂₂·7HCl were purchased from Bachem (Bubendorf, Switzerland) and used without further purification. Peptide composition and sample purity were confirmed by elemental analysis (Medac Ltd.) and NMR spectroscopy.

NMR spectra of the peptides were acquired using the 950 MHz spectrometer and methanol dried over magnesium (Section 5.2.3), in order to ensure an anhydrous solvent

as described in Section 3.2.1 at a peptide concentration of 50 μM .^a

5.2.2 CD Spectroscopy Sample Preparation and Measurement

The secondary structure of the KLL peptides was determined by CD spectroscopy. CD spectra were acquired (threefold measurements) using a Jasco J-815 CD Spectrometer, Hellma High Precision quartz Cells (1 mm pathlength), water (H_2O) and methanol (CH_3OH ; Sigma Aldrich).^b The peptides concentrations used were 50 μM and 35 μM for KLL₁₅ and KLL₂₂ respectively. CD samples were prepared by peptide weight using a Satorius scale accurate to 4 d.p. and by solvent volume using 100-1000 μl Gilson pipettes.

5.2.3 Neutron Diffraction and Sample Preparation

The samples for the neutron diffraction measurements were prepared by weight as described in Section 3.2.2 with the appropriately isotopically labelled peptide, H_2O , D_2O (99.98%) and isotopomers of methanol to a molecular ratio of 1:1000 (KLL₁₅ : water, pH 3.4), 1:333:667 (KLL₁₅ : methanol : water, pH 3.2), 1:3000 (KLL₂₂ : water, pH 4.0) and 1:264:3000 (KLL₂₂ : urea : water, pH 5.2). This corresponds to a peptide concentration of $\approx 38\text{ mM}$ for KLL₁₅ and $\approx 18\text{ mM}$ for KLL₂₂ without urea, $\approx 15\text{ mM}$ for KLL₂₂ with urea and a urea concentration of $\approx 4\text{ M}$. Unless otherwise stated, the level of deuteration is with respect to the exchangeable hydrogens in the sample (Hnk, Hnx, Ht, Hmo, Hu and Hw) and the non-exchangeable hydrogens on methanol (Hm).

All isotopomers of methanol, CH_3OH (Sigma Aldrich), CD_3OH (Sigma Aldrich, 99.9%D) and CD_3OD (VWR, 99.8%D), were dried over iodine (I_2) activated magnesium. Each degassed methanol isotopomer was refluxed over the activated magnesium at 330 K for 48 hrs. After reflux, the methanol isotopomers were cryogenically distilled onto 3 Å molecular sieves, which had been pre-dried under vacuum, to obtain anhydrous methanol. The methanol was subsequently checked by NMR spectroscopy to ensure the solvent was anhydrous.

^aNMR spectra were taken and assigned by Natasha Rhys.

^bCD measurements were performed by Samvid Kurlekar under my supervision.

Table 5.1: List of samples used for the collection of neutron diffraction data on KLL₁₅. The molecular ratio of KLL₁₅·5Cl⁻:water was 1:1000 and the molecular ratio of KLL₁₅·5Cl⁻:methanol:water was 1:333:667. If not otherwise stated, the level of deuteration refers to the exchangeable hydrogens on KLL₁₅, methanol and water (Hnk, Hnx, Ht, Hmo and Hw).

Label	KLL ₁₅ + water			KLL ₁₅ + methanol + water		
	Deuteration	μA	DCS lvl.	Sample	μA	DCS lvl.
I	100%H	120.3	94.0	100%H	212.2	97.7
II	64%H 36%D	120.1	95.9	64%H 36%D	165.5	97.2
III	50%H 50%D	120.0	95.0	100%H, CD ₃ OH	160.1	97.0
IV	25%H 75%D	80.0	97.0	50%H 50%D	161.6	97.7
V	100%D	79.9	97.2	50%H 50%D, CD ₃ OH _{0.5} D _{0.5}	170.0	98.3
VI	-	-	-	100% D	160.4	102.0

Table 5.2: List of samples used for the collection of neutron diffraction data on KLL₂₂. The molecular ratio of KLL₂₂·7Cl⁻:water was 1:3000 and the molecular ratio of KLL₂₂·7Cl⁻:urea:water was 1:264:3000. The level of deuteration refers to the exchangeable hydrogens on KLL₂₂, urea and water (Hnk, Hnx, Ht, Hu and Hw).

Label	Deuteration	KLL ₂₂ + water		KLL ₂₂ + urea + water	
		μA	DCS lvl.	μA	DCS lvl.
I	100% H	200.2	101.6	212.9	101.3
II	64% H 36% D	160.7	101.7	180.0	100.1
III	50% H 50% D	163.4	101.7	171.0	90.2
IV	36% H 64% D	160.2	102.7	180.8	91.7
V	100% D	170.0	106.2	167.8	106.8

Neutron diffraction data for isotopically substituted samples of KLL₁₅ in water or water/methanol (Table 5.1) and KLL₂₂ in water with, and without, urea (Table 5.2), were collected on the NIMROD diffractometer at the ISIS Facility (STFC, UK) for ≈ 3 hours per sample at 298 K and 1 atm pressure and as described in Section 3.2.2. For each data set, the measured scattering levels were within 10% of the expected theoretical values (DCS levels in Tables 5.1-5.2). The neutron scattering data were corrected using the GudrunN programme (Section 2.2.7.3). Due to the limitations of the inelastic scattering correction provided by GudrunN the neutron diffraction data had to be truncated for analysis using EPSR, discarding all of the diffraction data below $0.1 \text{ \AA}^{-1}$, which has been analysed separately as small angle scattering in Section 5.3.4.

5.2.4 Empirical Potential Structure Refinement and ANGULA

For the two KLL₁₅ simulations, the EPSR modeling box contained 5 KLL₁₅ peptides, 25 Cl⁻ and 5000 water molecules at the measured density of ($\rho = 1.024 \text{ g/ml} = 0.1049 \text{ atoms}\cdot\text{\AA}^{-3}$, $w = 9.2\%$ peptide) or 5 KLL₁₅ peptides, 25 Cl⁻, 1665 methanol and 3335 water molecules ($\rho = 0.973 \text{ g/ml} = 0.1023 \text{ atoms}\cdot\text{\AA}^{-3}$, $w = 7.4\%$ peptide) at a temperature of 298 K. The two KLL₂₂ simulations contained 2 KLL₂₂ peptides, 14 Cl⁻ and 6000 water molecules ($\rho = 0.997 \text{ g/ml} = 0.0997 \text{ atoms}\cdot\text{\AA}^{-3}$, $w = 4.7\%$ peptide). The urea containing simulation also contained 528 urea molecules ($\rho = 1.065 \text{ g/ml} = 0.1019 \text{ atoms}\cdot\text{\AA}^{-3}$, $w = 3.6\%$ peptide), in addition to the peptide, chloride and water molecules.

Starting potentials for water molecules were taken from the SPC/E [151] water model and parameters for the peptide molecules from the CHARMM forcefield [148, 206, 207, 208, 209] and modified in order to adjust for different atomic labeling in EPSR. These parameter are shown in Table 5.3, the labelling in Figure 5.2. All peptide starting conformations were helical. For each simulation 5000 snapshots were analysed using ANGULA, without adding coordinate systems or pseudo-atoms, as this exceeded the box-size which is currently required for ANGULA, yielding further analysis unpracticable.

5.3 Results

5.3.1 NMR Spectroscopy and Elemental Analysis

Table 5.4 shows the results of the elemental analysis compared with the theoretical elemental mass percentages of pure KLL₁₅·5HCl and KLL₂₂·7HCl. Although the chloride and carbon contents are slightly lower than expected, both peptides have approximately the expected elemental concentrations which is also confirmed by the measured DCS levels in Tables 5.1-5.2.

The highest resolution for NMR spectra of KLL₁₅ and KLL₂₂ was achieved using pure methanol as a solvent. These 1D ¹H NMR spectra of 50 μ M KLL₁₅ and KLL₂₂ in methanol are shown in Figure 5.3. Due to the redundancy of the amino acid sequence,

Table 5.3: The parameters used in the EPSR simulations of the KLL₁₅ and KLL₂₂ peptides, water, methanol and urea. The first 24 atom types are on the KLL peptides. For methanol, urea and water the second letter of the atom label (m, u, w) indicates the molecule to which the atom belongs, where m is for methanol, u is for urea and w is for water.

Atom Label	ϵ / kJ mol ⁻¹	σ / Å	q / e
H0	0.10042	2.3876	0.090
C0	0.32635	3.6349	-0.270
Cox	0.46024	3.5636	0.510
Ol	0.50208	3.0291	-0.510
Nl	0.83680	3.2963	-0.470
Ok	0.50208	3.0291	-0.510
Nk	0.83680	3.2963	-0.470
Hnx	0.19246	0.4001	0.310
Cka	0.13389	3.5636	0.070
Hka	0.09205	2.3520	0.090
Ck	0.23430	3.5814	-0.180
Hk	0.14226	2.3876	0.090
Cke	0.23430	3.5814	0.210
Hke	0.14226	2.3876	0.050
Nk	0.83680	3.2963	-0.300
H3n	0.19246	0.4001	0.330
Cla	0.13389	3.5636	0.070
Hla	0.09205	2.3520	0.090
Clb	0.23430	3.5814	-0.180
Hlb	0.14226	2.3876	0.090
Clg	0.13389	3.5636	-0.090
Hlg	0.18828	2.3876	0.090
Cld	0.32635	3.6349	-0.270
Hld	0.10042	2.3876	0.090
Nt	0.83680	3.2963	-0.620
Ht	0.19246	0.4001	0.310
Cu	0.29288	3.5636	0.600
Nu	0.83680	3.2963	-0.690
Hu	0.19246	0.4000	0.340
Ou	0.50208	3.0291	-0.580
Cm	0.32635	3.6527	-0.040
Hm	0.10042	2.3876	0.090
Om	0.80375	3.1360	-0.650
Hmo	0.19246	0.4000	0.420
Ow	0.65000	3.1660	-0.8476
Hw	0.00000	0.0000	0.4238

Table 5.4: Elemental analysis of KLL₁₅ and KLL₂₂. %_m (Analysis) is the result of the elemental analysis performed by MEDAC Ltd. %_m (theor.) corresponds to the expected mass percentages of KLL₁₅·5Cl[−] and KLL₂₂·7Cl[−].

Element	KLL ₁₅		KLL ₂₂	
	% _m (Analysis)	% _m (theor.)	% _m (Analysis)	% _m (theor.)
C	50.53	54.87	51.28	55.33
H	8.94	9.01	9.12	9.04
N	13.15	14.61	13.40	14.45
O	15.49	12.71	14.77	12.65
Cl	7.93	8.80	7.55	8.53
F	1.00	0.00	0.67	0.00

the individual residues are difficult to separate. Within a 1D ¹H spectrum, the downfield NH peaks (7-9 ppm) are those that show the most residue-specific separation.

Figure 5.4 shows the downfield 1D ¹H NMR spectra of KLL₁₅ (from 6.8 to 9.2 ppm) at various methanol:water ratios. The resolution decreases with increasing water content until individual residue peaks cannot no longer be assigned in the pure water solvent. All -NH peaks shift, to varying degrees, when the solvent becomes less apolar with decreasing methanol content. Most notably the L₃ -NH peak shifts downfield from 7.45 in pure methanol to 7.75 ppm in 30% methanol, while K₈ and K₁₂ shift upfield from 8.05 to 7.85 ppm and 7.9 to 7.7 ppm, respectively. This implies that the chemical environment of these residues is altered, either due to a change in the average peptide conformation or the residue specific solvation or combination of both.

5.3.2 CD Spectroscopy

Figure 5.5 shows example CD spectra of KLL₁₅ and KLL₂₂ in varied methanol:water mixtures. The recorded spectra were analysed using CONTIN/LL [276] to obtain an approximated helical content of these samples and this analysis is shown in Figure 5.6. While the shorter KLL₁₅ is almost completely non-helical in pure water, the longer KLL₂₂ is more than 50% helical in all water:methanol mixtures; yet both peptides become more helical with increasing methanol content in the solvent mixture.

Figure 5.7 shows example CD spectra of KLL₂₂ in pure water with varied urea concentrations and the respective helical content from CONTIN/LL. Under the experimen-

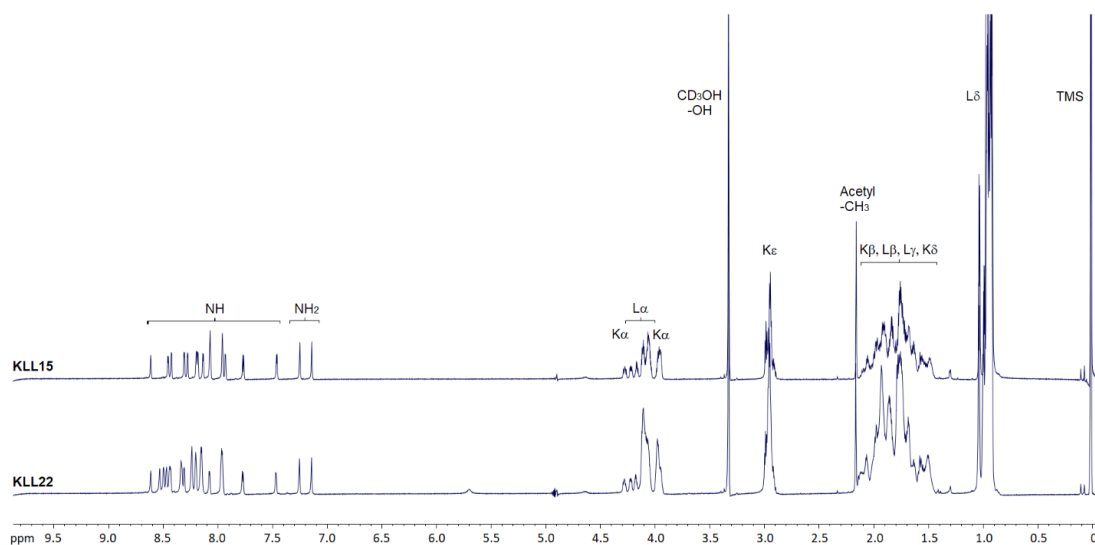


Figure 5.3: ^1H 1D NMR spectra of 50 μM KLL₁₅ and 50 μM KLL₂₂ in methanol with peak assignment.

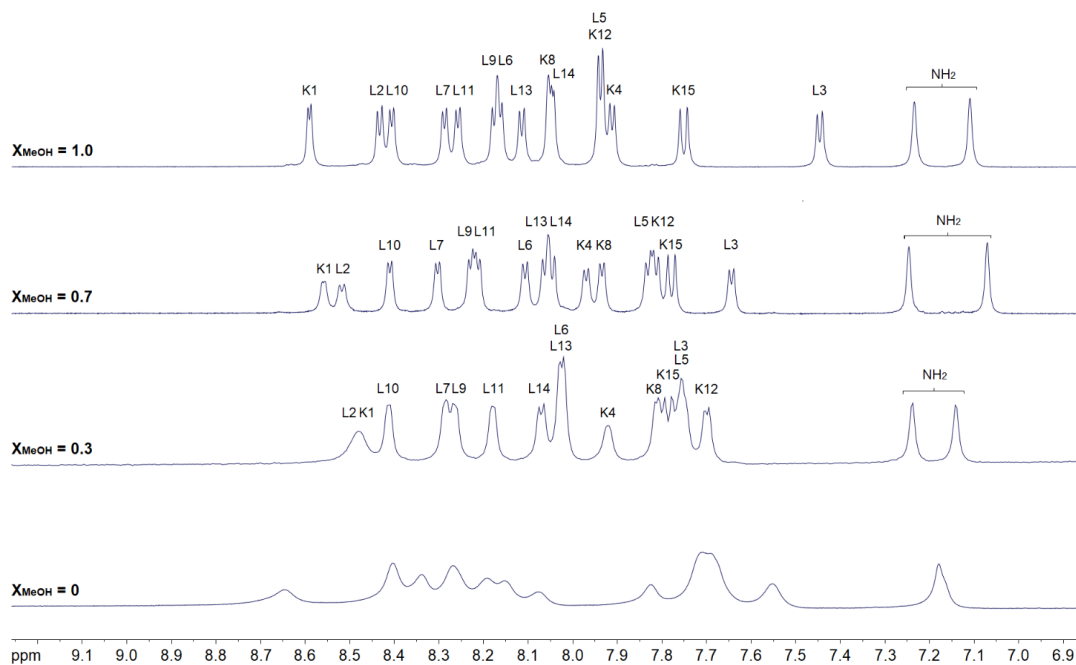


Figure 5.4: Downfield ^1H 1D NMR spectra of 50 μM KLL₁₅ in methanol:water mixtures with peak assignment, χ_{MeOH} refers to the molar fraction of methanol in the solvent mixture.

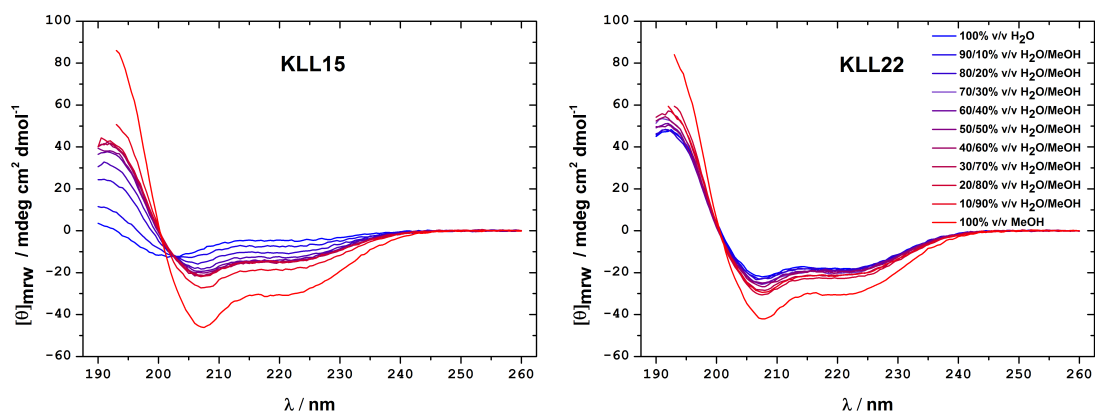


Figure 5.5: CD spectra of KLL₁₅ (50 μ M) and KLL₂₂ (35 μ M) in methanol:water mixtures (pathlength = 0.1 cm) as ellipticity per mean residue weight ($[\theta]_{\text{mwr}}$, cf. Section 2.1.2.2).

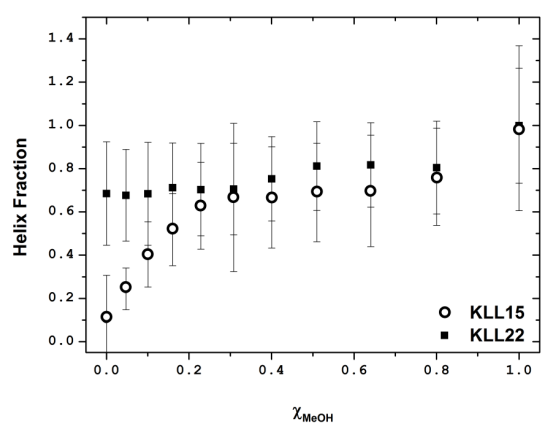


Figure 5.6: Helix fraction of KLL₁₅ and KLL₂₂ in methanol:water mixtures determined by CD spectra (Figure 5.5) using CONTIN/LL.

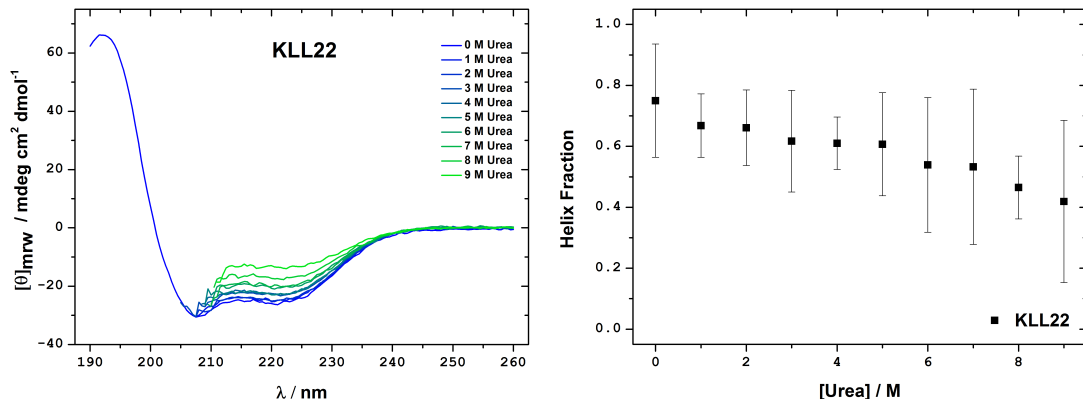


Figure 5.7: Left: CD spectra of KLL₂₂ (35 μ M) in water:urea mixtures (pathlength = 0.1 cm) as ellipticity per mean residue weight ($[\theta]_{\text{mrw}}$). Right: Helix fraction of KLL₂₂ in urea:water mixtures determined by CD spectra (left) using CONTIN/LL.

tal conditions it was not possible to obtain a complete CD spectrum for urea-containing samples below 210 nm because of the UV-absorbance of urea. Therefore, the estimated helical content for these samples might be less accurate than for the whole CD spectra in Figure 5.5. However, urea lowers the KLL₂₂ helical content in aqueous solution significantly from $\approx 75\%$ in pure water to $\approx 42\%$ in aqueous 9 M urea solution.

5.3.3 Diffraction Data and EPSR Fits

Measured neutron diffraction data ($F(Q)$), EPSR fits to these data and difference curves between data and simulation are shown in reciprocal space in Figures 5.8-5.9 (left, grey circles). The EPSR fits to the neutron data are generally good, with some small differences at lower Q values below $2 \text{ \AA}^{-1}$, especially for samples with a higher hydrogen content, due to the difficulty in correcting for inelastic scattering within this region. At larger Q values (above $2 \text{ \AA}^{-1}$) the light coloured difference curves are almost flat and the EPSR fits are good. The Fourier transformations of these data and EPSR fits in real space are also shown in Figures 5.8-5.9 (right), where an overall similar level of agreement between the experimental and simulated data is observed as in reciprocal space. Only samples with a medium hydrogen content (36-50%) differ slightly between the fourier transformed experimental diffraction data and the simulated from EPSR.

