International Tables for Crystallography (2006). Vol. F, Section 22.2.5.3.1, p. 548.

structures and other structures that permit effective hydrogen bonding in the folded molecule. Not surprisingly, the contribution of specific hydrogen bonds to stability depends on their location in the structure (Fersht & Serrano, 1993). Mutagenesis studies have shown that even the loss of a single hydrogen bond can be significantly destabilizing (Alber *et al.*, 1987) and that the energetic contribution can vary depending on whether or not the groups involved are charged (Fersht *et al.*, 1985).

22.2.5.2. *Saturation of hydrogen-bond potential*

A consistent conclusion from analyses of protein structures is that virtually all polar atoms either form explicit hydrogen bonds or are at least in contact with external water. The extent to which their *full* hydrogen-bond potential is fulfilled in a folded protein (for example, the potential of an Arg side chain to make five hydrogen bonds) has been examined in several studies. Baker & Hubbard (1984) considered the explicit hydrogen bonds made by main-chain and side-chain atoms in a number of refined protein structures and established general patterns for both, but did not differentiate buried and solvent-exposed atoms or allow for unmodelled solvent. Savage *et al.* (1993) used the solvent accessibilities of polar groups to estimate their assumed numbers of hydrogen bonds to external water. This supplemented the explicit hydrogen bonds that could be derived from the atomic coordinates and allowed an estimate of the extent to which potential hydrogen bonds are lost during protein folding. McDonald & Thornton (1994*a*) focused specifically on buried hydrogen-bond donors and acceptors in order to determine the extent to which the hydrogen-bond potential of these is utilized.

The results of these analyses can be summarized as follows. Almost all polar groups do in fact make at least one hydrogen bond. Hydrogen-bond donors are almost always hydrogen bonded; only 4% of NH groups 'lose' hydrogen bonds as a result of protein folding (Savage *et al.*, 1993). On the other hand, hydrogen-bond acceptors often do not exert their full hydrogen-bonding potential. For example, for main-chain C=O groups, which are expected to accept two hydrogen bonds, 24% of possible hydrogen bonds are estimated to be lost during folding (Savage *et al.*, 1993). Among buried C=O groups, although very few make no hydrogen bonds (as little as 2% if hydrogen-bonding criteria are relaxed), the majority fail to form a second hydrogen bond (McDonald & Thornton, 1994*a*). Steric factors, particularly in β-sheets or where Pro residues are adjacent, restrict hydrogen-bonding possibilities, although some of the 'lost' interactions may be recovered through C—H···O interactions (see Section 22.2.7.1). McDonald & Thornton also point out that failure to form a second hydrogen bond is less energetically expensive than failure to form the first. Among polar side chains, the ionizable side chains (Asp, Glu, Arg, Lys, His) show a very strong tendency to be fully hydrogen bonded or solvent exposed. Buried Arg side chains, for example, frequently form all five possible hydrogen bonds. The side chains that most often fail to fulfil their full hydrogen-bond potential are Ser, Thr and Tyr; these almost always donate one hydrogen bond but frequently fail to accept one.

22.2.5.3. *Secondary structures*

Secondary structures provide the means whereby the polar C=O and NH groups of the polypeptide chain can remain effectively hydrogen bonded when they are buried within a folded globular protein. In doing so, they provide the framework of folding patterns and account for the majority of hydrogen bonds within protein structures. The three secondary-structure classes (helices, β-sheets and turns) are each characterized by specific hydrogen-bonding patterns, which can be used for objective identification of these structures (Stickle *et al.*, 1992).

22.2.5.3.1. *Helices*

Helices have traditionally been defined in terms of their N—H···O=C hydrogen-bonding patterns as α-helices ($i \rightarrow i-4$), 3_{10}-helices ($i \rightarrow i-3$), or π-helices ($i \rightarrow i-5$); in an α-helix, for example, the peptide NH of residue 5 hydrogen bonds to the C=O of residue 1. In fact, the vast majority of helices in proteins are α-helices; 3_{10}-helices are rarely more than two turns (six residues) in length, and discrete π-helices have not been seen so far.

