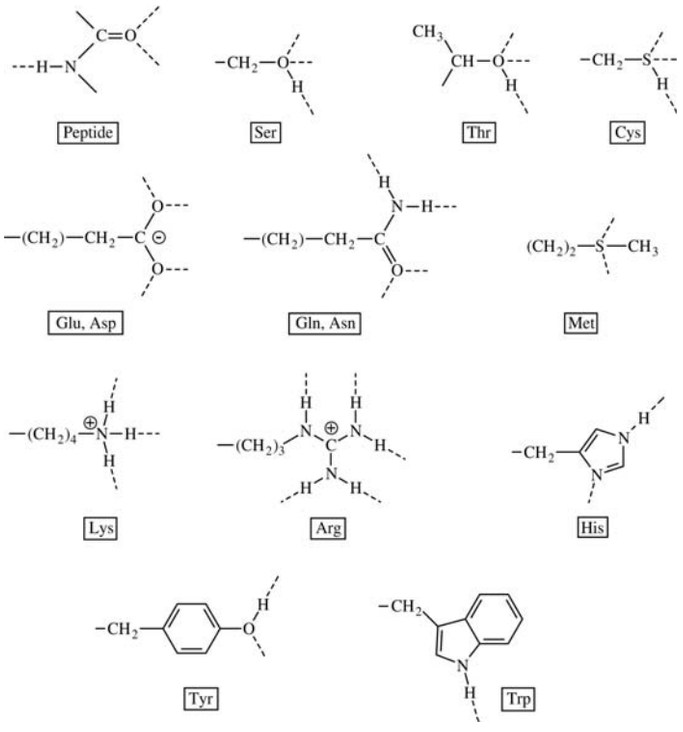

Fig. 22.2.3.1. Hydrogen-bonding potential of protein functional groups. Potential hydrogen bonds are shown with broken lines. Arg, Lys, Asp and Glu side chains are shown in their ionized forms.

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 —NH_2 and >NH groups act only as hydrogen-bond donors, and C=O only as acceptors, whereas the >N— centres are normally acceptors but at low pH can be protonated and act as hydrogen-bond donors.

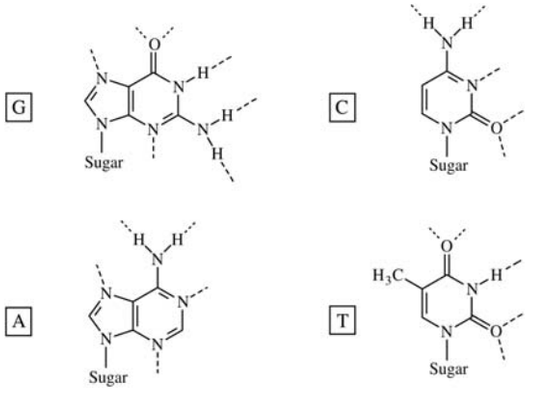

Fig. 22.2.3.2. Hydrogen-bonding potential of nucleic acid bases guanine (G), adenine (A), cytosine (C) and thymine (T) in their normal canonical forms.

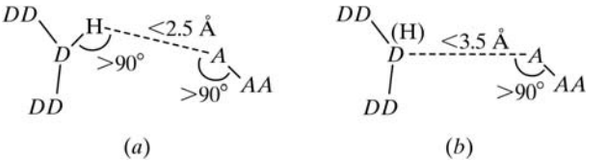

Fig. 22.2.4.1. Suggested criteria for identifying likely hydrogen bonds. *DD* and *AA* represent atoms covalently bonded to the donor atom, *D*, and acceptor atom, *A*, respectively. Here, (*a*) represents the criteria when the donor H atom can be placed, and (*b*) when it cannot be placed. Additional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

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 N—H···O angle of 90° and an H···O distance of 2.5 Å would be very unfavourable for hydrogen bonding, yet it translates to a N···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 >NH and 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, 1994*a*), 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 H···*A* is less than 2.5 Å (or *D*···*A* less than 3.5 Å if the donor is an —OH or —NH_3^+ group or a water molecule).

(2) The angle at the H atom, *D*—H···*A*, is greater than 90°.

(3) The angle at the acceptor, *AA*—*A*···H (or *AA*—*A*···*D* if the H-atom position is unreliable), is greater than 90°.

