b- and g-turns in proteins revisited: A new set of amino acid turn-type dependent positional preferences and potentials

KUNCHUR GURUPRASAD* and SASIDHARAN RAJKUMAR

Centre for DNA Fingerprinting and Diagnostics (CDFD), ECIL Road, Nacharam, Hyderabad 500 076, India

**Corresponding author (Fax, 91-40-7155610; Email, guru@www.cdfd.org.in).*

The number of b-turns in a representative set of 426 protein three-dimensional crystal structures selected from the recent Protein Data Bank has nearly doubled and the number of g-turns in a representative set of 320 proteins has increased over seven times since the previous analysis. b-turns (7153) and g-turns (911) extracted from these proteins were used to derive a revised set of type-dependent amino acid positional preferences and potentials. Compared with previous results, the preference for proline, methionine and tryptophan has increased and the preference for glutamine, valine, glutamic acid and alanine has decreased for b-turns. Certain new amino acid preferences were observed for both turn types and individual amino acids showed turn-type dependent positional preferences. The rationale for new amino acid preferences are discussed in the light of hydrogen bonds and other interactions involving the turns. Where main-chain hydrogen bonds of the type NH(i + 3) → CO(i) were not observed for some β-turns, other main-chain hydrogen bonds or solvent interactions were observed that possibly stabilize such b-turns. A number of unexpected isolated b-turns with proline at i + 2 position were also observed. The NH(i + 2) → CO(i) hydrogen bond was observed for almost all g-turns. Nearly 20% classic g-turns and 43% inverse g-turns are isolated turns.

1. Introduction

A b-turn consists of four consecutive residues defined by positions i, i + 1, i + 2, i + 3 which are not present in an a-helix; the distance between $C_\alpha(i)$ and $C_\alpha(i + 3)$ is less than 7 Å (Richardson 1981; Rose *et al* 1985) and the turn leads to reversal in the protein chain. b-turns may or may not be accompanied by the NH(i + 3)–CO(i) hydrogen bond connecting the main-chain atoms; CO of ith residue and NH of (i + 3)rd residue in the turn (Lewis *et al* 1973; Nemethy and Scheraga 1980), that was originally used to characterize b-turns (Venkatachalam 1968). b-turns have been classified into 9 different types (I, II, VIII, I′, II′, VIa1, VIa2, VIb, IV) based on the dihedral angle values (f, y) of the (i + 1)th and (i + 2)th position in the turn (Venkatachalam 1968; Richardson 1981; Hutchinson and Thornton 1994). A g-turn consists of three consecutive residues at positions i, i + 1 ,i + 2 defined by theexistence of a hydrogen bond between the CO group of (i)th residue and NH group of (i + 2)th residue (Rose *et al* 1985; Toniolo 1980). g-turns have been classified into two types (classic and inverse)based on the dihedral angle values of the (i + 1)th residue (Rose *et al* 1985). The classic g-turn gives rise to a 180° chain-reversal in proteins and is often observed at loop end of b-hairpins (Milner-White *et al* 1988). b-hairpins are commonly associated with type II′, I′ or type I b-turns at the loop ends (Sibanda and Thornton 1993; Sibanda *et al* 1989). The classic g-turns have been further classified into four classes based on the hydrogen bonding patterns (Milner-White *et al* 1988). The inverse g-turns includes a large proportion of weak hydrogen bonds according to the definition of hydrogen bonds (Kabsch and Sander 1983). Many inverse g-turns analysed earlier were found to occur within parallel or anti-parallel b-strands associated with b-sheets and the less weak inverse g-turns were observed to be frequently situated directly at either end of a-helices, or of strands of b-sheet or associated as part of the type VI b-turn. Inverse g-turns are also associated with other inverse g-turns and referred to as 'compound g-turn' (Milner-White 1990). Turns play an important role in protein structures; the b-turn, and the classic g-turn provide either a direction change of the polypeptide or as in the inverse g-turn give

Keywords. Amino acid preferences; amino acid potentials; b- and g-turns in proteins

rise to kink in the chain. Turns are often conserved during evolution and have been implicated in molecular recognition (Rose *et al* 1985) and protein folding (Rose 1978). Ninety per cent of turns in proteins constitute β-turns and the remaining are γ-turns. The knowledge of amino acid preferences at individual positions in β-turns and γ-turns have been used for turn prediction (Lewis *et al* 1973; Chou and Fasman 1974, 1979; Garnier *et al* 1978; Cohen *et al* 1986; Wilmot and Thornton 1988; McGregor *et al* 1989; Hutchinson and Thornton 1994; Chou 1997; Chou and Blinn 1997; Cai *et al* 1998; Shepherd *et al* 1999). In the present work we wanted to; (i) determine the statistically significant new amino acid positional preferences for turns in the enlarged representative data set of proteins, (ii) derive a revised set of positional potentials for β-turns and for the first time potentials for γ-turns, (iii) compare amino acid preferences with earlier work (Hutchinson and Thornton 1994) and seek rationale for new amino acid preferences and (iv) examine the situation of type VI β-turns with proline in $(i+2)$nd position. Slightly more stringent criteria was applied for selecting the representative protein chains from the Protein Data Bank compared with the earlier work. Our analysis may be useful for modeling and design of turns in proteins.

2 Materials and methods

The representative protein dataset for our analysis was obtained from the Protein Data Bank (Bernstein *et al* 1977; Abola *et al* 1997 and Sussman *et al* 1998) now maintained at the Research Collaboratory for Structural Bioinformatics (RCSB) (http://www.rcsb.org/pdb/), using the program PDB_SELECT (Hobohm and Sander 1994) and available under the filename recent.pdb_select (dated Thursday 1st October 1998), through ftp (ftp.embl-heidelberg.de/pub/databases/protein_extras/pdb_select).

The representative protein chains were selected so that no two chains had more than 25% sequence identity. Protein chains determined by X-ray crystallography at 2·0 Å resolution or better containing at least one β- or a γ-turn were used in the analysis. The PROMOTIF program (Hutchinson and Thornton 1996), which provides details of the location and types of structural motifs in proteins of known three-dimensional structure in the Protein Data Bank was used to extract the β-turns classified into the nine known types (I, II, VIII, I′, II′, VIa1, VIa2, VIb, IV) and the γ-turns classified into two types (inverse and classic). For definitions of β-turns and their classification see Venkatachalam (1968), Richardson (1981), Wilmot and Thornton (1988) and Hutchinson and Thornton (1994) and for the γ-turns see Rose *et al* (1985),Milner-White *et al* (1988) and Milner-White (1990). The (ϕ_{i+1}, ψ_{i+1}) and (ϕ_{i+2}, ψ_{i+2}) values for all the β-turns and the (ϕ_{i+1}, ψ_{i+1}) values for all the γ-turns were plotted in order to examine their distribution in the Ramachandran plot (Ramachandran and Sasisekharan 1968) and to ensure a reliable data set for our analysis. The conformational potentials, positional potentials and the turn-type dependent positional potentials for β- and γ-turns were calculated as described in Hutchinson and Thornton (1994). The potentials were examined for statistical significance by the *d*-test (based on normal distribution) described in Wilmot and Thornton (1988). We examined hydrogen-bond interactions described in Overington *et al* (1990), in order to deduce significant interactions for new amino acid preferences in β- and γ-turns.

3. Results and discussion

Protein chains (945) were selected by the PDB_SELECT program. Out of these only 426 protein chains represented crystal structures determined at 2·0 Å resolution or better and contained at least one β-turn and 320 protein chains which contained at least one γ-turn. The PDB codes corresponding to these representative protein chains used in our analysis are given in appendix A and B and may also be accessed from our website (http://www.cdfd.org.in). The distribution of β-turns and γ-turns in the present analysis and previous work (Hutchinson and Thornton 1994) are shown in table 1. The total number of amino acid residues constituting β-turns is nearly 30·4%. This corresponds to a 5% increase compared with the value reported earlier (25%) by Kabsch and Sander (1983). In contrast the number of amino acids in γ-turns is only 3·4%. The inverse γ-turns account for nearly 91% of the total γ-turns.

Table 1. Distribution of turns.

	Number of turns	
β-turn type	3,899 (earlier)	7,153 (present)
I	1,419	2,439
II	489	911
VIII	451	671
I′	142	304
II′	100	165
VIa1	17	44
VIa2	5	17
VIb	35	70
IV	1,241	2,532
γ-turn type	127 (earlier)	911 (present)
Inverse	115	833
Classic	12	78

'earlier' refers to work by Hutchinson and Thornton (1994) for β-turns and for the γ-turns by Milner-White *et al* (1988) and Milner-White (1990).

Table 2a. Positional and overall potentials for β-turns.

Amino acids	Positional				Overall		Difference (%)
	i	*i* + 1	*i* + 2	*i* + 3	present	earlier	
Ala	0· 78	0· 90	0· 50	0· 86	0· 76	0· 83	– 8· 43
Cys	1· 50	0· 61	0· 75	1· 23	**1· 02**	1· 08	– 5· 55
Asp	1· 66	1· 24	2· 01	0· 97	**1· 47**	1· 41	4· 26
Glu	0· 75	1· 26	0· 84	0· 78	0· 91	1· 01	– 9· 90
Phe	0· 98	0· 66	0· 78	0· 90	0· 83	0· 89	– 6· 74
Gly	1· 07	0· 97	2· 42	1· 77	**1· 56**	1· 48	5· 40
His	1· 16	0· 99	1· 20	1· 03	**1· 09**	1· 07	1· 87
Ile	0· 63	0· 58	0· 38	0· 68	0· 57	0· 59	– 3· 39
Lys	0· 80	1· 32	0· 84	1· 11	**1· 02**	1· 01	0· 99
Leu	0· 78	0· 57	0· 47	0· 75	0· 64	0· 66	– 3· 03
Met	0· 80	0· 47	0· 48	0· 79	0· 64	0· 57	12· 28
Asn	1· 47	1· 12	2· 31	1· 07	**1· 49**	1· 44	3· 47
Pro	1· 70	2· 73	0· 62	1· 16	**1· 55**	1· 38	12· 32
Gln	0· 71	0· 82	0· 79	0· 97	0· 82	0· 94	– 12· 76
Arg	0· 70	0· 97	0· 81	0· 97	0· 86	0· 82	4· 88
Ser	1· 28	1· 24	1· 13	1· 04	**1· 17**	1· 15	1· 74
Thr	1· 09	0· 90	0· 96	1· 17	**1· 03**	1· 00	3· 00
Val	0· 69	0· 63	0· 43	0· 78	0· 63	0· 70	– 10· 00
Trp	0· 90	0· 84	0· 81	0· 93	0· 87	0· 70	24· 29
Tyr	1· 00	0· 74	0· 86	0· 92	0· 88	0· 92	– 4· 35

'earlier' refers to work by Hutchinson and Thornton (1994).

