Protein Science (1994), *3*:2207–2216. Cambridge University Press. Printed in the USA.

A revised set of potentials for β-turn formation in proteins

E. GAIL HUTCHINSON AND JANET M. THORNTON

Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, Gower Street, London WC1E 6BT, United Kingdom

(RECEIVED August 17, 1994; ACCEPTED October 10, 1994)

Abstract

Three thousand eight hundred ninety-nine β-turns have been identified and classified using a nonhomologous data set of 205 protein chains. These were used to derive β-turn positional potentials for turn types I′ and II′ for the first time and to provide updated potentials for formation of the more common types I, II, and VIII. Many of the sequence preferences for each of the 4 positions in turns can be rationalized in terms of the formation of stabilizing hydrogen bonds, preferences for amino acids to adopt a particular conformation in ϕ,ψ space, and the involvement of turn types I′ and II′ in β-hairpins. Only 1,632 (42%) of the turns occur in isolation; the remainder have at least 1 residue in common with another turn and have hence been classified as multiple turns. Several types of multiple turn have been identified and analyzed.

Keywords: protein structure analysis; sequence preference

β-Turns are the most common type of nonrepetitive structure recognized in proteins and comprise, on average, 25% of the residues (Kabsch & Sander, 1983). Turns play an important part in proteins; they provide a direction change for the polypeptide chain and have been implicated in molecular recognition (Rose et al., 1985) and in protein folding. Thus, since they were first recognized (Venkatachalam, 1968), considerable effort has been devoted to their analysis (Lewis et al., 1973; Chou & Fasman, 1974; Richardson, 1981) and to prediction of turns from the amino acid sequence (Lewis et al., 1971; Chou & Fasman, 1974; Wilmot & Thornton, 1988, 1990). Such work is dependent on and limited by knowledge derived from protein structures that have been solved. In the most recent study (Wilmot & Thornton, 1990), only β-turn classes I, II, and VIII occurred sufficiently frequently to merit an analysis of sequence preferences. In this paper we describe an updated study of β-turns using more than 4 times the number of examples and provide a revised set of β-turn positional potentials, including, for the first time, potentials for type I′ and II′ turns. These potentials can be applied in prediction, modeling, and design studies. In addition, following the work of Isogai et al. (1980), we have identified and analyzed several types of multiple turns in our data set.

Reprint requests to: E. Gail Hutchinson, Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, Gower Street, London WC1E 6BT, United Kingdom; e-mail: gail@uk.ac.ucl.bioc.bsm.

Results and discussion

Turn types

A total of 3,899 β-turns was identified and classified using the 205 protein chains in the data set (Table 1 and files in the SUPLEMNT directory on the Diskette Appendix). This is more than 4 times the number of examples analyzed in the most up-to-date previous analysis (Wilmot & Thornton, 1990), demonstrating the importance of updating the work. In particular the I′ and II′ turn classes are now sufficiently populated to allow meaningful analysis of the sequence preferences for these turns. In addition, data from turn types I, II, and VIII, which were analyzed previously (Wilmot & Thornton, 1988, 1990), should now be more statistically significant. There are still, however, only very small numbers of types VIa and VIb β-turns.

Of the turns located, 1,666 (43%) did not fit into any of the classes using the stringent criteria of Lewis et al. (1973) and were therefore classified as type IV turns. When the ϕ,ψ angle constraints were relaxed by 10°, 425 of these were assigned to a class. However, the data obtained using the more stringent conditions were used for the calculation of average dihedral angles and the

Table 1. *Frequency and mean dihedral angles for standard β-turn types*

Turn type	Ramachandran nomenclature[a]	No. of turns		Mean dihedral angles[d]			
		[b]	[c]	$\phi(i+1)$	$\psi(i+1)$	$\phi(i+2)$	$\psi(i+2)$
I	$\alpha_R\alpha_R$	1,231	1,419	−64 (−60)	−27 (−30)	−90 (−90)	−7 (0)
II	$\beta\gamma_L$	405	489	−60 (−60)	131 (120)	84 (80)	1 (0)
VIII	$\alpha_R\beta$	325	451	−72 (−60)	−33 (−30)	−123 (−120)	121 (120)
I′	$\alpha_L\gamma_L$	127	142	55 (60)	38 (30)	78 (90)	6 (0)
II′	$\epsilon\alpha_R$	90	100	60 (60)	−126 (−120)	−91 (−80)	1 (0)
VIa1	$\beta\alpha_R$	15	17	−64 (−60)	142 (120)	−93 (−90)	5 (0)
VIa2	$\beta\alpha_R$	5	5	−132 (−120)	139 (120)	−80 (−60)	−10 (0)
VIb	$\beta\beta$	35	35	−135 (−135)	131 (135)	−76 (−75)	157 (160)
IV		1,666	1,241	−61	10	−53	17
TOTAL		3,899					

[a] Ramachandran nomenclature for turn type as in Wilmot and Thornton (1990). The nomenclature describes the regions of the Ramachandran plot occupied by residues $i+1$ and $i+2$ of the turn.
[b] Using normal cutoffs of 30° for deviation from standard angles, with one angle allowed to deviate by 45°.
[c] Allowing up to 40° deviation from standard angles, with one angle allowed to deviate by 50°.
[d] The idealized ϕ,ψ values as determined by Lewis et al. (1973) are given in parentheses after the averaged values determined from the data set. The values for the type VI turns are taken from Richardson (1981). Types VIa1 and VIa2 are the two subclasses of type VIa turns identified by Richardson (1981).

positional preferences. The averaged dihedral angles of residues $i+1$ and $i+2$ in each of the classes are rather similar to the ideal values (Table 1).