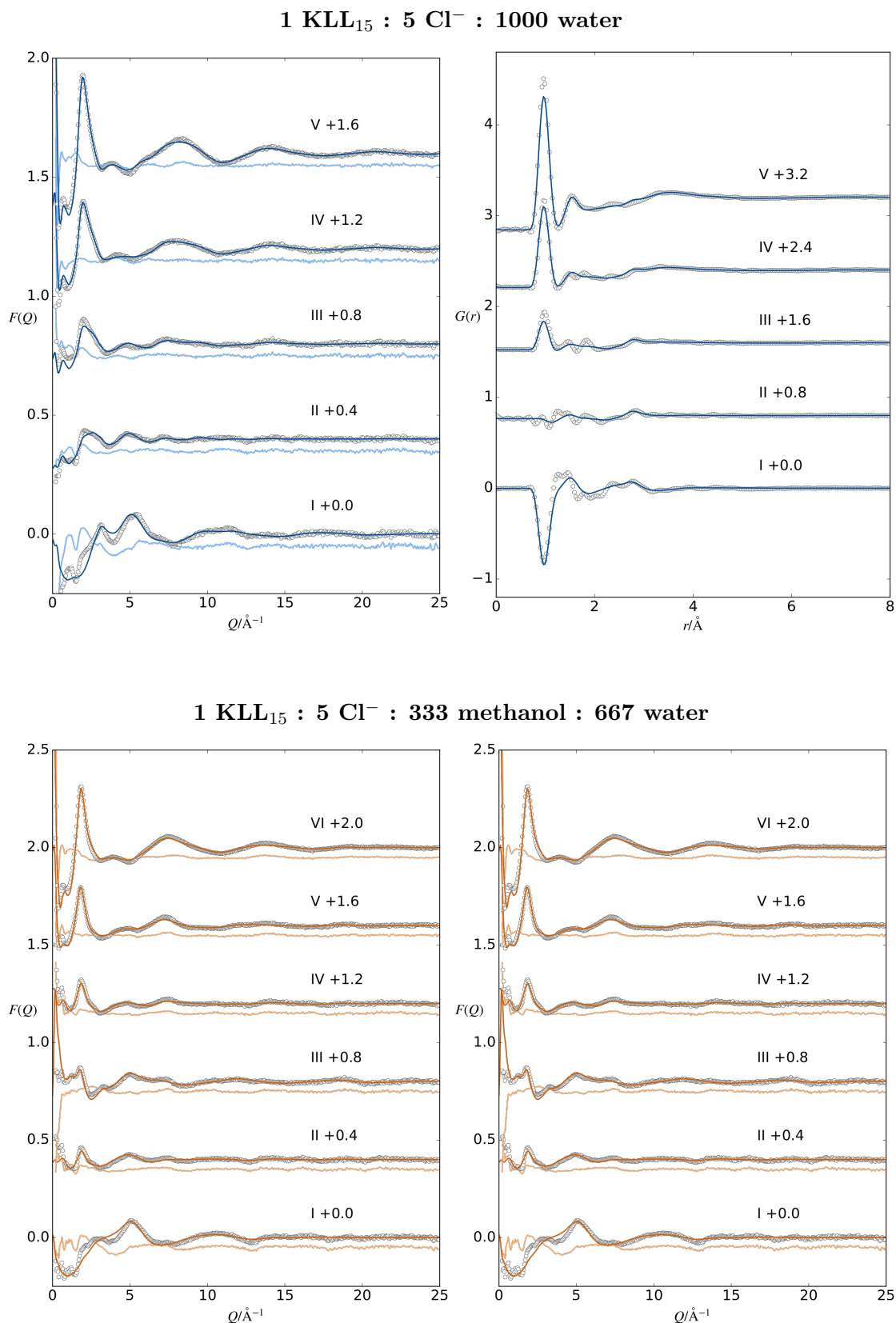


Figure 5.8: Measured diffraction data (grey dots) for KLL₁₅ in pure water (top, blue) or in methanol:water (bottom, orange) compared with the **EPSR** fits (dark coloured lines) in reciprocal space ($F(Q)$, left) and in real space ($G(r)$, right). The difference between simulation and data are shown as a light coloured lines. The data and corresponding fits have been shifted for clarity. The sample labels correspond to the level of deuteration in listed in Table 5.1.

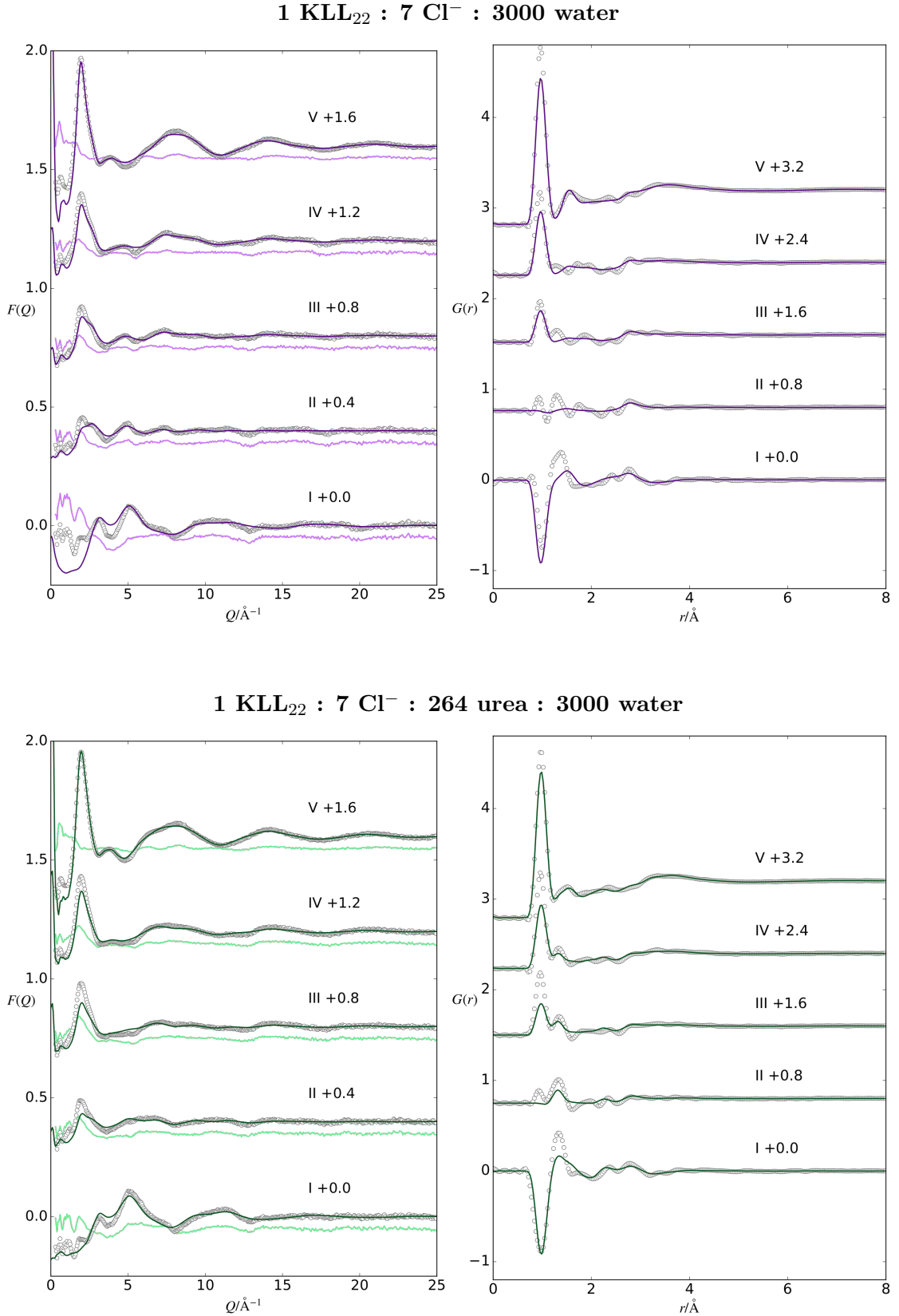


Figure 5.9: Measured diffraction data (grey dots) for KLL₂₂ in pure water (top, purple) or in aqueous urea solution (bottom, green) compared with the **EPSR** fits (dark coloured lines) in reciprocal space ($F(Q)$, left) and in real space ($G(r)$, right). The difference between simulation and data are shown as a light coloured lines. The data and corresponding fits have been shifted for clarity. The sample labels correspond to the level of deuteration in listed in Table

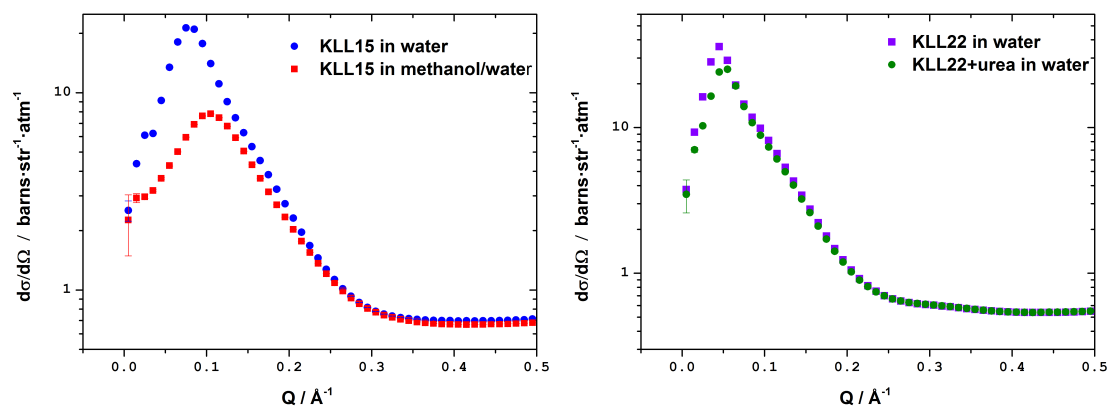


Figure 5.10: Differential scattering section of 100 % D samples

5.3.4 Small Angle Scattering

As the neutron scattering data below $0.1 \text{ \AA}^{-1}$ was omitted for the EPSR simulation in order to adequately correct the inelastic scattering background. This small angle scattering data (for $Q < 0.5 \text{ \AA}^{-1}$) of the deuterated peptide samples is analysed separately (Figure 5.10 shows this small angle scattering). This data was analysed using PRIMUS to determine the radius gyration (R_g) *via* Guinier approximation and GNOM to indirect Fourier transform the data to retrieve distance distribution functions (Figure 5.11).

The radii of gyration in Table 5.5 indicate that KLL₁₅ is, on average, in a more compact conformation compared to the peptide in pure water. A purely helical KLL₁₅ would, theoretically, have an R_g value of $9.7 \text{ \AA}$; therefore it is possible that the lower R_g value in methanol/water is indicative of a more helical peptide conformation. However, a similar effect of a more compact average peptide structure is observed for KLL₂₂ when urea is added to the aqueous peptide solution. A purely helical KLL₂₂ has a theoretical radius of gyration of $\approx 11.5 \text{ \AA}$. The addition of urea lowers the KLL₂₂ R_g from 16.8 – $11.7 \text{ \AA}$. This effect on the average peptide conformations is also observed in the distance distributions ($P(r)$) in Figure 5.11. The addition of urea or methanol to the aqueous peptide solutions of KLL₂₂ and KLL₁₅ respectively moves the $p(r)$ functions towards smaller distances and a more globular than helical structure.

Table 5.5: Radii of gyration determined by SANS data and the Guinier approximation using PRIMUS.

Sample	$R_g / \text{\AA}$
KLL ₁₅ in H ₂ O	14.8 ± 0.1
KLL ₁₅ in methanol/H ₂ O	12.6 ± 0.3
KLL ₂₂ in H ₂ O	16.8 ± 0.1
KLL ₂₂ +urea in H ₂ O	11.7 ± 0.1

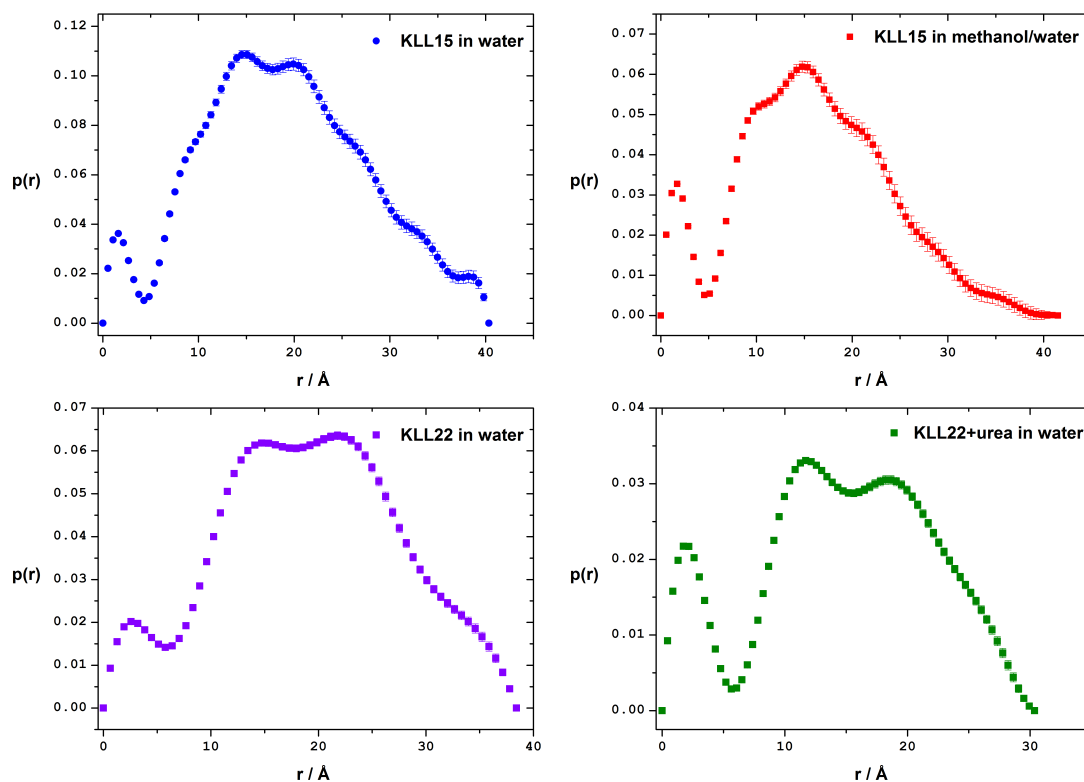


Figure 5.11: Distance distribution functions from SANS data in Figure 5.10

5.3.5 Peptide Conformations

The peptide starting conformations were helical throughout all of the performed simulations. These α -helices were generated using PyMOL [277] and had peptide backbone dihedral angles of all $\phi \approx -55^\circ$ and $\psi \approx -45^\circ$. Using ANGULA the final distributions of peptide dihedral angles were sampled over all peptides and residues. The associated ϕ/ψ -residue pairs were correlated in bivariate heat-maps and superimposed to generate Ramachandram plots for each peptide under the given conditions. These plots are shown in Figure 5.12 for KLL₁₅ in pure water (left, blue) and in methanol:water mixture (right, orange) and in Figure 5.13 for KLL₂₂ in pure water (left, purple) and in water:urea mixture (right, green). A comparison plot indicating secondary structure localisations within a Ramachandran plot is in Section 1.1.2.

Figure 5.12 shows that KLL₁₅ retains its helical starting conformation in purely aqueous solvent or the more apolar (2:1) water:methanol mixture. The corresponding ϕ/ψ -correlation is only slightly more localised at $\approx -50^\circ / -50^\circ$ (except for the N- and C-terminal residues), while in pure water the ϕ/ψ -correlation is more varied between -100 and 0° for both ϕ and ψ . This is not in accordance with the CD data where KLL₁₅ was mostly non-helical in pure water.

The Ramachandran plots of KLL₂₂ in Figure 5.12 show a less marked effect upon the addition of urea to the KLL₂₂ conformation. For both simulations with and without urea the ϕ/ψ -correlation has broader peak in an α -helical area, at $\approx -50^\circ / -50^\circ$, although with urea this peak is slightly less localised for both ϕ and ψ . Similar to KLL₁₅, the ϕ/ψ -pairs for the N- and C-terminal residues are not in the α -helical area of the plot and are unchanged by the addition of urea. Probably because these residues are less sterically hindered by the small acetyl/amide caps than residues within the peptide chain.

5.3.5.1 Residues most affected by a change in solvation

As aforementioned, the Ramachandran plots in Figures 5.12-5.13 are superpositions of residue-specific two-dimensional ϕ / ψ -correlations. These residue-specific heatmaps

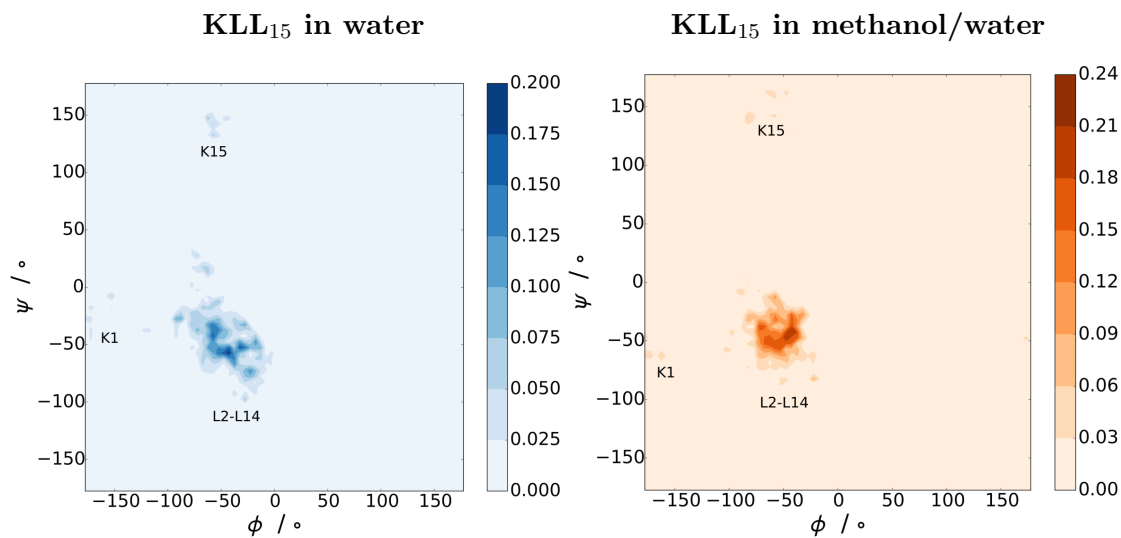


Figure 5.12: Ramachandran plots of all dihedral angle pairs ϕ / ψ for KLL₁₅ in pure water (left, blue) and methanol/water (right, orange).

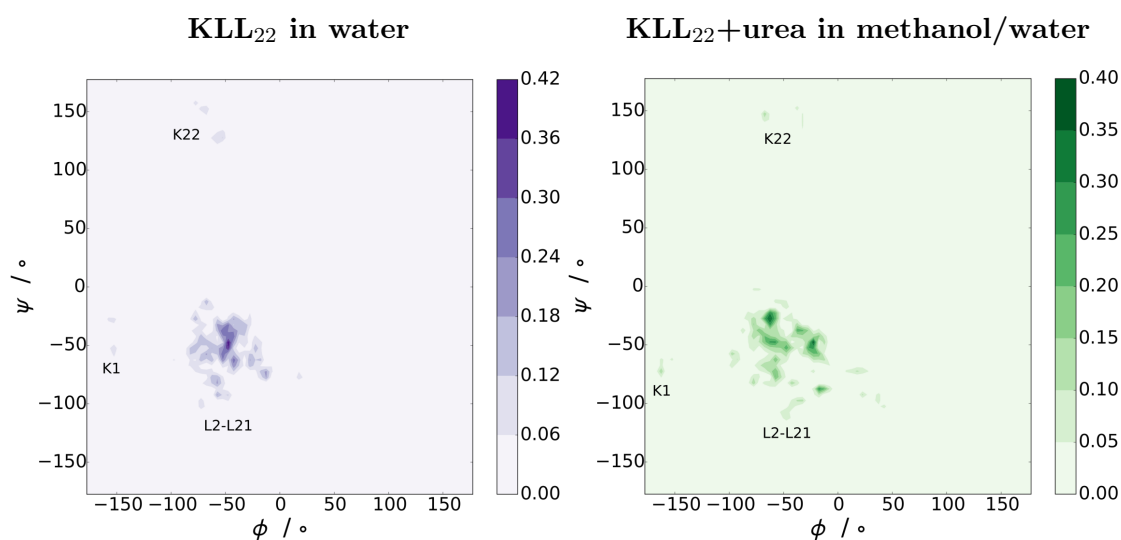


Figure 5.13: Ramachandran plots of all dihedral angle pairs ϕ / ψ for KLL₂₂ in pure water (left, purple) and aqueous urea solution (right, green).

were compared to determine the residues that were most affected by a change in solvent or urea concentration. For KLL₁₅, L₉ was found to be the most variant in terms of the ϕ / ψ angle correlation for all peptides in the simulation (Figure 5.14). For KLL₂₂, K₁₁ in all peptides was most affected by urea addition (Figure 5.15). The solvation and small molecule interactions of these residues will be analysed separately in the following Sections 5.3.6-5.3.7.