3.1 *Conformational, positional and overall potentials for β- and γ-turns*

The potentials for individual amino acids in β-turns are shown in table 2a. Notable among the increased preferences compared with earlier work (Hutchinson and Thornton 1994) are tryptophan (increased by ~ 24%), proline and methionine (~ 12%) and among the weak preferences are glutamine (decreased by ~ 12%), valine (~ 10%), glutamic acid (~ 9%) and alanine (~ 8%). Amino acid preferences for γ-turns are shown in table 2b.

Table 2b. Positional and overall potentials for γ-turns.

Amino acids	*i*	*i* + 1	*i* + 2	Overall
Ala	0· 80	0· 92	0· 58	0· 77
Cys	1· 04	0· 82	1· 56	**1· 14**
Asp	0· 74	2· 75	1· 10	**1· 53**
Glu	0· 95	0· 78	0· 58	0· 77
Phe	1· 03	0· 81	1· 03	0· 96
Gly	1· 53	0· 72	0· 76	1· 00
His	1· 03	1· 12	1· 22	**1· 12**
Ile	1· 27	0· 58	1· 17	1· 00
Lys	0· 80	0· 72	0· 91	0· 81
Leu	0· 86	0· 79	0· 88	0· 84
Met	1· 07	1· 33	0· 64	**1· 01**
Asn	1· 05	2· 45	1· 47	**1· 66**
Pro	1· 70	2· 13	0· 00	**1· 28**
Gln	0· 80	0· 80	1· 12	0· 91
Arg	1· 00	0· 74	1· 16	0· 97
Ser	0· 75	0· 59	1· 15	0· 83
Thr	0· 81	0· 38	1· 55	0· 91
Val	0· 82	0· 33	1· 47	0· 87
Trp	1· 13	1· 27	1· 48	**1· 29**
Tyr	1· 20	0· 76	1· 15	**1· 04**

Statistically significant overall potentials are in bold.

3.2 *Statistically significant β- and γ-turn type-dependent amino acid positional preferences and potentials*

The amino acid positional preferences and potentials vary depending upon the individual β- and γ-turn types. These are shown in table 3. The statistical significance of the potentials was assessed using a *d*-test, with the usually adopted 5% boundary ($d \geq 1\cdot 97$) as in Wilmot and Thornton (1988) and significant values were analysed to explain the observed potential trends. The significant amino acid positional preferences for type-dependent β-turns is schematically represented in figure 1A–H and in figure 1I, J for γ-turns. Five new amino acid preferences compared with earlier results were observed for type I, 7 for type II, 5 for type VIII and 5 for type I′. Type IV β-turns account for nearly 35· 4% of the total number of β-turns and these turns are least well- defined .For type VI

Table 3a. Type-dependent positional potentials for b-turns[a].

Amino acids	Type I (2439)			
	i	*i* + 1	*i* + 2	*i* + 3
Ala	0· 54(110)	1· 07(216)	0· 59(121)	0· 90(181)
Cys	**1· 43(54)**	0· 80(30)	1· 04(39)	**1· 48(55)***
Asp	**2· 76(405)**	**1· 21(175)***	**3· 22(440)**	1· 04(151)
Glu	0· 56(81)	**1· 38(195)**	1· 04(152)	0· 79(114)
Phe	0· 90(88)	0· 44(43)	0· 80(78)	0· 96(94)
Gly	1· 07(202)	0· 42(79)	0· 64(122)	**2· 57(491)**
His	**1· 48(81)**	0· 71(40)	1· 25(70)	1· 02(57)
Ile	0· 32(43)	0· 48(64)	0· 20(26)	0· 68(89)
Lys	0· 59(85)	**1· 33(190)***	0· 96(138)	1· 04(146)
Leu	0· 73(145)	0· 49(98)	0· 60(121)	0· 90(179)
Met	0· 48(24)	0· 32(16)	0· 47(23)	0· 88(43)
Asn	**2· 14(243)**	0· 71(82)	**2· 41(274)**	**1· 29(146)***
Pro	**1· 48(168)**	**4· 29(472)**	0· 14(16)	0· 00(0)
Gln	0· 51(47)	0· 76(69)	1· 07(97)	0· 71(66)
Arg	0· 70(78)	1· 04(117)	1· 02(114)	0· 90(101)
Ser	**1· 68(241)**	**1· 61(237)**	**1· 61(232)**	0· 94(140)
Thr	**1· 29(182)***	0· 94(133)	**1· 48(211)**	1· 04(146)
Val	0· 34(58)	0· 57(96)	0· 22(38)	0· 71(121)
Trp	0· 92(34)	0· 90(33)	1· 07(40)	1· 04(39)
Tyr	0· 77(70)	0· 59(54)	0· 96(87)	0· 88(79)

Amino acids	Type II (911)			
	i	*i* + 1	*i* + 2	*i* + 3
Ala	1· 07(81)	**1· 25(95)***	0· 13(10)	1· 09(83)
Cys	**1· 57(22)***	0· 28(4)	0· 43(6)	**1· 64(23)**
Asp	0· 29(16)	0· 71(39)	0· 60(33)	**1· 34(73)**
Glu	1· 11(60)	**1· 40(76)***	0· 24(13)	1· 03(56)
Phe	1· 01(37)	0· 77(28)	0· 33(12)	0· 88(32)
Gly	0· 97(69)	0· 31(22)	**9· 39(668)**	0· 89(63)
His	0· 81(17)	0· 96(20)	0· 67(14)	1· 10(23)
Ile	1· 08(53)	0· 59(29)	0· 00(0)	0· 73(36)
Lys	1· 25(66)	**1· 67(89)**	0· 24(13)	**1· 40(74)***
Leu	1· 12(85)	0· 56(42)	0· 11(8)	0· 52(39)
Met	0· 86(16)	0· 65(12)	0· 16(3)	1· 36(25)
Asn	0· 77(33)	0· 47(20)	**1· 57(67)***	0· 56(24)
Pro	**1· 91(83)**	**4· 92(213)**	0· 00(0)	0· 00(0)
Gln	1· 22(42)	1· 04(36)	0· 17(6)	**1· 52(52)***
Arg	0· 78(33)	0· 69(29)	0· 33(14)	0· 93(39)
Ser	0· 65(36)	0· 97(54)	0· 34(19)	**1· 52(84)**
Thr	0· 94(50)	0· 71(38)	0· 13(7)	**1· 38(73)***
Val	0· 98(62)	0· 65(41)	0· 02(1)	0· 91(58)
Trp	0· 64(9)	0· 57(8)	0· 36(5)	1· 15(16)
Tyr	1· 22(41)	0· 47(16)	0· 35(12)	1· 12(38)

Amino acids	Type VIII (671)			
	i	*i* + 1	*i* + 2	*i* + 3
Ala	1· 05(59)	0· 87(49)	0· 44(25)	0· 66(37)
Cys	**1· 82(19)***	0· 48(5)	0· 38(4)	0· 67(7)
Asp	0· 64(26)	**1· 85(74)**	**1· 85(74)**	0· 77(31)
Glu	0· 62(25)	1· 25(50)	0· 78(31)	0· 65(26)
Phe	0· 93(25)	0· 41(11)	1· 30(35)	0· 78(21)
Gly	**1· 40(73)**	0· 44(23)	0· 08(4)	0· 67(35)
His	0· 85(13)	1· 04(16)	**1· 63(25)***	0· 39(6)
Ile	0· 71(26)	0· 88(32)	1· 32(48)	0· 94(34)
Lys	0· 73(29)	**2· 16(84)***	1· 20(47)	0· 99(39)
Leu	0· 70(39)	0· 70(39)	0· 90(50)	0· 79(44)
Met	0· 88(12)	0· 66(9)	0· 81(11)	0· 58(8)
Asn	1· 01(32)	0· 57(18)	**1· 49(47)**	0· 63(20)
Pro	**3· 12(99)**	**1· 95(62)**	0· 00(0)	**4· 01(128)**
Gln	0· 71(18)	0· 94(24)	1· 30(33)	1· 10(28)
Arg	0· 71(22)	1· 29(40)	0· 87(27)	1· 00(31)
Ser	1· 23(50)	1· 23(50)	1· 00(41)	1· 00(41)
Thr	0· 79(31)	1· 00(39)	1· 07(42)	1· 28(50)
Val	0· 86(40)	0· 60(28)	**1· 54(72)**	**1· 38(64)***
Trp	0· 87(9)	0· 58(6)	0· 87(9)	0· 68(7)
Tyr	0· 96(24)	0· 48(12)	**1· 85(46)***	0· 56(14)