Positional potentials

Table 2 shows the positional potentials and the overall turn potentials for each residue. These indicate the preference for each residue to be in positions i to $i+3$ of a turn and the preference for occurring in a turn generally and are averaged over all turn types. The residues are ranked in order of decreasing hydrophobicity. Generally hydrophilic residues are more likely to occur in turns than hydrophobic residues, as turns tend to be on the solvent-exposed surface of proteins. Glycine, proline, asparagine, and aspartic acid are particularly favored in turns, and the reasons for this will be discussed later in terms of preferences for individual turn types. There are considerable differences between the potentials at each position; for example, proline is favored at the first 2 positions of turns, whereas glycine is favored at the last 2 positions, and asparagine and aspartic acid are most strongly favored at positions 1 and 3. Again, these can only be rationalized in terms of individual turn types. However, the data presented in Table 2 can be used for assessing the likelihood of a particular sequence to form a turn generally, independent of type.

Table 2. *Positional and overall turn potentials for each residue*

Residue type	Turn potential				
	i	$i+1$	$i+2$	$i+3$	Overall
Ile	0.66	0.61	0.42	0.68	0.59
Phe	0.98	0.66	0.96	0.95	0.89
Val	0.72	0.70	0.54	0.84	0.70
Leu	0.73	0.67	0.47	0.78	0.66
Trp	0.62	0.65	0.76	0.79	0.70
Met	0.70	0.48	0.41	0.68	0.57
Ala	0.81	0.96	0.66	0.89	0.83
Gly	1.09	1.04	2.14	1.64	1.48
Cys	1.42	0.73	0.98	1.20	1.08
Tyr	1.04	0.75	0.83	1.07	0.92
Pro	1.48	2.45	0.63	0.96	1.38
Thr	1.08	0.79	0.94	1.20	1.00
Ser	1.29	1.23	1.06	1.03	1.15
His	1.25	0.95	1.16	0.93	1.07
Glu	0.87	1.35	0.92	0.89	1.01
Asn	1.54	1.02	2.14	1.06	1.44
Gln	0.89	0.94	0.93	1.01	0.94
Asp	1.56	1.24	1.86	0.99	1.41
Lys	0.80	1.22	0.94	1.10	1.01
Arg	0.69	0.93	0.75	0.93	0.82

Residue frequencies and type-dependent positional potentials for positions i to $i+3$ of each of the turn classes, again ranked in order of decreasing hydrophobicity, are shown in Table 3A, B, C, D, E, and F. Values that are significant at the 5% level ($d \geq 1.97$) are indicated by asterisks (*), and significantly high values are in bold typeface. Potentials were not calculated for types VIa and VIb due to the small number of examples in each of these classes, and only the residue frequencies are shown in Table 3F.

For types I, II, and VIII turns the sequence preferences agree broadly with the very detailed previous analyses. Some weak preferences have disappeared or become stronger. Only the most significant preferences that can be rationalized are discussed below.

Type I turns

Previously we had observed a strong preference for side chains that can act as hydrogen bond acceptors (Asn, Asp, Cys, Ser) at position i. These stabilized the turn by forming a hydrogen

Table 3. *Type-dependent positional potentials for each amino acid at each of the 4 positions (i, i + 1, i + 2, and i + 3) of turns*[a]