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 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

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$\cdots$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$\cdots$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$\cdots$O $\sim$157° and C=O$\cdots$H $\sim$ 147°). The hydrogen-bond lengths in helices average 2.06 (16) Å (O$\cdots$H) or 2.99 (14) Å (O$\cdots$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$\cdots$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

also more linear (N—H···O ∼ 160°, compared with ∼157° in helices) (Baker & Hubbard, 1984). There also appears to be no difference between parallel and antiparallel β-sheets in the hydrogen-bond lengths and angles.

22.2.5.3.3. *Turns*

By far the most common type of turn is the β-turn, a sequence of four residues that brings about a reversal in the polypeptide chain direction. Hydrogen bonding does not seem to be essential for turn formation, but a common feature is a hydrogen bond between the C=O group of residue 1 and the NH group of residue 4, a 3_{10}-type, $i \rightarrow i-3$ interaction. Turns are also often associated with characteristic side-chain–main-chain hydrogen-bond configurations (see below). The hydrogen bonds in turns tend to be longer and less linear than those in helices and β-sheets; in particular, the angle at the acceptor oxygen atom C—O···H is around 120° (Baker & Hubbard, 1984).

In addition to β-turns, a small but significant number of γ-turns are found. In these three-residue turns, a hydrogen bond is formed between the C=O of residue 1 and the NH of residue 3, an $i \rightarrow i-2$ interaction. Although the approach to the acceptor oxygen atom is highly nonlinear (C—O···H ∼ 100°), the nonlinearity at the H atom is less pronounced (N—H···O ∼ 130–150°) (Baker & Hubbard, 1984). γ-turns are again of several types, depending on the configuration of the central residue. The classic γ-turn, first recognised by Matthews (1972) and Nemethy & Printz (1972), has a central residue with (φ, ψ) angles around (70°, −60°), which puts it in the normally disallowed region of the Ramachandran plot. More common, however, are structures in which an $i \rightarrow i-2$ hydrogen bond is associated with a central residue with a configuration around (90°, −70°) (Baker & Hubbard, 1984); these structures are not necessarily true turns in the sense of bringing about a sharp chain reversal, however.

22.2.5.3.4. *Aspects of in-plane geometry*

For hydrogen bonds involving sp^2 donors and/or acceptors, optimal interaction is expected to occur when the donor *D*—H group and the acceptor lone-pair orbital are coplanar (Taylor *et al.*, 1983). Analysis of 'in-plane' and 'out-of-plane' components of N—H···O hydrogen bonds in proteins shows that these have characteristic values for different secondary structures (Artymiuk & Blake, 1981; Baker & Hubbard, 1984). The out-of-plane component is tightly clustered at ∼25° for helices and ∼60° for the most common β-turns (type I and type III), but is widely scattered around a mean of 0° for β-sheets. The latter reflects different twists or curvature of β-sheets. The large out-of-plane component for turns is consistent with a relatively weak interaction.

22.2.5.4. *Side-chain hydrogen bonding*

An important concept in understanding the patterns of side-chain hydrogen bonding in proteins is that of local *versus* non-local interactions; local means that a side chain hydrogen bonds to another residue that is relatively close to it in the linear amino-acid sequence. Baker & Hubbard (1984) were first to introduce this distinction, with local defined as ±4 residues. Bordo & Argos (1994) define local as ±6 residues and Stickle *et al.* (1992) as ±10 residues. The distinction is not important, but the distributions in all three analyses show that ±5 would encompass all the significant populations of local hydrogen bonds. Local hydrogen bonds, in which side chains interact with nearby main-chain atoms or other side chains, are evidently critical for protein folding. Non-local hydrogen bonds, although fewer in number (see below), in turn can be very important for stabilization of the folded protein.

