

Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures

Narayanan Eswar and C.Ramakrishnan¹

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

¹To whom correspondence should be addressed
Email: ramki@crmbu2.mbu.iisc.ernet.in

A total of 19 835 polar residues from a data set of 250 non-homologous and highly resolved protein crystal structures were used to identify side-chain main-chain (SC-MC) hydrogen bonds. The ratio of the number of SC-MC hydrogen bonds to the total number of polar residues is close to 1:2, indicating the ubiquitous nature of such hydrogen bonds. Close to 56% of the SC-MC hydrogen bonds are local involving side-chain acceptor/donor (i) and a main-chain donor/acceptor within the window $i-5$ to $i+5$. These short-range hydrogen bonds form well defined conformational motifs characterized by specific combinations of backbone and side-chain torsion angles. (a) The Ser/Thr residues show the greatest preference in forming intra-helical hydrogen bonds between the atoms O^i and O_{i-4} . More than half the examples of such hydrogen bonds are found at the middle of α -helices rather than at their ends. The most favoured motif of these examples is $\alpha_R\alpha_R\alpha_R\alpha_R(g^-)$. (b) These residues also show great preference to form hydrogen bonds between O^i and O_{i-3} , which are closely related to the previous type and though intra-helical, these hydrogen bonds are more often found at the C-termini of helices than at the middle. The motif represented by $\alpha_R\alpha_R\alpha_R\alpha_R(g^+)$ is most preferred in these cases. (c) The Ser, Thr and Glu are the most frequently found residues participating in intra-residue hydrogen bonds (between the side-chain and main-chain of the same residue) which are characterized by specific motifs of the form $\beta(g^+)$ for Ser/Thr residues and $\alpha_R(g^-g^+t)$ for Glu/Gln. (d) The side-chain acceptor atoms of Asn/Asp and Ser/Thr residues show high preference to form hydrogen bonds with acceptors two residues ahead in the chain, which are characterized by the motifs $\beta(t')\alpha_R$ and $\beta(t)\alpha_R$, respectively. These hydrogen bonded segments, referred to as Asx turns, are known to provide stability to type I and type I' β -turns. (e) Ser/Thr residues often form a combination of SC-MC hydrogen bonds, with the side-chain donor hydrogen bonded to the carbonyl oxygen of its own peptide backbone and the side-chain acceptor hydrogen bonded to an amide hydrogen three residues ahead in the sequence. Such motifs are quite often seen at the beginning of α -helices, which are characterized by the $\beta(g^+)\alpha_R\alpha_R$ motif. A remarkable majority of all these hydrogen bonds are buried from the protein surface, away from the surrounding solvent. This strongly indicates the possibility of side-chains playing the role of the backbone, in the protein interiors, to satisfy the potential hydrogen bonding sites and maintaining the network of hydrogen bonds which is crucial to the structure of the protein.

Keywords: accessibility/conformation/dihedral angles/hydrogen bonds/motifs/side-chains

Introduction

The structure of a protein in solution is stabilized by a network of hydrogen bonds formed by the polar atoms of the polypeptide chain (Baker and Hubbard, 1984; Dill, 1990; McDonald and Thornton, 1994b). There are predominantly three types of hydrogen bond observed in protein structures, which are main-chain to main-chain, side-chain to main-chain and side-chain to side-chain, apart from the solvent mediated hydrogen bonds of any of these types. The majority of main-chain to main-chain hydrogen bonds are local in nature involving less than six consecutive residues of the sequence in the hydrogen bonded ring (Ippolito *et al.*, 1990; Stickle *et al.*, 1992). These hydrogen bonds show such characteristic features of regular secondary structures observed in proteins, namely the α -helices, β -sheets and the sharp turns (α -, β -, γ - and π -turns), that many algorithms use them as a diagnostic to identify these structures (Kabsch and Sander, 1983; Wilmot and Thornton, 1988; Milner-White, 1990; Nataraj *et al.*, 1995). Further, the relationships between the backbone dihedral angles of the intervening residues participating in these structures and the resulting hydrogen bonds are well characterized and are unique (Venkatachalam, 1968; Nataraj *et al.*, 1995), leading to the fact that identifications based on dihedral angles are as reliable as those based on the hydrogen bonds.

While the side-chain side-chain hydrogen bonds are a result of the tertiary arrangement of the structural components of proteins, the side-chain main-chain hydrogen bonds, by polar side-chains, present a case comparable with the main-chain main-chain hydrogen bonds. In other words, a majority of side-chain main-chain hydrogen bonds are also local in nature and involve less than six intervening residues on either side of the residue under consideration (Baker and Hubbard, 1984; Stickle *et al.*, 1992). Another factor of resemblance between these two types of hydrogen bond is the number of dihedral angles that define the conformation of the residues, which form the hydrogen-bonded ring and consequently the hydrogen bond itself. Though, in the latter case, a combination of the side-chain and the backbone torsion angles defines the hydrogen bond.

Side-chain main-chain hydrogen bonds have been observed as important stabilization motifs at the start sites of α -helices (Bordo and Argos, 1994; Doig *et al.*, 1997; Aurora and Rose, 1998). Many recent analyses have been reported in the literature (Penel *et al.*, 1999; Vijayakumar *et al.*, 1999; Wan and Milner-White, 1999a,b) in which the focus has been restricted to those motifs involved in the formation of secondary structures. Other reports have been on specific types of motifs like the Asx motifs (Richardson and Richardson, 1981; Richardson and Richardson, 1989; Questal *et al.*, 1993; Wilson and Finlay, 1997), backbone mimicry by selected residues (Eswar and Ramakrishnan, 1999), etc. While many analyses on the backbone dependence of side-chain rotamers have also been reported (McGregor *et al.*, 1987; Dunbrack and Karplus, 1993; Schrauber *et al.*, 1993; Stites and Pranata, 1995; Bower *et al.*, 1997; Chakrabarti and Pal, 1998), the results presented have

been derived from either a statistical treatment of observed rotamers vis-à-vis the backbone conformation or based on homologous structures.

The analysis described in this paper presents a comprehensive characterization of those side-chain main-chain hydrogen bonds that are frequently observed in protein structures in terms of the combinations of the torsion angles defining the side-chain conformation and those of the backbone. In this sense, this work is complementary to those of Dunbrack and Karplus (1993) since a relationship between the side-chain torsion angles and the backbone is involved. The results are presented in the light of the consequent hydrogen bonds that might be speculated to be the stabilizing motif for a particular rotamer of the polar residues.

Materials and methods

A data-set of 250 non-homologous and high resolution protein crystal structures derived from the Protein Data Bank (PDB, Bernstein *et al.*, 1971) was used for the analysis. The PDB codes for the proteins used are given below with the polypeptide chain identifiers indicated whenever homologous multiple chains are present.

1aan, 1aazA, 1abe, 1abk, 1acf, 1acx, 1afgA, 1ahc, 1ak3A, 1alc, 1ald, 1alkA, 1amp, 1ankA, 1aozA, 1apmE, 1arb, 1arp, 1ars, 1ast, 1bbhA, 1bbpA, 1bgc, 1bgh, 1bmdA, 1brsD, 1bsaA, 1byb, 1cbn, 1ccr, 1cewI, 1cgt, 1chmA, 1cmbA, 1cot, 1cpcA, 1cpcB, 1cpn, 1cseE, 1cse I, 1csh, 1ctf, 1cus, 1ddt, 1dfnA, 1dmb, 1dri, 1dsbA, 1eca, 1esl, 1ezm, 1fas, 1fdn, 1fgvH, 1fiaA, 1fkf, 1flp, 1flv, 1fna, 1frrA, 1fus, 1fxl, 1fxd, 1gd1O, 1gia, 1gky, 1glqA, 1glt, 1gog, 1gox, 1gp1A, 1gpr, 1hel, 1hip, 1hleA, 1hleB, 1hoe, 1hpi, 1hsbA, 1hsbB, 1hslA, 1huw, 1hvkA, 1hyp, 1iag, 1ifb, 1isaA, 1isuA, 1lcf, 1lec, 1lib, 1lis, 1lldA, 1ltsA, 1ltsC, 1ltsD, 1mba, 1mbd, 1mdc, 1mjc, 1molA, 1mpp, 1nar, 1nbaA, 1nlkR, 1npc, 1nscA, 1olbA, 1onc, 1opaA, 1ovaA, 1pda, 1pgb, 1phc, 1php, 1pii, 1pk4, 1pmy, 1poc, 1poh, 1ppa, 1ppbH, 1ppbL, 1ppfE, 1ppt, 1prn, 1ptf, 1ptsA, 1r69, 1rbp, 1rdg, 1rec, 1ris, 1rnh, 1ropA, 1sacA, 1sbp, 1sgt, 1shaA, 1shfA, 1shg, 1sim, 1sltA, 1smrA, 1srdA, 1stn, 1tca, 1ten, 1tfg, 1tgn, 1tgsI, 1tgxA, 1thbA, 1tml, 1ton, 1trb, 1trkA, 1ubq, 1utg, 1whtA, 1whtB, 1xib, 1ypiA, 256bA, 2acq, 2act, 2alp, 2apr, 2bbkH, 2bbkL, 2bmhA, 2cab, 2ccyA, 2cdv, 2chsA, 2ci2I, 2cmd, 2cpl, 2ctvA, 2cy3, 2cyp, 2end, 2fcr, 2gbp, 2gstA, 2had, 2hbq, 2hmqa, 2lh7, 2lhb, 2ltnA, 2ltnB, 2lzm, 2mcm, 2mltA, 2mnr, 2msbA, 2ohxA, 2ovo, 2pabA, 2pia, 2plt, 2por, 2prk, 2rhe, 2rspA, 2sarA, 2scpA, 2sga, 2sn3, 2spcA, 2trxA, 2tscA, 2wrpR, 2ztaA, 351c, 3app, 3b5c, 3bcl, 3blm, 3c2c, 3chy, 3cla, 3cox, 3dfr, 3dni, 3drcA, 3ebx, 3est, 3grs, 3il8, 3mdsA, 3psg, 3rp2A, 3rubL, 3rubS, 3sdhA, 3tgl, 4azuA, 4bp2, 4cpv, 4enl, 4fxn, 4gcr, 4ilb, 4icb, 4insC, 4insD, 4mt2, 4tnc, 5chaA, 5cpa, 5fd1, 5p21, 5pti, 5rubA, 6ldh, 7acn, 7rsa, 8dfr, 8fabA, 8fabB, 9wgaA.

