Hydrogen Bonds in Proteins: Role and Strength

Roderick E Hubbard, *University of York, York, UK*

Muhammad Kamran Haider, *University of York, York, UK*

Based in part on the previous version of this Encyclopedia of Life Sciences (ELS) article, Hydrogen Bonds in Proteins: Role and Strength by Roderick E Hubbard.

Advanced article

Article Contents

- Introduction
- What is a Hydrogen Bond?
- Role of Hydrogen Bonds in Protein Stability and Folding
- Strength of the Hydrogen Bond
- Geometry of Secondary Structures
- Role of Hydrogen Bonds in Ligand Binding and Specificity

Online posting date: 15th February 2010

Hydrogen bonds provide most of the directional interactions that underpin protein folding, protein structure and molecular recognition. The core of most protein structures is composed of secondary structures such as α helix and β sheet. This satisfies the hydrogen-bonding potential between main chain carbonyl oxygen and amide nitrogen buried in the hydrophobic core of the protein. Hydrogen bonding between a protein and its ligands (protein, nucleic acid, substrate, effector or inhibitor) provides a directionality and specificity of interaction that is a fundamental aspect of molecular recognition. The energetics and kinetics of hydrogen bonding therefore need to be optimal to allow the rapid sampling and kinetics of folding, conferring stability to the protein structure and providing the specificity required for selective macromolecular interactions.

Introduction

The hydrogen bond is one of the most important classes of molecular interaction in biology. The hydrophobic effect provides the thermodynamic drive for the overall structure of nucleic acids, proteins and membranes in water, through the burial of hydrophobic groups and exposure of the hydrophilic ones. However, hydrogen bonds confer directionality and specificity to the intramolecular interactions in these structures. This interaction is particularly important for proteins, where the hydrogen bond provides

ELS subject area: Structural Biology

How to cite:
Hubbard, Roderick E; and Kamran Haider, Muhammad (February 2010) Hydrogen Bonds in Proteins: Role and Strength. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0003011.pub2

the organization for distinct folds and also provides the selectivity in the protein–ligand interactions that underpin molecular recognition. **See also**: Hydrophobic Effect

What is a Hydrogen Bond?

A hydrogen bond is formed when a proton (H) covalently attached to one electronegative donor atom (D) is shared with another electronegative acceptor atom (A), as illustrated in **Figure 1a**. Although the exact nature of the bond (H–A) is still in debate (Martin and Derewenda, 1999), there has been some progress in developing a theoretical chemistry description (Scheiner, 1997). One of the widely used schemes was proposed by Morokuma (1977) in which *ab initio* calculations describe the interaction energy of a hydrogen bond in terms of electrostatic, charge transfer, polarization, exchange repulsion and coupling. Steiner has summarized the overall characteristics of these different components for different strengths of hydrogen bonding. In general, the electrostatic term dominates. Although van der Waals and charge transfer terms are always present, the dispersion term becomes significant in very weak hydrogen bonds. On the whole, the interaction can be thought of as partially covalent and partially electrostatic in character (Arnold and Oldfield, 2000; Tuttle *et al.*, 2004) but this does not provide an adequate understanding of very strong hydrogen bonds (Steiner, 2002), although these are not seen in protein structures.

Proteins contain many hydrogen bond donors and acceptors – the amide and carbonyl groups of the peptide backbone, as well as the polar functional groups (amides, acids, hydroxyls and amines) on the side-chains of all amino acids except for glycine, proline, alanine, valine, leucine, isoleucine and phenylalanine. Although cystine and methionine contain S–H groups, these form only weak hydrogen bonds. **See also**: Proteins: Fundamental Chemical Properties