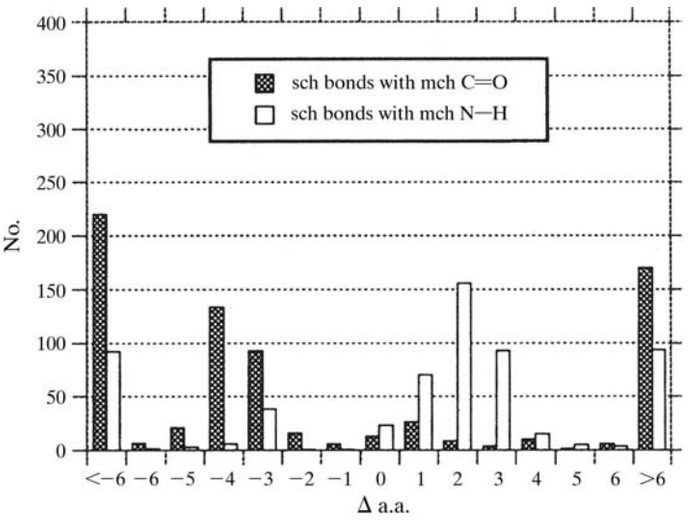

Fig. 22.2.5.1. Distribution of side-chain–main-chain hydrogen bonds as a function of the separation (Δ a.a.) along the polypeptide between the side-chain (sch) and main-chain (mch) groups involved, *i.e.* Δ a.a. = $-n$ means that a side chain interacts with a main-chain group n residues earlier in the polypeptide (towards the N-terminus). Reproduced with permission from Bordo & Argos (1994). Copyright (1994) Academic Press.

If hydrogen bonds with water are excluded, a rule of thirds applies. Approximately one-third of the hydrogen bonds made by side chains (sch's) are with main-chain (mch) C=O groups, one-third are with main-chain NH groups, and one-third with other side chains. Within these populations, however, there are significant differences. For sch–mch(C=O) hydrogen bonds, approximately 45% are local; for sch–mch(NH) hydrogen bonds, a much higher proportion is local (69%), and for sch–sch hydrogen bonds, the proportion is much less (35%) (Bordo & Argos, 1994).

The distribution of local sch–mch(NH) hydrogen bonds shows a marked positional preference (Fig. 22.2.5.1) that highlights consistent hydrogen-bonding motifs found in all proteins (Fig. 22.2.5.2). The major peak involves side chains that interact with an NH group two residues further on in the polypeptide, an n–NH($n+2$) hydrogen bond. This motif primarily involves Asp, Asn, Ser and Thr side chains and is most often found (i) in turns, where a side chain from position 1 hydrogen bonds to the NH of residue 3, (ii) in loop regions where it stabilizes the local structure

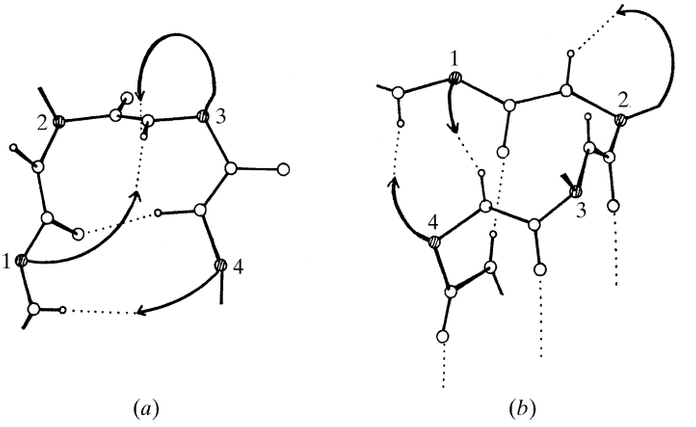

Fig. 22.2.5.2. Schematic representations of common classes of side-chain–main-chain hydrogen bonds (*a*) in turns and (*b*) at helix N-termini. Arrows represent side chains that hydrogen bond to main-chain CO or NH groups (NH identified by the small circle for H).

but is not necessarily associated with chain reversal, and (iii) at helix N-termini.

Helix N-termini are also the site of other characteristic local side-chain–NH hydrogen-bonding motifs (Baker & Hubbard, 1984; Presta & Rose, 1988; Richardson & Richardson, 1988; Harper & Rose, 1993; Bordo & Argos, 1994). Prominent among these are sch–NH($n+3$) hydrogen bonds involving Ser, Thr, Asp and Asn side chains, but sch–NH($n-3$) interactions, in which Glu or Gln side chains hydrogen bond back to a main-chain NH, form an important lesser category. Other motifs, such as that in which a Glu or Gln side chain bends round to hydrogen bond to its own NH group, are also found. Collectively, these contribute to helix capping motifs (Fig. 22.2.5.2*b*) that help satisfy the hydrogen bonding of the 'free' NH groups of the helix N-terminus and in effect extend the helix; the sch–mch(NH) hydrogen bond mimics the mch–mch C=O···HN hydrogen bonds of the helix. Helix N-capping by side chains is probably a very important influence in protein folding, acting as a stereochemical code for helix initiation (Presta & Rose, 1988; Harper & Rose, 1993).