Hydrogen bond criteria

Hydrogen bonds were identified using the criteria of the well known hydrogen bond length l (donor...acceptor) and angle θ (hydrogen-donor...acceptor) in the case of the N-H...O hydrogen bond, where the amide group comes from the peptide backbone (Ramakrishnan and Prasad, 1971). The position of the hydrogen atom was always fixed assuming standard geometry at the peptide nitrogen.

In the cases of hydrogen bonds involving (a) side-chain amides (as in Asn/Gln examples) and backbone oxygen atoms or (b) side-chain hydroxyl oxygens (as in Ser/Thr examples)

and backbone oxygen atoms, the lengths were calculated as described above. However, due to the ambiguity of positioning of the side-chain hydrogen atoms, the following angles were calculated: $\mu(\text{C}^\delta\text{-N}^\epsilon\text{...acceptor})$ for Asn residues, $\mu(\text{C}^\delta\text{-N}^\epsilon\text{...acceptor})$ for Gln residues and $\mu(\text{C}^\beta\text{-O}^\gamma\text{...acceptor})$ for Ser/Thr residues. The limits used for identifying the hydrogen bonds are $2.4 \text{ \AA} \leq l \leq 3.5 \text{ \AA}$ for all types of hydrogen bonds (Mitra and Ramakrishnan, 1981), $0^\circ \leq \theta \leq 40^\circ$ for N-H...O hydrogen bonds where N-H is the backbone amide, $80^\circ \leq \theta \leq 160^\circ$ for N-H...O hydrogen bonds where N-H is the side-chain amide and $70^\circ \leq \theta \leq 150^\circ$ for O-H...O hydrogen bonds (Mitra and Ramakrishnan, 1977).

Nomenclature

Throughout the text SC-MC hydrogen bonds have been denoted by the sequential positions of the residues providing donor and acceptor atoms respectively (for example, $i/i-4$, $i+2/i$, etc.) with respect to the residue ' i ', which always refers to the residue whose side-chain is involved in the hydrogen bond.

Backbone dihedral angles are denoted by the following convention when the (ϕ, ψ) s fall into the indicated ranges:

$$\alpha_R \rightarrow -140^\circ \leq \phi \leq -30^\circ, -90^\circ \leq \psi \leq 45^\circ;$$

$$\beta \rightarrow -180^\circ \leq \phi \leq -30^\circ, 60^\circ \leq \psi \leq 180^\circ \text{ or } -180^\circ \leq \psi \leq -150^\circ;$$

$$\alpha_L \rightarrow 20^\circ \leq \phi \leq 125^\circ, -45^\circ \leq \psi \leq 90^\circ.$$

Results and discussion

A data-set of 250 high resolution and non-homologous protein crystal structures derived from the Brookhaven Protein Data Bank (PDB, Bernstein *et al.*, 1977), consisting of 51 942 residues, was used for the analysis. Almost 49% ($N = 25\ 382$) of the residues in the data-set were polar in nature, comprising the residues Arg, His, Lys, Asn, Gln, Ser, Thr, Asp, Glu and Tyr. Since the analysis focuses on the hydrogen bonds made by the polar atoms of the side-chain, a B-factor cut-off (30 \AA^2) was applied to these polar residues, which finally resulted in a refined set of 19 835 polar residues which was then used for the analysis. A total of 20 different types of polar groups were identified from the side-chains of these 10 polar residues, which were capable of either donating or accepting a hydrogen bond from the polar groups of the main-chain. All such side-chain main-chain (SC-MC) hydrogen bonds were computed and the results are summarized in Table I.

At the outset, the data presented in Table I shows that the ratio of SC-MC hydrogen bonds to the number of polar residues is close to half (10 900 SC-MC hydrogen bonds for 19 835 polar residues considered for the analysis), indicating that almost every other polar residue engages in such hydrogen bonds. Among the polar residues, the indulgence of residues with shorter side-chains (Asp, Asn, Ser, Thr, Glu and Gln) in such hydrogen bonds is much higher than those with longer side-chains. It can also be calculated that about 56% (6140 out of 10 900) of all the SC-MC hydrogen bonds are local, involving a side-chain acceptor/donor (of residue i) and a main-chain donor/acceptor not farther than five residues on either side of it (between residues $i-5$ and $i+5$). The distribution of the number of hydrogen bonds in this window is also shown in Table I. In order to derive meaningful results, the analysis has been further restricted to those SC-MC hydrogen bonds made by a polar atom of a residue (i), which accounts for more than 10% of the total number of observed SC-MC

Table 1. The distribution of the side-chain main-chain (SC-MC) hydrogen bonds made by the polar atoms of the side-chain in the indicated amino acid (*i*) to the polar atoms of the backbone in the interval between *i*-5 and *i*+5

Residue	No. in data set	No. with B factors $\leq 30 \text{ \AA}^2$	Polar atoms in side chain ^a	Total no. of SC-MC H-bonds	No. of short range SC-MC H-bonds	No. of H-bonds made by the indicated polar atom of a side-chain to main-chain atoms of residues indicated by the relative position										
						-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Ser	3463	2924	OG ^d	1314	1021	14	242	247	14	49	359	24	44	11	9	8
Thr	3226	2890	OG ^a	671	462	5	6	11	0	0	2	27	227	172	11	1
Asp	3172	2493	OG1 ^d	1672	1292	15	366	204	34	118	482	19	25	6	14	9
Asn	2469	1974	OG1 ^a	517	369	2	3	6	0	2	0	43	169	127	14	3
			OD	1463	950	8	24	87	3	0	51	93	423	190	49	22
			OD1	641	431	6	2	6	0	0	33	34	214	108	20	8
			ND2	829	341	25	117	31	27	28	69	11	5	4	13	11
His	1089	955	ND1	160	98	3	37	3	18	2	5	22	1	1	5	1
			NE2	87	4	2	1	0	0	0	0	0	0	1	0	0
Glu	2920	1931	OE	667	355	8	22	99	12	5	164	9	21	5	7	3
Gln	1862	1368	OE1	262	110	0	10	53	2	0	33	0	6	2	4	0
			NE2	559	234	22	48	25	32	13	38	30	16	2	5	3
Lys	3135	2037	NZ	569	141	10	24	23	31	6	3	19	8	4	8	5
Arg	2112	1502	NE	196	56	9	4	15	6	2	2	4	3	1	5	5
			NH1	617	166	26	26	31	12	30	14	12	2	1	7	5
			NH2	525	87	16	11	14	10	2	4	6	7	4	10	3
Tyr	1934	1761	OH ^d	22	2	0	2	0	0	0	0	0	0	0	0	0
			OH ^a	129	21	9	4	4	0	0	0	1	0	1	2	0
Totals	25382	19835		10900	6140	175	945	856	199	256	1259	354	1171	640	183	87

^aSuperscript 'd' indicates that the atom acts as a donor and 'a' denotes an acceptor.^bNumbers in boldface indicate those SC-MC hydrogen bonds which account for more than 10% of the total SC-MC hydrogen bonds made by that atom.

Table II. Propensity of a polar atom of a residue (*i*) to form a SC-MC hydrogen bond of the indicated kind^a

Residue	Atoms involved		Type	No. of examples	Propensity
	Don/Acc	Acc/Don			
Ser	$O^{\gamma}_i(H)$	O_{i-4}	<i>ii-4</i>	242	1.82
		O_{i-3}	<i>ii-3</i>	247	2.04
		O_i	<i>ii</i>	359	1.94
	O^{γ}_i	N_{i+2}	<i>i+2/i</i>	227	1.43
		N_{i+3}	<i>i+3/i</i>	172	1.96
Thr	$O^{\gamma}_i(H)$	O_{i-4}	<i>ii-4</i>	366	2.85
		O_{i-3}	<i>ii-3</i>	204	1.77
		O_i	<i>ii</i>	482	2.65
	O^{γ}_i	N_{i+2}	<i>i+2/i</i>	169	1.07
		N_{i+3}	<i>i+3/i</i>	127	1.45
Asp	O^{δ}_i	N_{i+2}	<i>i+2/i</i>	423	2.93
		N_{i+3}	<i>i+3/i</i>	190	2.42
Asn	$O^{\delta 1}_i$	N_{i+2}	<i>i+2/i</i>	214	1.88
		N_{i+3}	<i>i+3/i</i>	108	1.77
	$N^{\delta 2}_i$	O_{i-4}	<i>ii-4</i>	117	1.29
His	$N^{\delta 1}_i$	O_{i-4}	<i>ii-4</i>	37	0.95
		O_{i-2}	<i>ii-2</i>	18	1.68
		O_{i+1}	<i>ii+1</i>	22	1.30
Glu	O^{ϵ}_i	N_{i-3}	<i>ii-3</i>	99	1.03
		N_i	<i>ii</i>	164	1.33
Gln	O^{ϵ}_i	N_{i-3}	<i>ii-3</i>	53	0.83
		N_i	<i>ii</i>	33	0.38
Tyr	O^{η}_i	O_{i-4}	<i>ii-4</i>	2	0.02

^aOnly the SC-MC hydrogen bonds shown in boldface in Table I are considered.

hydrogen bonds made by that atom. All such hydrogen bonds are shown in boldface in Table I.