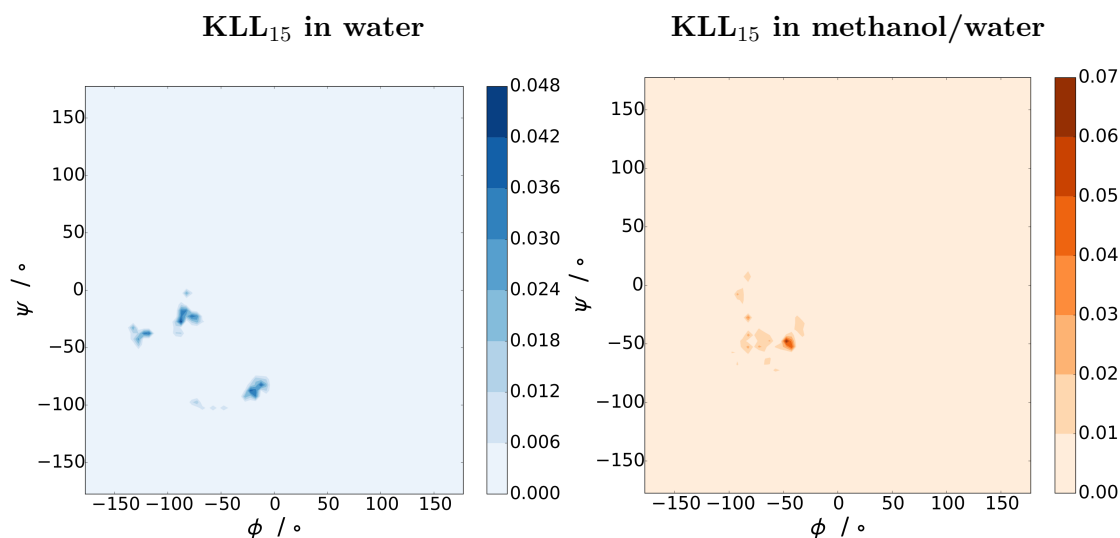


Figure 5.14: Ramachandran plots of the dihedral angle pair ϕ_{L_9} / ψ_{L_9} for KLL₁₅ in pure water (left, blue) and methanol/water (right, orange).

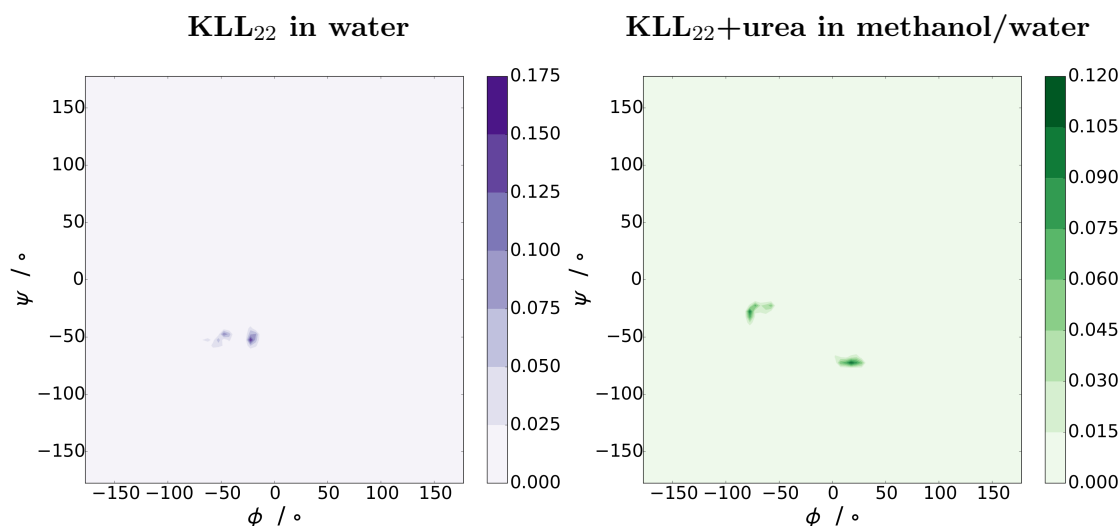


Figure 5.15: Ramachandran plots of the dihedral angle pair $\phi_{K_{11}} / \psi_{K_{11}}$ for KLL₂₂ in pure water (left, purple) and aqueous urea solution (right, green)

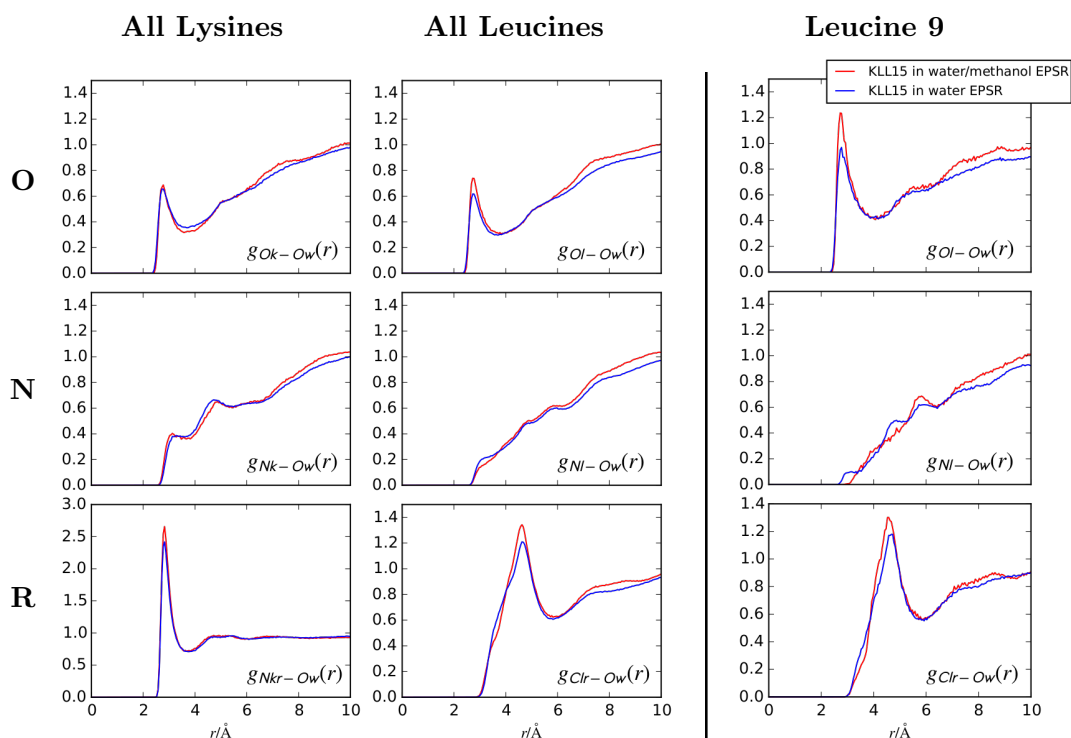


Figure 5.16: RDFs of KLL₁₅ peptide bond carbonyl oxygens (Ok and Ol), nitrogens (Nk and Nl) and residue ends (Nkr and Clr) of leucine and lysines with water oxygens (Ow).

5.3.6 KLL₁₅: Amphiphilicity

5.3.6.1 Peptide Water Interaction

Figure 5.16 shows the hydration RDFs of KLL₁₅ for all of the backbone nitrogens and oxygens and lysine and leucine side-chain atoms (Clr on leucine, Nkr on lysine, Figure 5.2). As L₉ was the most affected from a change in solvent environment, the residue specific RDFs for L₉ are also shown as a comparison. The corresponding coordination numbers are listed in Table 5.6. The change from pure water to a (1:2) methanol:water mixture lowers the water oxygen (Ow) molar fraction in the simulation by $\approx$ -51%. As the coordination numbers are normalised by the the molar fraction of the neighbouring atoms (Ow), these coordination numbers are expected to be on average 51% lower in the methanol:water KLL₁₅ simulation compared with KLL₁₅ in pure water solvent. For L₉, all sites lose hydrating water, especially the leucine backbone oxygens (Ol; $\approx$ -62% water) and nitrogens (Nl; $\approx$ -85% water) which both lose more than the expected 51% of the water oxygens around the rest of the molecule. L₉ oxygen (Ol) is more hydrated

Table 5.6: Coordination numbers for the KLL₁₅-Ow/Cl RDFs in Figures 5.16 and 5.17. Also shown are the hydration differences (HD).

	$r_{max} / \text{\AA}$	KLL ₁₅		
		in water	in water/methanol	HD / %
Ok-Ow	3.8	2.5	1.1	-56
Nk-Ow	3.5	1.2	0.7	-42
Nkr-Ow	3.7	5.0	2.6	-49
Nkr-Cl	4.0	0.4	0.5	-
Ol-Ow	3.8	2.0	0.8	-62
Nl-Ow	3.5	0.4	0.1	-85
Clr-Ow	5.9	17.8	9.1	-49
Ol-Ow (L₉)	3.8	3.0	2.0	-36
Nl-Ow (L₉)	3.5	0.2	0.0	-85
Clr-Ow (L₉)	5.9	16.3	8.3	-50

than on the average leucine residue ($\text{CN} \approx 3.0$) but also retains more water molecules in the methanol:water mixture ($\text{CN} \approx 2.0$, $\approx -36\%$ water). Overall, these results indicate that the hydrophobic leucine residues are more affected by a solvent-change and that this effect is residue specific.

5.3.6.2 Peptide Chloride Interaction

The KLL₁₅-chloride interactions are shown in the RDFs in Figure 5.17. As only the positively charged lysine side-chain nitrogens (Nkr) exhibit an observable chloride interaction, the respective coordination numbers for the Nkr-Cl RDFs are further listed in Table 5.6. Interestingly, this salt-interaction increases slightly in the more apolar methanol:water solvent from a coordination number of ≈ 0.4 in pure water to 0.5 in methanol:water, despite the molar fraction of chloride ions being lower in the methanol:water simulation.

5.3.6.3 Peptide Methanol Oxygen Interaction

The salient methanol oxygen (Om) interactions with KLL₁₅ are shown in Figure 5.18. While the backbone nitrogens (Nk and Nl) only show weak interactions with the methanol oxygen (Nk shows more contacts than Nl), the carbonyl oxygens (Ok and Ol) show small peaks at 2.8 $\text{\AA}$ without an observable difference in solvation between Ol and Ok.

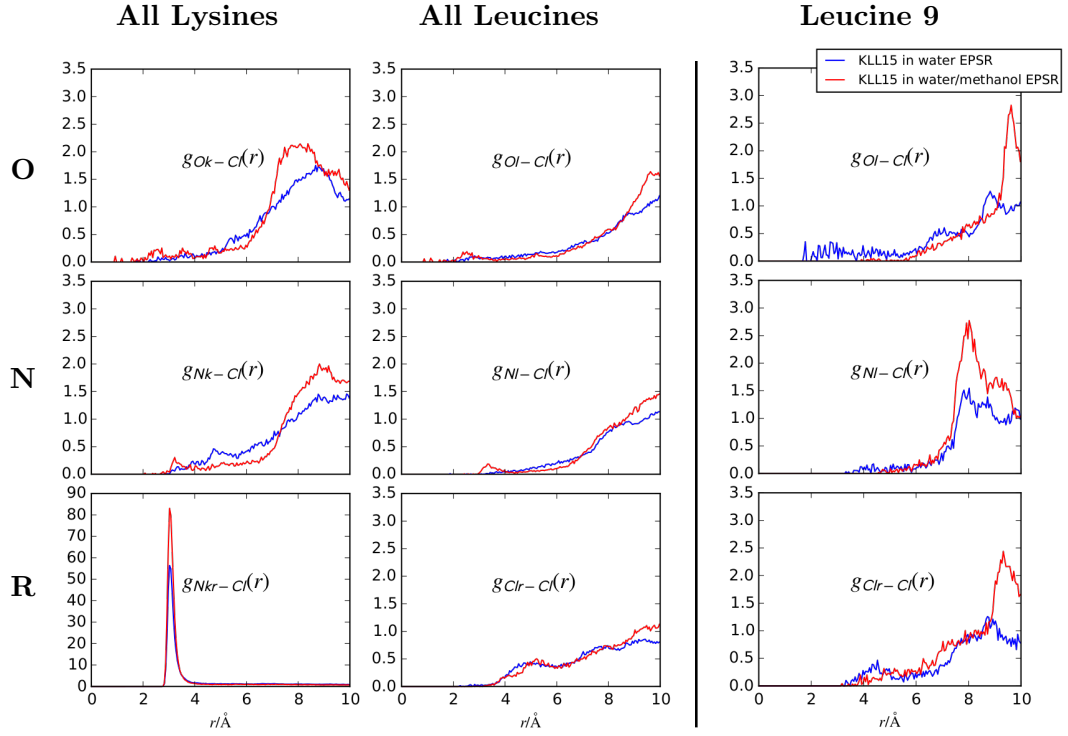


Figure 5.17: RDFs of KLL₁₅ peptide bond carbonyl oxygens (Ok and Ol), nitrogens (Nk and Nl) and residue ends (Nkr and Clr) of leucine and lysines with chloride (Cl).

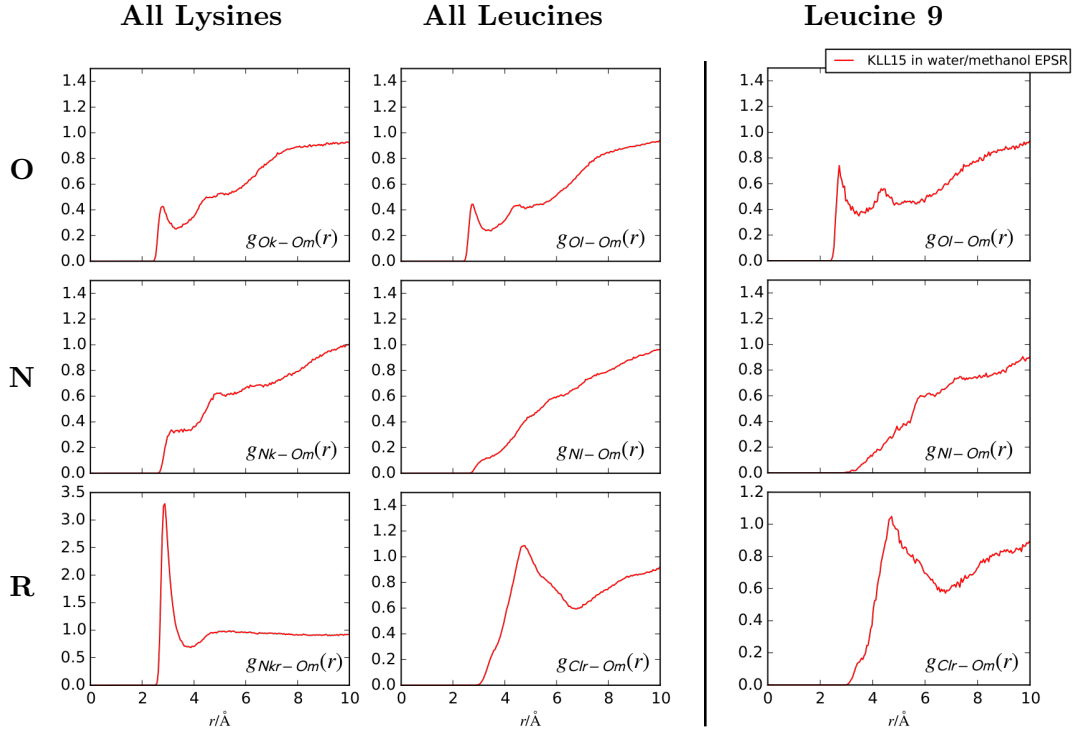


Figure 5.18: RDFs of KLL₁₅ peptide bond carbonyl oxygens (Ok and Ol), nitrogens (Nk and Nl) and residue ends (Nkr and Clr) of leucine and lysines with methanol oxygen (Om).

Table 5.7: Coordination numbers for the KLL₁₅-Om RDFs in Figure 5.18.

	$r_{max} / \text{\AA}$	KLL₁₅ in water/methanol
Ok-Om	3.3	0.2
Nk-Om	3.8	0.5
Nkr-Om	3.8	1.6
Ol-Om	3.3	0.2
Nl-Om	3.8	0.1
Clr-Om	6.7	5.7
Ol-Om (L₉)	3.3	0.3
Nl-Om (L₉)	3.8	0.2
Clr-Om (L₉)	6.7	6.1

Although the corresponding coordination numbers in Table 5.7 are small throughout ($\text{CN} \approx 0.2\text{-}0.3$), it is clear from Figure 5.18 that L₉ oxygen (Ol) on the average interacts slightly more with methanol oxygen compared with the other leucine residues.

5.3.6.4 Peptide Methanol Carbon Interaction

Figure 5.19 shows the RDFs between the backbone and central side-chain atoms of KLL₁₅ with the methanol carbon (Cm). The corresponding coordination numbers are listed in Table 5.8. From this Table it is clear that lysine backbone nitrogens interact significantly more with methanol carbon ($\text{CN} \approx 1.2$), than leucine backbone nitrogens ($\text{CN} \approx 0.3$), similar to the interaction with methanol Om. Another interesting observation is that L₉ oxygen (Ol), again, interacts more with methanol than all the other leucines.

In summary for KLL₁₅, the leucine residues are more affected by a change in solvent polarity compared with the lysines. In a more apolar solvent, leucine residues lose the most hydrating water and interact more with methanol. Some leucine residues, such as L₉, experience a larger conformational change, compared with the other residues when the solvent is changed, from a more variant ϕ/ψ -correlation in pure water to a clearly helical conformation in methanol:water. The corresponding L₉ backbone atoms interact more with water and methanol compared to other leucines.

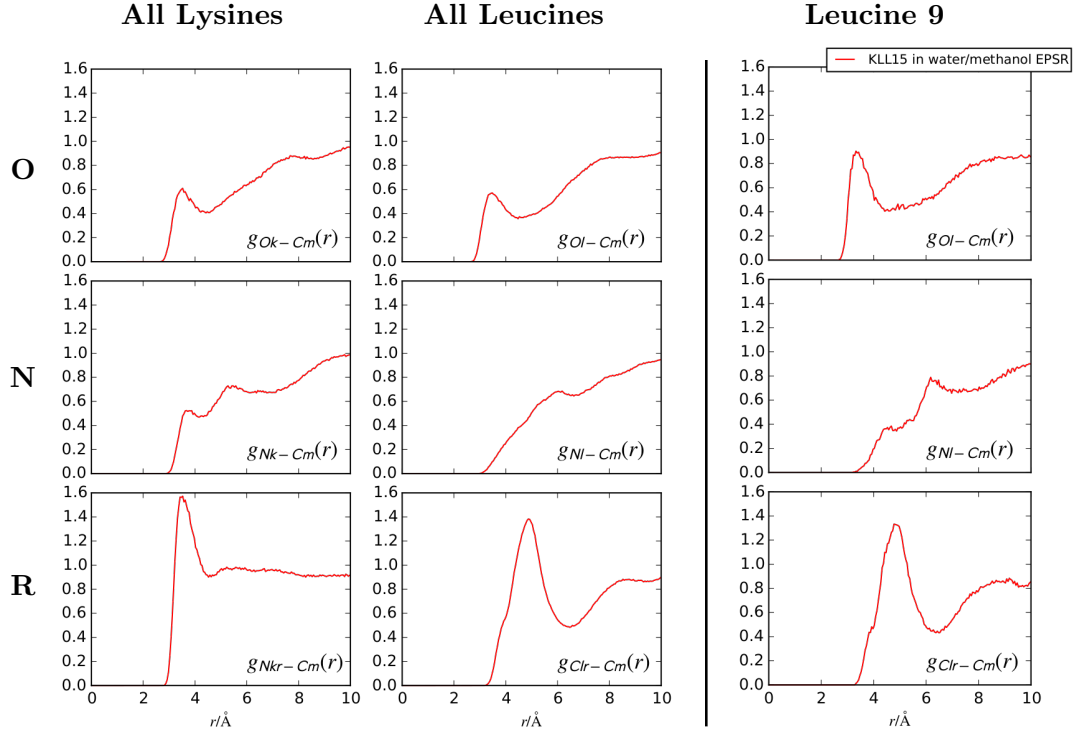


Figure 5.19: RDFs of KLL₁₅ peptide bond carbonyl oxygens (Ok and Ol), nitrogens (Nk and Nl) and residue ends (Nkr and Clr) of leucine and lysines with methanol carbon (Cm).

Table 5.8: Coordination numbers for the KLL₁₅-Cm RDFs in Figure 5.19.

	$r_{max} / \text{\AA}$	KLL ₁₅ in water/methanol
Ok-Cm	4.5	1.1
Nk-Cm	4.5	1.2
Nkr-Cm	4.6	2.5
Ol-Cm	4.5	1.0
Nl-Cm	4.5	0.3
Clr-Cm	6.5	5.3
Ol-Cm (L ₉)	4.5	1.4
Nl-Cm (L ₉)	4.5	0.4
Clr-Cm (L ₉)	6.5	5.7

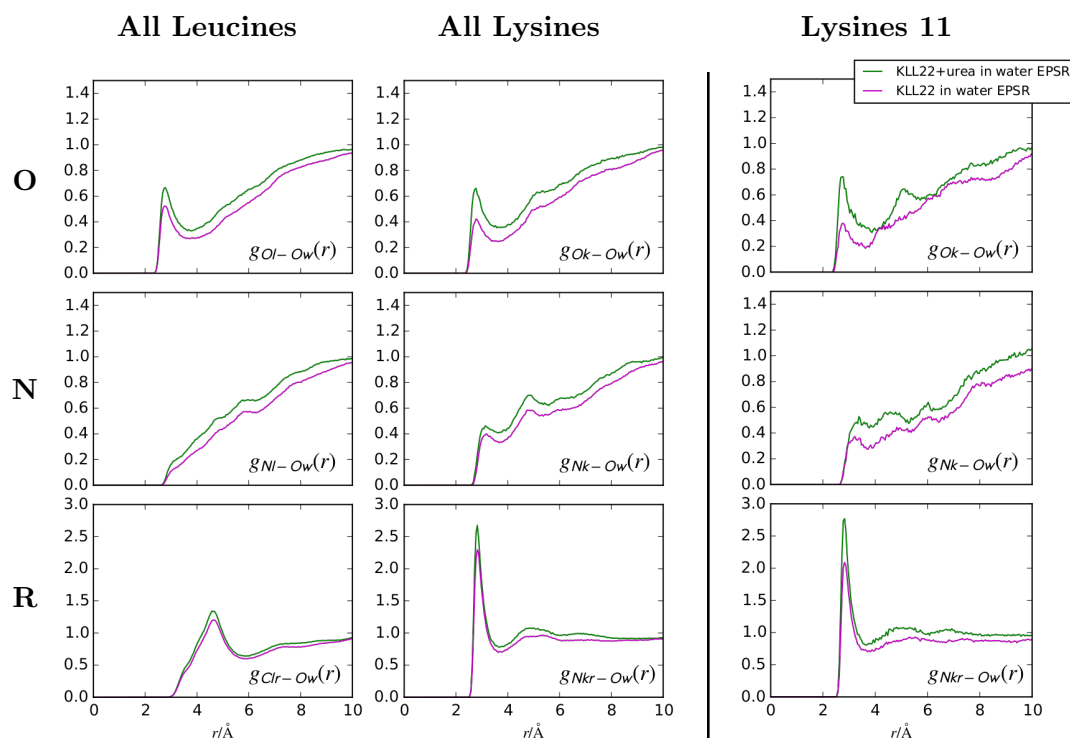


Figure 5.20: RDFs of KLL₂₂ peptide bond carbonyl oxygens (Ok and Ol), nitrogens (Nk and Nl) and residue ends (Nkr and Clr) of leucine and lysines with water oxygens (Ow).