Amino acids	Type I′ (304)			
	i	*i* + 1	*i* + 2	*i* + 3
Ala	0· 78(20)	0· 59(15)	0· 08(2)	0· 74(19)
Cys	1· 06(5)	0· 42(2)	0· 00(0)	0· 64(3)
Asp	1· 09(20)	**2· 79(51)**	0· 76(14)	0· 71(13)
Glu	0· 83(15)	0· 72(13)	0· 06(1)	1· 16(21)
Phe	**1· 97(24)***	0· 57(7)	0· 82(10)	0· 41(5)
Gly	0· 46(11)	**1· 73(41)**	**9· 38(223)**	0· 13(3)
His	1· 14(8)	**2· 44(17)**	0· 29(2)	1· 28(9)
Ile	1· 45(24)	0· 00(0)	0· 00(0)	0· 73(12)
Lys	1· 18(21)	0· 99(16)	0· 22(4)	**3· 64(65)**
Leu	0· 75(19)	0· 36(9)	0· 04(1)	0· 67(17)
Met	1· 13(7)	0· 48(3)	0· 48(3)	0· 48(3)
Asn	0· 91(13)	**6· 07(87)**	1· 11(16)	0· 70(10)
Pro	0· 34(5)	0· 00(0)	0· 00(0)	0· 00(0)
Gln	0· 69(8)	0· 87(10)	0· 43(5)	**2· 17(25)***
Arg	0· 85(12)	0· 64(9)	0· 43(6)	**1· 85(26)***
Ser	0· 86(16)	0· 81(15)	0· 54(10)	0· 59(11)
Thr	0· 62(11)	0· 06(1)	0· 00(0)	0· 84(15)
Val	**1· 89(40)***	0· 00(0)	0· 00(0)	1· 13(24)
Trp	0· 64(3)	0· 00(0)	0· 21(1)	0· 86(4)
Tyr	**1· 95(22)**	0· 71(8)	0· 53(6)	**1· 67(19)***

Amino acids	Type II′ (165)			
	i	*i* + 1	*i* + 2	*i* + 3
Ala	1· 08(15)	0· 07(1)	0· 94(13)	0· 72(10)
Cys	1· 17(3)	0· 00(0)	1· 17(3)	0· 39(1)
Asp	0· 80(8)	1· 00(10)	**3· 02(30)**	0· 70(7)
Glu	1· 22(12)	0· 41(4)	1· 12(11)	1· 02(10)
Phe	1· 21(8)	0· 45(3)	0· 15(1)	1· 06(7)
Gly	1· 01(13)	**9· 51(122)**	1· 01(13)	1· 32(17)
His	1· 05(4)	0· 26(1)	0· 53(2)	1· 58(6)
Ile	1· 22(11)	0· 00(0)	0· 33(3)	0· 56(5)
Lys	0· 72(7)	0· 41(4)	0· 51(5)	1· 24(12)
Leu	0· 51(7)	0· 07(1)	0· 36(5)	0· 95(13)
Met	0· 30(1)	0· 00(0)	0· 3 (1)	0· 89(3)
Asn	0· 64(5)	0· 77(6)	**3· 21(25)**	1· 28(10)
Pro	0· 51(4)	0· 00(0)	0· 76(6)	0· 00(0)
Gln	0· 96(6)	0· 48(3)	0· 8 (5)	1· 60(10)
Arg	0· 78(6)	0· 39(3)	0· 39(3)	1· 44(11)
Ser	1· 39(14)	0· 30(3)	**2· 09(21)**	1· 39(14)
Thr	0· 83(8)	0· 00(0)	1· 14(11)	**1· 65(16)**
Val	1· 39(16)	0· 17(2)	0· 43(5)	0· 35(4)
Trp	1· 58(4)	0· 00(0)	0· 39(1)	1· 18(3)
Tyr	**2· 11(13)**	0· 32(2)	0· 16(1)	0· 97(6)

Amino acids	Type VIa1 (44) *i*	*i* + 1	*i* + 2	*i* + 3
Ala	0· 27(1)	1· 90(7)	0· 00(0)	0· 27(1)
Cys	1· 46(1)	1· 46(1)	0· 00(0)	0· 00(0)
Asp	0· 75(2)	0· 38(1)	0· 00(0)	1· 50(4)
Glu	1· 14(3)	1· 52(4)	0· 00(0)	0· 38(1)
Phe	1· 70(3)	0· 57(1)	0· 00(0)	**3· 40(6)**
Gly	0· 87(3)	0· 00(0)	0· 00(0)	0· 58(2)
His	0· 00(0)	2· 96(3)	0· 00(0)	1· 97(2)
Ile	0· 42(1)	0· 00(0)	0· 00(0)	0· 42(1)
Lys	**2· 32(6)**	0· 77(2)	0· 00(0)	0· 77(2)
Leu	**2· 19(8)**	1· 09(4)	0· 00(0)	0· 82(3)
Met	1· 11(1)	1· 11(1)	0· 00(0)	0· 00(0)
Asn	0· 48(1)	0· 48(1)	0· 00(0)	1· 92(4)
Pro	1· 42(3)	**4· 27(9)**	**20· 98(44)**	0· 00(0)
Gln	1· 79(3)	0· 60(1)	0· 00(0)	0· 60(1)
Arg	0· 00(0)	1· 47(3)	0· 00(0)	1· 47(3)
Ser	0· 74(2)	0· 37(1)	0· 00(0)	0· 37(1)
Thr	0· 78(2)	0· 00(0)	0· 00(0)	1· 55(4)
Val	0· 65(2)	0· 00(0)	0· 00(0)	0· 33(1)
Trp	1· 48(1)	0· 00(0)	0· 00(0)	2· 96(2)
Tyr	0· 61(1)	**3· 04(5)**	0· 00(0)	**3· 04(5)**

Amino acids	Type VIa2 (17) *i*	*i* + 1	*i* + 2	*i* + 3
Ala	0· 00(0)	0· 70(1)	0· 00(0)	1· 40(2)
Cys	0· 00(0)	0· 00(0)	0· 00(0)	0· 00(0)
Asp	1· 95(2)	0· 97(1)	0· 00(0)	0· 00(0)
Glu	0· 98(1)	1· 97(2)	0· 00(0)	0· 00(0)
Phe	**4· 39(3)**	**4· 39(3)**	0· 00(0)	1· 46(1)
Gly	1· 50(2)	0· 00(0)	0· 00(0)	**3· 01(4)**
His	2· 55(1)	0· 00(0)	0· 00(0)	2· 55(1)
Ile	1· 08(1)	1· 08(1)	0· 00(0)	1· 08(1)
Lys	0· 00(0)	1· 00(1)	0· 00(0)	1· 00(1)
Leu	1· 42(2)	0· 00(0)	0· 00(0)	0· 00(0)
Met	0· 00(0)	2· 88(1)	0· 00(0)	2· 88(1)
Asn	0· 00(0)	1· 24(1)	0· 00(0)	0· 00(0)
Pro	1· 23(1)	0· 00(0)	**20· 90(17)**	0· 00(0)
Gln	1· 55(1)	1· 55(1)	0· 00(0)	1· 55(1)
Arg	0· 00(0)	1· 27(1)	0· 00(0)	2· 54(2)
Ser	0· 00(0)	0· 96(1)	0· 00(0)	0· 00(0)
Thr	0· 00(0)	0· 00(0)	0· 00(0)	1· 00(1)
Val	2· 53(3)	0· 00(0)	0· 00(0)	0· 84(1)
Trp	0· 00(0)	**7· 65(2)**	0· 00(0)	3· 82(1)
Tyr	0· 00(0)	1· 57(1)	0· 00(0)	0· 00(0)

Amino acids	Type VIb (70) *i*	*i* + 1	*i* + 2	*i* + 3
Ala	0· 34(2)	0· 51(3)	0· 00(0)	1· 53(9)
Cys	0· 92(1)	2· 76(3)	0· 00(0)	0· 00(0)
Asp	0· 71(3)	0· 95(4)	0· 00(0)	0· 00(0)
Glu	0· 24(1)	0· 72(3)	0· 00(0)	0· 48(2)
Phe	1· 78(5)	2· 13(6)	0· 00(0)	1· 42(4)
Gly	0· 91(5)	0· 00(0)	0· 00(0)	0· 91(5)
His	**3· 10(5)**	1· 86(3)	0· 00(0)	1· 86(3)
Ile	0· 79(3)	0· 79(3)	0· 00(0)	0· 79(3)
Lys	0· 73(3)	1· 21(5)	0· 00(0)	0· 97(4)
Leu	0· 34(2)	0· 34(2)	0· 00(0)	0· 34(2)
Met	1· 40(2)	0· 70(1)	0· 00(0)	0· 00(0)
Asn	0· 30(1)	1· 51(5)	0· 00(0)	0· 91(3)
Pro	**2· 39(8)**	0· 00(0)	**21· 01(70)**	**2· 68(9)**
Gln	0· 75(2)	0· 75(2)	0· 00(0)	1· 50(4)
Arg	1· 23(4)	0· 62(2)	0· 00(0)	0· 92(3)
Ser	1· 87(8)	0· 70(3)	0· 00(0)	1· 17(5)
Thr	1· 22(5)	**2· 19(9)**	0· 00(0)	1· 22(5)
Val	1· 23(6)	1· 02(5)	0· 00(0)	0· 61(3)
Trp	0· 93(1)	2· 78(3)	0· 00(0)	0· 93(1)
Tyr	1· 15(3)	**3· 06(8)**	0· 00(0)	1· 91(5)