The distribution of sch–mch(CO) hydrogen bonds also shows a striking preference, this time for positions −3 and −4. These sch–CO($n-3$) or sch–CO($n-4$) hydrogen bonds account for the vast majority of local hydrogen bonds between side chains and main-chain C=O groups. Almost all (∼85%) are in helices, with most of the remainder in turns. They involve predominantly (∼80%) Ser and Thr side chains but other side chains (Asn, His, Arg) can also participate. These local hydrogen bonds can occur at any point along a helix, where they are often associated with helix bending or kinking (Baker & Hubbard, 1984). However, they are most frequently found at helix C-termini (Bordo & Argos, 1994) and may constitute a termination motif.

Local side-chain–side-chain hydrogen bonds, although common, do not seem to fit into any obvious patterns; the only recurring interaction identified so far is between side chains on succeeding turns of helices, *i.e.* separated by approximately four residues. These frequently involve charged side chains, which can form hydrogen-bonded ion pairs. In sections of extended chain, side chains that are two residues apart may similarly interact.

Non-local hydrogen bonding by side chains is less easy to categorize but is no less significant; more than 50% of side-chain–main-chain(C=O) hydrogen bonds are non-local, as are ∼65% of side-chain–side-chain hydrogen bonds. In most proteins, a small number of polar side chains with multiple hydrogen-bonding capability act as the centre for networks of hydrogen bonds; these appear to be particularly important for stabilizing non-repetitive polypeptide chain structures (coil, loops). Examples are given in Baker & Hubbard (1984). Most often these involve larger side chains with more than one hydrogen-bonding centre (Asn, Asp, Gln, Glu, Arg, His) which cross-link different sections of the polypeptide. Arg side chains interacting with main-chain C=O groups seem to be particularly effective; Ser and Thr, on the other hand, are seldom used, even though both have the potential to form three hydrogen bonds.

The geometry of side-chain hydrogen bonding has been analysed by Baker & Hubbard (1984) and, more extensively, by Ippolito *et al.* (1990). The former concentrate on hydrogen-bond lengths and angles and show that the preferred angles fit well with stereochemical expectations. Ippolito *et al.* examine the preferences for the various hydrogen-bonding sites around each side-chain type by means of scatter plots (Fig. 22.2.5.3) from which probability densities are computed. These show that well defined preferences exist, determined by both steric and electronic effects.

22.2.5.5. *Hydrogen bonds with water molecules*

Water molecules, with their small size and double-donor, double-acceptor hydrogen-bonding capability, are ideal for completing intramolecular hydrogen-bonding networks, *e.g.* by linking two proton acceptor atoms, or two protein donor atoms, that cannot otherwise interact. Thus, buried water molecules, making multiple hydrogen bonds, help satisfy the hydrogen-bond potential of internal polar atoms and contribute to protein stability; internal waters average about three hydrogen bonds each (Baker & Hubbard, 1984; Williams *et al.*, 1994). From the survey of Williams *et al.* (1994), most (58%) occupy discrete cavities, while 22% are in clusters housing two waters and 20% are in larger clusters; some examples of larger clusters are given in Baker & Hubbard (1984). Buried waters are often conserved between homologous proteins (Baker, 1995), and each buried water–protein hydrogen bond is estimated to stabilize a folded protein by, on average, 0.6 kcal mol^{-1} (1 kcal mol^{-1} = 4.184 kJ mol^{-1}) (Williams *et al.*, 1994). More loosely bound external waters exchange much more rapidly and presumably contribute less energetically.