The generally used parameters for hydrogen bond geometry have been derived from analyses of the vast amount of crystal structure data available for organic compounds and for protein–ligand complexes. Analysis of
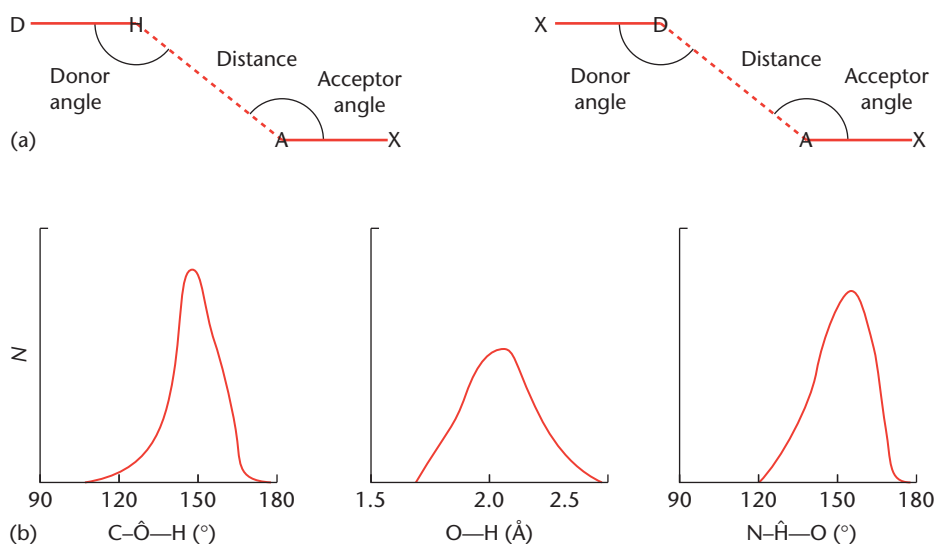
ENCYCLOPEDIA OF LIFE SCIENCES © 2010, John Wiley & Sons, Ltd. www.els.net   1

Figure 1 (a) Schematic representation of the geometry of a hydrogen bond. On the left is the definition of geometry when proton positions are defined; on the right when they are not. D, donor atom; A, acceptor and H, hydrogen. (b) Distribution of geometry for hydrogen bonds in α helices. The plots show approximate distributions (number of occurrences, N) for the angle at the carbonyl oxygen (O) acceptor, distance between the carbonyl oxygen acceptor and amide proton (H) donor and the angle at the amide proton donor and carbonyl oxygen acceptor (Baker and Hubbard, 1984).

small molecule crystal structures of amides shows the distance between the amide group (donor) and the carbonyl oxygen of another molecule (acceptor) to be between 1.85 and 2 Å (O–H). Broader distributions are observed in proteins, as seen in the schematic diagram of the geometry of inter main chain amide hydrogen bonds in α helices in **Figure 1b**. There is a broad spread of observed donor-to-acceptor distances of between 1.7 and 2.4 Å. The angle at the proton is between 130° and 170° and that at the acceptor is more tightly defined approximately 150° (Baker and Hubbard, 1984).

A more detailed analysis of hydrogen-bonding geometry can be found in Baker and Hubbard (1984). The most striking feature of this type of analysis is that the distance between donated protons and acceptors (O or N) is always greater than 1.6 Å and less than 2.5 Å. Also, the angle at both the acceptor and donor is quite tightly defined, only rarely falling below 120°.

McDonald and Thornton (1994) showed in their study on the satisfaction of hydrogen-bonding potential in proteins that the distance criterion <2.5 Å and the angle criterion of greater than 90° covers most of the classical hydrogen bonds in proteins.

In a recent study, Liu *et al.* (2008) derived the parameters for various types of hydrogen bonds in protein–ligand complexes, using potential of mean force (PMF) analysis and quantum mechanics (QM) calculations. They showed that distance- and angle-dependent PMF curves for various atom pairs give optimal distances and angles. The preferred interaction region for a hydrogen bond between carbonyl oxygen and amide nitrogen was shown to be 2.5–3.5 Å, with a sharp peak at 2.9 Å. They also showed that these values roughly correspond to the ones obtained by QM calculations (Liu *et al.*, 2008).