Amino acids	Type IV (2532) *i*	*i* + 1	*i* + 2	*i* + 3
Ala	0· 86(182)	0· 73(154)	0· 63(132)	0· 83(172)
Cys	**1· 62(62)**	0· 59(23)	0· 79(31)	1· 24(48)
Asp	**1· 55(237)**	**1· 20(180)**	**1· 86(281)**	0· 93(139)
Glu	0· 81(121)	**1· 28(192)**	1· 00(152)	0· 70(106)
Phe	0· 89(90)	0· 86(88)	0· 86(88)	0· 89(89)
Gly	1· 13(223)	**1· 33(258)**	**1· 69(323)**	**1· 96(370)**
His	1· 06(62)	1· 09(63)	**1· 49(85)**	1· 13(64)
Ile	0· 62(85)	0· 70(96)	0· 53(73)	0· 61(84)
Lys	0· 83(122)	1· 09(163)	0· 98(145)	0· 86(127)
Leu	0· 74(155)	0· 69(144)	0· 44(93)	0· 72(148)
Met	1· 06(54)	0· 49(25)	0· 56(29)	0· 64(33)
Asn	**1· 43(171)**	**1· 38(160)**	**3· 10(354)**	**1· 24(145)**
Pro	**1· 77(208)**	**1· 49(180)**	0· 51(61)	**2· 19(261)**
Gln	0· 70(67)	0· 83(78)	0· 74(71)	0· 81(77)
Arg	0· 68(78)	1· 03(120)	0· 91(106)	0· 91(105)
Ser	**1· 28(193)**	1· 16(177)	1· 09(169)	1· 03(158)
Thr	1· 16(171)	1· 09(158)	0· 89(130)	**1· 24(182)**
Val	0· 68(118)	0· 81(141)	0· 55(98)	0· 65(114)
Trp	0· 98(38)	1· 03(40)	0· 86(33)	0· 76(29)
Tyr	1· 00(95)	0· 98(92)	0· 83(77)	0· 86(80)

Statistically significant potential values are in bold and the amino acid preferences observed in the present study are marked with an asterisk.

[a]The statistically significant potentials are selected according to *d*-value (> 1· 97) as in Hutchinson and Thornton (1994).

Table 3b. Type-dependent positional potentials for g-turns[a].

Amino acids	Classic (78) *i*	*i* + 1	*i* + 2
Ala	0· 90(6)	0· 30(2)	0· 15(1)
Cys	0· 00(0)	0· 87(1)	2· 61(3)
Asp	0· 85(4)	1· 48(7)	1· 27(6)
Glu	0· 87(4)	1· 09(5)	0· 87(4)
Phe	0· 00(0)	0· 32(1)	0· 95(3)
Gly	**1· 80(11)**	**4· 42(27)**	0· 82(5)
His	1· 64(3)	0· 00(0)	**2· 73(5)**
Ile	0· 70(3)	1· 64(7)	0· 94(4)
Lys	0· 44(2)	0· 00(0)	**2· 44(11)**

Contd. . .

Table 3b. (*Contd*)

Amino acids	Classic (78)		
	i	*i* + 1	*i* + 2
Leu	0· 92(6)	0· 77(5)	0· 15(1)
Met	0· 62(1)	0· 62(1)	0· 62(1)
Asn	1· 87(7)	0· 80(3)	1· 07(4)
Pro	1· 85(7)	0· 00(0)	0· 00(0)
Gln	0· 69(2)	0· 35(1)	1· 38(4)
Arg	0· 55(2)	1· 11(4)	0· 83(3)
Ser	1· 05(5)	1· 47(7)	**2· 09(10)**
Thr	1· 32(6)	0· 66(3)	0· 88(4)
Val	0· 37(2)	0· 18(1)	1· 11(6)
Trp	1· 65(2)	0· 83(1)	0· 83(1)
Tyr	1· 72(5)	0· 69(2)	0· 69(2)

Amino acids	Inverse (833)		
	i	*i* + 1	*i* + 2
Ala	0· 79(56)	0· 98(70)	0· 62(44)
Cys	1· 14(14)	0· 81(10)	1· 47(18)
Asp	0· 73(37)	**2· 87(145)**	1· 09(55)
Glu	0· 96(47)	0· 75(37)	0· 55(27)
Phe	1· 12(38)	0· 86(29)	1· 04(35)
Gly	**1· 50(98)**	0· 37(24)	0· 75(49)
His	0· 97(19)	1· 23(24)	1· 07(21)
Ile	**1· 32(60)**	0· 48(22)	1· 19(54)
Lys	0· 83(40)	0· 79(38)	0· 77(37)
Leu	0· 85(59)	0· 79(55)	0· 95(66)
Met	1· 11(19)	1· 40(24)	0· 64(11)
Asn	0· 98(39)	**2· 61(104)**	**1· 50(60)**
Pro	**1· 68(68)**	**2· 32(94)**	0· 00(0)
Gln	0· 81(25)	0· 84(26)	1· 10(34)
Arg	1· 05(40)	0· 70(27)	1· 20(46)
Ser	0· 73(37)	0· 51(26)	1· 06(54)
Thr	0· 76(37)	0· 35(17)	**1· 61(78)**
Val	0· 86(50)	0· 35(20)	**1· 50(87)**
Trp	1· 08(14)	1· 31(17)	**1· 55(20)**
Tyr	1· 16(36)	0· 77(24)	1· 19(37)

b-turns and the g-turns, there are now sufficient numbers that allow calculation of potentials and analysis of their sequence preferences. Figure 2 shows the f, y, Ramachandran plot connecting average f, y values of (i + 1)th residue to the values of the (i + 2)th residue in b-turns. The average f, y values for b-turns were determined from the present data set shown in table 4. We examined atom–atom interactions in the protein (Overington *et al* 1990) of types; main-chain main-chain, main-chain side-chain and side-chain side-chain or interactions involving the main-chain or the side-chain atoms with water or other heterogen atoms in the protein. The hydrogen bond interactions may be one of the factors contributing to turn stability.

3.3 *Type I b-turns*

The new amino acid preferences observed were: threonine at position i, aspartic acid and lysine at i + 1, cysteine and asparagine at i + 3. Majority of type I b-turns with threonine at ith position are stabilized due to the main-chain main-chain hydrogen bond of both types observed, i.e., NH(i + 3)–CO(i) and NH(i + 2)–CO(i). Threonine side-chain at i makes hydrogen bond with main-chain nitrogen at i + 2 as in the "Asx-turn" (Rees *et al* 1983) shown in figure 3A. Further, threonine side-chain also makes hydrogen bond with main-chain NH(i + 3) or with main-chain CO(i + 3) or interactions with other amino acid side-chains at i + 2 (mainly aspartic acid) or at i + 3 (mainly with serine). Out of 423 type I b-turn sequences with either serine or threonine at i, 41 were of the type SPXX and 25 of type TPXX. These may be the (S or T) PXX motifs of DNA-binding proteins which were shown to fold into a compact b-turn stabilized by a side-chain main-chain interaction (Suzuki and Yagi 1991). Some type I b-turns with serine or threonine at ith position may also correspond to the ST-motifs identified recently (Wan and Milner-White 1999b), in which the hydrogen bond arrangement is a four to five residue motif and the first residue is a serine or threonine forming two characteristic hydrogen bonds to residues ahead of it in sequence. Some of the type I b-turns with an aspartic acid or asparagine at ith position may correspond to the two-hydrogen-bond four residue Asx-motif (Wan and Milner-White 1999a). Nearly a quarter of turns with aspartic acid or lysine at i + 1, had hydrogen bonds of the type NH(i + 3)–CO(i + 1). A majority of the interactions involving the side-chain or main-chain or both, for these amino acids were mainly with water and to a lesser extent with other main-chain or side-chain hydrogen bonds in the protein (see figure 4A). In the present analysis, we also observed a preference for asparagine and cysteine at i + 3 position. There was an equal preference for main-chain hydrogen bonds of both types, NH(i + 3)–CO(i + 1) and NH(i + 3)–CO(i) or only of the type NH(i + 3)–CO(i) for a majority of the type I b-turns with cysteine at i + 3. Some of these turns had side-chain side-chain interactions with other amino acid residues at i (mainly aspartic acid) (see figure 3C) or with the main-chain oxygen CO(i) or with main-chain nitrogen NH(i). On the contrary, more than half the number of turns with asparagine at i + 3, had main-chain main-chain hydrogen bond only of the type NH(i + 3)–CO(i) and nearly a quarter of the turns had both hydrogen bonds; NH(i + 3)–CO(i + 1) and NH(i + 3)–CO(i). A number of these turns had side-chain hydrogen bonds with main-chain oxygen CO at i or with main-chain nitrogen at i or side-chain interactions with other amino acids, mainly aspartic acid at i. These turns with asparagine at i + 3 are different from the Asx-motifs. An example of this turn type is shown in figure 3D.