The residues within helices have characteristic main-chain torsion angles, (φ, ψ), of around ($-63°$, $-40°$) that cause the C=O groups to tilt outwards by about 14° from the helix axis (Baker & Hubbard, 1984). This results in somewhat less linear hydrogen bonding than in the original Pauling model (Pauling *et al.*, 1951), with a degree of distortion towards 3_{10}-helix geometry. Thus, weak $i \rightarrow i-3$ interactions are often made in addition to the more favourable $i \rightarrow i-4$ hydrogen bonds, giving hydrogen-bond networks that may enhance helix elasticity (Stickle *et al.*, 1992). Tilting outwards also makes the C=O groups more accessible for additional hydrogen bonds from side chains or water molecules. For the α-type, $i \rightarrow i-4$ interactions, the hydrogen-bond angles at both donor and acceptor atoms are quite tightly clustered (N—H···O $\sim$157° and C=O···H $\sim$ 147°). The hydrogen-bond lengths in helices average 2.06 (16) Å (O···H) or 2.99 (14) Å (O···N) (Baker & Hubbard, 1984).

Few helices are regular throughout their length. Many are curved or kinked such that one side (often the outer, solvent-exposed side) of the helix is opened up a bit and has longer hydrogen bonds (Blundell *et al.*, 1983; Baker & Hubbard, 1984). The bends are often associated with additional hydrogen bonds from water molecules or side chains to C=O groups that are tilted out more than usual. Curved helices are normal in coiled-coil structures and can enable long helices to pack more effectively in globular structures. Sometimes a kink can be functionally important, as in manganese superoxide dismutase, where a kink in a long helix, incorporating a π-type ($i \rightarrow i-5$) hydrogen bond, enables the optimal positioning of active-site residues (Edwards *et al.*, 1998).

The beginnings and ends of helices are sites of hydrogen-bonding variations which can be seen as characteristic 'termination motifs'. At helix N-termini, 3_{10}-type $i \rightarrow i-3$ (or bifurcated $i \rightarrow i-3$ and $i \rightarrow i-4$) hydrogen bonds are often found. At C-termini, two common patterns occur. In one, labelled α_{C1} by Baker & Hubbard (1984), there is a transition from α-type, $i \rightarrow i-4$ to 3_{10}-type, $i \rightarrow i-3$ hydrogen bonding, often with genuine bifurcated hydrogen bonds, as in Fig. 22.2.2.1(*b*), at the transition point. The other, labelled α_{C2} (Baker & Hubbard, 1984) or referred to as the 'Schellman motif' (Schellman, 1980), has a π-type, $i \rightarrow i-5$ hydrogen bond coupled with a 3_{10}-type, $i-1 \rightarrow i-4$ hydrogen bond; residue $i-1$ has a left-handed α configuration and is often Gly. The beginnings and ends of helices are also the sites of specific side-chain hydrogen-bonding patterns, referred to as N-caps and C-caps (Presta & Rose, 1988; Richardson & Richardson, 1988); these are described below.

22.2.5.3.2. *β-sheets*

β-sheets consist of short strands of polypeptide (typically 5–7 residues) running parallel or antiparallel and cross-linked by N—H···O=C hydrogen bonds. Although the (φ, ψ) angles of residues within β-sheets can be quite variable, the hydrogen-bonding patterns within these segments tend to be quite regular, as in the original Pauling models (Pauling & Corey, 1951). Occasional β-bulges in the middle of β-strands can interrupt the hydrogen-bonding pattern (Richardson *et al.*, 1978), but otherwise disruptions occur only at the ends of strands. The hydrogen bonds in β-sheets appear to be slightly shorter than those in helices, by $\sim$0.1 Å, and

Copyright © 2006 International Union of Crystallography

references