These observations provide a working definition of what constitutes a hydrogen bond in proteins. For most protein structures, the resolution of experimentally determined structures is insufficient to observe proton positions. A hydrogen bond is therefore inferred when the distance between donor (N or O) and acceptor (usually O) is less than 3.5 Å and the angles at the donor and acceptor are greater than 90°.

Detailed inspection of protein structures has led to the suggestion that there are hydrogen bonds in proteins where a C–H group is the donor (Derewenda *et al.*, 1995). There will be weak polarization of the C–H bond in the presence of a suitable acceptor atom and the geometry observed in protein structures suggests that this type of interaction has some significance. However, a C–H bond will have only minor contributions to make to the overall stability of a protein structure. Likewise, N–H–S hydrogen bonds are of significance in some proteins, especially metalloproteins. Other nonconventional hydrogen bonds such as those involving π ring systems have also been implicated in protein structure (Steiner and Koellner, 2001) and ligand recognition (Tóth *et al.*, 2007). The discussion presented here, however, concentrates only on the more conventional hydrogen bonds made by protons attached to the atoms O and N. **See also**: Proline Residues in Proteins

Role of Hydrogen Bonds in Protein Stability and Folding

The folded, native structure of proteins at physiological temperatures and solvent conditions is a delicate balance between many competing interactions. The structures are

at a thermodynamic minimum, whereby the total energy of all the interactions between all the different components of the protein and the solvent is more favourable in one particular conformation than in any other. There are thousands of interactions between different parts of a protein structure to be considered, including van der Waals forces, hydrophobic interactions, salt bridges, electrostatic interactions and hydrogen bonds. **See also**: Hydrophobic Interactions in Proteins; Peptide Bonds, Disulfide Bonds and Properties of Small Peptides; Protein Folding *In Vivo*

The difference in energy between an unfolded protein and its native conformation is quite small, approximately $20–30 \, kJ \, mol^{-1}$. It is important to remember that the stability of a folded protein structure is the difference between the interactions made by all the atoms in the folded protein and the interactions made with the solvent when in an unfolded state. The major driving force for the folding of soluble, globular proteins appears to be the burial and clustering of hydrophobic side-chains to minimize their contact with water. Given the chemical nature of the groups found in proteins, the basic requirements for folding are, therefore, that the structures are compact and so minimize the area of hydrophobic surfaces that are exposed to the solvent, and that the buried hydrogen-bonding groups are all paired. When looking at the details of the interactions made within folded protein structures, it is striking how the sequence of amino acids that make up the protein have evolved to achieve this delicate balance of interactions. **See also**: Amino Acid Side-chain Hydrophobicity; Molten Globule; Protein Denaturation and the Denatured State; Protein Tertiary Structures: Prediction from Amino Acid Sequences

Our understanding of the mechanism or pathway for protein folding is slowly developing. Most models rely on the early formation of regions of secondary structure, such as α helix or β sheet, which then come together to form the final folded structure. The main focus of all these models is to discover the mechanism whereby a protein structure is able to find a folded native state among the extremely large number of possible conformations. In addition to requiring that the maximal number of hydrogen-bonding groups makes suitable interactions in the folded structure, an important idea that has emerged is that the making and breaking of hydrogen bonds has to be very rapid and fluid, to allow the correct fold to be found. **See also**: Amino Acid Substitutions: Effects on Protein Stability; Protein Folding: Overview of Pathways

Strength of the Hydrogen Bond

The strength of the hydrogen bonds in proteins has been an issue of much debate. Based on experimental and theoretical studies, it is estimated that an individual hydrogen bond provides between 20 and $25 \, kJ \, mol^{-1}$ of energy (Fleming and Rose, 2005). The differences in the energy of interaction depend on the geometry. In general, it is thought that the main determinant of the strength of

a hydrogen bond is its length. There is only a small energetic consequence for a hydrogen bond being distorted away from the linear, and this is reflected in the observed distributions of hydrogen-bond geometry (**Figure 1b**). This is carried through into the empirical descriptions of noncovalent interactions employed in many molecular mechanics programmes. It is possible to provide an adequate description of the interaction without an explicit hydrogen bond term in the force field. Instead, the van der Waals parameters of the donor and acceptor atoms are modified to allow and reward closer approach. **See also**: Thermodynamics in Biochemistry