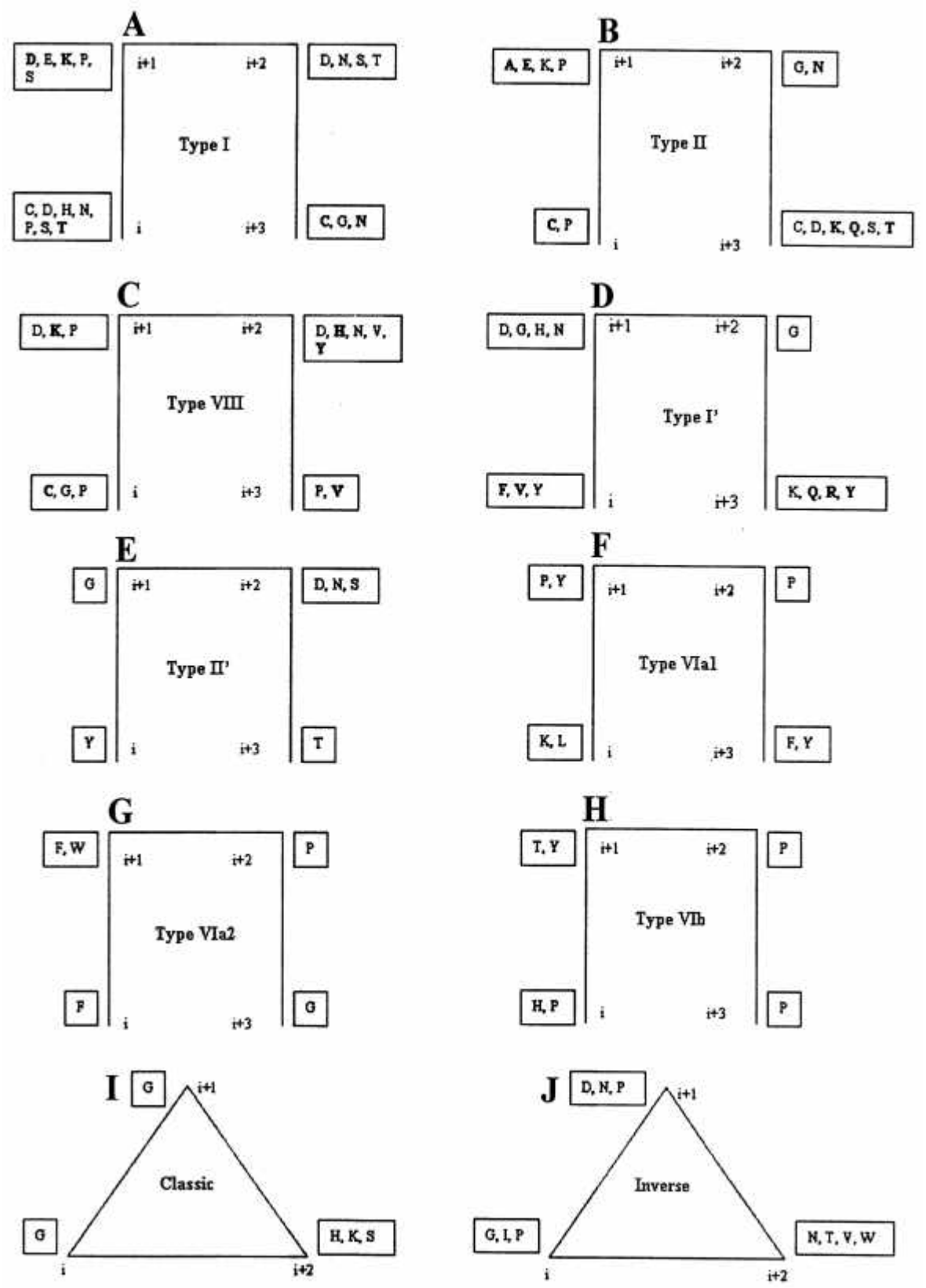

Figure 1. Statistically significant new amino acid type-dependent positional preferences for b-turn types I, II, VIII and I′ observed in the present study compared with earlier work of Hutchinson and Thornton (1994) are indicated in bold. The other statistically significant amino acid preferences at individual positions are also indicated in the figure for each turn type. For b-turn type II′ amino acid preferences are same as observed in the earlier work. For types VIa1, VIa2, VIb b-turns and the classic and inverse g-turns amino acid potentials are calculated for the first time and the preferences are reported based on the present work.

3.4 *Type II* b*-turns*

Cysteine was preferred at *i* and most of these turns have both types of hydrogen bonds; NH(*i* + 3)–CO(*i*) and NH(*i* + 2)–CO(*i*). Alanine and glutamic acid were observed at *i* + 1; alanine makes predominant interactions with water or other main-chain atoms (see figure 4B). The glutamic acid also makes solvent interactions through its side-chain. A number of these turns with alanine or glutamic acid at *i* + 1 had main-chain main-chain hydrogen bonds of the NH(*i* + 3)–CO(*i* + 1) type. A large number of turns with asparagine at *i* + 2 had hydrogen bond of the type NH(*i* + 2)–CO(*i*). We observed lysine, glutamine and threonine at the *i* + 3 position. These turns had either both hydrogen bond types; NH(*i* + 3)–CO(*i* + 1) and NH(*i* + 3) –CO(*i*) or only the regular b-turn hydrogen bond NH(*i* + 3)–CO(*i*) or only hydrogen bond NH(*i* + 3)–CO(*i* + 1). Further, a majority of the type II b-turns with threonine at *i* + 3, make side-chain hydrogen-bond with main-chain oxygen CO(*i*) (see figure 3E) and relatively few turns had hydrogen bond with main-chain nitrogen of residue at *i*. Threonine at *i* in type I b-turns and threonine at *i* + 3 in type II b-turns show a distinction in the type and number of side-chain main-chain, main-chain main-chain hydrogen bonds and side-chain side-chain interactions. One-third type II b-turns with glutamine at *i* + 3 make side-chain hydrogen bond with main-chain nitrogen at *i*. Few turns with glutamine or lysine at *i* + 3 have side-chain main-chain hydrogen bonds with CO(*i* +1) or with side chains at *i*.

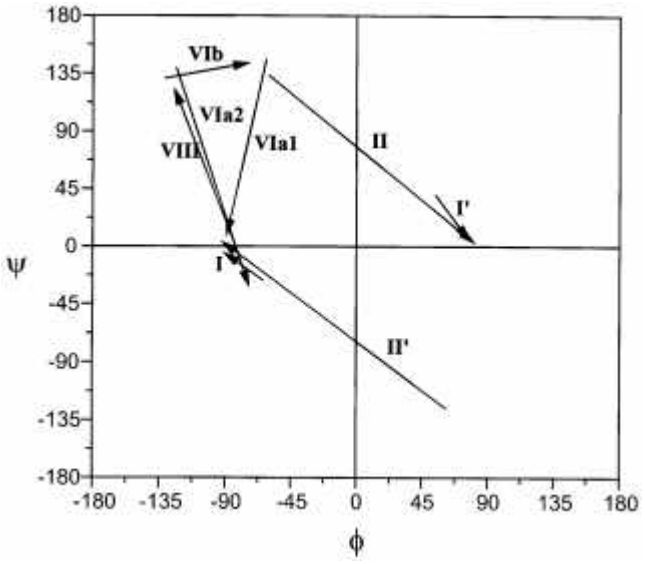

Figure 2. The average f, y *i* + 1 values connecting to *i* + 2 values for b-turns as in table 4.

3.5 *Type VIII* b*-turns*

Type VIII b-turns with cysteine at *i* had hydrogen bond of the type NH(*i* + 2)–CO(*i*). A number of these turns make side-chain main-chain hydrogen bond with main-chain oxygen CO at *i* + 2 (see figure 3F). Lysine preferred at *i* + 1 has main-chain and side-chain interactions with solvent or with other main-chain and side-chain atoms in the protein. Valine interacts mainly with water or other main-chain atoms in the protein. Histidine and tyrosine at *i* + 2 position also interact with water through their main-chain or side-chain atoms or make hydrogen bonds with other main-chain atoms in the protein. Nearly a third of turns with tyrosine at *i* + 2 and half thenumberofturns with histidine at *i* + 2 make hydrogen bond of the type

Table 4. Average dihedral angles for b- and g-turns.

b	f (*i* + 1)	Diff.	Ψ (*i* + 1)	Diff.	f (*i* + 2)	Diff.	Ψ (*i* + 2)	Diff.
I	– 63· 88 (– 60)	3· 88	– 26· 62 (– 30)	– 3· 38	– 91· 30 (– 90)	1· 3	– 4· 24 (0)	4· 24
II	– 58· 88 (– 60)	– 1· 12	132· 14 (120)	– 12· 14	82· 55 (80)	– 2· 55	1· 42 (0)	– 1· 42
VIII	– 73· 00 (– 60)	13	– 32· 42 (– 30)	2· 42	– 124· 66 (– 120)	4· 66	123· 19 (120)	– 3· 19
I′	54· 77 (60)	5· 23	39· 54 (30)	– 9· 54	78· 10 (90)	11· 9	4· 04 (0)	– 4· 04
II′	61· 54 (60)	– 1· 54	– 125· 61 (– 120)	5· 61	– 92· 59 (– 80)	12· 59	4· 08 (0)	– 4· 08
VIa1	– 62· 31 (– 60)	2· 31	144· 96 (120)	– 24· 96	– 89· 26 (– 90)	– 0· 74	6· 98 (0)	– 6· 98
VIa2	– 123· 12 (– 120)	3· 12	138· 86 (120)	– 18· 86	– 80· 77 (– 60)	20· 77	– 10· 13 (0)	10· 13
VIb	– 130· 48 (– 135)	– 4· 52	130· 56 (135)	5· 56	– 72· 50 (– 75)	– 2· 5	143· 29 (160)	16· 71
IV	– 69· 18 (– 61)	8· 61	17· 69 (10)	– 7· 69	– 52· 51 (– 53)	– 0· 49	25· 08 (17)	– 8· 08

g	f (*i* + 1)	Diff.	Ψ (*i* + 1)	Diff.
Classic	70· 97 (75)	5· 97	– (– 64)	– 6· 37
Inverse	– 82· 92 (– 79)	3· 92	57· 63 75· 59 (69)	– 6· 59