Inclusion of the above said criteria in the analysis further shows that the short range SC-MC hydrogen bonds are mainly restricted to the window ranging from *i-4* to *i+3* (indicated by the bold-faced numbers in Table I occurring only in this region). It is noteworthy that the frequency of occurrence of short-range hydrogen bonds made by the small polar residues is remarkably higher than those of the others. Another interesting point arising out of Table I is that for these residues (Asn, Asp, Ser and Thr), the prominent hydrogen bonds (shown in boldface) made to the first half of the window (*i-5* to *i*) are all those in which the side-chain atom of residue *i* is the donor, in contrast to those in the window (*i+1* to *i+5*) where they are acceptors. This fact presents a case of directional hydrogen bonds, similar to the main-chain main-chain hydrogen bonds (preponderance of N→1 hydrogen bonds over 1→N), and are probably analogously restricted by the local stereochemistry.

In order to assess the statistical significance of the numbers shown in boldface in Table I, the propensity of a polar atom of residue *i* to form the indicated hydrogen bonds was calculated and is shown in Table II, which shows that the types of hydrogen bonds shown in boldface in Table I, are indeed statistically significant and are not the effect of a random distribution.

Finally, the hydrogen bonds are grouped on the basis of the type (spatial disposition of the atom from the main-chain) of the atom in the side-chain and the predominant hydrogen bonds are discussed in detail in each of the following subsections with emphasis on the stereochemistry and defining torsion angles, conformational motifs arising out of the clusters of examples and the accessibility of the residue *i*.

ii-4 hydrogen bonds

Hydrogen bonds of the type *ii-4*, when formed between the main-chain atoms (5→1 hydrogen bonds) are typical of α -helices which bring residue *i* and *i-4* proximal to each other. However, these types of hydrogen bonds are also found to occur between the side-chain and main-chain atom of these residues (Baker and Hubbard, 1984; Lyu *et al.*, 1990; Armstrong and Baldwin, 1993; Aurora *et al.*, 1997). There are a total of 945 examples of such hydrogen bonds in the data-set and as shown in Table I more than 10% of the SC-MC hydrogen bonds made by the residues Ser, Thr, Asn and His (shown in boldface in Table I) fall into this type. On the other hand, propensity of these residues to form these hydrogen bonds (Table II) indicates that only the N^{δ} atom of Asn and the donor O^{γ} atom of Ser/Thr show high preference to form such hydrogen bonds, with Thr showing the maximum ($P = 2.85$) preference to participate in such hydrogen bonds.

Ser/Thr residues

These *ii-4* hydrogen bonds involving Ser/Thr residues are of a O(H)...O type with their γ -hydroxyl oxygen playing the role of a donor atom. These form 16-membered closed ring structures which are conformationally defined by the (ϕ, ψ)s at residues *i-3*, *i-2* and *i-1* and the ϕ of residue *i* along with its χ^1 angle. In order to ease the interpretation of the distributions of these dihedral angles, the ψ of residue *i* is also considered, although a rotation about it does not affect the hydrogen bonded ring. There are 608 examples of such hydrogen bonds in the data-set, of which 242 examples are of Ser and 366 examples are of Thr. The distribution of backbone dihedral angles of residues in the window between residues *i-3* and *i* (data not shown) reveals that there is an almost complete clustering of the angles in the α_R -region. It is found that almost 93% ($N = 567$) of the examples are indeed part of α -helices. In order to investigate the position of these hydrogen bonds within the helix, the secondary structure flanking the residue *i* was analysed. Results indicate that in more than 50% ($N = 326$) of the examples the residue *i* is flanked by at least one complete turn of an α -helix on either side, indicating that these hydrogen bonds are predominantly found in the middle of α -helices. Also, just over one-third (~35%) of the examples occurring in α -helices, 199 examples out of 567, occur at the last turn at the C-terminal end of the helices. On the other hand, there are only 28 examples where they occur in the first turn of the α -helix. Figure 1a shows a representative example of an intra-helical *ii-4* hydrogen bond. The distribution of χ^1 angles in these residues show populations in only the g^- and g^+ conformation, with none in the *t*-conformation (Figure 1b). Close to 72% ($N = 439$) of the examples have the g^- rotamer and 27% ($N = 163$) have the g^+ rotamer leading to two types of conformational motifs, namely, $\alpha_R\alpha_R\alpha_R(g^-)$ and $\alpha_R\alpha_R\alpha_R(g^+)$, respectively.

Asn/His residues

The δ -nitrogen atoms of the Asn and His side-chains too show a tendency to form *ii-4* hydrogen bonds. Compared with the earlier examples of Ser/Thr, these residues have an extra atom (the δ -nitrogen atom) and consequently result in 17-membered SC-MC hydrogen bonded rings. There are a total of 154 examples of such hydrogen bonds and the (ϕ, ψ) distribution of the residues in the window (*i-3* to *i*) again show near complete clustering at the α_R -region (data not shown). A majority of the examples, 132 out of the 154, are indeed part of α -helices. Investigation of the intra-helical position of these

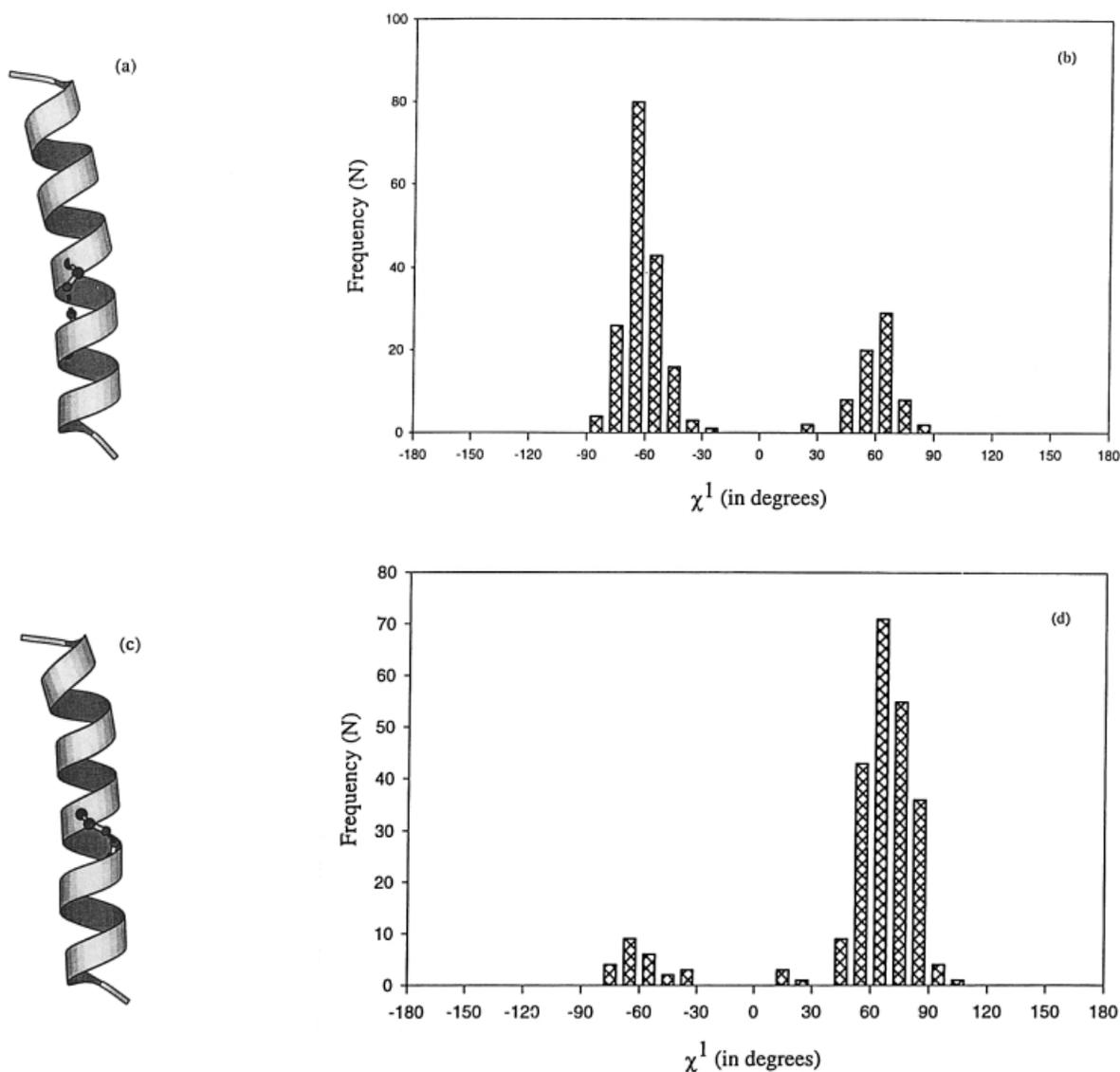


Fig. 1. Representative examples of *i/i-4* and *i/i-3* intra-helical hydrogen bonds from the data-set and the side-chain distributions of the residues at *i*. (a) *i/i-4* SC-MC hydrogen bond from 1gia (helical segment 70–90 with *i* = 80 and *i-4* = 76); (b) the distribution of χ^1 angles of Ser/Thr residues participating in *i/i-4* hydrogen bonds; (c) *i/i-3* SC-MC hydrogen bond 1hsbA (helical segment 61–81 with *i* = 71 and *i-3* = 68) and (d) the distribution of χ^1 angles of the examples of type (c) showing the complementary location of the peaks. Figures (a) and (c) were prepared using MOLSCRIPT (Kraulis, 1991).