5.3.7 KLL₂₂: Urea Influence

5.3.7.1 Peptide Water Interaction

The RDFs for the site-specific KLL₂₂ hydration are shown in Figure 5.20, including the residue specific hydration of K₁₁, which was found to be the most conformationally affected by urea addition (Figure 5.15). The corresponding coordination numbers for these functions are listed in Table 5.9. Due to the change in molar concentrations when urea is added, the molar fraction of Ow in the solution is lowered by $\approx 18.3\%$. This hydration difference between the urea-free and urea-containing solution would be expected for every water interaction if no site specificity of water solvation was observed. However, this level of hydration loss is not observed for KLL₂₂ sites. The hydration of the average leucine residues is, despite higher RDFs in Figure 5.20 for the urea containing simulation, largely unaffected by the addition of urea. The hydration of lysine residue backbone atoms Nk and Ok, on the other hand, increases by $\approx 30\%$ while the lysine side-chain nitrogens (Nkr) loses water contacts upon urea addition. The hydration of

Table 5.9: Coordination numbers for the KLL₂₂-Ow/Cl RDFs in Figures 5.20 and 5.21.

	$r_{max} / \text{\AA}$	KLL ₂₂		HD / %
		in water	+urea in water	
Ol-Ow	3.8	2.0	1.9	-6
Nl-Ow	3.5	0.2	0.2	-1
Clr-Ow	5.9	16.6	15.3	-8
Ok-Ow	3.8	1.5	1.9	+28
Nk-Ow	3.5	1.5	2.0	+31
Nkr-Ow	3.7	4.8	4.5	-7
Nkr-Cl	4.0	0.4	0.2	-
Ok-Ow (K₁₁)	3.8	1.2	1.9	+59
Nk-Ow (K₁₁)	3.5	1.3	1.9	+47
Nkr-Ow (K₁₁)	3.7	4.5	4.6	+2
Nkr-Cl (K₁₁)	4.0	0.7	0.1	-

K₁₁ in the presence of urea is similar to the average lysine. Without urea, K₁₁ is less hydrated compared to the other lysines.

5.3.7.2 Peptide Chloride Interaction

Similar to the chloride interaction of KLL₁₅, only the lysine side-chain nitrogen (Nkr) shows a significant nearest neighbour peak in the respective RDFs in Figure 5.21. Longer range chloride interactions with the peptide, for example of lysine oxygens (Ok), are probably due to this Nkr-Cl contact. This charged Nkr-Cl interaction (Table 5.9) is significantly lowered when urea is added (CN $\approx$ 0.4 to CN $\approx$ 0.2) to the solution. This effect is even more visible for K₁₁ where the coordination is lowered from a CN $\approx$ 0.7 in pure water to CN $\approx$ 0.1 in 4 M urea. Not only is K₁₁ the only residue that interacts with chloride more than the average lysine, but this K₁₁-Cl interaction is also more affected by the addition of urea compared with the other lysines on KLL₂₂.

5.3.7.3 Peptide Urea Interaction

Figure 5.22 shows the urea-peptide interaction for KLL₂₂ with the central urea carbon atom (Cu). For all of the backbone sites on KLL₂₂ there is almost no interaction with urea molecules; this is also visible in the coordination numbers listed in Table 5.10. Only the charged lysine side chains and the apolar leucine side chain interact significantly with

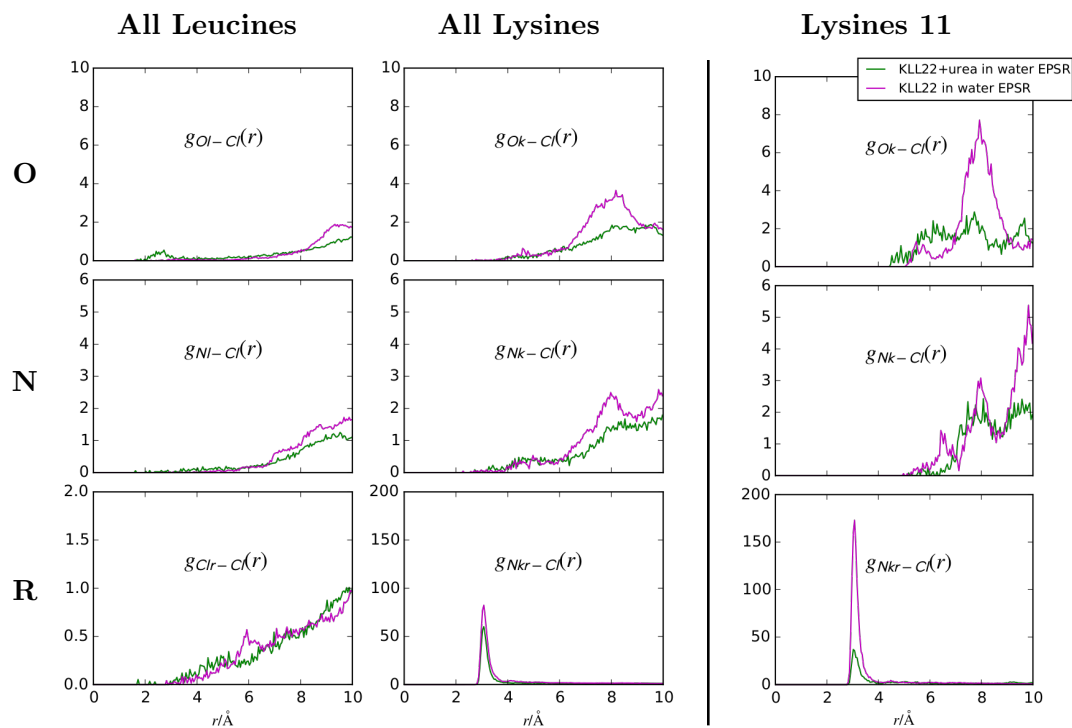


Figure 5.21: RDFs of KLL₂₂ peptide bond carbonyl oxygens (Ok and Ol), nitrogens (Nk and Nl) and residue ends (Nkr and Clr) of leucine and lysines with chloride (Cl).

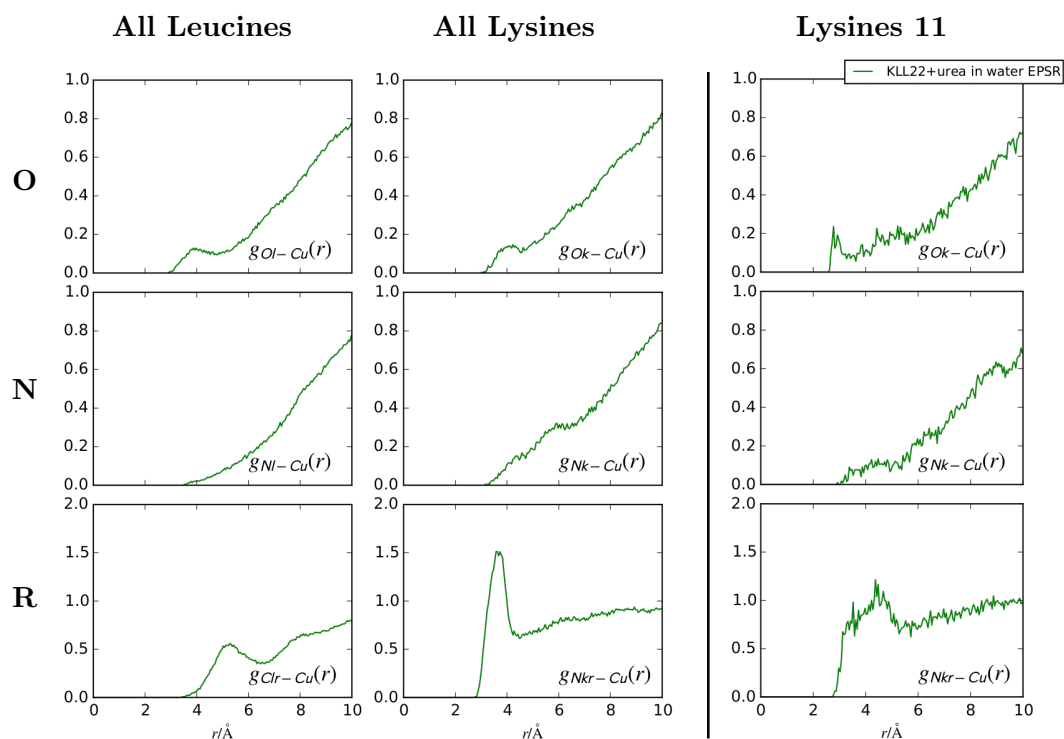


Figure 5.22: RDFs of KLL₂₂ peptide bond carbonyl oxygens (Ok and Ol), nitrogens (Nk and Nl) and residue ends (Nkr and Clr) of leucine and lysines with urea carbon (Cu).

Table 5.10: Coordination numbers for the KLL₂₂-Cu RDFs in Figure 5.22.

	$r_{max} / \text{\AA}$	KLL₂₂ Leucines +urea in water
Ol-Cu	4.8	0.1
Nl-Cu	4.8	<0.1
Clr-Cu	6.5	0.7
Ok-Cu	4.8	0.1
Nk-Cu	4.8	0.2
Nkr-Cu	4.4	0.5
Ok-Cu (K₁₁)	4.8	0.1
Nk-Cu (K₁₁)	4.8	0.1
Nkr-Cu (K₁₁)	4.4	0.4

the urea molecules in the solution. Interestingly, K₁₁ does not interact more with urea than other lysine residues so a replacement of chloride ions with urea appears unlikely.

To determine if urea or water hydrogen bonds with KLL₂₂ are preferred, a preference ratio (PR, Section 3.2.6) was determined by weighting the respective coordination number ratios by the molar water/urea ratio. Similar to the urea preference ratio in Chapters 3 and 4 a ratio above 1.0 indicates a urea preference over water and a ratio below 1.0 indicates a water preference over urea at the specific peptide site. Figure 5.23 shows the hydrogen bond RDFs from water or urea with any peptide backbone hydrogen (Hnx) and oxygen (Ox) and the side-chain hydrogens of leucine (Hld) and lysine (Hnk). The respective coordination numbers and preference ratios are listed in Table 5.11.

The KLL₂₂ peptide backbone oxygens (Ox) form hydrogen bonds with water (*via* Hw) and urea (*via* Hu) visible in a small RDF peak at $\approx 1.8 \text{\AA}$. Similarly, the peptide backbone hydrogens (Hnx) form hydrogen bonds with water and urea oxygens (Ow and Ou) at $\approx 2 \text{\AA}$. Although both hydration and urea interaction peaks are small, the preference ratio between urea and water to form hydrogen bonds with the peptide backbone are both at $\text{PR} \approx 0.2$ indicating that water is strongly preferred at these sites. The apolar leucine side-chain hydrogens (Hld) do not hydrogen bond towards water or urea, yet there is a small interaction peak with Ow and Ou visible at $\approx 2.5 \text{\AA}$. However, this apolar interaction is also preferentially with water over urea, as indicated by $\text{PR} \approx 0.3$. Only the lysine side-chain hydrogens (Hnk) interact with water and urea (at $\approx 1.8 \text{\AA}$) without a clear preference ($\text{PR} \approx 1.0$). These results are, overall, not in agreement with

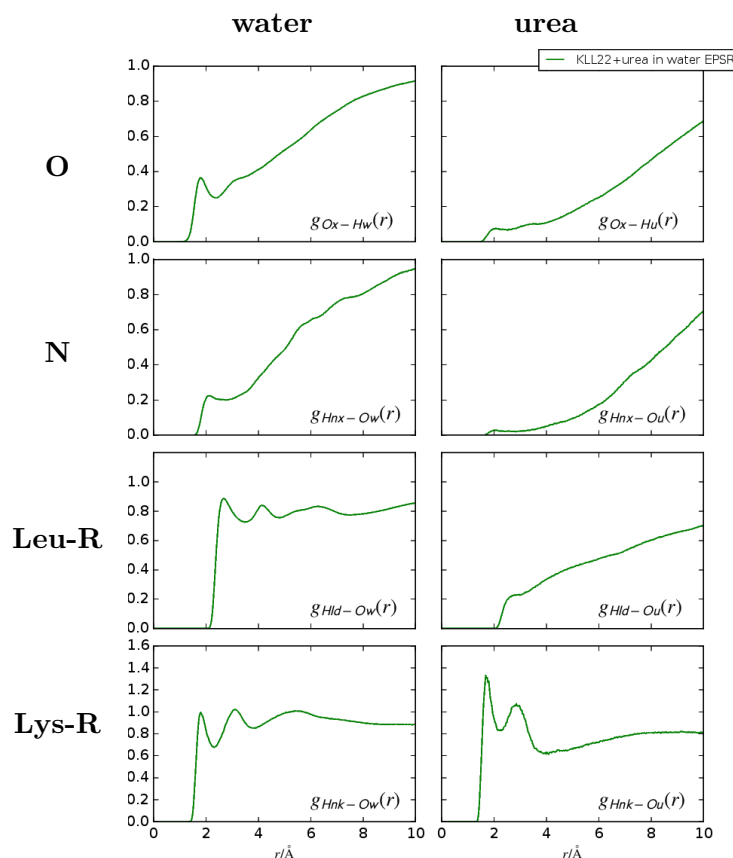


Figure 5.23: Hydrogen bond RDFs of KLL₂₂ peptide bond carbonyl oxygens (Ox), backbone hydrogens (Hnx) and side-chain lysine (Hnk) and leucine (Hld) hydrogens.

the GPG-NH₂-urea interaction presented in Chapter 3, where urea was found to be preferred over water in peptide backbone nitrogen contacts. However, a similar effect of lowered ion interactions upon the addition of urea has been observed for GPG-NH₂-chloride contacts (Chapter 3) and KGPGK-TFA contacts (Chapter 4).

5.4 Conclusions

From the results presented in this chapter, conclusions can be drawn regarding the influence of an amphiphilic solvent and urea on α -helices and amphiphilic helices, in particular.

CD spectra on KLL₁₅ in methanol:water solutions showed that this peptide is almost non-helical in pure water and over 60% helical in 30% methanol:70%water and 100% helical in pure methanol. This solvent dependent folding/unfolding is also reflected in the NMR spectroscopy data where, due to the lack of order, a clear peak assignment in pure

Table 5.11: Coordination numbers for the KLL₂₂ hydrogen bond RDFs in Figure 5.23 including a preference ratio of urea vs. water contacts.

	$r_{max} / \text{\AA}$	KLL ₂₂ Leucines +urea in water	PR
Ox-Hw	2.5	0.8	-
Hnx-Ow	2.5	0.4	-
Hld-Ow	3.5	0.7	-
Hnk-Ow	2.5	1.2	-
Ox-Hu	2.5	<0.1	0.2
Hnx-Ou	2.5	<0.1	0.2
Hld-Ou	3.5	<0.1	0.3
Hnk-Ou	2.5	0.2	1.0

water for any peptide was impossible suggesting that the peptide was largely unfolded. However, the -NH peak assignments for KLL₁₅ in varying methanol concentrations also indicated that the proton environment between 30%-100% methanol changes depending on the peptide residue suggesting a residue dependent folding.

For all KLL peptides EPSR simulations, helical peptides were used as the initial conformations. Although CD and NMR data suggests that KLL₁₅ (at a lower μM concentration) is mostly unhelical in pure water, the Ramachandran plot of KLL₁₅ in pure water, derived from the EPSR simulation (at a higher mM KLL₁₅ concentration), still shows the ϕ/ψ -pairs mostly in the helical area of the Ramachandran plot (cf. Figure 1.5). This could either be an artefact of the initial KLL₁₅ conformation or a peptide concentration effect, where the peptide itself, at higher concentrations, contributes to a more apolar peptide environment. However, the KLL₁₅ conformation in 30% methanol shows, as expected, a more ordered α -helical structure. For future work, it would be useful to either run the EPSR simulation for longer time, use unfolded KLL₁₅ peptides as starting conformations, run additional MD simulations or to determine the concentration effect of the peptide on its conformation by a method that can access this peptide concentration between 50 μM and 38 mM. As the KLL peptides were designed to have similar structural and functional properties as mellitin [267, 275], it is possible, that they also form multimers in solution. However, the small angle scattering data does not indicate larger KLL-multimers.

L₉ was shown to be one of the most solvent affected residues regarding the back-

bone dihedral conformation. Conversely, in NMR spectra this residue did not show a particularly large peak shift upon a change in solvent. However, it is important to note that, the actual transition between 30-0% methanol that was probed using neutron scattering and EPSR could not be resolved as below 30% methanol NMR spectrum peak assignment became impossible.

The atomic scale analysis on KLL₁₅ revealed that the leucine backbone hydration is more affected by a change in solvent compared to the lysine hydration. Some leucine residues, like L₉, are particularly sensitive to an increased water concentration and, conversely, interact more with methanol than an average leucine residue on KLL₁₅. However, the lysine residues showed that an increased methanol content slightly increased the charged interactions between the lysine side-chain and the chloride ions in solution. While it is difficult to determine if a change in solvation affects the residue conformation or *vice versa*, previous studies have shown that apolar transmembrane α -helices [278] and amphiphilic α -helices [279] can form their secondary structure upon membrane contact [278]. This might indicate that a lack of hydration is necessary for the formation of these structures.

The influence of urea was studied on the longer KLL₂₂ peptide. This peptide was $\approx 70\%$ helical in pure water, according to the CD measurements, and only showed a transition between 80-100% methanol to become completely helical, within the sensitivity of CD and the CONTIN/LL analysis. The urea concentration used for neutron diffraction experiments (4M) slightly decreased the helical content of KLL₂₂ in CD from $\approx 75\%$ to $\approx 60\%$. This is reflected in the Ramachandran plots of KLL₂₂ where the addition of urea leads to a slightly less helical ϕ/ψ -correlation. However, even at 9 M urea about 40% of the KLL₂₂ peptides were still helical according to the CD measurements. It appears that the effect of urea on secondary structures is ambiguous, where some hydrophobic secondary structures have been found to be preferred in urea-denatures protein states [280] although the secondary structure forming hydrogen bonds have been found to be weakened by urea [281].

The addition of urea showed an overall increase in peptide backbone hydration without a direct urea-backbone interaction, unlike the results in Chapter 3 where urea (at

a similar concentration of 4 M) was decreasing and replacing water-peptide contacts. Conversely, the hydration of the amino acid residues is slightly decreased when urea is added and urea is found to interact directly with the residue sites. It is possible, that this difference between GPG-NH₂ and KLL₂₂ regarding the urea-influence is due to the lack of a secondary structure in GPG-NH₂ which makes the peptide backbone more solvent-accessible. This might even suggest that turns and loop sequences without a repetitive secondary structure are the initial site of urea denaturation, because at these sites the peptide backbone is more accessible for a direct hydrogen bond urea-peptide interaction [282, 249].

Similar to the results presented in Chapter 4, there is a marked decrease in charged interactions, here between chloride and lysine residues, which is visible when urea is added. This is even more pronounced in K₁₁ than in the average lysine-chloride interaction, suggesting that there is a residue specific decrease in salt interactions that might, as an initial step in urea induced peptide denaturation destabilise the peptide conformation [233].

Chapter 6

Conclusions and Further Work

6.1 Conclusions

Protein folding and unfolding were investigated using an inductive approach by determining the structures and molecular interactions of small model peptides on the atomic scale in solution using neutron diffraction with isotopic substitution combined with Empirical Potential Structure Refinement (EPSR), a Monte Carlo-based simulation which uses diffraction data as a direct constraint to an atomic model, along with Molecular Dynamics (MD) and other biophysical analysis techniques.

Overall, it was found that neither the peptides nor urea or co-solvents and solutes, such as ions and methanol, perturbed the aqueous solvent structure significantly. Small differences in the water structure compared to pure water can be explained by differences in ionic strength [223]. This is in agreement with previous studies on peptides [248] and urea [225] in aqueous solution and, principally, dissents from the theories of indirect urea-induced protein denaturation [187]; and with the idea that water forms highly ordered structures around hydrophobic protein sites [57] to initiate the folding process [55].

6.1.1 Urea Influence on Peptides

Investigations on the structure of GPG-NH₂ in solution and how urea interacts with this small peptide (Chapter 3), have shown that, within this system, urea preferentially interacts with the polar peptide backbone atoms while urea is less preferred around apolar peptide sites. Urea was found to lower the peptide backbone hydration, effectively replacing water molecules, most notably at the polar -NH sites in the peptide backbone. However, urea also decreased the peptide hydration without directly hydrogen bonding to these sites, which might indicate a combination of direct and indirect urea influence on protein hydration.

While an indirect influence of urea on the solvent structure was not observed, the influence of urea on the investigated peptides, namely GPG-NH₂, KGPGK and KLL₂₂, has been found to be closely connected to the electrostatic interactions between the formally charged sites within the solution. All three peptides have been found to have fewer peptide-ion (Cl⁻ or TFA⁻) interactions in the urea-containing solutions compared with the urea-free peptide solutions.

The structure and the solvation interactions for KGPGK with and without urea (Chapter 4) have shown that, again, urea hydrogen bonds with the peptide backbone are preferred over water. At this lower urea and higher ionic concentration, compared to the investigations in Chapter 3, the site specific urea/water preference ratios (PR) indicate an even higher urea preference over water compared to GPG-NH₂ in Chapter 3, although the urea-interaction RDFs, within the closest distance range, and the corresponding CNs are small due to the lower urea concentration. There is a clear preference for urea to interact with the C- and N-terminal sites of the KGPGK peptide backbone, especially in the EPSR simulations. This might indicate that urea preferentially interacts with protein sites that are more solvent accessible.

How urea interacts with a larger secondary structure forming peptides was probed using the α -helical KLL₂₂ in aqueous urea solution in Chapter 5. While the GPG-NH₂ and KGPGK backbone atoms were found to prefer urea over water, KLL₂₂ had a strong preference to coordinate water molecules around the backbone atoms. Taken together, this suggests that urea, although it might prefer to interact with the peptide backbone,

the highly ordered secondary structure of the α -helical KLL₂₂ prevents this direct urea-contact; whereas the shorter and more flexible GPG-NH₂ and KGPGK peptides were directly accessible for urea-backbone contacts. In larger protein structures, this could lead to a structure specific urea-peptide interaction, where flexible and solvent accessible peptide backbone sites might nucleate urea induced protein unfolding at relatively low urea concentrations.

