

## 22.2. Hydrogen bonding in biological macromolecules

BY E. N. BAKER

### 22.2.1. Introduction

The hydrogen bond (Huggins, 1971) plays a critical role in the structure and function of biological macromolecules. This is because, uniquely among the non-covalent interactions that stabilize such structures, it combines a strong directional character with its energetic contributions. Thus, hydrogen-bonding patterns define the secondary structures that form the framework of proteins, are responsible for the specificity of base pairing in nucleic acids, shape the loops and irregular features that often determine molecular recognition, and provide for appropriately oriented functional groups in catalytic and/or binding sites.

Much of our present knowledge of hydrogen bonding in biological structures is foreshadowed in Linus Pauling's influential book (Pauling, 1960), and Jeffrey & Saenger (1991) have provided a comprehensive recent review. Other important reviews have covered hydrogen-bonding patterns in globular proteins (Baker & Hubbard, 1984; Stickle *et al.*, 1992), the satisfaction of hydrogen-bonding potential in proteins (McDonald & Thornton, 1994a), hydrogen-bonding patterns for side chains (Ippolito *et al.*, 1990) and side-chain hydrogen bonding in relation to secondary structures (Bordo & Argos, 1994).

### 22.2.2. Nature of the hydrogen bond

Hydrogen bonds are attractive electrostatic interactions of the type  $D-H \cdots A$ , where the H atom is formally attached to a donor atom,  $D$  (assumed to be more negative than H), and is directed towards an acceptor,  $A$ . The acceptor  $A$  is normally an electronegative atom, usually O or N, but occasionally S or Cl, with a full or partial negative charge and a lone pair of electrons directed towards the H atom. Although most of the hydrogen bonds in proteins and nucleic acids are  $N-H \cdots O$  or  $O-H \cdots O$  (less often,  $N-H \cdots N$ ), it is important to be aware that other possibilities exist, including  $N-H \cdots S$ ,  $O-H \cdots S$  and  $C-H \cdots O$ , and that these can be very important in specific cases (Adman *et al.*, 1975; Derewenda *et al.*, 1995). Likewise, the  $\pi$ -electron clouds of aromatic rings can also act as acceptors for appropriately oriented  $D-H$  groups (Legon & Millen, 1987; Mitchell *et al.*, 1994).

In an ideal hydrogen bond, the donor heavy atom, the H atom, the acceptor lone pair and the acceptor heavy atom should all lie in a straight line (Legon & Millen, 1987), as illustrated in Fig. 22.2.2.1(a). The strength of the interaction is also expected to depend on the electronegativities of the atoms involved. Hydrogen bonds are said to be bifurcated when a single  $D-H$  group interacts with two acceptors in a three-centred hydrogen bond (Fig. 22.2.2.1b); these hydrogen bonds are necessarily nonlinear and weaker. However, the term bifurcated is also sometimes applied to the quite different situation where a donor atom with two H atoms or an acceptor atom with two lone pairs makes two hydrogen bonds, as in Figs. 22.2.2.1(c) and (d). These interactions can be strong and linear. Some hydrogen-bonding arrangements are said to be cooperative; for example, hydrogen bonding by a peptide  $C=O$  group should enhance the polarity of the whole peptide unit and hence the acidity of the amide proton and the strength of its hydrogen bonding (Jeffrey & Saenger, 1991).

### 22.2.3. Hydrogen-bonding groups

#### 22.2.3.1. Proteins

The hydrogen-bonding capacities of the various hydrogen-bonding groups in proteins are shown in Fig. 22.2.3.1. All, with

the exception of the peptide NH and Trp side-chain NH groups, can participate in more than one hydrogen-bond interaction. Peptide and side-chain  $C=O$  groups, for example, can act as acceptors for two hydrogen bonds by using both lone pairs of electrons on the  $sp^2$ -hybridized oxygen. Likewise, the  $-OH$  groups of Ser or Thr can act as donors through their single H atom, and acceptors through their two lone pairs. In Tyr side chains, the  $C-O$  bond has some double-bond character, and the phenolic  $-OH$  is thus likely to prefer only two hydrogen bonds, both in the ring plane. The carboxylate groups of Asp and Glu are normally ionized above pH 4 and their  $C-O$  bonds also have partial double-bond character; each carboxylate oxygen should then be able to accept two hydrogen bonds, although the restriction to two may be less severe than for  $C=O$ .

Several uncertainties exist. Crystallographically, it is not usually possible to distinguish the amide oxygen and nitrogen atoms of Asn and Gln, and the decision as to which is which has to be made on environmental grounds by considering what hydrogen bonds would be made in each of the two possible arrangements. Likewise, two possibilities exist for His side chains by rotating  $180^\circ$  about  $C^\beta-C^\gamma$ . This problem has been analysed by McDonald & Thornton (1994b), and corrections can be made with *HBPLUS*.