In most instances, the contribution a particular hydrogen bond makes to the stability of a protein structure or to the strength of interaction of a ligand is much less than $10–40 \, kJ \, mol^{-1}$. Consider the difference between the overall energy of an amide and a carbonyl group within a polypeptide when they are randomly structured in solution, and the energy when the two groups are hydrogen-bonding to each other. The strength of the hydrogen bond between the carbonyl and amide will be approximately the same as the strength of the hydrogen bond between the solvent and the individual groups. The main difference is the increase in entropy gained as the solvent is displaced from the groups. The subtlety of these interactions and the number of hydrogen-bonding interactions involved make it difficult to calculate directly the impact of individual hydrogen bonds to stability. For these reasons, a useful approximation that has arisen is to make an inventory of the hydrogen bonding in structures (McDonald and Thornton, 1994). In general, if the number of hydrogen bonds in a structure or complex is increased, then it will have a greater stability. **See also**: Protein Stability

The strength of hydrogen bonds is an important amount of energy, as one of the key requirements for protein folding and recognition is for the different interactions to be sampled rapidly. The hydrogen bond provides sufficient energy for specificity in recognition to be at a reasonable energy minimum, but low enough so that the activation energy for making or breaking the bonds will allow rapid kinetics – essential for exploring conformation in folding and in recognition. **See also**: Cell Biophysics

Geometry of Secondary Structures

One consequence of burying hydrophobic groups in the core of globular proteins is that the amide and carbonyl groups of the peptide backbone can no longer bond hydrogen to the solvent. The solution for these groups is to hydrogen-bond to each other, and early work by Pauling in 1951 established that the most stable conformations for a polypeptide chain were the α helix and the β sheet (Pauling and Corey, 1951; Pauling *et al.*, 1951). These structures maximize the pairing of the hydrogen-bonding groups of the peptide backbone, allowing protein chains to be buried in hydrophobic cores and, importantly, providing the main scaffold to support the functional architecture of the
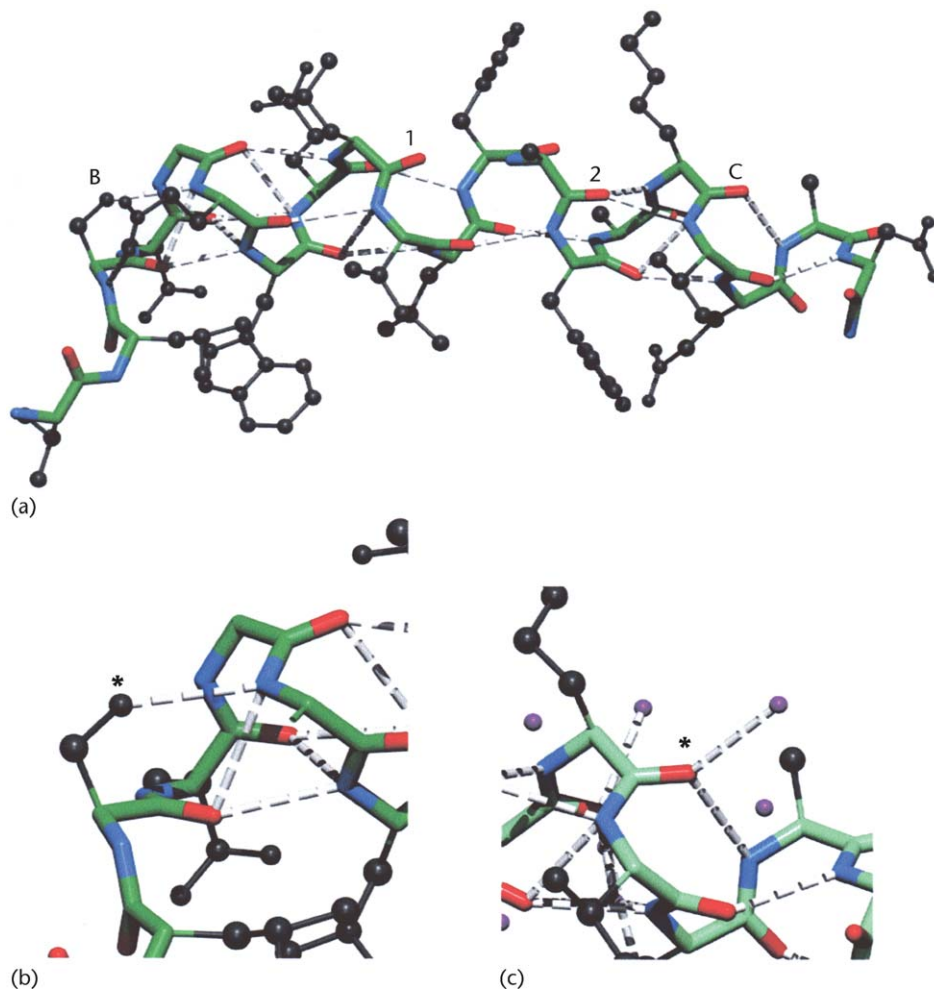
Figure 2 An α helix from the structure of oxygenated human myoglobin (Phillips, 1980). (a) The complete helix with main chain atoms shown as liquorice bonds (nitrogen blue, oxygen red and carbon green), side-chains shown as balls and sticks in black, and hydrogen bonds as white dashed lines. (b) Detail of the *N*-terminal region of the helix, marked B in (a). The asterisk marks the serine oxygen that caps the helix. (c) Detail of the *C*-terminal portion of the helix, including water positions observed in the structure. The asterisk highlights the carboxyl oxygen making a bifurcated hydrogen bond. Coordinates from Protein Data Bank entry 1MBO.