In summary, it was found that urea, generally at any probed concentration, decreases ionic interactions with the peptides in solution. At a higher urea concentrations and solvent accessible peptide backbone sites, urea interacts directly with the polar peptide backbone and replaces water molecules. Therefore, urea induced protein denaturation appears to be dependent on (ionic) co-solute concentrations and the three-dimensional protein structure.

6.1.2 Peptide Folding and Stability

Within Chapters 4 and 5, results on the interplay between hydration, conformation and folding of peptides were presented for the model peptides KGXGK and KLL₁₅. The results for the β -turn model peptides KGXGK in Chapter 4 showed that proline has a unique effect on the preferred peptide backbone conformation facilitating a specific hydration and water bridging pattern between peptide residues that might precede the formation of a direct intra peptide hydrogen bond.

How water and methanol solvents influences the conformation and hydration of the amphiphilic KLL₁₅ peptide was shown in Chapter 5, where the helicity of KLL₁₅ increased with methanol content in the solvent and, conversely, the disorder in the helical peptide conformation increased with an increased water content in the solvent. This conformational transition was found to be residue specific, mostly occurring at the hydrophobic leucine residues, as these residues lost more water contacts compared with the rest of the peptide, when the methanol content of the solvent was increased.

While it is difficult to infer from these findings a conclusive theory on protein folding and stability, these results indicate an ambiguous influence of water on peptide folding

vs. conformational stability. The formation of β -turn like structures, and possibly intra-peptide hydrogen bonds in general, appears to be dependent on a specific backbone hydration pattern, facilitated by the local amino acid sequence. Conversely, hydrophobic residues in folded protein structures are particularly sensitive, in terms of conformation and hydration, to an increased peptide backbone hydration.

6.2 Further Work

The work described in this thesis addresses the structure and interactions of specific small model peptides in solution to investigate the folding and unfolding of peptides and proteins. Determining the solvent impact on these peptides on the atomic-scale can be used as a starting point for further investigations as more work is needed to develop a consistent and universal theory on protein folding.

The presented results indicate that protein folding and unfolding processes are influenced by and specific to many factors such as individual amino acids, their sequence, the resultant native folded structure and co-solute concentrations, including pH. In fact, all of these factors might be specific to distinct folding mechanisms. As the peptides studied in this thesis contained only four distinct amino acids (glycine, leucine, proline and lysine), that were studied using neutron diffraction and EPSR, atomic scale investigations on the other proteinogenic amino acids is necessary to derive at a more conclusive answer of the role of solvation in protein folding. Other factors such as pH, temperature and salt concentrations could be varied in order to probe their influence on folding and urea induced unfolding.

Other structural protein motifs, such as β -sheets or proteins with a tertiary (or quaternary) structure and hydrophobic core might show different folding behaviour than the small peptides investigated. An interplay of hydrophobic collapse [283, 284] and water mediated folding nucleation [198] depending on the primary structure is possible and cannot be disproved by only studying small peptides without the presence of a longer range structure. However, the influence of individual atomic interactions on the neutron diffraction data decreases with increasing size of the studied molecules. The structural

investigation of larger proteins folding and unfolding can therefore only be meaningful with additional methods employed, such as for example Molecular Dynamics, NMR and other spectroscopic techniques including CD, small angle scattering or analytical ultracentrifugation.

Another important aspect of protein folding and unfolding, that has to be neglected using the described neutron diffraction experiments, is kinetics. As the neutron diffraction data collection takes several hours, the fast transitioning of folding cannot be elucidated by EPSR alone. It is therefore necessary to either study peptides and proteins that are not completely folded or unfolded (Chapters 3-4) or to employ other methods such Molecular dynamics simulations (Chapter 4) or time dependent spectroscopic methods, such as stopped-flow to overcome this obstacle.

Appendix A

EPSR Input Parameters

A.1 Water EPSR Input Parameters

Table A.1: List of water atom labels and coordinates.

Atom No	Label	TIP3P			SPC/E		
		x	y	z	x	y	z
001	Ow	-2.6991999	3.1742001	0.0000000	-2.6991999	3.1742001	0.0000000
002	Hw	-1.7292000	3.1742001	0.0000000	1.7292000	3.1742001	0.0000000
003	Hw	-3.0225999	3.4698000	-0.865400	-3.0225999	3.4698000	-0.865400

Table A.2: List of water bond lengths.

Atom No 1	Atom No 2	TIP3P	SPC/E
		bond length / Å	bond length / Å
001	002	0.97600	0.97600
001	003	0.97600	0.97600

Table A.3: List of water bond angles.

Atom No 1	Atom No 2	Atom No 3	TIP3P	SPC/E
			angle / °	angle / °
002	001	003	104.52000	104.50000

A.2 Methanol EPSR Input Parameters

Table A.4: List of methanol atom labels and coordinates.

Atom No	Label	x	y	z
001	Cm	-1.50000000	1.37899995	0.03200000
002	Hm	-1.43400002	2.10400009	0.856999993
003	Hm	-1.37000000	0.360000014	0.428999990
004	Hm	-0.703999996	1.58800006	-0.698000014
005	Om	-2.77200007	1.49100006	-0.616999984
006	Hmo	-3.47600007	1.31400001	-0.00300000

Table A.5: List of methanol bond lengths.

Atom No 1	Atom No 2	bond length / Å
001	005	1.43239
001	002	1.10028
001	003	1.10130
001	004	1.10009
005	006	0.95076

Table A.6: List of methanol bond angles.

Atom No 1	Atom No 2	Atom No 3	angle / °
002	001	005	109.96698
003	001	005	109.90753
004	001	005	109.08920
002	001	003	109.40891
002	001	004	109.20641
003	001	004	109.24348
001	005	006	110.51110

Table A.7: List of methanol rotational groups.

Atom No 1	Atom No 2
1	5

A.3 Urea EPSR Input Parameters

Table A.8: List of urea atom labels and coordinates.

Atom No	Label	x	y	z
001	Nu	-5.02099991	3.85700011	0.00000000
002	Hu	-4.14400005	3.35899997	0.00000000
003	Hu	-5.02699995	4.86499977	0.00000000
004	Cu	-6.17399979	3.16899991	0.00000000
005	Ou	-6.21299982	1.95799994	0.00000000
006	Nu	-7.28000021	3.93000007	0.00000000
007	Hu	-7.20900011	4.93599987	0.00000000
008	Hu	-8.18799973	3.49000001	0.00000000

Table A.9: List of urea bond lengths.

Atom No 1	Atom No 2	bond length / Å
001	002	1.00000
001	003	1.00000
001	004	1.36000
004	005	1.23000
006	004	1.36000
006	007	1.00000
006	008	1.00000
002	007	3.45500
002	008	4.06400
003	007	2.20300
003	008	3.45500

Table A.10: List of urea bond angles.

Atom No 1	Atom No 2	Atom No 3	angle / °
002	001	003	120.00000
002	001	004	120.00000
003	001	004	120.00000
001	004	005	122.50000
001	004	006	115.00000
006	004	005	122.50000
007	006	004	120.00000
008	006	004	120.00000
007	006	008	120.00000

Table A.11: List of urea dihedral angles.

Atom No 1	Atom No 2	Atom No 3	Atom No 4	dihedral angle / °
2	1	4	5	0.00000
2	1	4	6	180.00000
3	1	4	5	180.00000
3	1	4	6	0.00000
1	4	6	7	0.00000
1	4	6	8	180.00000
5	4	6	7	180.00000
5	4	6	8	0.00000

A.4 TFA EPSR Input Parameters

Table A.12: List of TFA atom labels and coordinates.

Atom No	Label	x	y	z
001	Ct1	-3.73900008	2.26399994	0.08299999
002	Ct2	-2.24200010	2.52399993	-0.03200000
003	Ft	-4.36800003	3.20000005	0.82099998
004	Ft	-4.32399988	2.24300003	-1.12899995
005	Ft	-3.98699999	1.07599998	0.66699996
006	Ot	-1.94000006	3.78699994	-0.21099995
007	Ot	-1.42100000	1.64300001	-0.20000003

Table A.13: List of TFA bond lengths.

Atom No 1	Atom No 2	bond length / Å
001	002	1.52376
001	003	1.34773
001	004	1.34596
001	005	1.34681
002	006	1.31088
002	007	1.21591

Table A.14: List of TFA bond angles.

Atom No 1	Atom No 2	Atom No 3	angle / °
002	001	003	112.41663
002	001	004	111.20488
002	001	005	111.35474
003	001	004	107.52213
003	001	005	106.81155
004	001	005	107.25735
001	002	006	113.64305
001	002	007	123.37768
006	002	007	121.57900

Table A.15: List of TFA rotational groups.

Atom No 1	Atom No 2
1	2

A.5 GPG-NH₂ EPSR Input Parameters

Table A.16: List of GPG-NH₂ atom labels and coordinates.

Atom No	Label	GPG-NH ₂ (<i>trans</i>)			GPG-NH ₂ (<i>cis</i>)		
		x	y	z	x	y	z
001	N1	10.4442997	23.2173004	42.8475990	35.7630005	14.3577003	46.8908005
002	Hn1	9.6233997	23.7381992	43.2332993	36.3485985	14.3599997	47.7579002
003	Hn1	10.6138000	23.5016003	41.8549004	35.7327995	15.3181000	46.4771996
004	Hn1	10.2335997	22.1935005	42.8717003	34.7900009	14.0712004	47.1441994
005	Cp	11.6352997	23.4792004	43.6582985	36.3041000	13.4052000	45.9188995
006	Hc	11.4438000	23.1436005	44.7010994	36.2924004	12.3907003	46.3700981
007	Hc	12.4863997	22.8938007	43.2473984	35.6492004	13.4000998	45.0206985
008	Co	12.0087996	24.9493008	43.6753006	37.7201996	13.7563000	45.4962997
009	O1	11.3310003	25.7509995	43.0460014	38.2580986	14.7488003	45.9692993
010	N2	13.0909004	25.3887997	44.3877983	38.3995018	12.9892998	44.5878983
011	Cn	13.9329004	24.5223999	45.1846008	39.7288017	13.3255997	44.1255989
012	Hc	14.3140001	23.6394005	44.6281013	40.4524994	13.4934998	44.9528008
013	Hc	13.3580999	24.2029991	46.0814018	39.6556015	14.2306004	43.4832001
014	Cn	13.4807997	26.7847004	44.4594994	37.8586006	11.8037996	43.9469986
015	Hc	12.6188002	27.4510002	44.6855011	36.8287010	11.9607000	43.5591011
016	Cm	14.4359999	26.7891006	45.6617012	38.8083992	11.6282997	42.7528992
017	Hc	15.1162996	27.6639996	45.7336998	38.8064003	10.6236000	42.2793007
018	Hc	13.8202000	26.7546005	46.5900002	38.5270996	12.3759003	41.9757004
019	Cm	15.1080999	25.4316998	45.5447998	40.1377983	12.0865002	43.3294983
020	Hc	15.7931995	25.4845009	44.6679993	40.4687996	11.3009996	44.0463982
021	Hc	15.6917000	25.1002998	46.4294014	40.9527016	12.2600002	42.5954018
022	Co	14.1515999	27.2401009	43.1750984	37.8838005	10.6044998	44.8805008
023	O2	14.2648001	26.4622993	42.2364998	38.2990990	10.7286997	46.0259018
024	N3	14.6163998	28.5174007	43.0736008	37.4432983	9.39360046	44.4393997
025	Hn3	14.5057001	29.1537991	43.8778992	37.0989990	9.33500004	43.4765015
026	Cn	15.2754002	29.0284004	41.8816986	37.4600983	8.22809982	45.3148003
027	Hc	16.2945004	29.3656998	42.1671982	38.5073013	8.05430031	45.6463013
028	Hc	15.3860998	28.2434998	41.1031990	36.8423004	8.45479965	46.2112999
029	Co	14.5347996	30.1923008	41.2578011	36.9478989	6.91389990	44.7358017
030	NT	13.3521996	30.6441994	41.7473984	36.4850006	6.77099991	43.4659996
031	HT1	12.9019003	30.2110004	42.5647011	36.4430008	7.53970003	42.7910004
032	HT2	12.8767996	31.4381008	41.2929993	36.1553001	5.84730005	43.1481018
033	O3	15.0200996	30.7339001	40.2761002	36.9589005	5.93450022	45.4665985

Table A.17: List of GPG-NH₂ bond lengths.

		GPG-NH₂ (<i>trans</i>)	GPG-NH₂ (<i>cis</i>)
Atom No 1	Atom No 2	bond length / Å	bond length / Å
001	002	1.04593	1.04632
001	003	1.04643	1.04611
001	004	1.04553	1.04548
001	005	1.46434	1.46446
005	006	1.11208	1.11037
005	007	1.11171	1.11161
005	008	1.51690	1.51895
008	009	1.22399	1.22398
008	010	1.36812	1.36928
011	010	1.44724	1.44702
010	014	1.45110	1.45215
011	012	1.11113	1.11184
011	013	1.11205	1.11223
011	019	1.52894	1.52854
015	014	1.11269	1.11165
014	016	1.53549	1.53584
022	014	1.51890	1.51999
016	019	1.51918	1.51978
017	016	1.11060	1.11073
018	016	1.11451	1.11448
020	019	1.11397	1.11379
021	019	1.11038	1.11044
022	023	1.22424	1.22468
022	024	1.36302	1.36194
025	024	1.03158	1.02428
026	024	1.45466	1.45774
026	027	1.11079	1.11209
026	028	1.11103	1.11211
026	029	1.51407	1.52470
029	033	1.22171	1.22205
029	030	1.35737	1.35908
031	030	1.02879	1.02386
032	030	1.03090	1.03101

Table A.18: List of GPG-NH₂ bond angles (Part 1).

			GPG-NH₂ (<i>trans</i>)	GPG-NH₂ (<i>cis</i>)
Atom No 1	Atom No 2	Atom No 3	angle / °	angle / °
002	001	003	109.97755	109.98585
002	001	004	108.72363	108.70014
002	001	005	110.18884	110.15853
003	001	004	108.69617	108.69590
003	001	005	110.17358	110.20606
004	001	005	109.04298	109.05595
001	005	006	108.97203	108.70422
001	005	007	108.89674	108.75667
001	005	008	112.32142	112.25763
006	005	007	108.63312	108.53819
006	005	008	108.92647	109.51577
007	005	008	109.02419	108.99635
005	008	009	119.51604	119.31615
005	008	010	120.79169	121.17195
009	008	010	119.69202	119.50935
008	010	014	123.18016	124.42291
011	010	008	123.68172	122.51426
011	010	014	113.10105	113.01122
010	011	019	102.75957	102.93238
012	011	010	113.54961	113.28784
012	011	013	110.66665	110.46751
012	011	019	109.08561	109.61769
013	011	010	108.37415	108.25058
013	011	019	112.21745	112.11078
010	014	016	102.03983	101.89256
015	014	010	112.20491	112.57341
015	014	016	108.73461	108.52848
015	014	022	109.53307	109.93913
022	014	010	111.42594	111.52498
022	014	016	112.73540	112.15303
014	016	019	102.26343	102.21185
017	016	014	115.72624	115.70638
017	016	018	107.98933	107.99719
017	016	019	115.96792	115.85247
018	016	014	107.95818	108.02596
018	016	019	106.31258	106.44656
011	019	016	102.09116	102.00206
011	019	020	108.38958	108.42082
011	019	021	114.46707	114.45726
020	019	016	106.87592	106.94979
020	019	021	108.53962	108.51884
021	019	016	115.96960	115.99054
014	022	024	119.63914	119.81766
023	022	014	119.91163	119.99976
023	022	024	120.44792	120.18256
025	024	022	118.91196	117.64895
026	024	022	123.00355	120.83951
025	024	026	118.08264	121.51116

Table A.19: List of GPG-NH₂ bond angles (Part 2).

			GPG-NH₂ (<i>trans</i>)	GPG-NH₂ (<i>cis</i>)
Atom No 1	Atom No 2	Atom No 3	angle / °	angle / °
027	026	024	108.16208	108.34953
028	026	024	111.78356	108.34344
029	026	024	112.71075	117.20836
027	026	028	107.65038	108.34087
027	026	029	108.73874	107.14718
028	026	029	107.64245	107.17010
026	029	033	118.52846	117.43081
033	029	030	119.23082	118.52388
026	029	030	122.23946	124.04527
031	030	029	121.85027	123.43520
031	030	032	118.20278	117.14528
032	030	029	119.94505	119.41951

Table A.20: List of GPG-NH₂ dihedral angles.

				GPG-NH₂ (<i>trans</i>)	GPG-NH₂ (<i>cis</i>)
Atom No 1	Atom No 2	Atom No 3	Atom No 4	dihedral angle / °	dihedral angle / °
5	8	10	11	0.00000	180.00000
5	8	10	14	180.00000	0.00000
9	8	10	11	180.00000	0.00000
9	8	10	14	0.00000	180.00000
14	22	24	25	0.00000	0.00000
14	22	24	26	180.00000	180.00000
23	22	24	25	180.00000	180.00000
23	22	24	26	0.00000	0.00000
33	29	30	31	180.00000	180.00000
33	29	30	32	0.00000	0.00000
26	29	30	31	0.00000	0.00000
26	29	30	32	180.00000	180.00000

Table A.21: List of GPG-NH₂ rotational groups.

Atom No 1	Atom No 2
1	5
5	8
14	22
24	26
26	29

A.6 KGPGK EPSR Input Parameters

Table A.22: List of KGPGK atom labels and coordinates (Part 1).

Atom No	Label	x	y	z
001	N1	-1.216000	7.381000	-3.413000
002	H3n	-1.746000	7.217000	-4.284999
003	H3n	-1.178000	8.399999	-3.246000
004	H3n	-0.251000	7.035999	-3.542999
005	Ck	-1.876000	6.688000	-2.255000
006	Hka	-1.137000	7.019999	-1.519999
007	Ck	-3.278000	7.230999	-1.904999
008	Hk	-4.001999	6.886000	-2.650000
009	Hk	-3.256000	8.321999	-1.957999
010	Ck	-3.733999	6.807000	-0.4920000
011	Hk	-3.816999	5.718999	-0.4379999
012	Hk	-2.983999	7.118999	0.2399999
013	Ck	-5.093999	7.440999	-0.1330000
014	Hk	-5.022999	8.529999	-0.2000000
015	Hk	-5.853000	7.113999	-0.8479999
016	Ck	-5.527999	7.048999	1.289999
017	Hk	-5.635000	5.964000	1.371999
018	Hk	-4.797999	7.403999	2.023000
019	Nk1	-6.823999	7.652999	1.741000
020	H3n	-7.044000	7.359000	2.707000
021	H3n	-6.769000	8.684000	1.717000
022	H3n	-7.596000	7.353000	1.123000
023	Co	-1.845000	5.175000	-2.348999
024	O1	-2.367000	4.565999	-3.260999
025	N2	-1.199000	4.543000	-1.347000
026	H2	-0.7929999	5.080999	-0.5950000
027	Cg	-1.018000	3.089999	-1.289000
028	Hg	-1.993000	2.595999	-1.347000
029	Hg	-0.4050000	2.783999	-2.140000
030	Co	-0.3330000	2.618999	-0.016000
031	O2	-0.1490000	3.364000	0.9269999
032	N3	0.015000	1.307999	0.048000
033	Cp	-0.1749999	0.3120000	-1.019000

Table A.23: List of KGPGK atom labels and coordinates (Part 2).

Atom No	Label	x	y	z
034	Hp	-1.241000	0.088000	-1.108999
035	Hp	0.2140000	0.6769999	-1.973000
036	Cp	0.6830000	0.6890000	1.205999
037	Hpa	0.1169999	0.9139999	2.114000
038	Cp	0.6650000	-0.8280000	0.9430000
039	Hp	-0.2370000	-1.261000	1.384999
040	Hp	1.536999	-1.342000	1.356999
041	Cp	0.5920000	-0.9449999	-0.5830000
042	Hp	0.1000000	-1.865000	-0.9070000
043	Hp	1.603000	-0.9190000	-0.9990000
044	Co	2.101000	1.197999	1.404999
045	O3	2.838000	1.488000	0.4810000
046	N4	2.519999	1.213000	2.687000
047	H4	1.878999	0.9330000	3.414000
048	Cg	3.838999	1.677999	3.121999
049	Hg	4.584000	1.455999	2.352000
050	Hg	3.809999	2.756999	3.295000
051	Co	4.261000	1.011999	4.414000
052	O4	3.529999	0.9520000	5.382999
053	N5	5.449999	0.3779999	4.368999
054	H5	5.997000	0.4289999	3.522000
055	Ck	6.004000	-0.4420000	5.453000
056	Hka	5.649000	-0.034000	6.407000
057	Ck	5.544000	-1.917000	5.341000
058	Hk	4.454000	-1.944000	5.276000
059	Hk	5.933000	-2.355000	4.419000
060	Ck	5.991000	-2.765000	6.553999
061	Hk	5.583000	-2.332999	7.471000
062	Hk	7.080999	-2.750999	6.637000
063	Ck	5.517000	-4.227000	6.414999
064	Hk	4.426000	-4.257999	6.349999
065	Hk	5.916999	-4.656000	5.493000
066	Ck	5.981999	-5.073999	7.612999
067	Hk	7.073999	-5.084000	7.673999
068	Hk	5.564000	-4.681000	8.543999
069	Nk5	5.611999	-6.526000	7.539000
070	H3n	4.585000	-6.633999	7.502999
071	H3n	6.018000	-6.958000	6.692999
072	H3n	5.967000	-7.032000	8.366999
073	Co	7.518000	-0.3109999	5.453000
074	N6	8.001000	0.7369999	6.143000
075	H6b	8.991000	0.9330000	6.125999
076	H6a	7.376999	1.330000	6.665999
077	O5	8.246999	-0.9710000	4.742000

Table A.24: List of KGPGK bond lengths (Part 1).