hydrogen bonds reveals that these hydrogen bonds are more often found at the C-terminal end of helices than at the middle or the N-terminus. It is found that close to 54% ($N = 72$) of the examples occur at the C-terminus against just 52 examples in the middle and just five examples at the N-terminus. The distribution of χ^1/χ^{21} of residue *i* in these examples shows that the χ^1 angle is almost always restricted to the g^- conformation and the χ^{21} angles are clustered around $\pm 60^\circ$. A closer examination of the results shows that, surprisingly, all except one example of His have positive values for the χ^{21} angle.

i/i-3 hydrogen bonds

The *i/i-3* hydrogen bonds are similar to the *i/i-4* hydrogen bonds in that, at first sight, these also appear to be typical intra-helical hydrogen bonds owing to the fact that the periodicity of α -helices (3.6 residues per turn) roughly brings residues *i-4* and *i-3* spatially close to the residue *i*. From Table I it can be seen that a large number of SC-MC hydrogen bonds made by the residues Ser/Thr and Glu/Gln are of this type. Notwithstanding the h-bonds marked in boldface in Table I, Table II shows

that only the Ser/Thr residues show a really high preference to make such hydrogen bonds. Glu shows a propensity of $P \approx 1$ which neither indicates a preference nor a disfavour and Gln has a propensity $P < 1$, indicating a disfavour to form such hydrogen bonds. Consequently, only the Ser/Thr residues have been chosen for further analysis. There are a total of 451 examples of Ser/Thr residues participating in such hydrogen bonds out of which 247 examples are of Ser and 204 of Thr. These residues form closed ring structures comprising 13 atoms through a O(H)...O hydrogen bond, which has been known to be a mimic of the main-chain main-chain 5 \rightarrow 1 hydrogen bonded α -turn (Eswar and Ramakrishnan, 1999). Investigation of the distribution of the backbone dihedral angles of the residues in the window *i-2* to *i*, which defines the conformation of the hydrogen bond along with the side-chain dihedral angles at residue *i*, shows a clustering very similar to that of the *i/i-4* hydrogen bonds. Out of the 451 examples, almost close to 75% ($N = 338$) of the examples fall into the α_R -region of the Ramachandran map (Ramachandran *et al.*,

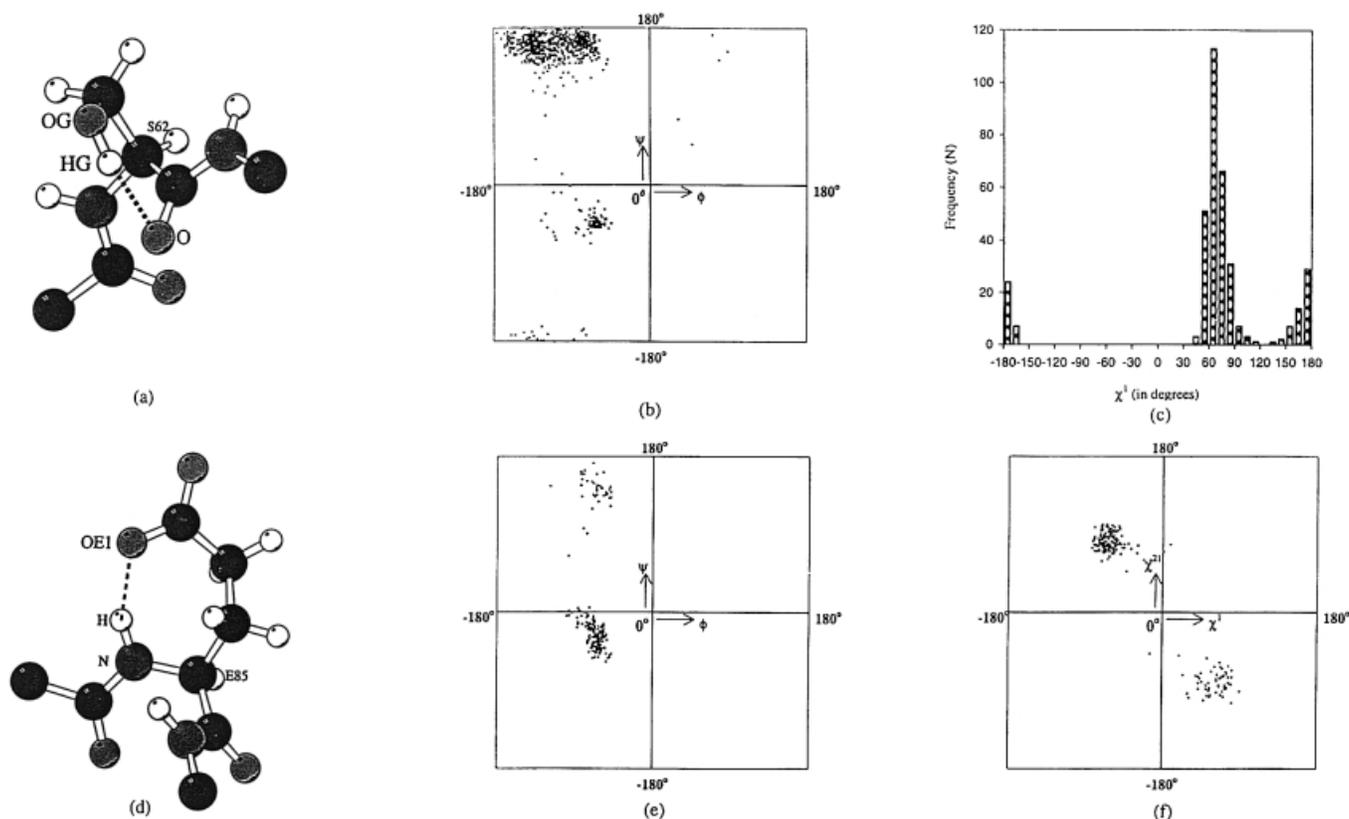


Fig. 2. Representative examples of *i/i* type hydrogen bonds made by (a) Ser/Thr examples, (9wgaA, Ser62), (b) plot of backbone dihedral angles (ϕ, ψ) of the examples of type shown in (a), and (c) the distribution of their side-chain dihedral angle χ^1 . (d) An example of *i/i* hydrogen bond made by Glu/Gln examples (1dsbA, Glu85) and (e) the plots of the backbone dihedral angles and (f) the plot of (χ^1, χ^2) angles of these examples. It can be seen that the hydrogen bonded conformation is independent of the ϕ -angle in (a) and the ψ -angle in (d). Figures (a) and (d) were prepared using MOLSCRIPT (Kraulis, 1991).

1963; Ramakrishnan and Ramachandran, 1965) and 106 examples into the β -region. It was found that close to two-thirds of the total number of examples (301 out of 451) and also 89% (301 out of 338) of the examples occurring in the α_R -region, are indeed part of α -helices. Investigation of the position of these hydrogen bonds within the helix showed that there are only 69 examples in which the *i/i*-3 hydrogen bond is in the middle of a helix. In contrast, more than 77% ($N = 234$) of the examples occur at the C-terminal end of α -helices and the remaining 104 examples occur at the N-terminal end. This enforces the fact that a majority of examples occur at the ends of α -helices.

Investigation of the χ^1 rotamer distribution of the residues involved in the *i/i*-3 SC-MC hydrogen bonds reveals that the prominent clustering is at the g^+ conformation (Figure 1d). Almost close to 85% ($N = 383$) of the examples have the χ^1 values in the range of the g^+ conformation. Figure 1c shows an example of an *i/i*-3 hydrogen bond from the middle of a helix and it is evident from Figure 1a-d that, in comparison, the *i/i*-4 and *i/i*-3 SC-MC hydrogen bonds show peaks at complimentary χ^1 positions (g^- and g^+ respectively) owing to the periodicity of the helix. A combination of the side-chain rotamers and the main-chain angles of the residues in the window $i-2$ to i yields two conformational motifs of this hydrogen bond which are $\alpha_R\alpha_R\alpha_Rg^+$ and $\alpha_R\alpha_R\alpha_Rg^-$.

i/i type hydrogen bonds

These types of hydrogen bond are unique since they occur between the side-chain and peptide backbone of the same

residue. From Table I it can be seen that there are a total of 1259 examples of such intra-residue hydrogen bonds in the data-set. A major fraction of the examples comes from Thr (38% or $N = 482$ out of 1259), Ser (29% or $N = 359$) and Glu residues (13% or $N = 164$). It can also be seen that 28.8% ($N = 482$ out of 1672) of the SC-MC hydrogen bonds made by the Thr side-chain, 27.3% ($N = 359$ out of 1314) of Ser, 24.6% ($N = 164$ out of 667) of Glu and 12.6% ($N = 33$ out of 262) of Gln are of this type. Propensities of the residues to form such hydrogen bonds can be seen from Table II, which highlights the fact that only Thr, Ser and Glu show high propensities to form such hydrogen bonds. Though more than 10% of the SC-MC hydrogen bonds made by the Gln side-chain are of the *i/i* type (shown in boldface in Table I), the propensity of the residue to form such hydrogen bonds is very low ($P = 0.38$).