protein. **See also**: Pauling, Linus Carl; Protein Secondary Structures: Prediction

Figure 2a shows an α helix from the structure of myoglobin, together with details of the *N*-terminus of the helix and of a section of the helix showing the observed position of solvent (water) molecules. This example demonstrates many of the features of hydrogen bonding in helices. Although most of the helix contains hydrogen bonds between the carbonyl of residue *n* and the amide of residue *n* + 4, there are irregularities. For example, at position 1 the carbonyl is too far from an amide in the main chain to make a strong hydrogen bond, but is interacting with the solvent. This type of distortion is often seen in the middle of the long helices and reflects the curvature of the helix as it wraps around the surface of a globular protein. At position 2 another form of distortion, particularly common at the beginning and end of helices, where the carbonyl of residue

n is hydrogen bonded to amides not only of residue *n* + 4, but also of *n* + 3 in a type of structure called a 3_{10} helix. Less common is hydrogen bonding with amide *n* + 5 to give a type of structure called a π helix. On the whole, this strain in the observed α helices results in an average length of hydrogen bonds that is slightly more (2.05 Å) than that seen in small molecule structures and in β sheet structures. **See also**: Myoglobin; Protein Structure Classification

Figure 2b also shows the details of another type of hydrogen-bonding interaction that is important for a secondary structure. The formation of a helix leaves unmatched hydrogen-bonding potential in the peptide backbone at both ends, and this is often satisfied by hydrogen-bonding to an amino acid side-chain, capping the helix. In this figure, it can be seen how the hydroxyl of a serine amino acid hydrogen bonds to the free amide at the end of the helix.

In principle, carbonyl groups can take part in two hydrogen bonds – often called a bifurcated hydrogen bond. This was seen in the irregularity at position 2 in **Figure 2a**, but is also seen in the organization of water molecules on the hydrophilic side of a helix. **Figure 2c** shows details of position C where a water molecule is hydrogen-bonding to a carbonyl involved in a 3_{10} helix interaction.