Atom No 1	Atom No 2	bond length / Å
001	002	1.03353
001	003	1.03329
001	005	1.50227
001	004	1.03303
005	007	1.54368
005	023	1.51623
005	006	1.09388
008	007	1.09464
009	007	1.09251
007	010	1.54411
010	013	1.54287
011	010	1.09250
012	010	1.09347
014	013	1.09337
013	016	1.53849
015	013	1.09281
018	016	1.09372
019	016	1.49928
017	016	1.09334
019	020	1.03344
019	021	1.03275
019	022	1.03340
023	024	1.21454
023	025	1.34935
027	025	1.46538
026	025	1.00984
028	027	1.09454
027	030	1.52039
029	027	1.09252
030	032	1.35791
030	031	1.21578
036	032	1.47321
033	032	1.47194
035	033	1.09301
041	033	1.53572
034	033	1.09299
036	038	1.53973
036	037	1.09336
036	044	1.51967
038	041	1.53222
039	038	1.09383
040	038	1.09361
042	041	1.09245
043	041	1.09355

Table A.25: List of KGPGK bond lengths (Part 2).

Atom No 1	Atom No 2	bond length / Å
044	046	1.34882
044	045	1.21698
047	046	1.00887
048	046	1.46465
049	048	1.09417
048	051	1.51357
050	048	1.09317
051	053	1.34822
051	052	1.21529
055	053	1.46778
054	053	1.00956
055	057	1.54912
055	056	1.09442
055	073	1.51966
059	057	1.09236
057	060	1.54605
058	057	1.09227
060	063	1.54319
061	060	1.09269
062	060	1.09324
065	063	1.09276
063	066	1.53910
064	063	1.09337
069	066	1.50023
068	066	1.09359
067	066	1.09375
069	070	1.03329
069	071	1.03304
069	072	1.03327
073	077	1.21349
073	074	1.34450
076	074	1.00725
075	074	1.00936

Table A.26: List of KGPGK bond angles (Part 1).

Atom No 1	Atom No 2	Atom No 3	angle / °
002	001	003	108.16194
002	001	005	110.60162
002	001	004	108.65484
003	001	005	110.27316
003	001	004	108.38102
004	001	005	110.69215
001	005	007	114.30025
001	005	023	113.80013
001	005	006	94.65328
007	005	023	112.55878
006	005	007	110.76074
006	005	023	109.31146
005	007	008	109.60458
005	007	009	108.78286
005	007	010	112.27796
008	007	009	107.15993
008	007	010	109.93385
009	007	010	108.93844
007	010	013	111.12499
011	010	007	109.94562
012	010	007	109.37108
011	010	013	109.31507
012	010	013	109.36488
011	010	012	107.64851
014	013	010	109.74186
010	013	016	111.04957
015	013	010	109.69545
014	013	016	109.19444
014	013	015	107.64038
015	013	016	109.45056
018	016	013	110.41988
019	016	013	114.79866
017	016	013	110.47680
019	016	018	104.15768
018	016	017	109.70400
019	016	017	107.01545
020	019	016	110.52354
021	019	016	110.43559
022	019	016	110.42110
020	019	021	108.48561
020	019	022	108.50465
021	019	022	108.40315
005	023	024	122.55574
005	023	025	115.53905
024	023	025	121.90502

Table A.27: List of KGPGK bond angles (Part 2).

Atom No 1	Atom No 2	Atom No 3	angle / °
027	025	023	123.56912
026	025	023	119.73082
026	025	027	116.68734
028	027	025	109.59641
030	027	025	113.32592
029	027	025	108.43288
028	027	030	107.81144
028	027	029	109.39747
029	027	030	108.21756
027	030	032	117.00135
027	030	031	121.85554
031	030	032	121.08189
036	032	030	123.98018
033	032	030	125.87625
036	032	033	110.12150
035	033	032	111.14930
041	033	032	106.47401
034	033	032	108.92068
035	033	041	110.08312
035	033	034	110.10112
034	033	041	110.04371
038	036	032	105.92649
037	036	032	109.36495
044	036	032	112.66197
037	036	038	109.78838
038	036	044	111.30118
037	036	044	107.77975
036	038	041	104.23530
036	038	039	109.30712
036	038	040	112.89734
039	038	041	109.44672
040	038	041	112.27961
039	038	040	108.57217
038	041	033	104.12555
042	041	033	112.34435
043	041	033	109.52853
042	041	038	112.40439
043	041	038	109.45132
042	041	043	108.87878
036	044	046	114.71127
036	044	045	123.05678
045	044	046	122.06419

Table A.28: List of KGPGK bond angles (Part 3).

Atom No 1	Atom No 2	Atom No 3	angle / °
047	046	044	118.97678
048	046	044	124.44164
047	046	048	116.50496
049	048	046	109.86157
051	048	046	111.40186
050	048	046	109.66247
049	048	051	108.75948
049	048	050	109.25034
050	048	051	107.85590
048	051	053	115.10704
048	051	052	122.31931
052	051	053	122.26694
055	053	051	124.81503
054	053	051	118.82123
054	053	055	116.31637
057	055	053	111.49769
056	055	053	108.42933
073	055	053	109.13998
056	055	057	108.44820
057	055	073	112.20470
056	055	073	106.94791
055	057	059	109.69795
055	057	060	112.31341
055	057	058	108.91497
059	057	060	109.83615
059	057	058	107.17132
058	057	060	108.76295
057	060	063	111.11004
061	060	057	109.48878
062	060	057	109.92603
061	060	063	109.60018
062	060	063	108.97856
061	060	062	107.66820
065	063	060	109.60256
060	063	066	111.00545
064	063	060	109.79793
065	063	066	109.27514
065	063	064	107.69563
064	063	066	109.39863
069	066	063	114.81712
068	066	063	110.45087
067	066	063	110.49243

Table A.29: List of KGPGK bond angles (Part 4).

Atom No 1	Atom No 2	Atom No 3	angle / °
069	066	068	107.18985
069	066	067	103.89462
068	066	067	109.71996
070	069	066	110.36544
071	069	066	110.37759
072	069	066	110.46925
070	069	071	108.56502
070	069	072	108.55488
071	069	072	108.44743
055	073	077	123.47853
055	073	074	115.15674
077	073	074	120.58526
076	074	073	120.18652
075	074	073	119.67412
076	074	075	120.13575

Table A.30: List of KGPGK dihedral angles.

Atom No 1	Atom No 2	Atom No 3	Atom No 4	dihedral angle / °
5	23	25	27	180.00000
5	23	25	26	0.00000
24	23	25	27	0.00000
24	23	25	26	180.00000
27	30	32	36	180.00000
27	30	32	33	0.00000
31	30	32	36	0.00000
31	30	32	33	180.00000
36	44	46	47	0.00000
36	44	46	48	180.00000
45	44	46	47	180.00000
45	44	46	48	0.00000
48	51	53	55	180.00000
48	51	53	54	0.00000
52	51	53	55	0.00000
52	51	53	54	180.00000
55	73	74	76	0.00000
55	73	74	75	180.00000
77	73	74	76	180.00000
77	73	74	75	0.00000

Table A.31: List of KGPGK rotational groups.

Atom No 1	Atom No 2
1	5
5	7
5	23
7	10
10	13
13	16
16	19
25	27
27	30
32	36
32	33
33	41
36	38
36	44
38	41
46	48
48	51
53	55
55	57
55	73
57	60
60	63
63	66
66	69

Appendix B

KGPGK Trajectory Analysis

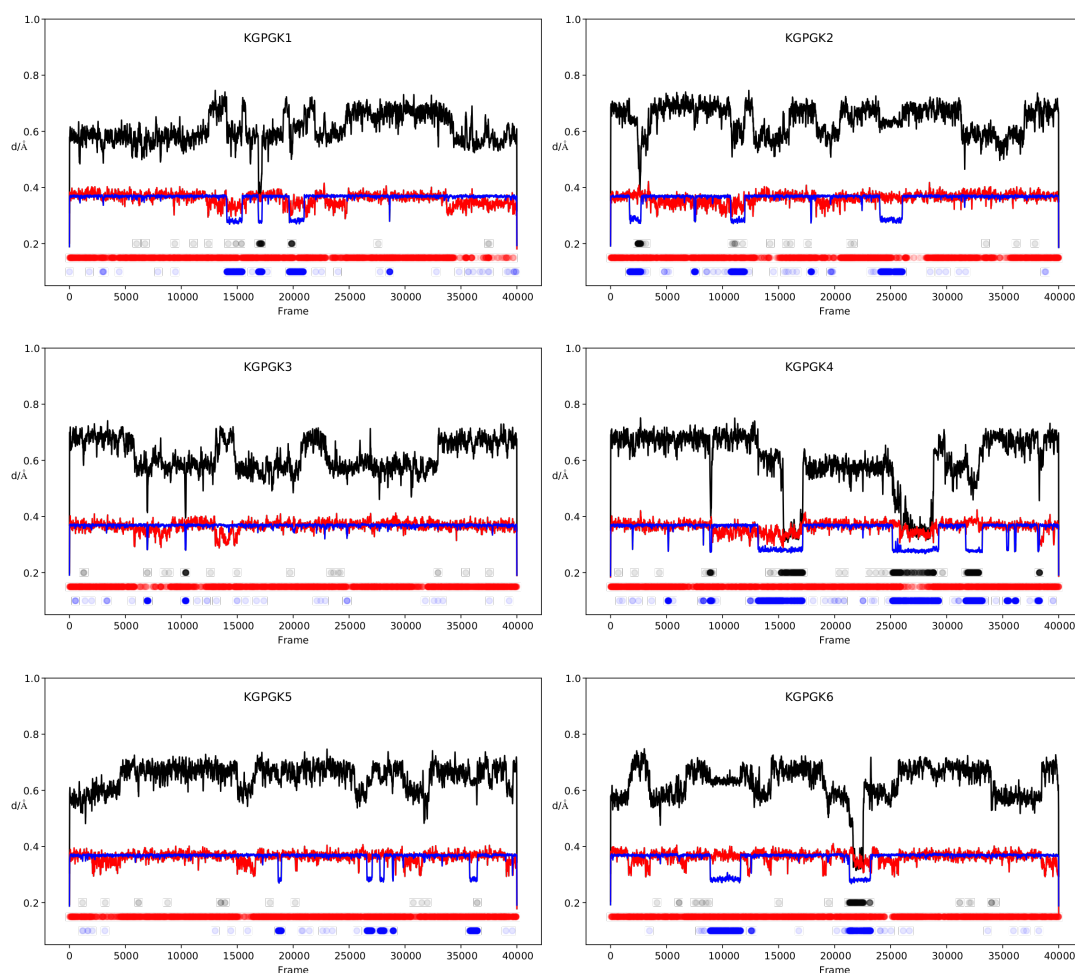


Figure B.1: Analysis of KGP GK peptides in the MD trajectory. The top graph shows the intramolecular distances N4...N5 (blue), O2...N5 (black) and O2...N4 (red). The bottom graph illustrates the water molecule mediating this interaction (Part 1).

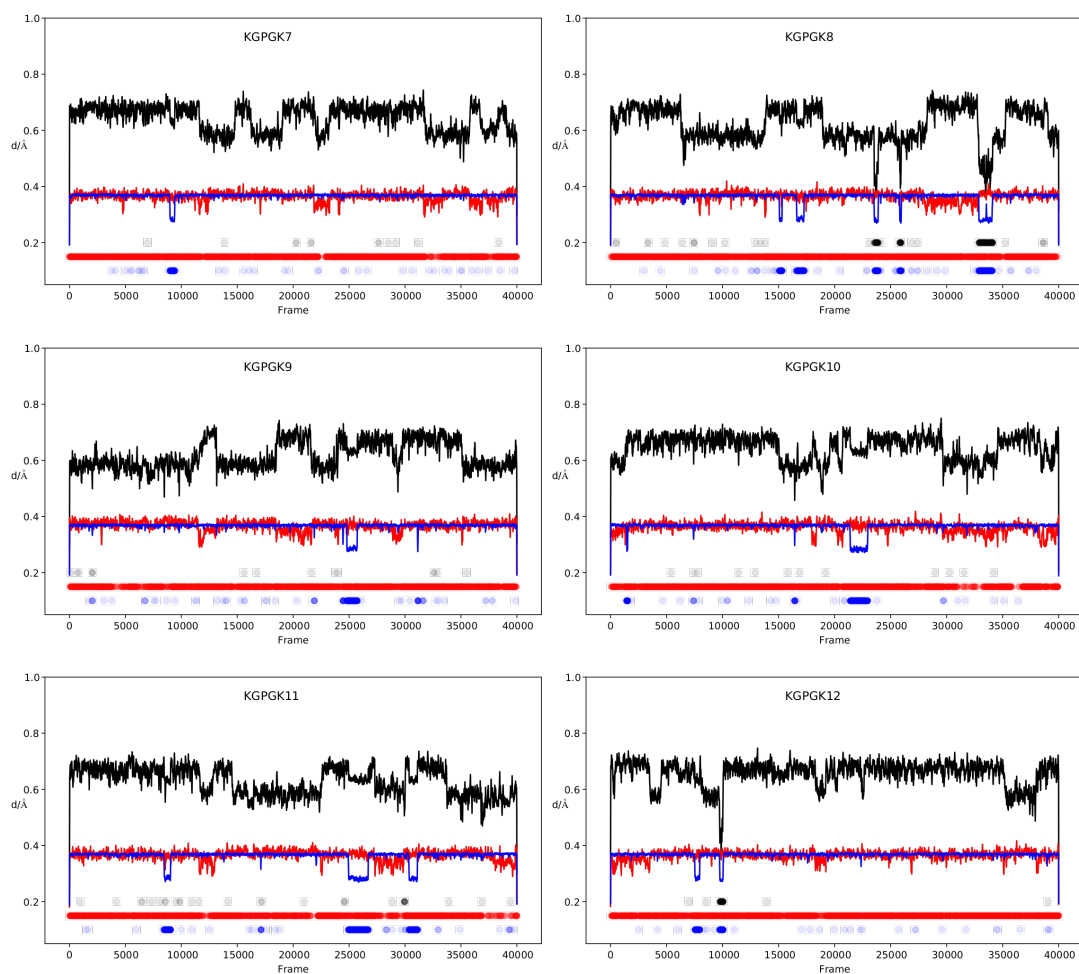


Figure B.2: Analysis of KGPGK peptides in the MD trajectory. The top graph shows the intramolecular distances $N4 \cdots N5$ (blue), $O2 \cdots N5$ (black) and $O2 \cdots N4$ (red). The bottom graph illustrates the water molecule mediating this interaction (Part 2).

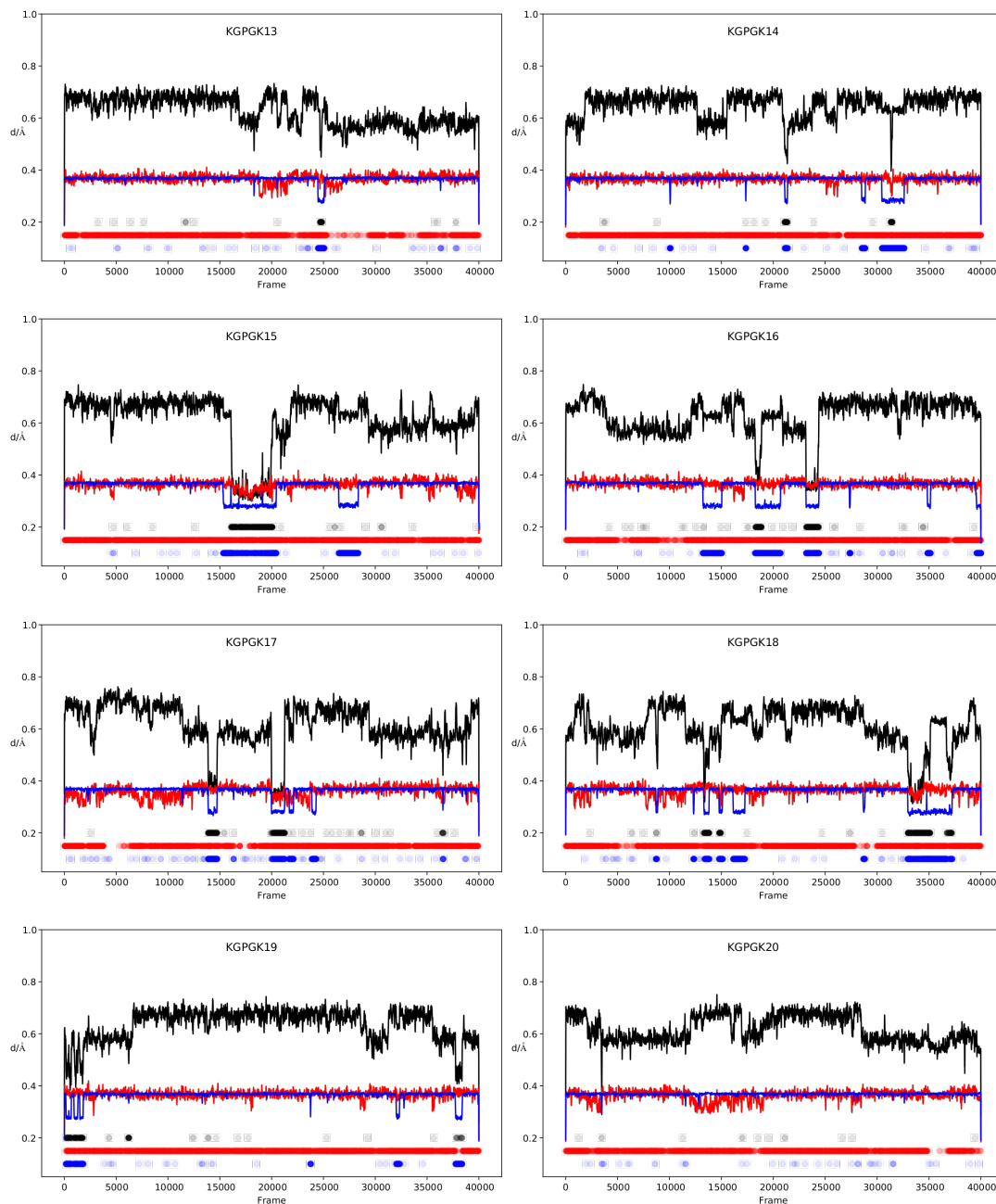


Figure B.3: Analysis of KGP GK peptides in the MD trajectory. The top graph shows the intramolecular distances $N4 \cdots N5$ (blue), $O2 \cdots N5$ (black) and $O2 \cdots N4$ (red). The bottom graph illustrates the water molecule mediating this interaction (Part 3).

References

- [1] CB Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973. 1, 2, 14, 17
- [2] C Levinthal. Are there pathways for protein folding? *Journal de chimie physique*, 65:44–45, 1968. 1, 15
- [3] ML Anson and AE Mirsky. On some general properties of proteins. *The Journal of general physiology*, 9(2):169, 1925. 1
- [4] PA Alexander, Y He, Y Chen, J Orban, and PN Bryan. The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proceedings of the National Academy of Sciences*, 104(29):11963–11968, 2007. 1
- [5] RB Russell, MAS Saqi, RA Sayle, PA Bates, and MJE Sternberg. Recognition of analogous and homologous protein folds: analysis of sequence and structure conservation. *Journal of molecular biology*, 269(3):423–439, 1997. 1
- [6] C Orengo, DT Jones, and JM Thornton. Protein superfamilies and domain superfolds. *Nature*, 372(6507):631, 1994. 1
- [7] A Cavalli, X Salvatella, CM Dobson, and M Vendruscolo. Protein structure determination from nmr chemical shifts. *Proceedings of the National Academy of Sciences*, 104(23):9615–9620, 2007. 1, 26
- [8] D Wishart. Nmr spectroscopy and protein structure determination: applications to drug discovery and development. *Current Pharmaceutical Biotechnology*, 6(2):105–120, 2005. 1, 26

-
- [9] X Bai, G McMullan, and SHW Scheres. How cryo-EM is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40(1):49–57, 2015. 2, 25
- [10] A Bartesaghi, A Merk, S Banerjee, D Matthies, X Wu, JLS Milne, and S Subramaniam. 2.2 Å resolution cryo-EM structure of β -galactosidase in complex with a cell-permeant inhibitor. *Science*, 348(6239):1147–1151, 2015. 2, 23
- [11] The Protein Data Base. <http://www.rcsb.org>. Accessed: 20-10-2017. 2
- [12] Redundancy in the Protein Data Bank. <https://www.rcsb.org/pdb/statistics/clusterStatistics.do>. Accessed: 20-10-2017. 2
- [13] The UniProt Reference Clusters. <http://www.uniprot.org/uniref>. Accessed: 20-10-2017. 2
- [14] The Universal Protein Database. <http://www.uniprot.org>. Accessed: 20-10-2017. 2
- [15] VN Uversky, CJ Oldfield, and AK Dunker. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Annual Review of Biophysics*, 37:215–246, 2008. 2
- [16] EF Garman. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica Section D*, 66(4):339–351, 2010. 2
- [17] JC Bischof and X He. Thermal stability of proteins. *Annals of the New York Academy of Sciences*, 1066(1):12–33, 2006. 2
- [18] A Fischer. Heat denaturation of proteins as a chain reaction. *Nature*, 137(3466):576–577, 1936. 2
- [19] M Bucciantini, E Giannoni, F Chiti, F Baroni, L Formigli, J Zurdo, N Taddei, G Ramponi, CM Dobson, and M Stefani. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature*, 416(6880):507–511, 2002. 3
- [20] J Luo, J Maréchal, S Wärmländer, A Gräslund, and A Perálvarez-Marín. In silico analysis of the apolipoprotein E and the amyloid β -peptide interaction: misfolding

- induced by frustration of the salt bridge network. *PLoS computational biology*, 6(2):e1000663, 2010. 3
- [21] PAJ Muller and KH Vousden. p53 Mutations in cancer. *Nature Cell Biology*, 15(1):2–8, 2013. 3
- [22] WF DeGrado, CM Summa, V Pavone, F Natri, and A Lombardi. De novo design and structural characterization of proteins and metalloproteins. *Annual review of biochemistry*, 68(1):779–819, 1999. 3
- [23] M Levitt and A Warshel. Computer simulation of protein folding. *Nature*, 253(5494):694–698, 1975. 3
- [24] Y Song, F DiMaio, RY Wang, D Kim, C Miles, TJ Brunette, J Thompson, and D Baker. High-resolution comparative modeling with RosettaCM. *Structure*, 21(10):1735–1742, 2013. 3
- [25] S Ovchinnikov, DE Kim, RY Wang, Y Liu, F DiMaio, and D Baker. Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins: Structure, Function, and Bioinformatics*, 84(S1):67–75, 2016. 3
- [26] A Perez, JA Morrone, C Simmerling, and KA Dill. Advances in free-energy-based simulations of protein folding and ligand binding. *Current opinion in structural biology*, 36:25–31, 2016. 3
- [27] KA Dill and JL MacCallum. The protein-folding problem, 50 years on. *Science*, 338(6110):1042–1046, 2012. 3
- [28] M Rother and JA Krzycki. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea*, 2010, 2010. 3
- [29] C Walsh. *Posttranslational modification of proteins: expanding nature’s inventory*. Roberts and Company Publishers, 2006. 3
- [30] M Sugahara, T Nakane, T Masuda, M Suzuki, S Inoue, C Song, R Tanaka, T Nakatsu, E Mizohata, F Yumoto, et al. Hydroxyethyl cellulose matrix applied to serial crystallography. *Scientific Reports*, 7, 2017. 6