For some side chains, the ionization state is uncertain. Arg and Lys are assumed to be fully protonated, as in Fig. 22.2.3.1, and Asp and Glu are assumed to be fully ionized. Nevertheless, a survey by Flocco & Mowbray (1995) has shown that a small but significant number of short  $O \cdots O$  distances between Asp and Glu side chains must represent  $O-H \cdots O$  hydrogen bonds, with one carboxyl group protonated. His side chains, in addition to the orientational uncertainty, have a  $pK_a$  ( $\sim 6.5$ ) that implies that they may be in either their neutral or their protonated form, depending on pH and environment. In the neutral form, only one N atom is protonated (more often  $N^{\epsilon 2}$ , but sometimes  $N^{\delta 1}$ ), but in the protonated form both N atoms carry protons; again, the actual state has to be deduced from their environment.

#### 22.2.3.2. Nucleic acids

The three components of nucleic acids, *i.e.* phosphate groups, sugars and bases, all participate in hydrogen bonding to greater or lesser extent. The phosphate oxygen atoms can potentially act as acceptors of two or more hydrogen bonds and are frequently the

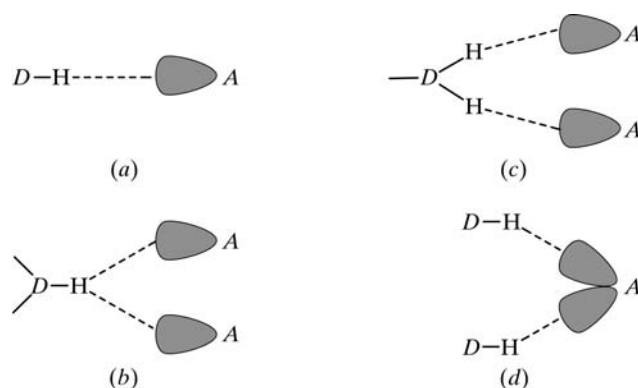


Fig. 22.2.2.1. Hydrogen-bonding configurations. (a) The standard two-centre hydrogen bond in which an H atom attached to a donor atom,  $D$ , is directed towards a lone pair of an acceptor,  $A$ . (b) A classic three-centre, or bifurcated, hydrogen bond, with a single H atom shared between the lone pairs of two acceptors. The situations shown in (c) and (d) are not true three-centre hydrogen bonds since they are essentially equivalent to that in (a).

## 22.2. HYDROGEN BONDING IN BIOLOGICAL MACROMOLECULES

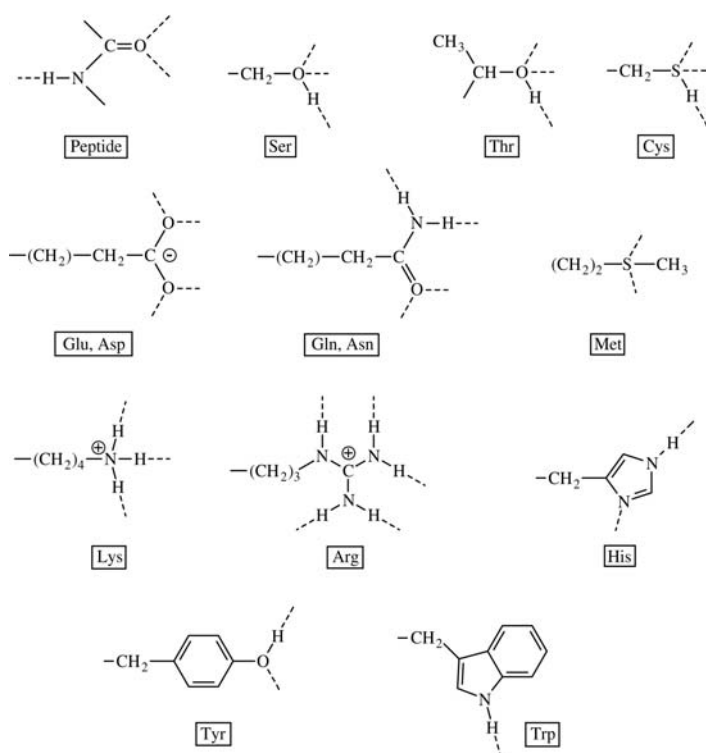


Fig. 22.2.3.1. Hydrogen-bonding potential of protein functional groups. Potential hydrogen bonds are shown with broken lines. Arg, Lys, Asp and Glu side chains are shown in their ionized forms.

recipients of hydrogen bonds from protein side chains in protein–DNA complexes. The sugar residues of RNA have a 2'-OH which can act as both hydrogen-bond donor and acceptor, and the 4'-O of both ribose and deoxyribose can potentially accept two hydrogen bonds.

It is the bases of DNA and RNA that have the greatest hydrogen-bonding potential, however, with a variety of hydrogen-bond donor or acceptor sites. Although each of the bases could theoretically occur in several tautomeric forms, only the canonical forms shown in Fig. 22.2.3.2 are actually observed in nucleic acids. This leads to clearly defined hydrogen-bonding patterns which are critical to both base pairing and protein–nucleic acid recognition. The  $\text{—NH}_2$  and  $\text{>NH}$  groups act only as hydrogen-bond donors, and  $\text{C=O}$  only as acceptors, whereas the  $\text{>N—}$  centres are normally acceptors but at low pH can be protonated and act as hydrogen-bond donors.