The other major class of secondary structure seen in proteins is the β sheet, predominantly found as the central core of protein structures making little interaction with the solvent. Here, the polypeptide chain is relatively extended with the hydrogen bonding of the peptide units knitting together the individual strands. As can be seen in **Figure 3**, there is essentially no difference in the geometry of hydrogen bonds within parallel and antiparallel sheets.

Another common feature in protein structures is the formation of tight turns, allowing changes of direction in the protein chain. A description of the different classes of turns in proteins can be found in Hutchinson and Thornton (1994). An important feature for many of the different categories of turns is the satisfaction of the hydrogen-bonding potential of the peptide backbone. **See also**: Protein Structure Prediction
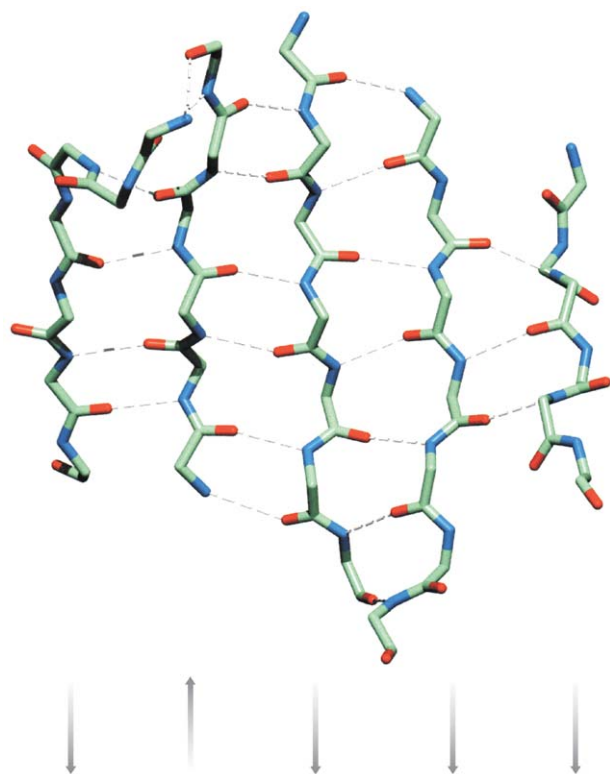

Figure 3 A β sheet from the structure of thioredoxin (Weichsel *et al.*, 1996), showing just the main chain atoms as liquorice bonds (nitrogen blue, oxygen red and carbon green) and hydrogen bonds as white dashed lines. The arrows show the direction of the polypeptide chain, emphasizing that both parallel and antiparallel strands are present in this structure. Coordinates from Protein Data Bank entry 1ERT.

Role of Hydrogen Bonds in Ligand Binding and Specificity

The function of most proteins requires the specific recognition and binding of other molecules, be they other proteins, small organic ligands, nucleic acids or lipids. The central requirement is specificity, binding particular molecule(s) in an environment where there are many other molecules present. As with protein folding, a large number of different molecular interactions is involved and the resulting specificity comes from an appropriate match of shape and chemical functionality between the ligand and the protein-binding site. **See also**: Protein–Ligand Interactions: General Description; Protein–Ligand Interactions: Molecular Basis

In general, most small ligands bind to a distinct binding cleft on the surface of a protein. This is particularly true of enzymes, where the binding of the ligand produces an appropriate chemical environment in which specific chemical reactions can be catalysed. **Figure 4** is a stereo figure of detail taken from a high-resolution structure of the interaction between a carbohydrate molecule and a cellulase enzyme (Varrot *et al.*, 1999). Much of the interaction energy between the protein and ligand comes from the close van der Waals contact between the ligand and the protein surface, with significant entropic contributions from the release of the bound water that surrounds both the free ligand and the unoccupied protein-binding site. At the same time, much of the binding selectivity comes from the provision of the directional, hydrogen-bonding groups on the protein surface, to partner the hydrogen-bonding potential of the ligand. **See also**: Enzyme Specificity and Selectivity; Substrate Binding to Enzymes