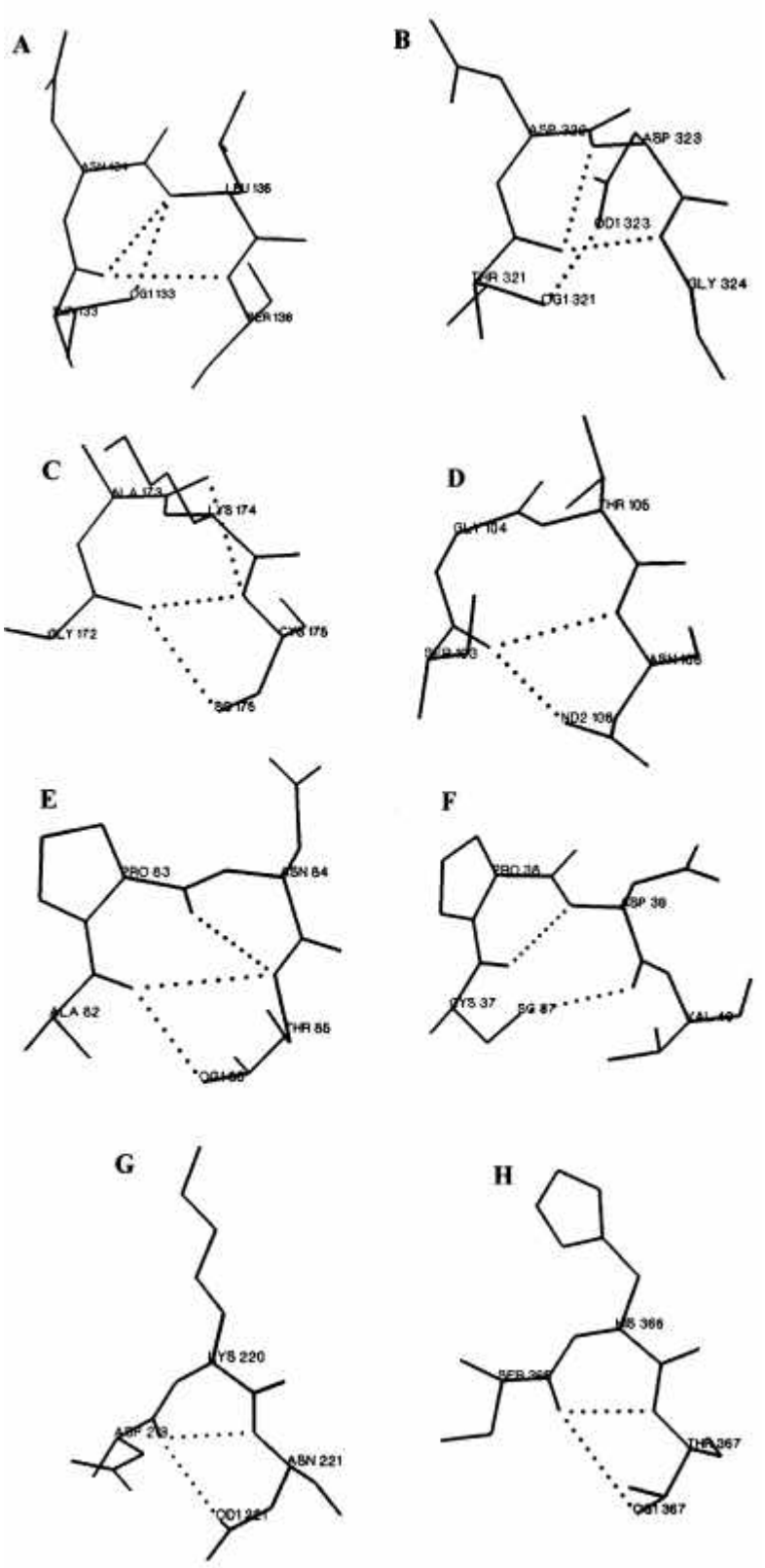

Figure 3. Hydrogen bond and side-chain interactions in b-turns (**A F**) and g-turns (**G, H**) described in text.

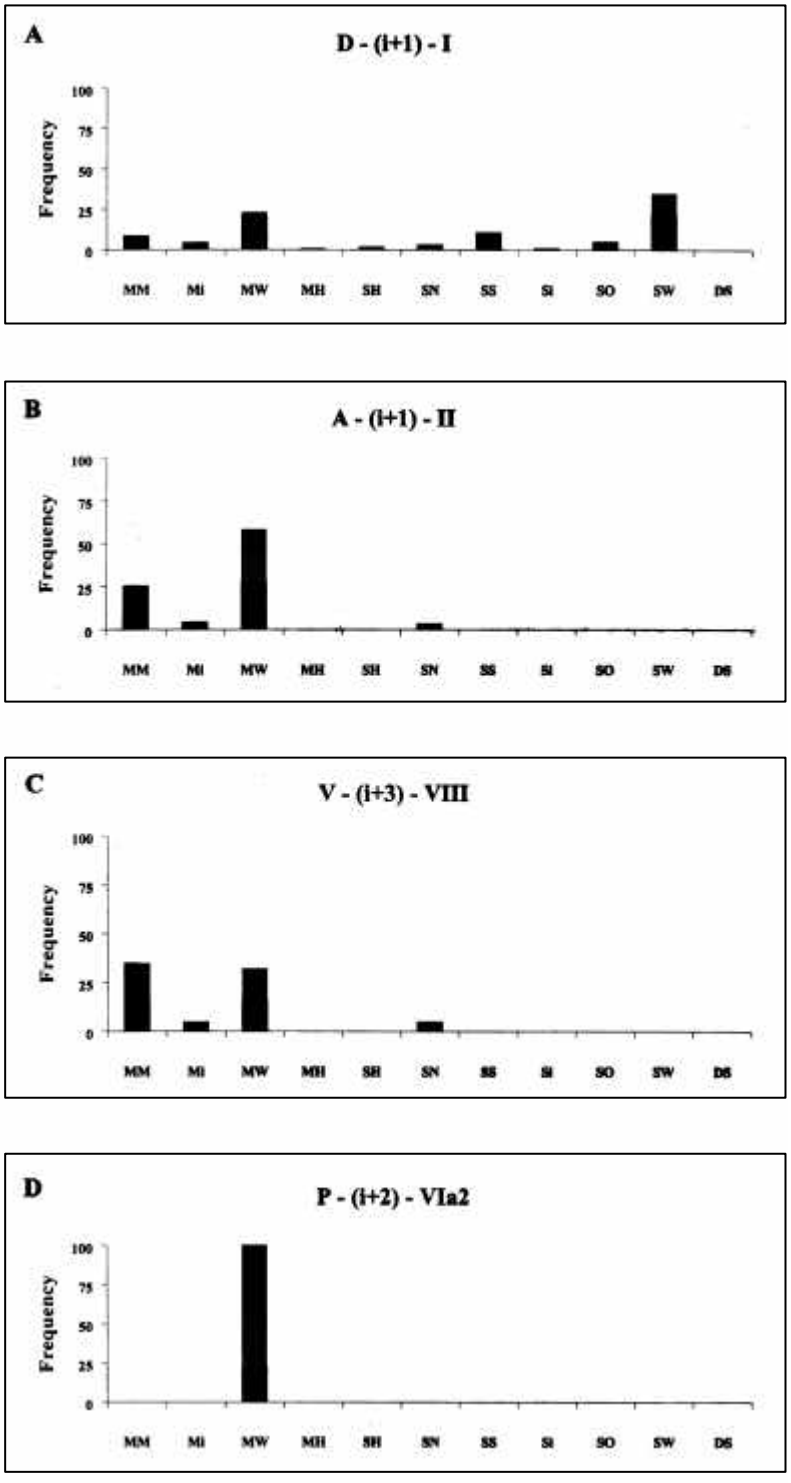

Figure 4. Hydrogen bond and other interactions for amino acids at specific positions in a particular turn type. MM, Main-chain in turn with main-chain atom in the protein hydrogen bond; Mi, main-chain main-chain hydrogen bond within the turn; MW, main-chain in turn with solvent interaction (< 4· 0 Å); MH, main-chain in turn with heterogen atom interaction in the protein; SH, side-chain in turn with heterogen atom interaction in the protein; SN, side-chain and main-chain nitrogen hydrogen bond within the turn; SS, side-chain in turn with side-chain atom in the protein interaction; Si, side-chain side-chain interaction in turn; SO, sidechain and main-chain oxygen hydrogen bond within the turn; SW, side-chain in turn with solvent interaction; DS, disulphide interaction.

NH(i + 2)–CO(i). Valine is preferred at i + 3 but 61 out of 64 of these turns did not have any main-chain hydrogen bonds. These make hydrogen bond with other main-chain atoms in the protein or solvent interactions (see figure 4C).

3.6 *Type I′ β-turns*

Usually, β-sheet preferring residues are observed at ith position in type I′ β-turns. Phenylalanine and valine were observed to occur frequently at i. These turns have both hydrogen bonds of type; NH(i + 3)–CO(i) and NH(i + 2)–CO(i). We observed a preference for glutamine, arginine and tyrosine at i + 3. Ten out of the 19 turns with tyrosine at i + 3 had both types of hydrogen bonds; NH(i + 3)–CO(i + 1) and NH(i + 3)–CO(i). Majority of turns with glutamine or arginine at i + 3 had hydrogen bonds of both the types; NH(i + 3)–CO(i + 1) and NH(i + 3)–CO(i). Nearly a third of these turns with glutamine or arginine at i + 3 make side-chain hydrogen bond with main-chain oxygen of residues either at i or i + 1.

3.7 *Type II′ β-turns*

Almost all amino acid preferences were consistent with previous observations.

3.8 *Type VIa1 and VIa2 β-turns*

Type VI β-turns are fewer compared with other β-turn types. Seven amino acid positional preferences for type VIa1 β-turns and 5 for type VIa2 β-turns were observed. Lysine has main-chain and side-chain hydrogen bonds with other protein atoms or solvent. Leucine has mainly main-chain hydrogen bonds with other main-chain atoms in the protein. Proline and tyrosine are preferred at i + 1. Proline may be preferred due to the restricted conformation of the dihedral angle. These turns have been suggested to be stabilized by the formation of multiple turns (Hutchinson and Thornton 1994). Turns with phenylalanine and tyrosine at i + 3 have the regular NH(i + 3)–CO(i) hydrogen bond. For type VIa2 β-turns with phenylalanine at i and i + 1, main-chain and solvent interactions were mainly observed and with tryptophan at i + 1 side-chain interactions were also observed.