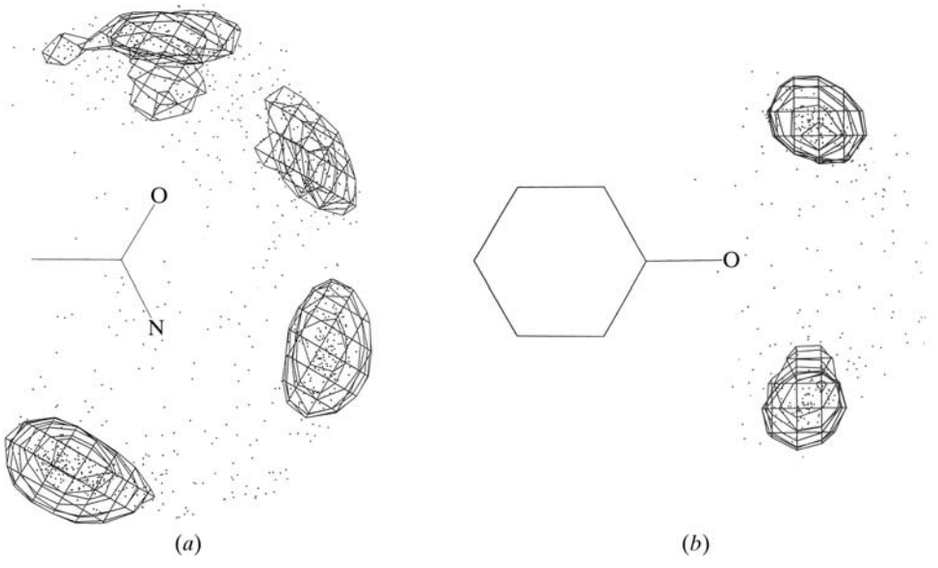

Fig. 22.2.5.3. Typical scatter plots showing the distribution of hydrogen-bonding partners around protein side chains, shown for (*a*) Asn or Gln and (*b*) Tyr. Reproduced with permission from Ippolito *et al.* (1990). Copyright (1990) Academic Press.

Several patterns of hydrogen bonding are consistently observed. Water molecules are most often seen interacting with oxygen atoms rather than nitrogen atoms and acting as hydrogen-bond donors rather than acceptors. Possible reasons include the greater number of acceptor sites in proteins and the fewer geometrical restrictions imposed by acceptors (Baker & Hubbard, 1984; Baker, 1995). There is also a predominance of interactions with main-chain atoms rather than side-chain atoms: on average ∼40% with main-chain C=O groups, 15% with main-chain NH and 45% with side-chain groups (Baker & Hubbard, 1984; Thanki *et al.*, 1988). Favoured main-chain binding sites include the N- and C-termini of helices, C=O groups on the solvent-exposed sides of helices, the edge strands of β-sheets, and the ends of strands where they add extra inter-strand hydrogen bonds at the position where the strands diverge (Thanki *et al.*,

1991). Among side chains, the most highly hydrated appear to be Asp and Glu, whose COO^- groups bind, on average, two water molecules each (Baker & Hubbard, 1984; Thanki *et al.*, 1988). On the other hand, the best-ordered water sites are created by residues whose side chains simultaneously make hydrogen bonds to other protein atoms (His, Asp, Asn, Arg) or may be sterically restricted (Tyr, Trp).

The distributions of water molecules around protein groups follow the geometrical patterns expected from simple bonding ideas (Baker & Hubbard, 1984; Thanki *et al.*, 1988). Interactions with NH groups are linear, and those with C=O groups show a preferred angle of ~130° at the oxygen-atom acceptor, consistent with interaction with an oxygen-atom lone pair; restriction to the peptide plane is not very strong, however. Although the distributions around polar side chains generally follow the expected patterns (Thanki *et al.*, 1988), there is little evidence of ordered water clusters around non-polar groups. This may be because water clusters need to be 'anchored' by hydrogen bonding to polar groups to be seen crystallographically.

22.2.6. Hydrogen bonding in nucleic acids

Hydrogen bonding by purine and pyrimidine bases is, together with base stacking, a major determinant of nucleic acid structure. With so many hydrogen-bonding groups, there are many potential modes of interaction between bases (Jeffrey & Saenger, 1991). Those that are actually found in DNA and RNA structures are, however, much more restricted in number, at least based on presently available experimental data.

22.2.6.1. *DNA*

DNA structure is dominated by the prevalence of duplex structures and hence by the classic Watson–Crick hydrogen-bonding pattern of A–T and G–C base pairs. This hydrogen-bonding pattern is not affected by whether the double helix has A-form, B-form, or Z-form geometry. Other hydrogen-bonding modes in DNA are probably very rare, arising only as a result of mutations (which produce mismatches), chemical modifications, such as methylation, or other disturbances, such as the binding of drugs or proteins so as to alter DNA conformation. Mismatches can give stable hydrogen bonding but at the expense of local perturbations of the DNA structure.

22.2.6.2. *RNA*

In contrast to DNA, RNA molecules generally form single-stranded structures, which are correspondingly much more complex and less regular. This means that catalytic and other activities can be generated in addition to their information-carrying roles. Current knowledge of detailed RNA three-dimensional structure is limited to transfer RNAs and several ribozymes, including a large ribosomal RNA domain (Cate *et al.*, 1996). Even from this small sample, however, it is clear that a great diversity of hydrogen-bonding interactions exists; RNA molecules contain regions of double-helical structure, often with classical Watson–Crick A–U and G–C base pairing, but these regions are interspersed with loops and bulges and tertiary interactions between the various secondary-structural (double-helical) elements. These interactions include many unconventional base pairings (*e.g.* see Fig. 22.2.6.1).