The Ser and Thr residues form these *i/i* hydrogen bonds through a $O^{\gamma}(H)\dots O$ type of interaction resulting in a six-membered closed ring conformation, while the Glu and Gln residues form a $N(H)\dots O^{\epsilon}$ type of hydrogen bond, consisting of seven atoms in the ring, through their side-chain ϵ -carbonyl oxygens. Accordingly, the number of degrees of freedom that defines the hydrogen bonded conformation varies between the two sets of residues, Ser/Thr and Glu/Gln. A representative example from each of these two sets is shown in Figure 2a and d. It can be seen that in the Ser/Thr examples, the hydrogen bonded conformation is independent of the value of ϕ , while for the Glu/Gln examples they are independent of the value of ψ at residue i .

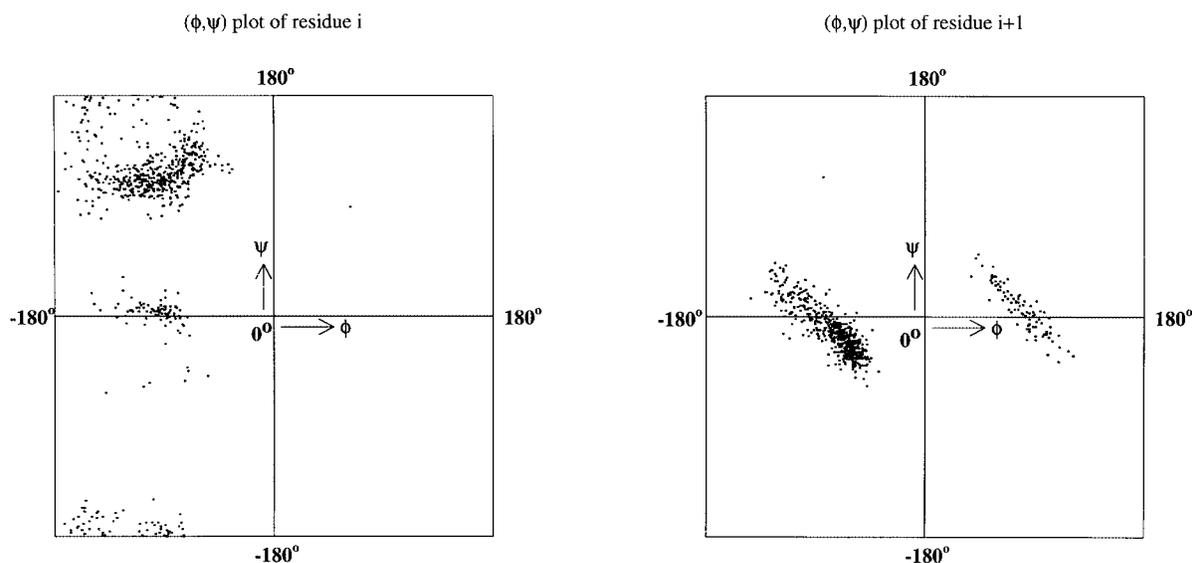


Fig. 3. The (ϕ, ψ) plot of the backbone dihedral angles of residues i and $i+1$ of the examples of Asp/Asn participating in $i+2/i$ SC-MC hydrogen bonds.

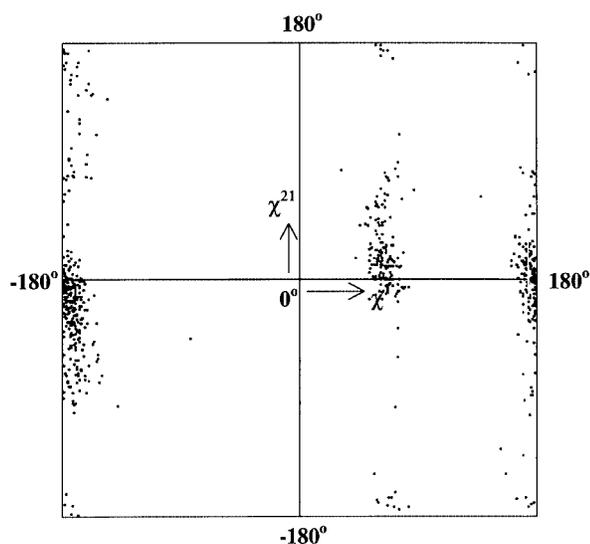


Fig. 4. The distribution of the χ^1/χ^{12} angles of Asp/Asn examples participating in $i+2/i$ type hydrogen bonds.

Ser/Thr residues

There are a total of 841 examples of Ser/Thr (359 examples of Ser and 482 examples of Thr) i/i SC-MC hydrogen bonds. The (ϕ, ψ) and χ^1 distributions for these examples are shown in Figure 2b and c. It can be seen that the (ϕ, ψ) points of these examples are clustered around the usual β - or the α_R -regions of the Ramachandran map. The ψ -values form two distinct clusters, one close to $\psi = +150^\circ$ and the other close to $\psi = -40^\circ$. The dependence on the ψ -value is evident here since there are almost no points near the left-handed α -helical (α_L) region, which is characterized by $\psi \approx +40^\circ$, indicating that only those backbone conformations which have ψ in one of the two clusters favours formation of this hydrogen bond. Close to 87% ($N = 733$) of the examples fall into the β -region and 12% ($N = 99$) into the α_R -region. More than 50% of the examples in the two clusters are part of standard secondary structures; 401 examples out of the 733 which occur in the β -region are part of β -strands and 77 out of the 99 which occur in the α_R -region are part of α -helices. The distribution

of the χ^1 angles also shows a distinct clustering only around the t or the g^+ rotamers. Investigation of the two distinct clusters of the backbone dihedral angles and the side-chain torsion angles revealed that, only two kind of motifs, namely, $\alpha_R(t)$ and $\beta(g^+)$, favour the formation of such a hydrogen bonded conformation.

Glu/Gln residues

The Glu/Gln residues also exhibit a preference to participate in i/i hydrogen bonds. These residues form seven-membered rings that are closed by a hydrogen bond between its own backbone amide and the ϵ -carbonyl oxygen of the side-chain. As can be seen from Figure 2d, in contrast to examples of the Ser/Thr residues, the i/i hydrogen bonded motif of Glu/Gln side-chains is dependent on the ϕ -angle instead of the ψ as is in the Ser/Thr examples. There are 197 examples of such motifs in the data-set of which 164 are of Glu and 33 are of Gln. Though the Gln residues do not show a propensity to form such hydrogen bonds (Table II) they are considered here since the mode of hydrogen bonding, namely the acceptor ϵ -oxygen, is the same as Glu residues. The (ϕ, ψ) distribution of these examples, shown in Figure 2e and f, clearly shows the restriction of the ϕ -angle to the interval $-90^\circ \leq \phi \leq -45^\circ$. Almost all of the examples have their ϕ -values confined to this interval. Further, there are only two regions of clustering for the backbone dihedral angles. 159 out of the 197 examples have (ϕ, ψ) s in the α_R region and 38 fall into the β -region. The side-chain torsion angles χ^1 and χ^2 are confined to either of the g^+ or g^- conformation with the t -conformation completely absent (Figure 2f). The χ^{31} angle is almost completely restricted to the interval $-90^\circ \leq \chi^{31} \leq +90^\circ$ (indicated as the t' -conformation) with only 26 out of the 197 examples having the χ^{31} near the *trans* region. The 197 examples can be grouped into well defined motifs. The most populated conformational motif is the $\alpha_R(g^-g^+t')$ motifs with 47% ($N = 91$) examples followed by the $\alpha_R(g^+g^-t')$ having 24% ($N = 47$) examples. Majority of the examples falling under the β -region of the backbone dihedral angle space (24 out of 38 examples), belong to the $\beta(g^-g^+t')$ motif.