3.9 *Type VIb β-turns*

Main-chain hydrogen bonds in the turn were not observed for most type VIb β-turns. Proline and histidine are preferred at i. Proline may be preferred due to the restricted φ conformation. Histidine makes main-chain or side-chain solvent interactions or main-chain hydrogen bonds with other main-chain atoms in the protein. Threonine and tyrosine are preferred at i + 1. These residues are mainly involved in solvent interactions. Proline is the only residue preferred in all these turns at i + 2 and is also observed in nine type VIb β-turns at i + 3. Proline at i, i + 2 and i + 3 interact mainly with solvent.

3.10 *Proline at i + 2 in type VI β-turns*

It has earlier been suggested by Hutchinson and Thornton (1994), that the formation of a double turn (equivalent to having a proline at the i + 1 position in the second overlapping turn, which is the most favoured position for proline in a β-turn) is likely to stabilize the relatively unfavourable type VI β-turn. In fact, proline is the only amino acid observed at the i + 2 position in all 131 type VI β-turns. However, we also observed 16 type VIa1 β-turns, 5 type VIa2 β-turns, 45 type VIb β-turns, 3 type I β-turns, 4 type II′ β-turns and 27 type IV β-turns with proline at i + 2 as isolated β-turns. Proline is mainly surrounded by water molecules in all these turns. Preliminary analysis did not reveal any specific preference for either, the number or preferred orientation of water molecules around proline at i + 2 in isolated β-turns.

3.11 *Type IV β-turns*

Type IV β-turns constitute 35% of the total number of β-turns. Significant amino acid preferences are shown in table 3a. The restricted conformational space for proline in the Ramachandran plot and a relatively larger conformational space available for glycine, asparatic acid or asparagine (Ramachandran and Sasisekharan 1968) probably favour the choice of these amino acid residues. Proline and glycine interact mainly with solvent and aspartic acid and asparagine make hydrogen bonds with protein atoms outside the turn.

3.12 *Inverse and classic γ turns*

The potentials for amino acid preferences in γ turns is shown in table 2b and the type-dependent positional potentials in table 3b. Glycine is preferred at ith and i + 1th positions for classic γ turns and glycine has mainly main-chain solvent interactions. Histidine, lysine and serine are preferred at i + 2th position, which are stabilized through solvent interactions or due to the side-chain main-chain hydrogen bonds with protein atoms in addition to the NH(i + 2)–CO(i) hydrogen bond. Nearly a quarter of inverse γ turns with asparagine at i + 2 make side-chain hydrogen bond with main-chain oxygen of ith residue (see figure 3G) and in 4 inverse γ turns interactions with the side-chain at i, either with glutamic acid or asparagine. In four inverse γ turns with threonine at i + 2, the side-chain makes hydrogen

bond with main-chain oxygen at *i* (see figure 3H) that resembles a b-turn with 9 atoms in the ring enclosed by the main-chain hydrogen bond of the type NH(i + 3)–CO(i). In all classic g-turns analysed, proline is always at position *i*, whereas, in the inverse g-turn proline is preferred either at i + 1 or *i* or both. However, in b-turns, proline can be at the end; there are 9 type VIb turns, 128 type VIII turns and 261 type IV turns with a proline at the i + 3rd position.

3.13 *Hydrogen bonds in turns*

The main-chain main-chain hydrogen bonds in b-turns can be one or more of 6 types; NH(i + 3)–CO(i), NH(i + 2)–CO(i), NH(i)–CO(i + 3), NH(i)–CO(i + 2), NH(i + 3)–CO(i + 1) and NH(i + 1) – CO(i + 3). Of these, NH(i + 3) –CO(i) is expected for b-turns and NH(i + 2)–CO(i) is expected for g-turns. In earlier analysis on a set of 23 proteins (Nemethy and Scheraga 1980), it was observed that the NH(i + 3) → CO(i) hydrogen bond was not present for many of the b-turns. These authors pointed out that the b-turn dihedral angles proposed by Venkatachalam (1968) do not lead to optimal hydrogen bonding, especially for type I turns. Based on this observation Rose *et al* (1985) suggested that hydrogen bonding is not of major importance to the stability of b-turns. We have not carried out a detailed analysis of main-chain main-chain hydrogen bonds in all b-turns. However, from the present analysis we observed that with a few exceptions, generally the hydrogen bonds of one of the 6 types may be possible for most b-turns. For instance, out of 182 type I b-turns with threonine at the *i*th position only 29 turns did not have any main-chain hydrogen bonds. The same is true for 28 out of 146 type I b-turns with asparagine at i + 3rd position. However, in type VIII b-turns with valine in the i + 3rd position, 61 out of 64 turns did not have any main-chain hydrogen bonds described above. Such turns are usually stabilized through hydrogen bond interactions with other main-chain atoms in the protein or through solvent interactions.

4. Conclusions

The b- or g-turn type-dependent amino acid positional preferences and potentials derived from the present study may be useful for modelling turns in proteins. Where NH(i + 3) → CO(i) type hydrogen bond is not observed, such b-turns may be stabilized due to other main-chain hydrogen bonds in the turn or through main-chain or side-chain interactions with other atoms in the protein or through solvent interactions. It is not uncommon to find isolated b-turns with proline at i + 2 contrary to expectations. These turns are possibly stabilized due to predominant interactions with solvent surrounding the proline residue.

Acknowledgements

We acknowledge the technical help of Mr Ranjit Kumar, Mr Pavan Kumar and Mr Shiva Prasad.

Appendix A. Protein chains (426) from Protein Data Bank used for b-turn analysis.

119l	153l	1a1iA	1a1x	1a28B	1a2pA	1a2yA	1a2zA	1a34A
1a62	1a68	1a6q	1a7tA	1a8e	1a8i	1a9s	1aac	1aba
1ad2	1adoA	1af7	1afwA	1agjA	1agqD	1ah7	1aho	1aj2
1ajj	1ajsA	1ak0	1ak1	1ako	1akz	1al3	1alo	1alu
1alvA	1aly	1amm	1amp	1amuA	1amx	1anf	1aocA	1aohA
1aol	1aop	1aoqA	1aozA	1apyB	1aq0A	1aq6A	1aqb	1aqzB
1arb	1arv	1at0	1atlA	1atzB	1avmA	1awd	1awsA	1axn
1ayl	1azo	1ba1	1bbpA	1bdmB	1bdo	1bebA	1benB	1bfd
1bfg	1bftA	1bgc	1bgp	1bkf	1bkrA	1brt	1btkB	1btn
1bv1	1byb	1c52	1cbn	1cem	1ceo	1cewI	1cex	1cfb
1chd	1chmA	1ckaA	1clc	1cnv	1cpcB	1cpo	1cseE	1cseI
1csh	1csn	1ctj	1cydA	1dad	1dkzA	1dokA	1dorA	1dosA
1dun	1dupA	1dxy	1eca	1ecl	1ecpA	1ede	1edg	1edmB
1edt	1erv	1ezm	1fdr	1fds	1fit	1fleI	1fmtB	1fna
1fua	1furA	1fus	1fvkA	1fwcA	1g3p	1gai	1garA	1gd1O
1gdoA	1gifA	1gky	1gnd	1gotB	1gotG	1gsa	1guqA	1gvp
1ha1	1havA	1hcrA	1hfc	1hgxA	1hoe	1hsbA	1htrP	1hxn
1iakA	1idaA	1idk	1ido	1ifc	1igd	1iibA	1iso	1isuA
1ixh	1jdw	1jer	1jetA	1jfrA	1jpc	1kid	1knb	1kpf
1kptA	1kuh	1kveA	1kveB	1kvu	1kwaB	1lam	1latB	1lbu
1lcl	1lis	1lit	1lki	1lkkA	1lmb3	1lml	1lt5D	1ltsA
1lucB	1mai	1mbd	1mkaA	1mldA	1mml	1molA	1mpgA	1mrj
1mrp	1msc	1msi	1msk	1mtyB	1mtyD	1mtyG	1mucA	1mugA
1mwe	1mzm	1nar	1nbaB	1nbcA	1nciB	1neu	1nfn	1nif
1nls	1nox	1np1A	1npk	1nulB	1nwpA	1nxb	1ois	1onc
1onrA	1opd	1opy	1orc	1ospO	1ovaA	1oyc	1pcfA	1pda
1pdo	1pgs	1phe	1phnA	1php	1pii	1plc	1pmi	1pne

1pnkB	1poa	1poc	1pot	1ppn	1ppt	1prxB	1ptq	1pty
1pud	1qba	1qnf	1r69	1ra9	1rcf	1rec	1regY	1reqD
1rgeA	1rhs	1rie	1rmg	1rro	1rss	1rsy	1rvaA	1ryp1
1ryp2	1rypF	1rypI	1rypJ	1sbp	1sfp	1sftB	1sgpI	1skz
1sltA	1sluA	1smd	1spuA	1sra	1stmA	1svb	1svpA	1tadC
1tca	1tfe	1thv	1thx	1tib	1tif	1tml	1trkA	1tsp
1tvxA	1tys	1uae	1ubi	1uch	1unkA	1urnA	1uxy	1v39
1vcaA	1vcc	1vhh	1vid	1vif	1vin	1vjs	1vls	1vpsA
1vsd	1vwlB	1wab	1wba	1wdcA	1wer	1whi	1who	1whtB
1wpoB	1xgsA	1xikA	1xjo	1xnb	1xsoA	1xyzA	1yaiC	1yasA
1ycc	1yer	1ytbA	1yveI	1zin	256bA	2a0b	2abk	2acy
2arcA	2ayh	2baa	2bbkH	2bbkL	2bopA	2cba	2ccyA	2chsA
2ctc	2cyp	2dri	2end	2eng	2erl	2fdn	2fha	2fivA
2gdm	2hbg	2hft	2hmzA	2hpdA	2hts	2i1b	2ilk	2kinA
2kinB	2lbd	2mcm	2msbB	2nacA	2pgd	2phy	2pia	2pii
2plc	2por	2pspA	2pth	2rn2	2rspB	2sak	2scpA	2sicI
2sil	2sn3	2sns	2tgi	2tysA	2vhbB	2wea	3b5c	3chy
3cla	3cox	3cyr	3daaA	3grs	3lzt	3nul	3pcgM	3pte
3sdhA	3seb	3tss	3vub	4bcl	4mt2	4pgaA	4xis	5csmA
5hpgA	5icb	5p21	5pti	5ptp	6cel	6gsvA	7ahlA	7rsa
8abp	8rucI	8rxnA						

Appendix B. Protein chains (320) from Protein Data Bank used for g-turn analysis.