- [31] S Park, T Yokoyama, N Shibayama, Y Shiro, and JRH Tame. 1.25 Å resolution crystal structures of human haemoglobin in the oxy, deoxy and carbonmonoxy forms. *Journal of Molecular Biology*, 360(3):690–701, 2006. 6
- [32] L Pauling, RB Corey, and HR Branson. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*, 37(4):205–211, 1951. 4
- [33] BW Matthews. How planar are planar peptide bonds? *Protein Science*, 25(4):776–777, 2016. 4
- [34] WJ Wedemeyer, E Welker, and HA Scheraga. Proline *cis-trans* isomerization and protein folding. *Biochemistry*, 41(50):14637–14644, 2002. 7, 108
- [35] GN Ramachandran, C Ramakrishnan, and V Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, 1963. 7, 146
- [36] D Voet, JG Voet, CW Pratt, et al. *Fundamentals of Biochemistry*. Wiley New York, 1999. 7
- [37] NC Price and RA Dwek. *Principles and Problems in Physical Chemistry for Biochemists*. Clarendon Press, 1974. 8, 11, 13, 14
- [38] A Fersht. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Macmillan, 1999. 9, 27
- [39] A Cooper, CM Johnson, JH Lakey, and M Nöllmann. Heat does not come in different colours: entropy–enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophysical Chemistry*, 93(2):215–230, 2001. 9
- [40] M Sela, FH White Jr, and CB Anfinsen. Reductive cleavage of disulfide bridges in ribonuclease. *Science*, 125(3250):691–692, 1957. 10
- [41] JM Berg, JL Tymoczko, and L Stryer. *Biochemistry (5th ed.)*. WH Freeman, 2002. 11

- [42] H Lodish, A Berk, SL Zipursky, P Matsudaira, D Baltimore, and J Darnell. Protein structure and function. *Molecular Cell Biology*, 2000. 11
- [43] KE Van Holde, WC Johnson, PS Ho, et al. *Principles of Physical Biochemistry*. Pearson/Prentice, 2006. 11
- [44] MK Gilson and BH Honig. The dielectric constant of a folded protein. *Biopolymers*, 25(11):2097–2119, 1986. 11
- [45] CN Schutz and A Warshel. What are the dielectric constants of proteins and how to validate electrostatic models? *Proteins: Structure, Function, and Bioinformatics*, 44(4):400–417, 2001. 11
- [46] T Simonson. What is the dielectric constant of a protein when its backbone is fixed? *Journal of Chemical Theory and Computation*, 9(10):4603–4608, 2013. 11
- [47] L Li, C Li, Z Zhang, and E Alexov. On the dielectric constant of proteins: smooth dielectric function for macromolecular modeling and its implementation in Delphi. *Journal of Chemical Theory and Computation*, 9(4):2126–2136, 2013. 11
- [48] WH Keesom. The second virial coefficient for rigid spherical molecules, whose mutual attraction is equivalent to that of a quadruplet placed at their centre. *Proceedings of the Royal Netherlands Academy of Arts and Sciences*, 18:636–646, 1915. 12
- [49] P Debye. Die van-der-Waalsschen Kohäsionskräfte. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen*, 1920:55–73, 1920. 12
- [50] F London. Zur Theorie und Systematik der Molekularkräfte. *Zeitschrift für Physik*, 63(3):245–279, 1930. 12
- [51] F Cordier, M Rogowski, S Grzesiek, and A Bax. Observation of through-hydrogen-bond ${}^{2h}J_{HC'}$ in a perdeuterated protein. *Journal of Magnetic Resonance*, 140(2):510–512, 1999. 13
- [52] NT Southall, KA Dill, and ADJ Haymet. A view of the hydrophobic effect. *The Journal of Physical Chemistry B*, 106(3):521–533, 2002. 13

- [53] F Biedermann, WM Nau, and H Schneider. The hydrophobic effect revisited? studies with supramolecular complexes imply high-energy water as a noncovalent driving force. *Angewandte Chemie International Edition*, 53(42):11158–11171, 2014. 13
- [54] W Kauzmann. Some factors in the interpretation of protein denaturation. *Advances in Protein Chemistry*, 14:1–63, 1959. 14
- [55] D Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, 2005. 14, 193
- [56] F Franks, editor. *Water, A Comprehensive Treatise*, volume 4. 1975. 14
- [57] B Widom, P Bhimalapuram, and K Koga. The hydrophobic effect. *Physical Chemistry Chemical Physics*, 5(15):3085–3093, 2003. 14, 193
- [58] SR Durell and A Ben-Naim. Hydrophobic-hydrophilic forces in protein folding. *Biopolymers*, 2017. 14
- [59] D Baker and DA Agard. Kinetics versus thermodynamics in protein folding. *Biochemistry*, 33(24):7505–7509, 1994. 14
- [60] CN Pace, S Trevino, E Prabhakaran, and JM Scholtz. Protein structure, stability and solubility in water and other solvents. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 359(1448):1225–1235, 2004. 14
- [61] DJ Barlow and JM Thornton. Ion-pairs in proteins. *Journal of molecular biology*, 168(4):867–885, 1983. 15
- [62] A Szilágyi and P Závodszky. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*, 8(5):493–504, 2000. 15
- [63] CN Pace, BA Shirley, M McNutt, and K Gajiwala. Forces contributing to the conformational stability of proteins. *The FASEB journal*, 10(1):75–83, 1996. 15
- [64] KA Dill, HS Chan, et al. From Levinthal to pathways to funnels. *Nature Structural Biology*, 4(1):10–19, 1997. 15

- [65] C Frieden, SD Hoeltzli, and IJ Ropson. NMR and protein folding: Equilibrium and stopped-flow studies. *Protein Science*, 2(12):2007–2014, 1993. 15
- [66] PN Lewis, D Kotelchuck, HA Scheraga, et al. Helix probability profiles of denatured proteins and their correlation with native structures. *Proceedings of the National Academy of Sciences*, 65(4):810–815, 1970. 15, 17
- [67] TL Religa, JS Markson, U Mayor, SMV Freund, and AR Fersht. Solution structure of a protein denatured state and folding intermediate. *Nature*, 437(7061):1053, 2005. 15
- [68] FM Hughson, D Barrick, and RL Baldwin. Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. *Biochemistry*, 30(17):4113–4118, 1991. 15
- [69] RL Baldwin and GD Rose. Molten globules, entropy-driven conformational change and protein folding. *Current opinion in structural biology*, 23(1):4–10, 2013. 15
- [70] CL Brooks, M Gruebele, JN Onuchic, and PG Wolynes. Chemical physics of protein folding. *Proceedings of the National Academy of Sciences*, 95(19):11037–11038, 1998. 15
- [71] AR Panchenko, Z Luthey-Schulten, and PG Wolynes. Foldons, protein structural modules, and exons. *Proceedings of the National Academy of Sciences*, 93(5):2008–2013, 1996. 15
- [72] Y Shiu, U Jeng, Y Huang, Y Lai, H Lu, C Liang, I Hsu, C Su, I Chao, et al. Global and local structural changes of cytochrome c and lysozyme characterized by a multigroup unfolding process. *Biophysical Journal*, 94(12):4828–4836, 2008. 15
- [73] O Bilsel, JA Zitzewitz, KE Bowers, and CR Matthews. Folding mechanism of the α -subunit of tryptophan synthase, an α/β barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry*, 38(3):1018–1029, 1999. 15
- [74] SW Englander and L Mayne. The nature of protein folding pathways. *Proceedings of the National Academy of Sciences*, 111(45):15873–15880, 2014. 15

- [75] PG Wolynes, JN Onuchic, D Thirumalai, et al. Navigating the folding routes. *Science*, pages 1619–1619, 1995. 16
- [76] A Dwevedi. *Protein folding: Examining the challenges from synthesis to folded form*. Springer, 2014. 16, 17
- [77] S Rackovsky and HA Scheraga. Hydrophobicity, hydrophilicity, and the radial and orientational distributions of residues in native proteins. *Proceedings of the National Academy of Sciences*, 74(12):5248–5251, 1977. 16
- [78] KA Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509, 1985. 16
- [79] GD Rose and JE Dworkin. The hydrophobicity profile. In GD Fasman, editor, *Prediction of Protein Structure and the Principles of Protein Conformation*, pages 625–633. Springer, 1989. 16
- [80] GD Rose and R Wolfenden. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annual review of biophysics and biomolecular structure*, 22(1):381–415, 1993. 16
- [81] KH Mok, LT Kuhn, M Goez, IJ Day, JC Lin, NH Andersen, and PJ Hore. A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature*, 447(7140):106, 2007. 16
- [82] GI Makhatadze and PL Privalov. Energetics of protein structure. *Advances in protein chemistry*, 47:307–425, 1995. 16
- [83] CB Anfinsen and HA Scheraga. Experimental and theoretical aspects of protein folding. *Advances in protein chemistry*, 29:205–300, 1975. 16
- [84] H Roder, GA Elöve, and SW Englander. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature*, 335(6192):700–704, 1988. 17
- [85] KA Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133–7155, 1990. 17

-
- [86] DB Wetlaufer. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proceedings of the National Academy of Sciences*, 70(3):697–701, 1973. 17
- [87] M Karplus and DL Weaver. Protein-folding dynamics. *Nature*, 260(5550):404–406, 1976. 17
- [88] H Wu. Studies on denaturation of proteins. A theory of denaturation. *Chinese Journal of Physiology*, 5(4):321–344, 1931. 17
- [89] C Rimington et al. Protein structure and denaturation. *Nature*, 127:440–441, 1931. 17
- [90] Y Zhang and PS Cremer. Interactions between macromolecules and ions: The Hofmeister series. *Current opinion in chemical biology*, 10(6):658–663, 2006. 17
- [91] F Hofmeister. Zur Lehre von der Wirkung der Salze. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 25(1):1–30, 1888. 17
- [92] W Ramsden. Some new properties of urea. *The Journal of Physiology*, 28:23–27, 1902. 17
- [93] FG Hopkins. Denaturation of proteins by urea and related substances. *Nature*, 126(3174):328–330, 1930. 17
- [94] GI Makhatadze and PL Privalov. Protein interactions with urea and guanidinium chloride: a calorimetric study. *Journal of molecular biology*, 226(2):491–505, 1992. 17
- [95] Q Zou, BJ Bennion, V Daggett, and KP Murphy. The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea. *Journal of the American Chemical Society*, 124(7):1192–1202, 2002. 17
- [96] SS Cho, G Reddy, JE Straub, and D Thirumalai. Entropic stabilization of proteins by TMAO. *The Journal of Physical Chemistry B*, 115(45):13401–13407, 2011. 17
- [97] JM Sanz and AR Fersht. Rationally designing the accumulation of a folding intermediate of barnase by protein engineering. *Biochemistry*, 32(49):13584–13592, 1993. 17

- [98] VN Uversky and OB Ptitsyn. Further evidence on the equilibrium ‘pre-molten globule state’: four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *Journal of Molecular Biology*, 255(1):215–228, 1996. 17
- [99] M Auton and DW Bolen. Application of the transfer model to understand how naturally occurring osmolytes affect protein stability. *Methods in enzymology*, 428:397–418, 2007. 18
- [100] SN Timasheff. Solvent stabilization of protein structure. *Protein Stability and Folding: Theory and Practice*, pages 253–269, 1995. 18
- [101] SN Timasheff. A physicochemical basis for the selection of osmolytes by nature. In *Water and life*, pages 70–84. Springer, 1992. 18
- [102] C Tanford. Protein denaturation. *Advances in Protein Chemistry*, 23:121–282, 1968. 18
- [103] RF Greene and CN Pace. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α -chymotrypsin, and β -lactoglobulin. *Journal of Biological Chemistry*, 249(17):5388–5393, 1974. 18
- [104] M Auton, LMF Holthauzen, and DW Bolen. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proceedings of the National Academy of Sciences*, 104(39):15317–15322, 2007. 19, 105, 156
- [105] B Moeser and D Horinek. Unified description of urea denaturation: backbone and side chains contribute equally in the transfer model. *The Journal of Physical Chemistry B*, 118(1):107–114, 2013. 19, 105, 156
- [106] MJ Lindberg, R Byström, N Boknäs, PM Andersen, and M Oliveberg. Systematically perturbed folding patterns of amyotrophic lateral sclerosis (ALS)-associated SOD1 mutants. *Proceedings of the National Academy of Sciences*, 102(28):9754–9759, 2005. 19
- [107] B Schuler, EA Lipman, and WA Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419(6908):743–747, 2002. 19

- [108] EA Roman and FL González Flecha. Kinetics and thermodynamics of membrane protein folding. *Biomolecules*, 4(1):354–373, 2014. 19
- [109] J Lohöfener, N Steinke, P Kay-Fedorov, P Baruch, A Nikulin, S Tishchenko, DJ Manstein, and R Fedorov. The activation mechanism of 2'-5'-oligoadenylate synthetase gives new insights into OAS/cGAS triggers of innate immunity. *Structure*, 23(5):851–862, 2015. 23
- [110] X Cheng and BP Schoenborn. Hydration in protein crystals. a neutron diffraction analysis of carbonmonoxymyoglobin. *Acta Crystallographica Section B: Structural Science*, 46(2):195–208, 1990. 23
- [111] DI Svergun, S Richard, MHJ Koch, Z Sayers, S Kuprin, and G Zaccai. Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proceedings of the National Academy of Sciences*, 95(5):2267–2272, 1998. 23
- [112] H Mansur, R Oréfice, M Pereira, Z Lobato, W Vasconcelos, and L Machado. FTIR and UV-vis study of chemically engineered biomaterial surfaces for protein immobilization. *Journal of Spectroscopy*, 16(3-4):351–360, 2002. 23
- [113] T Beitlich, K Kühnel, C Schulze-Briesse, RL Shoeman, and I Schlichting. Cryoradiolytic reduction of crystalline heme proteins: analysis by UV-vis spectroscopy and X-ray crystallography. *Journal of Synchrotron Radiation*, 14(1):11–23, 2007. 23
- [114] MR Eftink. Fluorescence techniques for studying protein structure. *Methods of Biochemical Analysis: Protein Structure Determination*, 35:127–205, 2006. 23
- [115] WC Johnson. Protein secondary structure and circular dichroism: a practical guide. *Proteins: Structure, Function, and Bioinformatics*, 7(3):205–214, 1990. 23
- [116] AJ Miles and BA Wallace. Circular dichroism spectroscopy of membrane proteins. *Chemical Society Reviews*, 45(18):4859–4872, 2016. 23
- [117] G Wagner, SG Hyberts, and TF Havel. NMR structure determination in solution: a critique and comparison with X-ray crystallography. *Annual review of biophysics and biomolecular structure*, 21(1):167–198, 1992. 23

-
- [118] MR Jensen, M Zweckstetter, J Huang, and M Blackledge. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chemical reviews*, 114(13):6632–6660, 2014. 23
- [119] DS Sivia. *Elementary scattering theory: For X-ray and neutron users*. Oxford University Press, 2011. 24, 25, 31
- [120] BTM Willis and CJ Carlile. *Experimental neutron scattering*. Oxford University Press, 2017. 24
- [121] AK Soper. The radial distribution functions of water as derived from radiation total scattering experiments: Is there anything we can say for sure? *International Scholarly Research Notices: Physical Chemistry*, 2013:1–67, 2013. 24, 76, 78, 138, 139
- [122] DA Skoog, FJ Holler, and SR Crouch. *Principles of instrumental analysis*. Cengage learning, 2017. 25
- [123] JT Pelton and LR McLean. Spectroscopic methods for analysis of protein secondary structure. *Analytical biochemistry*, 277(2):167–176, 2000. 25
- [124] T Schlick, R Collepardo-Guevara, LA Halvorsen, S Jung, and X Xiao. Biomolecular modeling and simulation: a field coming of age. *Quarterly reviews of biophysics*, 44(2):191–228, 2011. 25, 29
- [125] ER Hernández. Molecular Dynamics: from basic techniques to applications (A Molecular Dynamics Primer). *AIP Conference Proceedings*, 1077(1):95–123, 2008. 25, 29
- [126] MP Allen et al. Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins*, 23:1–28, 2004. 25, 29
- [127] MP Allen and DJ Tildesley. *Computer simulation of liquids*. Oxford University press, 1989. 25
- [128] DJ Earl and MW Deem. Monte Carlo simulations. *Molecular modeling of proteins*, pages 25–36, 2008. 25

- [129] JP Ulmschneider, MB Ulmschneider, and A Di Nola. Monte Carlo vs molecular dynamics for all-atom polypeptide folding simulations. *The Journal of Physical Chemistry B*, 110(33):16733–16742, 2006. 25
- [130] LJ Walport, R Morra, EJ Mancini, and C Redfield. ^1H , ^{13}C , and ^{15}N resonance assignments for the tandem PHD finger motifs of human CHD4. *Biomolecular NMR assignments*, 9(2):239–242, 2015. 26
- [131] JW Akitt and BE Mann. *NMR and Chemistry: An introduction to modern NMR spectroscopy*. CRC Press, 2000. 26
- [132] AE Derome. *Modern NMR techniques for chemistry research*, volume 6. Elsevier, 2013. 26
- [133] H Rattle. *An NMR primer for life scientists*. Partnership press, 1995. 26
- [134] PG Takis, KD Papavasileiou, LD Peristeras, VS Melissas, and AN Troganis. Probing micro-solvation in numbers: the case of neutral dipeptides in water. *Physical Chemistry Chemical Physics*, 15(19):7354–7362, 2013. 27
- [135] SM Kelly, TJ Jess, and NC Price. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1751(2):119–139, 2005. 27
- [136] G Garab and H van Amerongen. Linear dichroism and circular dichroism in photosynthesis research. *Photosynthesis research*, 101(2-3):135–146, 2009. 27
- [137] NJ Greenfield. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*, 1(6):2876, 2006. 27
- [138] JLS Lopes, AJ Miles, L Whitmore, and BA Wallace. Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: Applications in secondary structure analyses. *Protein Science*, 23(12):1765–1772, 2014. 27
- [139] LA Compton and WC Johnson. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Analytical biochemistry*, 155(1):155–167, 1986. 27

- [140] AJ Miles and BA Wallace. Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics. *Chemical Society Reviews*, 35(1):39–51, 2006. 27
- [141] S Brahms and J Brahms. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *Journal of molecular biology*, 138(2):149–178, 1980. 29
- [142] RO Dror, RM Dirks, JP Grossman, H Xu, and DE Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, 41:429–452, 2012. 29
- [143] ER Lindahl. Molecular dynamics simulations. *Molecular modeling of proteins*, pages 3–23, 2008. 29
- [144] K Binder, J Horbach, W Kob, W Paul, and F Varnik. Molecular dynamics simulations. *Journal of Physics: Condensed Matter*, 16(5):S429, 2004. 29
- [145] WF van Gunsteren and HJC Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition*, 29(9):992–1023, 1990. 29
- [146] BR Brooks, RE Brucoleri, BD Olafson, DJ States, S Swaminathan, and M Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983. 30, 50
- [147] BR Brooks, CL Brooks, AD MacKerell, L Nilsson, RJ Petrella, B Roux, Y Won, G Archontis, C Bartels, S Boresch, et al. CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614, 2009. 30, 50
- [148] AD MacKerell, D Bashford, M Bellott, RL Dunbrack, JD Evanseck, MJ Field, S Fischer, J Gao, H Guo, S Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998. 30, 50, 67, 69, 112, 113, 166
- [149] WD Cornell, P Cieplak, CI Bayly, IR Gould, KM Merz, DM Ferguson, DC Spellmeyer, T Fox, JW Caldwell, and PA Kollman. A second generation

- force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995. 30
- [150] WL Jorgensen, DS Maxwell, and J Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996. 30
- [151] HJC Berendsen, JR Grigera, and TP Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271, 1987. 31, 50, 166
- [152] WL Jorgensen, J Chandrasekhar, JD Madura, RW Impey, and ML Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926, 1983. 31, 50, 67, 69, 112, 113
- [153] IA Zaliznyak and S Lee. Magnetic neutron scattering. Technical report, Brookhaven National Laboratory, 2004. 31
- [154] R Pynn. Neutron scattering: a primer. *Los Alamos Science*, 19:1–31, 1990. 31, 43
- [155] GL Squires. *Introduction to the theory of thermal neutron scattering*. Cambridge University Press, 2012. 31
- [156] L Sanche. Low-energy electron scattering from molecules on surfaces. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 23(10):1597, 1990. 33
- [157] R Hofstadter. Electron scattering and nuclear structure. *Reviews of Modern Physics*, 28(3):214, 1956. 33
- [158] E Rutherford. Bakerian lecture: Nuclear constitution of atoms. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 97(686):374–400, 1920. 33
- [159] J Chadwick. Possible existence of a neutron. *Nature*, 129(3252):312, 1932. 33
- [160] J Chadwick. The existence of a neutron. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 136, pages 692–708, 1932. 33