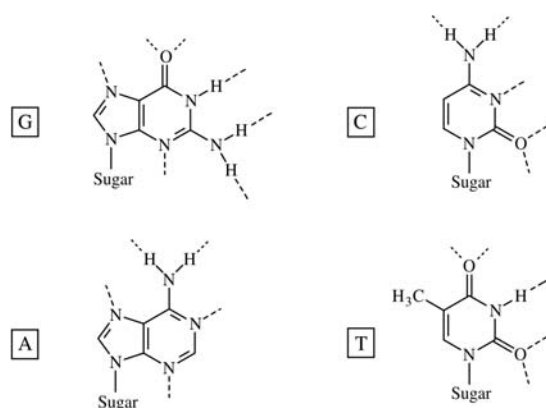


Fig. 22.2.3.2. Hydrogen-bonding potential of nucleic acid bases guanine (G), adenine (A), cytosine (C) and thymine (T) in their normal canonical forms.

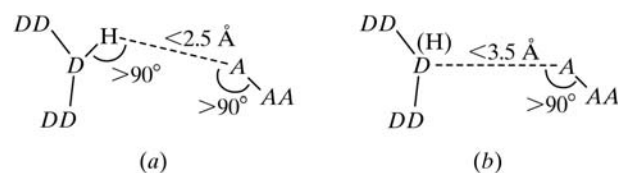


Fig. 22.2.4.1. Suggested criteria for identifying likely hydrogen bonds.  $DD$  and  $AA$  represent atoms covalently bonded to the donor atom,  $D$ , and acceptor atom,  $A$ , respectively. Here, (a) represents the criteria when the donor H atom can be placed, and (b) when it cannot be placed. Additional criteria based on the angle  $DD\text{—}D\cdots A$  could be incorporated with (b). Adapted from Baker & Hubbard (1984) and McDonald & Thornton (1994a).

### 22.2.4. Identification of hydrogen bonds: geometrical considerations

Because hydrogen bonds are electrostatic interactions for which the attractive energy falls off rather slowly (Hagler *et al.*, 1974), it is not possible to choose an exact cutoff for hydrogen-bonding distances. Rather, both distances and angles must be considered together; the latter are particularly important because of the directionality of hydrogen bonding. Inferences drawn from distances alone can be highly misleading. An approach with an  $\text{N—H}\cdots\text{O}$  angle of  $90^\circ$  and an  $\text{H}\cdots\text{O}$  distance of 2.5 Å would be very unfavourable for hydrogen bonding, yet it translates to a  $\text{N}\cdots\text{O}$  distance of 2.7 Å. This could (wrongly) be taken as evidence of a strong hydrogen bond.

For macromolecular structures determined by X-ray crystallography, problems also arise from the imprecision of atomic positions and the fact that H atoms cannot usually be seen. Thus, the geometric criteria must be relatively liberal. H atoms should also be added in calculated positions where this is possible; this can be done reliably for most NH groups (peptide NH, side chains of Trp, Asn, Gln, Arg, His, and all  $\text{>NH}$  and  $\text{NH}_2$  groups in nucleic acid bases).

The hydrogen-bond criteria used by Baker & Hubbard (1984) are shown in Fig. 22.2.4.1. Very similar criteria are used in the program *HBPLUS* (McDonald & Thornton, 1994a), which also adds H atoms in their calculated positions if they are not already present in the coordinate file. In general, hydrogen bonds may be inferred if an interatomic contact obeys *all* of the following criteria:

- (1) The distance  $\text{H}\cdots\text{A}$  is less than 2.5 Å (or  $\text{D}\cdots\text{A}$  less than 3.5 Å if the donor is an  $\text{—OH}$  or  $\text{—NH}_3^+$  group or a water molecule).
- (2) The angle at the H atom,  $\text{D—H}\cdots\text{A}$ , is greater than  $90^\circ$ .
- (3) The angle at the acceptor,  $\text{AA—A}\cdots\text{H}$  (or  $\text{AA—A}\cdots\text{D}$  if the H-atom position is unreliable), is greater than  $90^\circ$ .

Other criteria can be applied, for example taking into account the hybridization state of the atoms involved and the degree to which any approach lies in the plane of the lone pair(s). In all analyses of hydrogen bonding, however, it is clear that a combination of distance and angle criteria is effective in excluding unlikely hydrogen bonds.

### 22.2.5. Hydrogen bonding in proteins

#### 22.2.5.1. Contribution to protein folding and stability

The net contribution of hydrogen bonding to protein folding and stability has been the subject of much debate over the years. The current view is that although the hydrophobic effect provides the driving force for protein folding (Kauzmann, 1959), many polar groups, notably peptide NH and  $\text{C=O}$  groups, inevitably become buried during this process, and failure of these groups to find hydrogen-bonding partners in the folded protein would be strongly destabilizing. This, therefore, favours the formation of secondary