Some RNA structural motifs may prove to be of widespread general importance in RNA molecules. One example is a sharp turn with sequence CUGA in the hammerhead ribozyme that exactly matches turns in tRNAs (Pley *et al.*, 1994). Another is the GNRA tetraloop structure (N = any base, R = purine). This loop has a well defined structure, stabilized by hydrogen bonding and stacking involving its own bases, and it also presents further hydrogen-bonding groups that can dock into 'receptor' structures in other parts of the RNA molecule. This results in triple or quadruple base interactions (Fig. 22.2.6.1) that tie different parts of the RNA structure together; the parallel with hydrogen-bonding side chains in proteins is very strong. The 2′-hydroxyls of ribose groups are also used in some of these interactions (Fig. 22.2.6.1). Further ribose interactions involve interdigitated ribose groups that line the interfaces between adjacent helices such that pairs of riboses interact by hydrogen bonding through their 2′-hydroxyl groups, forming 'ribose zippers' As many more RNA structures are determined experimentally, it is likely that more hydrogen-bonding motifs will be recognized, and their full role in RNA structure can be better assessed than at our present, imperfect state of knowledge.

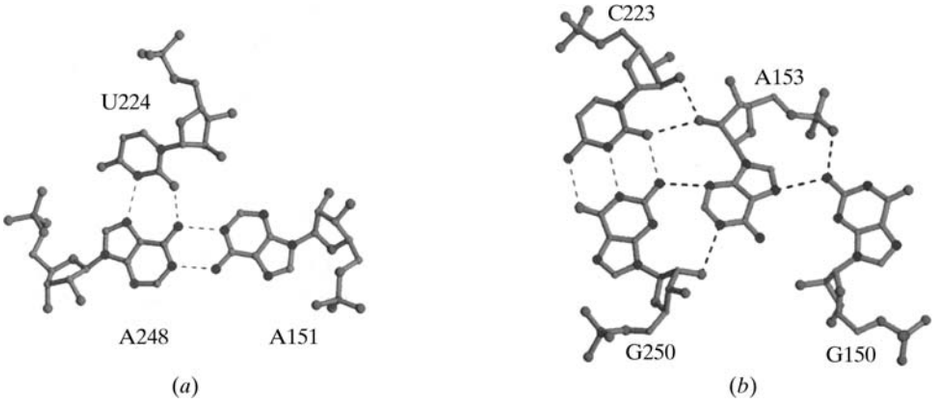

Fig. 22.2.6.1. Hydrogen-bonding interactions in RNA tertiary structure. In (*a*), a triple base interaction is shown. In (*b*), G150 and A153 of a GAAA tetraloop participate in multiple hydrogen-bond interactions involving bases, riboses and phosphate. Reprinted with permission from Cate *et al.* (1996). Copyright (1996) American Association for the Advancement of Science.

22.2.7. Non-conventional hydrogen bonds

The vast majority of hydrogen bonds in biological macromolecules involve nitrogen and oxygen donors exclusively. Nevertheless, several other interactions have all the characteristics of hydrogen bonds and clearly contribute to structure and stability where they occur.

22.2.7.1. *C—H···O hydrogen bonds*

Sutor (1962) first summarized evidence for C—H···O hydrogen bonds following earlier suggestions by Pauling (1960), and current evidence has been nicely summarized in several recent articles (Derewenda *et al.*, 1995; Wahl & Sundaralingam, 1997). The energy of C—H···O hydrogen bonds has been generally estimated as ~0.5 kcal mol^{-1} (about 10% of an N—H···O interaction) but may be higher, especially in hydrophobic environments. It also depends on the acidity of the C—H proton, with methylene (CH_2) and methyne (CH) groups being most favourable.

A number of examples of C—H···O hydrogen bonds can be found in nucleic acid structures (Wahl & Sundaralingam, 1997). The best known is that between the backbone O5′ oxygen and a purine C(8)—H or pyrimidine C(6)—H, when the bases are in the *anti* conformation. Another example is given by a U–U base pair, in which the two bases form a conventional N(3)—H···O(4) hydrogen bond and a C(5)—H···O hydrogen bond.

references