- [161] WC Röntgen. Über eine neue Art von Strahlen. *Annalen der Physik*, 300(1):1–11, 1898. 33
- [162] SE McLain, CJ Benmore, JE Siewenie, J Urquidi, and JFC Turner. On the structure of liquid hydrogen fluoride. *Angewandte Chemie International Edition*, 43(15):1952–1955, 2004. 36
- [163] AK Soper and RN Silver. Hydrogen-hydrogen pair correlation function in liquid water. *Physical Review Letters*, 49(7):471, 1982. 36
- [164] VF Sears. Neutron scattering lengths and cross sections. *Neutron News*, 3(3):26–37, 1992. 36, 38, 48
- [165] WH McMaster, NK Del Grande, JH Mallett, and JH Hubbell. Compilation of X-ray cross sections. Technical report, California University, Lawrence Radiation Laboratory, 1969. 38
- [166] MM Castellanos, A McAuley, and JE Curtis. Investigating structure and dynamics of proteins in amorphous phases using neutron scattering. *Computational and structural biotechnology journal*, 2016. 38
- [167] SM King. Small angle neutron scattering. *Modern Techniques for Polymer Characterisation*, 171, 1999. 41
- [168] C Benmore and A Soper. The SANDALS Manual. Technical report, ISIS Facility, Rutherford Appleton Laboratory, 1998. 42, 45
- [169] DT Bowron, AK Soper, K Jones, S Ansell, S Birch, J Norris, L Perrott, D Riedel, NJ Rhodes, SR Wakefield, et al. NIMROD: The Near and InterMediate Range Order Diffractometer of the ISIS second target station. *Review of Scientific Instruments*, 81(3), 2010. 42, 45
- [170] How ISIS works, in depth. <http://www.isis.stfc.ac.uk/about/how-isis-works---in-depth4371.html>. Accessed: 14-09-2017. 42
- [171] JJ Towey. *A structural approach to reveal the cryoprotective action of glycerol*. PhD thesis, University of Leeds, 2013. 44

- [172] AK Soper. GudrunN and GudrunX: programs for correcting raw neutron and X-ray diffraction data to differential scattering cross section. Technical report, Science & Technology Facilities Council, 2011. 46, 112
- [173] AK Soper, WS Howells, and AC Hannon. ATLAS – Analysis of Time-of-Flight Diffraction Data from Liquid and Amorphous Samples. Technical report, Rutherford Appleton Laboratory, 1989. 46
- [174] J Mayers. The use of vanadium as a scattering standard for pulsed source neutron spectrometers. *Nuclear Instruments and Methods in Physics Research*, 221(3):609–618, 1984. 46
- [175] AK Soper. Multiple scattering from an infinite plane slab. *Nuclear Instruments and Methods In Physics Research*, 212(1):337–347, 1983. 47
- [176] G Placzek. The scattering of neutrons by systems of heavy nuclei. *Physical review*, 86(3):377, 1952. 47
- [177] U Dahlborg and B Kunsch. On the Placzek correction for simple liquids and gases. *Physics and Chemistry of Liquids*, 12(3):237–254, 1983. 47
- [178] AK Soper. Empirical potential Monte Carlo simulation of fluid structure. *Chemical Physics*, 202(2):295–306, 1996. 48
- [179] AK Soper. Computer simulation as a tool for the interpretation of total scattering data from glasses and liquids. *Molecular Simulation*, 38(14-15):1171–1185, 2012. 48
- [180] AK Soper. Empirical potential structure refinement Version 18 ‘EPSRshell’: a user’s guide. Technical report, Rutherford Appleton Laboratory, 2011. 49
- [181] WL Jorgensen and J Tirado-Rives. Monte carlo vs molecular dynamics for conformational sampling. *The Journal of Physical Chemistry*, 100(34):14508–14513, 1996. 51
- [182] LC Pardo. ANGULA. <https://gcm.upc.edu/en/members/luis-carlos/angula/ANGULA>, note = Accessed: 01-08-2017. 54

- [183] S Busch, LC Pardo, WB O'Dell, CD Bruce, CD Lorenz, and SE McLain. On the structure of water and chloride ion interactions with a peptide backbone in solution. *Physical Chemistry Chemical Physics*, 15(48):21023–21033, 2013. 54, 56, 64, 66, 67, 78, 80, 81, 83, 85, 87, 88, 92, 95, 96, 97, 98, 103
- [184] S Busch, CD Lorenz, J Taylor, LC Pardo, and SE McLain. Short-range interactions of concentrated proline in aqueous solution. *The Journal of Physical Chemistry B*, 118(49):14267–14277, 2014. 56, 70
- [185] AJ Johnston, S Busch, LC Pardo, SK Callear, PC Biggin, and SE McLain. On the atomic structure of cocaine in solution. *Physical Chemistry Chemical Physics*, 18(2):991–999, 2016. 59
- [186] DB Watlaufer, SK Malik, L Stoller, and RL Coffin. Nonpolar group participation in the denaturation of proteins by urea and guanidinium salts. model compound studies. *Journal of the American Chemical Society*, 86(3):508–514, 1964. 63
- [187] HS Frank and F Franks. Structural approach to the solvent power of water for hydrocarbons; urea as a structure breaker. *Journal of Chemical Physics*, 48(10):4746–4757, 1968. 63, 193
- [188] YLA Rezus and HJ Bakker. Effect of urea on the structural dynamics of water. *Proceedings of the National Academy of Sciences*, 103(49):18417–18420, 2006. 63, 139
- [189] DR Canchi and AE García. Backbone and side-chain contributions in protein denaturation by urea. *Biophysical Journal*, 100(6):1526–1533, 2011. 63
- [190] MC Stumpe and H Grubmüller. Urea impedes the hydrophobic collapse of partially unfolded proteins. *Biophysical Journal*, 96(9):3744–3752, 2009. 64
- [191] EP O'Brien, RI Dima, B Brooks, and D Thirumalai. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *Journal of the American Chemical Society*, 129(23):7346–7353, 2007. 64
- [192] TJ Bowen. Physical studies on a soluble protein obtained by the degradation of elastin with urea. *Biochemical Journal*, 55(5):766, 1953. 64

- [193] RB Simpson and W Kauzmann. The kinetics of protein denaturation. the behavior of the optical rotation of ovalbumin in urea solutions. *Journal of the American Chemical Society*, 75(21):5139–5152, 1953. 64
- [194] T Raskar, S Khavnekar, and M Hosur. Time-dependent X-ray diffraction studies on urea/hen egg white lysozyme complexes reveal structural changes that indicate onset of denaturation. *Scientific Reports*, 6, 2016. 64, 155
- [195] L Hua, R Zhou, D Thirumalai, and BJ Berne. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proceedings of the National Academy of Sciences*, 105(44):16928–16933, 2008. 64, 105
- [196] PJ Rossky. Protein denaturation by urea: slash and bond. *Proceedings of the National Academy of Sciences*, 105(44):16825–16826, 2008. 64
- [197] BJ Bennion and V Daggett. The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences*, 101(17):6433–6438, 2003. 64
- [198] S Busch, CD Bruce, C Redfield, CD Lorenz, and SE McLain. Water mediation is essential to nucleation of β -turn formation in peptide folding motifs. *Angewandte Chemie International Edition*, 52(49):13091–13095, 2013. 64, 67, 76, 88, 109, 196
- [199] MS Cheung, AE García, and JN Onuchic. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences*, 99(2):685–690, 2002. 64, 108
- [200] PN Lewis, FA Momany, and HA Scheraga. Folding of polypeptide chains in proteins: A proposed mechanism for folding. *Proceedings of the National Academy of Sciences*, 68(9):2293–2297, 1971. 64, 108
- [201] PY Chou and GD Fasman. Prediction of protein conformation. *Biochemistry*, 13(2):222–245, 1974. 64, 108
- [202] CM Wilmot and JM Thornton. Analysis and prediction of the different types of β -turn in proteins. *Journal of Molecular Biology*, 203(1):221–232, 1988. 64, 108

- [203] K Guruprasad and S Rajkumar. Beta-and gamma-turns in proteins revisited: A new set of amino acid turn-type dependent positional preferences and potentials. *Journal of Bioscience*, 25(2):143–156, 2000. 64, 108
- [204] W Schlenk. The man behind the flask. *Angewandte Chemie, International Edition*, 40(2):331–337, 2001. 65
- [205] EJ Greenhow. Acetylation as a means of end-point indication in the catalytic thermometric titration of bases. *Analyst*, 102(1217):584–590, 1977. 66
- [206] RB Best, X Zhu, J Shim, PEM Lopes, J Mittal, M Feig, and AD MacKerell. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, 2012. 67, 112, 113, 166
- [207] AD MacKerell, M Feig, and CL Brooks. 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. *Journal of Computational Chemistry*, 25(11):1400–1415, 2004. 67, 69, 112, 113, 166
- [208] K Vanommeslaeghe, E Hatcher, C Acharya, S Kundu, S Zhong, J Shim, E Darian, O Guvench, P Lopes, I Vorobyov, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690, 2010. 67, 112, 113, 166
- [209] W Yu, X He, K Vanommeslaeghe, and AD MacKerell. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry*, 33(31):2451–2468, 2012. 67, 112, 113, 166
- [210] YK Kang and HY Choi. Cis–trans isomerization and puckering of proline residue. *Biophysical Chemistry*, 111(2):135–142, 2004. 68

- [211] WE Reiher. *Theoretical Studies of Hydrogen Bonding*. PhD thesis, Harvard University, 1985. 69, 113
- [212] J Ryckaert, G Ciccotti, and HJC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977. 69
- [213] B Hess, C Kutzner, D van der Spoel, and E Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008. 69
- [214] L Martínez, R Andrade, EG Birgin, and JM Martínez. Packmol: a package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, 2009. 69
- [215] S Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984. 69
- [216] WG Hoover. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985. 69
- [217] GJ Martyna, ME Tuckerman, DJ Tobias, and ML Klein. Explicit reversible integrators for extended systems dynamics. *Molecular Physics*, 87(5):1117–1157, 1996. 69
- [218] T Darden, D York, and L Pedersen. Particle mesh Ewald: An $n \cdot \log(n)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089, 1993. 69
- [219] U Essmann, L Perera, ML Berkowitz, T Darden, H Lee, and LG Pedersen. A smooth particle mesh Ewald method. *Journal of Chemical Physics*, 103(19):8577, 1995. 69
- [220] LC Pardo, M Rovira-Esteva, JL Tamarit, N Veglio, JF Bermejo, and GJ Cuello. A procedure to quantify the short range order of disordered phases. In R Sylwester, D Aleksandra, and M Victor, editors, *Metastable Systems under Pressure*, pages 79–91. Springer, 2010. 70

- [221] MC Stumpe and H Grubmüller. Interaction of urea with amino acids: implications for urea-induced protein denaturation. *Journal of the American Chemical Society*, 129(51):16126–16131, 2007. 70, 71, 105, 156
- [222] AK Soper. The atomic structure of water in 1:14 urea:water solutions. Unpublished data, courtesy of AK Soper, 2015. 76, 78, 79, 80, 83, 84
- [223] R Mancinelli, A Botti, F Bruni, MA Ricci, and A K Soper. Perturbation of water structure due to monovalent ions in solution. *Physical Chemistry Chemical Physics*, 9:2959–2967, 2007. 76, 193
- [224] AH Narten and HA Levy. Observed diffraction pattern and proposed models of liquid water. *Science*, 165(3892):447–454, 1969. 76
- [225] AK Soper, EW Castner Jr, and A Luzar. Impact of urea on water structure: a clue to its properties as a denaturant? *Biophysical Chemistry*, 105(2):649–666, 2003. 76, 103, 193
- [226] F Meersman, D Bowron, AK Soper, and MHJ Koch. An X-ray and neutron scattering study of the equilibrium between trimethylamine N-oxide and urea in aqueous solution. *Physical Chemistry Chemical Physics*, 13:13765–13771, 2011. 76, 103
- [227] P Mark and L Nilsson. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A*, 105(43):9954–9960, 2001. 76, 139
- [228] F Bruni, S Imberti, R Mancinelli, and MA Ricci. Aqueous solutions of divalent chlorides: ions hydration shell and water structure. *The Journal of Chemical Physics*, 136(6):064520, 2012. 77
- [229] AJ Johnston, Y Zhang, S Busch, LC Pardo, S Imberti, and SE McLain. Amphipathic solvation of indole: Implications for the role of tryptophan in membrane proteins. *The Journal of Physical Chemistry B*, 119(19):5979–5987, 2015. 99
- [230] V Amendola, L Fabbrizzi, and L Mosca. Anion recognition by hydrogen bonding: urea-based receptors. *Chemical Society Reviews*, 39:3889–3915, 2010. 103

- [231] PO Åstrand, A Wallqvist, G Karlström, and P Linse. Properties of urea-water solvation calculated from a new abinitio polarizable intermolecular potential. *The Journal of Chemical Physics*, 95(11):8419–8429, 1991. 103
- [232] F Ahmad and CC Bigelow. Inorganic salt denaturants stabilize ribonuclease against denaturation by urea. *Canadian Journal of Biochemistry*, 56(11):1003–1005, 1978. 107
- [233] S Muzammil, Y Kumar, and S Tayyab. Anion-induced stabilization of human serum albumin prevents the formation of intermediate during urea denaturation. *Proteins: Structure, Function, and Bioinformatics*, 40(1):29–38, 2000. 107, 156, 191
- [234] A Wang and DW Bolen. A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry*, 36(30):9101–9108, 1997. 107, 156
- [235] MS Dodia, HG Bhimani, CM Rawal, RH Joshi, and SP Singh. Salt dependent resistance against chemical denaturation of alkaline protease from a newly isolated haloalkaliphilic bacillus sp. *Bioresource technology*, 99(14):6223–6227, 2008. 107
- [236] J Chen, X Gong, C Zeng, Y Wang, and G Zhang. Mechanical insight into resistance of betaine to urea-induced protein denaturation. *The Journal of Physical Chemistry B*, 120(48):12327–12333, 2016. 107, 156
- [237] CN Pace, DV Laurents, and RE Erickson. Urea denaturation of barnase: pH dependence and characterization of the unfolded state. *Biochemistry*, 31(10):2728–2734, 1992. 107
- [238] CN Pace, DV Laurents, and JA Thomson. pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. *Biochemistry*, 29(10):2564–2572, 1990. 107
- [239] AC Dumetz, AM Snellinger-O’Brien, EW Kaler, and AM Lenhoff. Patterns of protein–protein interactions in salt solutions and implications for protein crystallization. *Protein Science*, 16(9):1867–1877, 2007. 108

- [240] MW MacArthur and JM Thornton. Influence of proline residues on protein conformation. *Journal of Molecular Biology*, 218(2):397–412, 1991. 108
- [241] P Craveur, AP Joseph, P Poulain, AG de Brevern, and J Rebehmed. Cis–trans isomerization of omega dihedrals in proteins. *Amino Acids*, 45(2):279–289, 2013. 108
- [242] C Grathwohl and K Wüthrich. NMR studies of the rates of proline cis–trans isomerization in oligopeptides. *Biopolymers*, 20(12):2623–2633, 1981. 108
- [243] SCR Lummis, DL Beene, LW Lee, HA Lester, et al. Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature*, 438(7065):248, 2005. 108
- [244] MP Williamson. The structure and function of proline-rich regions in proteins. *Biochemical Journal*, 297(Pt 2):249, 1994. 108
- [245] AL Morris, MW MacArthur, EG Hutchinson, and JM Thornton. Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function, and Bioinformatics*, 12(4):345–364, 1992. 108
- [246] PDBeMotif. <http://www.ebi.ac.uk/pdbe-site/pdbemotif/>. Accessed: 11-08-2017. 108
- [247] AK Soper. Inelasticity corrections for time-of-flight and fixed wavelength neutron diffraction experiments. *Molecular Physics*, 107(16):1667–1684, 2009. 119
- [248] SE McLain, AK Soper, I Daidone, JC Smith, and A Watts. Charge-based interactions between peptides observed as the dominant force for association in aqueous solution. *Angewandte Chemie International Edition*, 47(47):9059–9062, 2008. 141, 193
- [249] N Steinke, RJ Gillams, LC Pardo, CD Lorenz, and SE McLain. Atomic scale insights into urea–peptide interactions in solution. *Physical Chemistry Chemical Physics*, 18(5):3862–3870, 2016. 155, 156, 191
- [250] Z Su and CL Dias. Molecular interactions accounting for protein denaturation by urea. *Journal of Molecular Liquids*, 228:168–175, 2017. 156

- [251] W Li, R Zhou, and Y Mu. Salting effects on protein components in aqueous NaCl and urea solutions: toward understanding of urea-induced protein denaturation. *The Journal of Physical Chemistry B*, 116(4):1446–1451, 2012. 156
- [252] PH Yancey, ME Clark, SC Hand, RD Bowlus, and GN Somero. Living with water stress: evolution of osmolyte systems. *Science*, 217(4566):1214–1222, 1982. 156
- [253] K Campbell and L Kurgan. Sequence-only based prediction of β -turn location and type using collocation of amino acid pairs. *Open Bioinformatics Journal*, 2:37–49, 2008. 158
- [254] H Hsu, H Chang, H Peng, S Huang, M Lin, and A Yang. Assessing computational amino acid β -turn propensities with a phage-displayed combinatorial library and directed evolution. *Structure*, 14(10):1499–1510, 2006. 158
- [255] HJ Dyson, M Rance, RA Houghten, RA Lerner, and PE Wright. Folding of immunogenic peptide fragments of proteins in water solution. *Journal of Molecular Biology*, 201(1):161–200, 1988. 158
- [256] AMC Marcelino and LM Gierasch. Roles of β -turns in protein folding: From peptide models to protein engineering. *Biopolymers*, 89(5):380–391, 2008. 158
- [257] B Rost. Protein secondary structure prediction continues to rise. *Journal of Structural Biology*, 134(2-3):204–218, 2001. 161
- [258] C Cole, JD Barber, and GJ Barton. The Jpred 3 secondary structure prediction server. *Nucleic acids research*, 36(2):W197–W201, 2008. 161
- [259] J Zhou and O Troyanskaya. Deep supervised and convolutional generative stochastic network for protein secondary structure prediction. In *International Conference on Machine Learning*, pages 745–753, 2014. 161
- [260] JM Scholtz and RL Baldwin. The mechanism of α -helix formation by peptides. *Annual review of biophysics and biomolecular structure*, 21(1):95–118, 1992. 161
- [261] S Akiyama, S Takahashi, K Ishimori, and I Morishima. Stepwise formation of $[\alpha]$ -helices during cytochrome c folding. *Nature Structural & Molecular Biology*, 7(6):514, 2000. 161

- [262] JM Scholtz, D Barrick, EJ York, JM Stewart, and RL Baldwin. Urea unfolding of peptide helices as a model for interpreting protein unfolding. *Proceedings of the National Academy of Sciences*, 92(1):185–189, 1995. 161
- [263] JK Myers, CN Pace, and MJ Scholtz. Denaturant m -values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science*, 4(10):2138–2148, 1995. 161
- [264] OB Ptitsyn, RH Pain, GV Semisotnov, E Zerovnik, and OI Razgulyaev. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS letters*, 262(1):20–24, 1990. 161
- [265] IK Lednev, AS Karnoup, MC Sparrow, and SA Asher. α -Helix peptide folding and unfolding activation barriers: a nanosecond UV resonance Raman study. *Journal of the American Chemical Society*, 121(35):8074–8086, 1999. 161
- [266] KV Soman, A Karimi, and DA Case. Unfolding of an α -helix in water. *Biopolymers*, 31(12):1351–1361, 1991. 161
- [267] I Cornut, E Thiaudiere, and J Dufourcq. The amphipathic helix in cytotoxic peptides. In *The amphipathic helix*, pages 173–219. CRC Press, 1993. 161, 189
- [268] KT O’Neil and WF DeGrado. How calmodulin binds its targets: sequence independent recognition of amphiphilic α -helices. *Trends in Biochemical Sciences*, 15(2):59–64, 1990. 161
- [269] G von Heijne. Mitochondrial targeting sequences may form amphiphilic helices. *The EMBO journal*, 5(6):1335, 1986. 161
- [270] J Segrest, MK Jones, H De Loof, CG Brouillette, YV Venkatachalapathi, and GM Anantharamaiah. The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function. *Journal of Lipid Research*, 33(2):141–166, 1992. 161
- [271] JD Lear, ZR Wasserman, and WF DeGrado. Synthetic amphiphilic peptide models for protein ion channels. *Science*, 240(4856):1177–1181, 1988. 161

- [272] N Wiradharma, U Khoe, CAE Hauser, SV Seow, S Zhang, and Y Yang. Synthetic cationic amphiphilic α -helical peptides as antimicrobial agents. *Biomaterials*, 32(8):2204–2212, 2011. 161
- [273] KA Brogden. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology*, 3(3), 2005. 161
- [274] I Cornut, K Büttner, J Dasseux, and J Dufourcq. The amphipathic α -helix concept: application to the de novo design of ideally amphipathic leu, lys peptides with hemolytic activity higher than that of melittin. *FEBS letters*, 349(1):29–33, 1994. 162
- [275] H Vogel and F Jaehnig. The structure of melittin in membranes. *Biophysical Journal*, 50(4):573–582, 1986. 162, 189
- [276] N Sreerama and RW Woody. Protein secondary structure from circular dichroism spectroscopy: Combining variable selection principle and cluster analysis with neural network, ridge regression and self-consistent methods. *Journal of Molecular Biology*, 242(4):497–507, 1994. 168
- [277] WL DeLano. The PyMOL molecular graphics system. <http://pymol.org>, 2002. 176
- [278] H Nymeyer, TB Woolf, and AE Garcia. Folding is not required for bilayer insertion: Replica exchange simulations of an α -helical peptide with an explicit lipid bilayer. *Proteins: Structure, Function, and Bioinformatics*, 59(4):783–790, 2005. 190
- [279] J Bürck, S Roth, P Wadhwani, S Afonin, N Kanithasen, E Strandberg, and AS Ulrich. Conformation and membrane orientation of amphiphilic helical peptides by oriented circular dichroism. *Biophysical journal*, 95(8):3872–3881, 2008. 190
- [280] CN Pace, BMP Huyghues-Despointes, H Fu, K Takano, JM Scholtz, and GR Grimsley. Urea denatured state ensembles contain extensive secondary structure that is increased in hydrophobic proteins. *Protein Science*, 19(5):929–943, 2010. 190
- [281] LMF Holthauzen, J Rösger, and DW Bolen. Hydrogen bonding progressively strengthens upon transfer of the protein urea-denatured state to water and protecting osmolytes. *Biochemistry*, 49(6):1310, 2010. 190

-
- [282] LB Sagle, Y Zhang, VA Litosh, X Chen, Y Cho, and PS Cremer. Investigating the hydrogen-bonding model of urea denaturation. *Journal of the American Chemical Society*, 131(26):9304–9310, 2009. 191
- [283] M Sadqi, LJ Lapidus, and V Muñoz. How fast is protein hydrophobic collapse? *Proceedings of the National Academy of Sciences*, 100(21):12117–12122, 2003. 196
- [284] R Zhou, X Huang, CJ Margulis, and BJ Berne. Hydrophobic collapse in multidomain protein folding. *Science*, 305(5690):1605–1609, 2004. 196