On the Role of Hydration in the Folding of Ordered Structures in Peptides
Abstract

Proteins with the same sequence are found to fold into the same 3D structure reliably in solution, yet the exact processes underpinning the interactions between peptide and solvent in the folding process are not well understood. Previous study on the glycyl-prolyl-glycinamide (GPG–NH$_2$) tripeptide has indicated that hydrogen bonding between backbone amide groups and solvent water molecules may mediate the protein fold. This study extends the previous work to a longer peptide, tyrosyl-prolyl-glycyl-serinamide (YPGS–NH$_2$) which is a sequence designed to have $\beta$-turn forming propensity. Whole-molecule and specific interactions between water and hydrogen bonding groups on this peptide are compared using complementary experimental and computational techniques: solution nuclear magnetic resonance spectroscopy, molecular dynamics and empirical potential structure refinement modelling. This multi-faceted approach aims to produce a unified and experimentally consistent model of the system. Understanding the folding process, in particular the role of water, is crucial in understanding clinically how mutations in protein encoding DNA may lead to mutated, and therefore misfolded, proteins which are the root of various diseases.

Keywords: NMR, MD, EPSR, Neutron Diffraction, $\beta$-turn, water, hydration

Abbreviations:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amino acid</td>
<td>L- $\alpha$ amino carboxylic acid found in nature</td>
</tr>
<tr>
<td>EPSR</td>
<td>Empirical Potential Structure Refinement</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>ND</td>
<td>Neutron Diffraction</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
</tbody>
</table>
On the Role of Hydration in the Folding of Ordered Structures in Peptides

Protein folding is an essential life process. In all living cells, proteins are synthesised as linear strands, but most must exist as folded globular structures in order to carry out their biological function. The fold a protein adopts is consistent for a given sequence of amino acids (as encoded by DNA). Since proteins often have very specific biological functions, determined by their folded structure in vivo, cells have the ability to influence which functions are carried out on both the intracellular and extracellular level by tightly controlling DNA transcription and translation of RNA to protein. However, the mechanics behind this reproducibility of folding are not yet completely understood. In particular the specific role of water remains largely unknown, and a cause of scientific contention. (Ben-Naim, 2013)

DNA is the blueprint from which proteins are synthesized, but mutations arise in DNA. If these mutations arise on portions of DNA which encode a protein, the sequence of amino acids that make up the protein may be altered. This in turn can cause the protein to fold differently, adopting a different 3-dimensional structure, which may then have different functionality as compared to a non-mutated variant. Whilst the accumulation of mutations and subsequent environmental selection is the driving force for evolution, mutations are also the cause of a vast range of diseases. Mutated DNA in the germ line may lead to inherited genetic conditions where mutant proteins may undergo; a gain-of-function resulting in an autosomal dominant disease, e.g. hyperthyroidism (Weinstein, 2006); or a loss-of-function resulting in an autosomal recessive condition, e.g. cystic fibrosis (Kelley et al., 1997). Alternatively, mutations acquired in somatic cells may also cause diseases, primarily cancers. Many mutations in proteins however have little effect on protein function, only effecting minimal changes to the folded structure, but there is currently no way to predict which mutations will cause significant protein misfolding detrimental to function.
The ‘hydrophobic effect’ is usually cited as the principal entropic path for protein folding; where it is thought that hydrophobic residues cluster together to eliminate unfavourable water-hydrophobic species interactions. However, recent studies have suggested that water has a more active role in this process, where specifically water is thought to stabilise folded proteins (Huggins, 2016) and aid in nucleating the folding sites in order to guide the protein into its fully folded functional structure. (Blanco et al., 1994; Busch et al., 2013)

My research is focused on the $\beta$-turn; a common 4-residue structural motif found in loop regions in folded proteins (Rose et al., 1985). Misfolding of this structure within proteins has implications in forming plaques, the mechanism of diseases such as Alzheimer’s. (Hsiao et al., 1996) The proline-glycine (PG) motif (Hsu et al., 2006) is known to promote $\beta$-turn formation in loop regions. Short polypeptides containing this motif are ideal for investigating how water contributes to protein folding in solution. In solution, short PG containing peptides exist in an equilibrium between closed and open conformations in solution, allowing the effect of water on the folding process to be probed. Previous study (Busch et al., 2013) on the glycyl-prolyl-glycinamide (GPG-NH$_2$) tripeptide in water suggested that water has an active role in mediating the turn, where a single water molecule engages in multiple hydrogen bonds with the peptide. This may therefore also be true for longer polypeptides and proteins, including in vivo. I am using both experimental and theoretical techniques to study this, with the aim of marrying together molecular dynamic modelling of systems in silico with models which are consistent with experimental data.

I am using a range of experimental and theoretical techniques. Nuclear Magnetic Resonance experimentation is used to gain information on the atomic chemical environments, which are influenced by the folded state of the peptide in solution. Molecular Dynamics is used to model the peptide solutions in silico to study associations of water
molecules to the peptide in a time-dependent manner, including mapping the changes in hydration as a result of peptide folding and unfolding. By combining information from neutron scattering on the peptide in solution with Empirical Potential Structure Refinement (EPSR) modelling, (Soper, 2012) the atomic scale interactions of the peptide with both solvent water molecules and itself can be modelled. Importantly, using the analysis program ANGULA, (Busch et al., 2014) the peptides can be categorised from the EPSR model into different folding states, which allows the detailed assessment of how water interacts and the hydration changes around different atomic sites of peptides which adopt different conformations in solution. (Steinke et al., 2017) For these systems, the information gained by using neutrons is crucial, as it provides insight to interactions on the atomic scale, particularly for hydrogen atoms which are invisible to many other techniques.

The hope is that all these methods will lead to a unified and consistent model, which could have the longer term impact of predicting results for one method from another, thus understanding better the structure and hydration of protein structures, and their implications in causing disease.
Bibliography