119l	1a1x	1a2yA	1a2zA	1a34A	1a68	1a6q	1a8e
1a8i	1a9s	1ad2	1adoA	1af7	1afwA	1agjA	1ah7
1aj2	1ajsA	1ak0	1ak1	1ako	1akz	1al3	1alo
1alvA	1aly	1amm	1amp	1amuA	1anf	1aocA	1aop
1aoqA	1aozA	1apyB	1aq0A	1aqzB	1arb	1arv	1atlA
1atzB	1avmA	1awsA	1axn	1ayl	1ba1	1bbpA	1bdmB
1bdo	1bebA	1bfd	1bfg	1bgc	1bgp	1bkf	1bkrA
1brt	1btkB	1btn	1byb	1c52	1cem	1ceo	1cex
1cfb	1chd	1chmA	1clc	1cnv	1cpo	1cseE	1csh
1csn	1ctj	1cydA	1dad	1dorA	1dosA	1dun	1dupA
1dxy	1eca	1ecl	1ecpA	1ede	1edg	1edt	1erv
1ezm	1fdr	1fds	1fit	1fleI	1fmtB	1fna	1fua
1furA	1fus	1g3p	1gai	1garA	1gd1O	1gifA	1gky
1gnd	1gotB	1gotG	1gsa	1guqA	1ha1	1hcrA	1hgxA
1iakA	1idaA	1idk	1ido	1iibA	1iso	1ixh	1jdw
1jer	1jetA	1jfrA	1knb	1kpf	1kptA	1kveA	1kvu
1kwaB	1lam	1lbu	1lcl	1lis	1lit	1lki	1lml
1ltsA	1lucB	1mai	1mbd	1mldA	1mml	1molA	1mpgA
1mrj	1mrp	1msc	1msi	1msk	1mtyB	1mtyD	1mucA
1mwe	1mzm	1nar	1nbaB	1neu	1nif	1nls	1nox
1npk	1nulB	1onc	1onrA	1opd	1opy	1ovaA	1oyc
1pda	1pgs	1phe	1php	1pii	1plc	1pmi	1pnkB
1poa	1poc	1pot	1ppn	1prxB	1ptq	1pty	1qba
1qnf	1ra9	1rec	1regY	1reqD	1rhs	1rie	1rmg
1rro	1rsy	1rvaA	1ryp1	1ryp2	1rypF	1rypI	1rypJ
1sbp	1sfp	1sftB	1skz	1sltA	1sluA	1smd	1spuA
1sra	1stmA	1svb	1svpA	1tadC	1tca	1thv	1tib
1tml	1trkA	1tsp	1tys	1uae	1uch	1v39	1vcaA
1vhh	1vid	1vin	1vjs	1vls	1vpsA	1vsd	1vwlB
1wab	1wba	1wer	1whi	1who	1whtB	1wpoB	1xgsA
1xjo	1xsoA	1xyzA	1yaiC	1yasA	1ycc	1yer	1ytbA
1yveI	256bA	2a0b	2abk	2arcA	2baa	2bbkH	2bbkL
2cba	2chsA	2ctc	2cyp	2dri	2end	2eng	2fha
2hbg	2hft	2hmzA	2hpdA	2hts	2i1b	2ilk	2kinB
2lbd	2nacA	2pgd	2phy	2pia	2plc	2por	2pth
2rn2	2rspB	2scpA	2sicI	2sil	2sns	2spcA	2tgi

Contd. . .

2tysA	2vhbB	2wea	3b5c	3chy	3cla	3cox	3cyr
3daaA	3grs	3lzt	3pcgM	3pte	3sdhA	3seb	3tss
3vub	4bcl	4mt2	4pgaA	4xis	5csmA	5hpgA	5icb

Appendix B. (*Contd*)

5p21	5pti	5ptp	6cel	7ahlA	7rsa	8abp	8rucI

References

Abola E E, Sussman J L, Prilusky J and Manning N O 1997 Protein Data Bank archives of three-dimensional macromolecular structures; *Methods Enzymol.* **277** 556–571

Bernstein F C, Koetzle T F, Williams G J, Meyer E E Jr, Brice M D, Rodgers J R, Kennard O, Shimanouchi T and Tasumi M 1977 The Protein Data Bank: a computer-based archival file for macromolecular structures; *J. Mol. Biol.* **112** 535–542

Cai Y D, Yu H and Chou K C 1998 Prediction of beta-turns; *J. Protein Chem.* **4** 363–376

Chou K C and Blinn J R 1997 Classification and prediction of beta-turn types; *J. Protein Chem.* **16** 575–595

Chou K C 1997 Prediction of beta-turns; *J. Pept. Res.* **49** 120–144

Chou P Y and Fasman G D 1974 Prediction of protein conformation; *Biochemistry* **13** 222–245

Chou P Y and Fasman G D 1979 Prediction of beta-turns; *Biophys. J.* **26** 367–373

Cohen F E, Abarbanel R M, Kuntz I D and Fletterick R J 1986 Turn prediction in proteins using a pattern-matching approach; *Biochemistry* **25** 266–275

Garnier J, Osguthorpe D J and Robson B 1978 Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins; *J. Mol. Biol.* **120** 97–120

Hobohm U and Sander C 1994 Enlarged representative set of protein structures; *Protein Sci.* **3** 522–524

Hutchinson E G and Thornton J M 1996 PROMOTIF – a program to identify and analyze structural motifs in proteins; *Protein Sci.* **5** 212–220

Hutchinson E G and Thornton J M 1994 A revised set of potentials for beta-turn formation in proteins; *Protein Sci.* **3** 2207–2216

Kabsch W and Sander C 1983 Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features; *Biopolymers* **22** 2577–2637

Lewis P N, Momany F A and Scheraga H A 1973 Chain reversals in proteins; *Biochim. Biophys. Acta* **303** 211–229

McGregor M J, Flores T P and Sternberg M J 1989 Prediction of beta-turns in proteins using neural networks; *Protein Eng.* **2** 521–526

Milner-White E J, Ross B M, Ismail R, Belhadj-Mostefa K and Poet R 1988 One type of gamma-turn, rather than the other gives rise to chain-reversal in proteins; *J. Mol. Biol.* **204** 777–782

Milner-White E J 1990 Situations of gamma-turns in proteins. Their relation to alpha-helices, beta-sheets and ligand binding sites; *J. Mol. Biol.* **216** 386–397

Nemethy G and Scheraga H A 1980 Stereochemical requirements for the existence of hydrogen bonds in beta-bends; *Biochem. Biophys. Res. Commun.* **95** 320–327

Overington J, Johnson M S, Sali A and Blundell T L 1990 Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction; *Proc. R. Soc. London B Biol. Sci.* **241** 132–145

Ramachandran G N and Sasisekharan V 1968 Conformation of polypeptides and proteins; *Adv. Protein Chem.* **23** 283–438

Rees D C, Lewis M and Lipscomb W N 1983 Refined crystal structure of carboxypeptidase A at 1.54 A resolution; *J. Mol. Biol.* **168** 367–387

Richardson J S 1981 The anatomy and taxonomy of protein structure; *Adv. Protein Chem.* **34** 167–339

Rose G D 1978 Prediction of chain turns in globular proteins on a hydrophobic basis; *Nature (London)* **272** 586–590

Rose G D, Gierasch L M and Smith J A 1985 Turns in peptides and proteins; *Adv. Protein Chem.* **37** 1–109

Shepherd A J, Gorse D and Thornton J M 1999 Prediction of the location and type of beta-turns in proteins using neural networks; *Protein Sci.* **8** 1045–1055

Sibanda B L, Blundell T L and Thornton J M 1989 Conformation of beta-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering; *J. Mol. Biol.* **206** 759–777

Sibanda B L and Thornton J M 1993 Accommodating sequence changes in beta-hairpins in proteins; *J. Mol. Biol.* **229** 428–447

Sussman J L, Lin D, Jiang J, Manning N O, Prilusky J, Ritter O and Abola E E 1998 Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules; *Acta. Crystallogr. D* **54** 1078–1084

Suzuki M and Yagi N 1991 Structure of the SPXX motif; *Proc. R. Soc. London B Biol. Sci.* **246** 231–235

Toniolo C 1980 Intramolecularly hydrogen-bonded peptide conformations; *CRC Crit. Rev. Biochem.* **9** 1–44

Venkatachalam C M 1968 Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units; *Biopolymers* **6** 1425–1436

Wan W Y and Milner-White E J 1999a A recurring two-hydrogen- bond motif incorporating a serine or threonine residue is found both at alpha-helical N termini and in other situations; *J. Mol. Biol.* **286** 1651–1662

Wan W Y and Milner-White E J 1999b A natural grouping of motifs with an aspartate or asparagine residue forming two hydrogen bonds to residues ahead in sequence: their occurrence at alpha-helical N termini and in other situations; *J. Mol. Biol.* **286** 1633–1649

Wilmot C M and Thornton J M 1988 Analysis and prediction of the different types of beta-turn in proteins; *J. Mol. Biol.* **203** 221–232

MS received 15 November 1999; accepted 8 February 2000

Corresponding editor: SEYED E HASNAIN