$i+2/i$ hydrogen bonds

In contrast to the previously discussed types of SC-MC hydrogen bonds, the $i+2/i$ hydrogen bonds are different in the

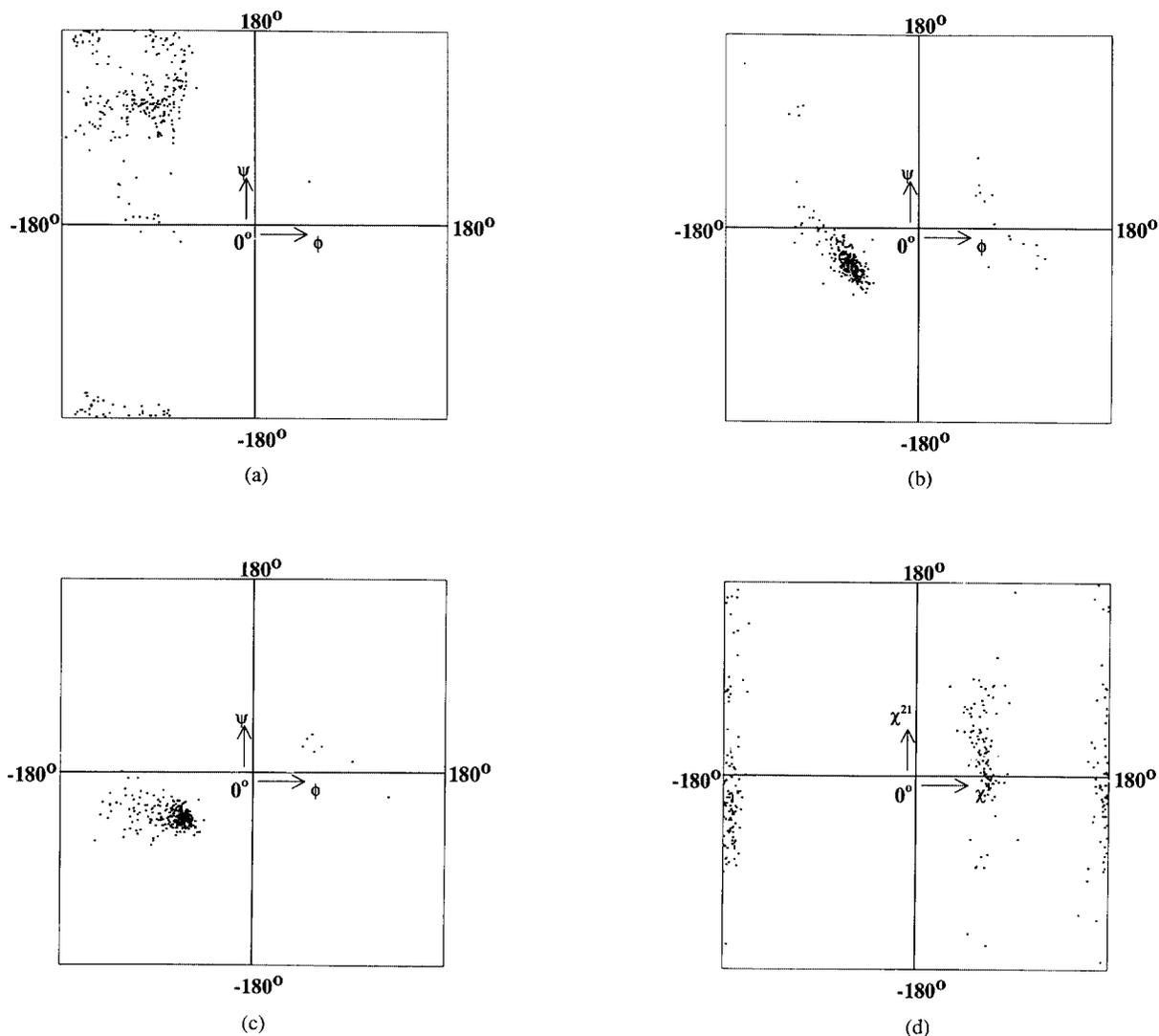


Fig. 5. Plot of (ϕ, ψ) angles at residues (a) i , (b) $i+1$ and (c) $i+2$ of the Asn/Asp examples engaged in $i+3/i$ hydrogen bonds as well as (d) the plot of (χ^1, χ^2) .

sense that these hydrogen bonds are directed towards the main-chain polar groups succeeding residue i , unlike in the former cases where they were all towards residues earlier in the chain. The most predominant of these hydrogen bonds are the ones made by the δ -carboxyl groups of Asp/Asn or the γ -hydroxyl groups of Ser/Thr. Another remarkable point of difference between these hydrogen bonds to the successive residues in the chain and those towards the preceding residues, is the fact that the polar atoms of the Asp/Asn and Ser/Thr side-chains act as an acceptor for hydrogen bonds. These four residues contribute to a total of 1033 examples of such hydrogen bonds out of which 637 examples comprise the δ -carboxyl group of Asn/Asp and the remaining 396 examples involve the acceptor γ -hydroxyl group of Ser/Thr residues. It can also be seen from Table II that out of these four residues, Asp, Asn and Ser show a high propensity to form such hydrogen bonds, for which the Thr residues neither show a preference nor a disfavour with $P \approx 1$.

Asn/Asp residues

The hydrogen bonds between the δ -carboxyl oxygen of Asn/Asp of residue i to the amide hydrogen of residue $i+2$ results

in a hydrogen bonded ring of 10 atoms. This hydrogen bonded conformation has particularly evoked a lot of interest since it mimics the type II' β -turn while also stabilizing the type I and type I' β -turns (Richardson and Richardson, 1989; Ramakrishnan *et al.*, 1996; Eswar and Ramakrishnan, 1999) and is also believed to be crucial at the initiating sites of α -helices (Wan and Milner-White, 1999a). The distribution of the 637 examples of such hydrogen bonds between Asp and Asn is 423:214, the former being almost double that of the latter. The backbone dihedral angle distribution of the residues i and $i+1$, which provides the conformational definition of the hydrogen bond, is shown in Figure 3 and it can be seen that the conformation of residue i is restricted to mostly the β -region with almost 86% ($N = 545$) of the examples occurring in this region, with only a small fraction of the examples ($N = 87$) occurring in the bridge region of the Ramachandran map ($\psi = 0^\circ$). The clustering of (ϕ, ψ) points of residue $i+1$, on the other hand, shows that they are restricted to the bridge regions of the Ramachandran map on either side of the ϕ -axis. Out of the 637 examples, almost as many as 545 examples are followed by an α_R -conformation at $i+1$ and 90 examples are followed by an α_L -conformation.

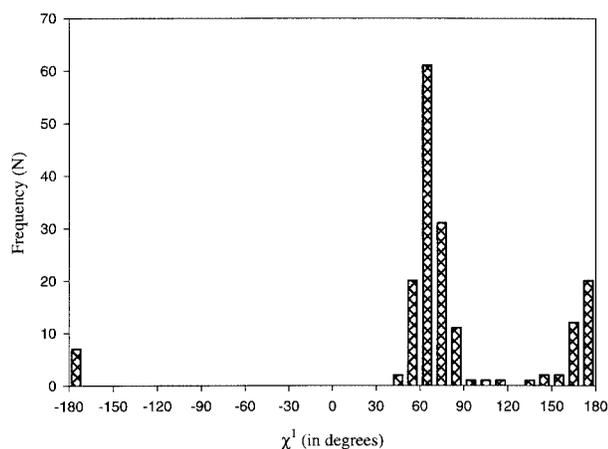


Fig. 6. The distribution of the χ^1 angles of Ser/Thr examples participating in $i+3/i$ SC-MC hydrogen bonds. Clustering at the g^+ and t rotamers is evident.

Table III. Predominant motifs made by polar residues through SC-MC hydrogen bonds

Residues	SC-MC h-bond	No. of indicated SC-MC h-bond	Predominant motifs ^a	Frequency
Ser/Thr	$i/i-4$	608	$\alpha_R\alpha_R\alpha_R\alpha_R(g^-)$	423
			$\alpha_R\alpha_R\alpha_R\alpha_R(g^+)$	142
Asn	$i/i-4$	117	$\alpha_R\alpha_R\alpha_R\alpha_R(g^-g^-)$	97
His	$i/i-4$	37	$\alpha_R\alpha_R\alpha_R\alpha_R(g^-g^+)$	32
Ser/Thr	$i/i-3$	451	$\alpha_R\alpha_R\alpha_R\alpha_R(g^+)$	297
			$\alpha_R\alpha_R\beta(g^+)$	31
			$\beta\alpha_L\beta(g^+)$	36
Ser/Thr	i/i	841	$\beta(g^+)$	721
			$\alpha_R(t)$	94
Glu/Gln	i/i	197	$\alpha_R(g^-g^+t')$	91
			$\alpha_R(g^+g^+t')$	47
Asp/Asn ^b	$i+2/i$	637	$\beta(tt')\alpha_R$	446
			$\beta(g^+t')\alpha_R$	89
			$\alpha_R(g^+t')\alpha_L$	80
Ser/Thr	$i+2/i$	396	$\beta(t)\alpha_R$	203
			$\beta(g^+)\alpha_R$	117
			$\alpha_R(g^+)\alpha_L$	69
Asp/Asn ^b	$i+3/i$	298	$\beta(tt')\alpha_R\alpha_R$	154
			$\beta(g^+t')\alpha_R\alpha_R$	96
Ser/Thr	$i+3/i$	299	$\beta(g^+)\alpha_R\alpha_R$	248
			$\beta(t)\alpha_R\alpha_R$	45

^aSide-chain conformations are shown in parentheses.

^bWhen the χ^{21} values of Asp/Asn side-chains lie within the range $-90^\circ \leq \theta \leq +90^\circ$, it is indicated as t' .

Investigation of the distribution of the χ^1/χ^{21} angles of these examples (Figure 4) shows that the χ^1 angle is restricted to either the g^+ or t rotamer, with the g^- rotamer completely absent. Further, there seems to be a preference for the t rotamer indicated by the fact that almost close to 72% ($N = 439$) of the examples possess this conformation. The χ^{21} angles of these examples cluster either around the t region ($\chi^{21} = 180^\circ$) or in the t' -region (between $-90^\circ \leq \chi^{21} \leq +90^\circ$), with the majority of the examples falling into the latter region. Based on the backbone dihedral angles and the side-chain dihedral angles, the examples can be classified into three predominant motifs; $\alpha_R(g^+t')\alpha_L$ (80 examples); $\beta(g^+t')\alpha_R$ (89 examples); $\beta(tt')\alpha_R$ (446 examples).