This particular structure of a cellulase–carbohydrate complex has been determined at a sufficiently high resolution such that it is possible to define the orientation of the individual hydrogen atoms, and these are shown for the ligand. All of the hydrogen-bonding potential of the ligand is satisfied: appropriate hydrogen bond donors and acceptors are provided for both the proton and the oxygen of each hydroxyl group on the carbohydrate. For example, the proton of hydroxyl A is hydrogen-bonding to a carboxylate group of an aspartic acid and the hydroxyl oxygen is bonding to the imide of a tryptophan. Discrete solvent positions also play an important role. The water molecule B forms an intricate network of hydrogen-bonding interactions, bridging between the ligand and the protein. Bridging water molecules are often seen mediating protein–ligand interactions and can be a central feature in protein–nucleic acid recognition. This complex also provides a striking example of how aromatic residues such as tryptophan (C on the figure) confer specificity on the interaction by forming a hydrophobic patch on the protein surface, recognizing the hydrophobic face of the carbohydrate. **See also**: Protein–DNA Interactions; Protein–RNA Interactions; Water: Structure and Properties

The formation of protein–ligand complexes is driven by geometric and electrostatic complementarity between the
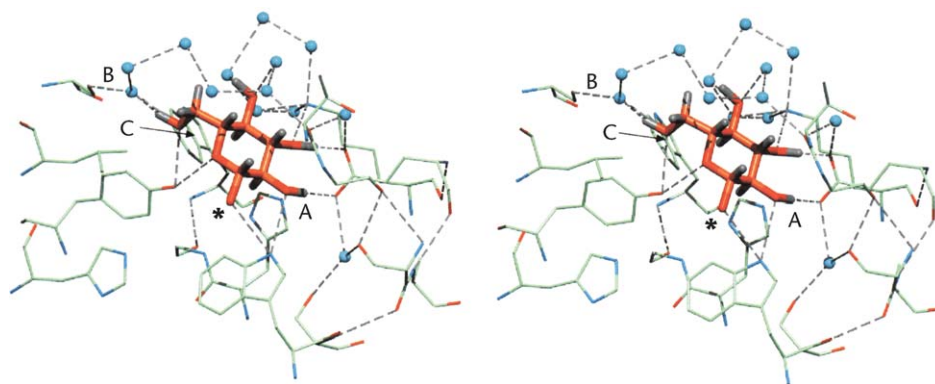
Figure 4   Side-by-side stereo figure showing the details of the interaction between a portion of an inhibitor binding to the cellulase, Cel5A (Varrot *et al.*, 1999). This structure is of sufficient resolution (0.95 Å) so that proton positions can be modelled; for clarity, they are shown only on the ligand. Water molecules are shown as blue spheres and hydrogen bonds as white dashed lines. The nitrogen atoms are in blue, hydrogen grey and oxygen red, with the carbon atoms of the protein in green and ligand orange. The asterisk shows the site of linkage to the rest of the inhibitor. Details about A, B and C are discussed in the text.

interacting parts of the molecules. The satisfaction of hydrogen bond donors and acceptors within the binding site provides an important energetic contribution towards the formation of protein–ligand complexes. However, form a stable complex, this should also be accompanied by increased van der Waals interactions due to shape complementarity. Together, this combination of subtle interactions ensures that the correct molecular recognition can occur as a result of the combined effect of a large number of small interactions. **See also**: Protein–Ligand Interactions: Energetic Contributions and Shape Complementarity

The same general principles underlie all other examples of molecular recognition. In protein–protein recognition, the interacting surfaces of the molecules tend to be larger. With such large surfaces, the recognition tends to be less dictated by the precise shape of the interacting molecules, but relies more on matching of hydrophobic patches and the complementarity of key hydrogen-bonding groups. **See also**: Protein–Protein Interactions

References

Arnold WD and Oldfield E (2000) The chemical nature of hydrogen bonding in proteins via NMR: J-couplings, chemical shifts, and AIM theory. *Journal of the American Chemical Society* **122**(51): 12835–12841.

Baker EN and Hubbard RE (1984) Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology* **44**: 97–179.