The $i+2/i$ hydrogen bond has been described in the literature

as an important motif at the initiating site of α -helices. Investigation of the secondary structure of the residues following the residue i indicates that only 22% ($N = 139$ out of 637) of the examples initiate a helix. The preferred motif at this helix initiating sites is the $\beta(t)\alpha_R$ motif accounting for almost 130 of the 139 examples.

Ser/Thr residues

The $i+2/i$ hydrogen bonds involving the γ -hydroxyl oxygen of the Ser/Thr residues results in a 12-membered closed hydrogen bonded ring with the side-chain oxygen atom acting as an acceptor to the hydrogen bonds from backbone amide groups ahead in the sequence (Wan and Milner-White, 1999b). There are a total of 396 examples of such hydrogen bonds out of which 227 are by Ser and 169 are by Thr residues. An examination of the backbone conformations of residues i and $i+1$ of these examples reveals that a majority (81%, $N = 321$) of the examples adopt the β -conformation at i and at residue $i+1$, the majority occurs in the bridge/ α_R region of the Ramachandran map with as many as 323 examples adopting this conformation and 71 examples adopting the α_L conformation.

The χ^1 values of the examples show clear clustering at only the g^+ and t rotamers. 209 examples out of 396 adopt the t rotamer and 187 adopt the g^+ rotamer. And investigation of the secondary structures following residues i indicate that only 45 examples function at helix initiators.

The fact that the clustering of the backbone dihedral angles of residue i at the β -conformation and the side-chain torsion angle of these examples at g^+ is similar to the clustering of these angles in the case of the i/i SC-MC hydrogen bonds prompted us to investigate the existence of any additional hydrogen bonds through the γ -hydroxyl atom. It was found that in 110 examples, the $i+2/i$ hydrogen bonds had an additional i/i hydrogen bond in which γ -hydroxyl oxygen of Ser/Thr residues additionally act as a donor to the carbonyl oxygen of its own peptide backbone. In order to ascertain if these $i+2/i-i/i$ hydrogen bonds are characteristic of the helix initiating examples, the secondary structures following the motif were analysed. It was found that there are only 18 examples out of the 45, which function as α -helix start sites, which had this additional hydrogen bond.

$i+3/i$ hydrogen bonds

The $i+3/i$ hydrogen bonds are closely related to the previous $i+2/i$ type hydrogen bonds in that the side chain atoms function as an acceptor to polar groups of the backbone ahead in the sequence. From Table I it can be seen that the Asp/Asn and Ser/Thr residues contribute the maximum number of examples to this type of hydrogen bond. In total, there are 640 examples of such hydrogen bonds out of which the Asp/Asn account for 298 examples and the Ser/Thr residues account for a further 299 examples. Table II also shows that these are the only four residues which show high propensity to form such hydrogen bonds.

Asp/Asn residues

Asp/Asn residues form $i+3/i$ hydrogen bonds through the δ -carboxyl oxygen atoms and the backbone amide hydrogens three residues ahead in the chain. They form 13-membered hydrogen bonded rings which mimics the 13-membered main-chain main-chain hydrogen bonded conformation of an α -turn (Eswar and Ramakrishnan, 1999). There are 298 examples of Asp/Asn participating in such hydrogen bonds in the data-set

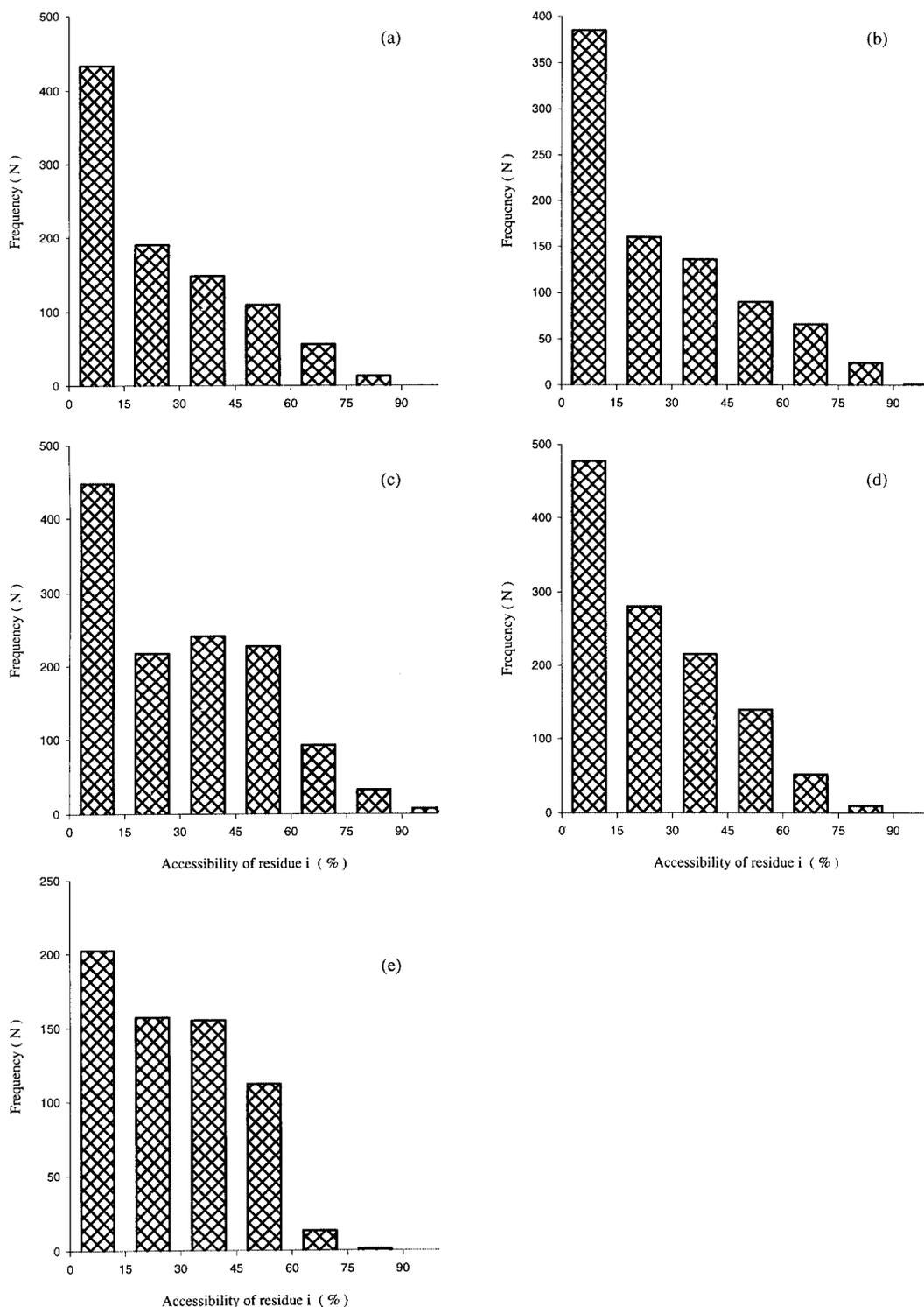


Fig. 7. The distribution of the accessibility of the residue i in the SC-MC hydrogen bonded examples (a) $i/i-4$, (b) $i/i-3$, (c) i/i , (d) $i+2/i$ and (e) $i+3/i$.

out of which 190 examples are of Asp and 108 examples are of Asn.

Figure 5 shows the (ϕ, ψ) plots of the backbone dihedral angles at residues i , $i+1$ and at $i+2$. It can be seen that the residue i is mostly clustered at the β -region and there are only very few examples elsewhere. Indeed, 276 examples out of the 298 have a β -conformation at the residue i and only 18 examples possess an α_R -conformation. Inspection of the plots

for the residue at $i+1$ indicates that there is a very strong clustering near the α_R -region. And the residue at $i+2$ shows almost complete clustering in the same region. Analysis of the examples reveals that almost 93% ($N = 276$) of the examples possess the main-chain motif of $\beta\alpha_R\alpha_R$. This repeated occurrence of clusters at the α_R -region prompted us to investigate the secondary structure of the residues following i . Close to 180 examples were indeed at the initiating positions of

α -helices with the hydrogen bond serving to satisfy the unsatisfied polar amide groups of the N-terminus of the helices (Penel *et al.*, 1999; Vijayakumar *et al.*, 1999; Wan and Milner-White, 1999a).

Analysis of the side-chain torsion angles of the residue i (as shown in Figure 5d) shows a distinct clustering of the χ^1 at the g^+ and the t rotamers with the g^- rotamer completely absent. The χ^{21} angles are restricted to the area between the t' -region.

Ser/Thr residues

Out of a total of 640 examples of such $i+3/i$ hydrogen bonds, 299 examples correspond with those made by Ser/Thr residues. There are 172 examples of Ser and 127 examples of Thr residues which participate in such hydrogen bonds. These hydrogen bonds result in a 12-membered closed ring with the hydrogen bond between the γ -hydroxyl oxygen and the amide hydrogen three residues away. The backbone dihedral angle distribution of the residue i shows that an overwhelming majority of the examples (295 out of 299) possess the β -conformation and only four examples possess the α_R -conformation. The secondary structure of the following residues ($i+1$, $i+2$, etc.) reveals that in 80% ($N = 240$) of the examples the following residues are part of α -helices. This confirms the fact that this hydrogen bonded motif is a strong helix initiator.

Distribution of the side-chain rotamers of the examples, shown in Figure 6, reveals a preference for the g^+ and the t rotamers. Almost close to 84% ($N = 251$) of the examples adopt the g^+ rotamer while there are only 48 examples which adopt the t rotamer. An investigation similar to the one described earlier (in the case of $i+2/i$) for an additional i/i hydrogen bond revealed that 66% ($N = 202$ out of the 299) of the examples form an $i+3/i-i/i$ type hydrogen bond. Interestingly, a large majority of 74% ($N = 178$ out of 240) of the examples which are found at the start sites of α -helices possessed this motif.

Accessibility of the residue i in the various hydrogen bonds

The accessibility of the residue i participating in each of these hydrogen bonds was calculated using the Lee and Richard algorithm (Lee and Richards, 1971). Since this analysis involved only polar residues it was expected that these hydrogen bonds will all be surface features in globular proteins. But contrary to expectations, it was found that the majority of each of these hydrogen bonded interactions were in fact buried from the solvent. Figure 7 shows the distribution of the accessibilities of the residue i in each of the five hydrogen bonds analysed. It can be seen that in all the cases, the accessibility profile drops from the left to the right indicating that a majority of the residues participating in these hydrogen bonds are indeed buried (Chothia, 1976; Miller *et al.*, 1987). It can be calculated that 81.5% of the residues participating in $i/i-4$ hydrogen bonds, 79.6% of those participating in $i/i-3$, 71.9% of those engaged in i/i , 83% of those in $i+2/i$ and 80.3% of those in $i+3/i$ hydrogen bonds have accessibilities of less than 45%.

Summary

In summary, the analysis of the side-chain main-chain hydrogen bonds made by the polar side-chains reveals that these hydrogen bonds are well ordered and can indeed be characterized by the torsion angles enclosed by the hydrogen bonded ring similar to the case of the regular main-chain main-chain hydrogen

bonds. The predominant motifs identified by this analysis are given in Table III.

The fact that these hydrogen bonds are buried from the surface in contact with the solvent brings out the fact that the residues participating in these well defined structures are probably involved in satisfying the potential main-chain hydrogen bonding sites which are buried upon folding. In a sense, these polar side-chain can be envisaged as providing the role of a backbone at the buried locations. Some of these hydrogen bonds (the $i+2/i$ and $i+3/i$) are known to be at the initiating sites of α -helices where they indeed play the role of the main-chain by allowing the side-chain to mimic the first turn of the α -helix. The intra-peptide i/i hydrogen bonds present a classic case of the need to compulsorily satisfy the hydrogen bonding potential of polar groups in the buried interior of proteins. The absence of other polar groups in the vicinity, suitable of hydrogen bonding, could be the force behind the formation of these strained (Karplus, 1996) hydrogen bonds.

Conclusions

The analysis seeks to establish the fact that side-chain main-chain hydrogen bonds do indeed form regular patterns of conformational motifs. These can be compared with the main-chain main-chain hydrogen bonds like the γ -, β - or the α -turns which are identified based on the clustering of the backbone dihedral angles. Fixing of the side-chain rotamers in both theoretically generated models as well as unrefined X-ray crystal structures is largely based on the principle of maximizing the hydrogen bonds (McDonald and Thornton, 1994a). The results of this analysis, which establish a strong conformational relationship between the backbone and side-chain dihedral angles, can be used to identify the rotamers based on the backbone (ϕ, ψ) angles. The analysis also establishes the fact that SC-MC hydrogen bonds are crucial elements at the start sites of α -helices where they play the role of the protein backbone to initiate the secondary structure. And the fact that these hydrogen bonds are predominantly buried indicates that they play very important roles in the interior of the protein where the side-chains probably contribute to satisfy the hydrogen bonding potential of the backbone polar groups.

References

- Armstrong, K.M. and Baldwin, R.L. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11337–11340.
- Aurora, R. and Rose, G.D. (1998) *Protein Sci.*, **7**, 21–38.
- Aurora, R., Creamer, T.P., Srinivasan, R. and Rose, G.D. (1997) *J. Biol. Chem.*, **272**, 1413–1416.
- Baker, E.N. and Hubbard, R.E. (1984) *Prog. Biophys. Mol. Biol.*, **44**, 97–179.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.*, **112**, 535–542.
- Bordo, D. and Argos, P. (1994) *J. Mol. Biol.*, **243**, 504–519.
- Bower, M.J., Cohen, F.E. and Dunbrack, R.L., Jr (1997) *J. Mol. Biol.*, **267**, 1268–1282.
- Chakrabarti, P. and Pal, D. (1998) *Protein Engng*, **11**, 631–647.
- Chothia, C. (1976) *J. Mol. Biol.*, **105**, 1–14.
- Dill, K.A. (1990) *Biochemistry*, **29**, 7133–7155.
- Doig, A.J., MacArthur, M.W., Stapley, B.J. and Thornton, J.M. (1997) *Protein Sci.*, **6**, 147–155.
- Dunbrack, R.L., Jr and Karplus, M. (1993) *J. Mol. Biol.*, **230**, 543–574.
- Eswar, N. and Ramakrishnan, C. (1999) *Protein Engng*, **12**, 447–455.
- Gunasekaran, K., Nagarajaram, H.A., Ramakrishnan, C. and Balam, P. (1998) *J. Mol. Biol.*, **275**, 917–932.
- Ippolito, J.A., Alexander, R.S. and Christianson, D.W. (1990) *J. Mol. Biol.*, **215**, 457–471.
- Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.
- Karplus, P.A. (1996) *Protein Sci.*, **5**, 1406–1420.
- Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

- Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.
- Lyu,P.C., Liff,M.I., Marky,L.A. and Kallenbach,N.R. (1990) *Science*, **250**, 669–673.
- McDonald,I.K. and Thornton,J.M. (1994a) *Protein Engng.*, **8**, 217–224.
- McDonald,I.K. and Thornton,J.M. (1994b) *J. Mol. Biol.*, **238**, 777–793.
- McGregor,M.J., Islam,S.A. and Sternberg,M.J.E. (1987) *J. Mol. Biol.*, **198**, 295–310.
- Miller,S., Janin,J., Lesk,A.M. and Chothia,C. (1987) *J. Mol. Biol.*, **196**, 641–656.
- Milner-White,E.J. (1990) *J. Mol. Biol.*, **216**, 385–397.
- Mitra,J. and Ramakrishnan,C. (1977) *Int. J. Peptide Protein Res.*, **9**, 27–48.
- Mitra,J. and Ramakrishnan,C. (1981) *Int. J. Peptide Protein Res.*, **17**, 401–411.
- Nataraj,D.V., Srinivasan,N., Sowdhamini,R. and Ramakrishnan,C. (1995) *Curr. Sci. (India)*, **69**, 434–447.
- Penel,S., Hughes,E. and Doig,A.J. (1999) *J. Mol. Biol.*, **287**, 127–143.
- Questal,J.L., Morris,D.G., Maccallum,P.H., Poet,R. and Milner-White,E.J. (1993) *J. Mol. Biol.*, **231**, 888–896.
- Ramachandran,G.N., Ramakrishnan,C. and Sasisekharan,V. (1963) *J. Mol. Biol.*, **7**, 95–99.
- Ramakrishnan,C. and Ramachandran,G.N. (1965) *Biophys. J.*, **5**, 909–933.
- Ramakrishnan,C. and Prasad,N. (1971) *Int. J. Peptide Protein Res.*, **3**, 209–231.
- Ramakrishnan,C., Srinivasan,N. and Nataraj,D.V. (1996) *Int. J. Peptide Protein Res.*, **48**, 420–428.
- Richardson,J.S. and Richardson,D.C. (1989) In Fasman,G.D. (ed.), *Prediction of Protein Structure and the Principles of Protein Conformation*. Plenum, New York, pp. 1–98.
- Richardson,J.S. and Richardson,D.C. (1981) *Adv. Protein Chem.*, **34**, 167–339.
- Schrauber,H., Eisenhaber,F. and Argos,P. (1993) *J. Mol. Biol.*, **230**, 592–612.
- Stickle,D.F., Presta,L.G., Dill,K.A. and Rose,G.D. (1992) *J. Mol. Biol.*, **226**, 1143–1159.
- Stites,W.E. and Pranata,J. (1995) *Proteins Struct. Funct. Genet.*, **22**, 132–140.
- Venkatachalam,C.M. (1968) *Biopolymers*, **6**, 1425–1436.
- Vijayakumar,M., Qian,H. and Zhou,H. (1999) *Proteins Struct. Funct. Genet.*, **34**, 497–507.
- Wan,W. and Milner-White,E.J. (1999a) *J. Mol. Biol.*, **286**, 1633–1649.
- Wan,W. and Milner-White,E.J. (1999b) *J. Mol. Biol.*, **286**, 1651–1662.
- Wilmot,C.M. and Thornton,J.M. (1988) *J. Mol. Biol.*, **203**, 221–232.
- Wilson,D.R. and Finlay,B.B. (1997) *Protein Engng.*, **10**, 519–529.

Received October 5, 1999; revised January 12, 2000; accepted January 19, 2000