Derewenda ZS, Lee L and Derewenda U (1995) The occurrence of C–H–O hydrogen bonds in proteins. *Journal of Molecular Biology* **252**: 248–262.

Fleming PJ and Rose GD (2005) Do all backbone polar groups in proteins form hydrogen bonds? *Protein Science* **14**(7): 1911–1917.

Hutchinson EG and Thornton JM (1994) A revised set of potentials for b-turn formation in proteins. *Protein Science* **3**: 2207–2216.

Liu Z, Wang G, Li Z and Wang R (2008) Geometrical preferences of the hydrogen bonds on protein–ligand binding interface derived from statistical surveys and quantum mechanics calculations. *Journal of Chemical Theory and Computation* **4**: 1959–1973.

Martin TW and Derewenda ZS (1999) The name is bond – H bond. *Nature Structural Biology* **6**: 403–406.

McDonald IK and Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *Journal of Molecular Biology* **238**: 777–793.

Morokuma K (1977) Why do molecules interact? The origin of electron donor-acceptor complexes, hydrogen bonding and proton affinity. *Accounts of Chemical Research* **10**(8): 294–300.

Pauling L and Corey RB (1951) Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proceedings of the National Academy of Sciences of the USA* **37**: 729–740.

Pauling L, Corey RB and Branson HR (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the USA* **37**: 205–211.

Phillips SEV (1980) Structure and refinement of oxy-myoglobin at 1.6 Angstrom resolution. *Journal of Molecular Biology* **142**: 531–554.

Scheiner S (1997) *Hydrogen Bonding: A Theoretical Perspective*. Oxford: Oxford University Press.

Steiner T (2002) The hydrogen bond in the solid state. *Angewandte Chemie International Edition* **41**(1): 48–76.

Steiner T and Koellner G (2001) Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *Journal of Molecular Biology* **305**(3): 535–557.

Tóth G, Bowers SG, Truong AP and Probst G (2007) The role and significance of unconventional hydrogen bonds in small molecule recognition by biological receptors of pharmaceutical relevance. *Current Pharmaceutical Design* **13**(34): 3476–3493.

Tuttle T, Grafenstein J, Wu A, Kraka E and Cremer D (2004) Analysis of the NMR spin-spin coupling mechanism across a H-bond: nature of the H-bond in proteins. *Journal of Physical Chemistry. B* **108**(3): 1115–1129.

Varrot A, Schulein M, Pipelier M, Vasella A and Davies GJ (1999) Lateral protonation of a glycosidase inhibitor. Structure of the *Bacillus agaradhaerens* Cel5A in complex with a cellobiose-derived imidazole at 0.97 Angstrom resolution. *Journal of the American Chemical Society* **121**: 2621–2622.

Weichsel A, Glasdaska JR, Powis G and Montfort WR (1996) Crystal structures of reduced, oxidised and mutated human thioredoxins: evidence for a regulatory homodimer. *Structure* **4**: 735–751.

Further Reading

Dunitz JD (1995) Win some, lose some: enthalpy–entropy compensation in weak intermolecular interactions. *Chemistry and Biology* **2**: 709–712.

Fersht AR (1987) The hydrogen bond in molecular recognition. *Trends in Biochemical Sciences* **12**: 301–304.

Fersht AR (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. New York: WH Freeman.

Fersht AR, Shi J-P, Knill-Jones J *et al*. (1985) Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**: 235–238.

Huggins ML (1971) 50 years of hydrogen bond theory. *Angewandte Chemie International Edition* **10**: 147–208.

Jeffrey GA (1997) *An Introduction to Hydrogen Bonding*. Oxford: Oxford University Press.

Karshikoff A (2006) *Non-covalent Interactions in Proteins*. London: Imperial College Press.

Mills JE and Dean PM (1996) Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *Journal of Computer-Aided Molecular Design* **10**(6): 607–622.

Ming-Hong Hao (2006) Theoretical calculation of hydrogen-bonding strength for drug molecules. *Journal of Chemical Theory and Computation* **2**(3): 863–872.