Residue	*i*	*i* + 1	*i* + 2	*i* + 3
A. Type I turns				
Ile	0.32* (21)	0.60* (40)	0.29• (19)	0.75* (50)
Phe	0.81 (40)	0.42* (21)	1.13 (56)	1.01 (50)
Val	0.39* (33)	0.72* (61)	0.50* (42)	0.77* (65)
Leu	0.68* (68)	0.70* (70)	0.52* (52)	0.74* (74)
Trp	0.54* (10)	0.65 (12)	1.30 (24)	1.14 (21)
Met	0.65 (16)	0.53* (13)	0.41* (10)	0.61 (15)
Ala	0.63* (68)	1.07 (115)	0.80* (86)	0.87 (93)
Gly	1.07 (109)	0.40* (41)	0.61* (62)	**2.38* (242)**
Cys	**1.57* (36)**	0.87 (20)	0.96 (22)	1.35 (31)
Tyr	0.66* (30)	0.61* (28)	0.90 (41)	0.85 (39)
Pro	**1.31* (76)**	**3.49* (203)**	0.21* (12)	0.03* (2)
Thr	1.11 (81)	0.94 (69)	**1.44* (105)**	1.11 (81)
Ser	**1.52* (120)**	**1.50* (119)**	**1.29* (102)**	1.02 (81)
His	**1.60* (45)**	0.53* (15)	0.99 (28)	0.92 (26)
Glu	0.74* (54)	**1.54* (112)**	1.21 (88)	1.02 (74)
Asn	**2.25* (125)**	0.68* (38)	**2.26* (126)**	1.08 (60)
Gln	0.72 (31)	0.81 (35)	1.18 (51)	0.58* (25)
Asp	**2.51* (180)**	1.20 (86)	**2.54* (182)**	1.11 (80)
Lys	0.70* (50)	1.20 (85)	1.01 (72)	1.10 (78)
rg	0.64* (36)	0.86 (48)	0.88 (49)	0.78 (44)
C. Type VIII turns				
Ile	0.92 (16)	0.97 (17)	1.26 (22)	0.63 (11)
Phe	0.69 (9)	0.31* (4)	1.54 (20)	1.00 (13)
Val	0.77 (17)	0.99 (22)	**1.58* (35)**	1.31 (29)
Leu	0.45* (12)	0.87 (23)	1.06 (28)	0.83 (22)
Trp	1.24 (6)	0.21 (1)	0.82 (4)	0.21 (1)
Met	0.46 (3)	0.46 (3)	0.62 (4)	0.31 (2)
Ala	0.88 (25)	1.06 (30)	0.53* (15)	0.89 (25)
Gly	**1.49* (40)**	0.30* (8)	0.15* (4)	0.67 (18)
Cys	1.16 (7)	0.99 (6)	0.99 (6)	0.33 (2)
Tyr	0.91 (11)	0.50 (6)	1.50 (18)	0.67 (8)
Pro	**2.74* (42)**	**2.09* (32)**	0.00* (0)	**4.06* (62)**
Thr	1.14 (22)	0.99 (19)	1.30 (25)	**1.46* (28)**
Ser	1.39 (29)	1.15 (24)	0.86 (18)	1.01 (21)
His	0.81 (6)	0.40 (3)	1.35 (10)	0.81 (6)
Glu	0.89 (17)	1.25 (24)	0.52* (10)	0.52* (10)
Asn	1.30 (19)	0.82 (12)	**1.91* (28)**	0.89 (13)
Gln	0.79 (9)	1.23 (14)	0.79 (9)	0.79 (9)
Asp	0.63 (12)	**1.96* (37)**	**1.75* (33)**	0.80 (15)
Lys	0.53* (10)	1.34 (25)	1.07 (20)	0.91 (17)
Arg	0.81 (12)	0.95 (14)	1.02 (15)	0.75 (11)

Residue	*i*	*i* + 1	*i* + 2	*i* + 3
B. Type II turns				
Ile	1.01 (22)	0.64 (14)	0.00* (0)	1.10 (24)
Phe	1.23 (20)	0.74 (12)	0.37* (6)	0.80 (13)
Val	1.12 (31)	0.43* (12)	0.00* (0)	1.12 (31)
Leu	0.73 (24)	0.70 (23)	0.21* (7)	0.70 (23)
Trp	0.49 (3)	0.66 (4)	0.33 (2)	0.99 (6)
Met	0.74 (6)	0.62 (5)	0.12* (1)	1.36 (11)
Ala	1.05 (37)	1.22 (43)	0.14* (5)	1.07 (38)
Gly	0.96 (32)	0.18• (6)	**9.17* (307)**	0.57* (19)
Cys	0.13* (1)	0.00* (0)	0.26* (2)	**1.72* (13)**
Tyr	**1.59* (24)**	0.73 (11)	0.20* (3)	1.26 (19)
Pro	**1.99* (38)**	**4.91* (94)**	0.00* (0)	0.00* (0)
Thr	0.91 (22)	0.67 (16)	0.12* (3)	1.12 (27)
Ser	0.88 (23)	0.96 (25)	0.31* (8)	**1.38* (36)**
His	1.51 (14)	0.97 (9)	0.65 (6)	0.97 (9)
Glu	1.00 (24)	1.21 (29)	0.13* (3)	1.21 (29)
Asn	1.15 (21)	0.65 (12)	1.25 (23)	0.55* (10)
Gln	1.19 (17)	1.26 (18)	0.56 (8)	1.40 (20)
Asp	0.34* (8)	1.02 (24)	0.51* (12)	0.89 (21)
Lys	1.07 (25)	**1.41* (33)**	0.26* (6)	**1.58* (37)**
Arg	0.70 (13)	0.81 (15)	0.16* (3)	1.03 (19)
D. Type I′ turns				
Ile	1.61 (11)	0.15* (1)	0.00* (0)	0.74 (5)
Phe	1.18 (6)	0.59 (3)	0.78 (4)	0.40 (2)
Val	1.50 (13)	0.00* (0)	0.00* (0)	1.28 (11)
Leu	0.87 (9)	0.29* (3)	0.00* (0)	0.68 (7)
Trp	1.05 (2)	0.00 (0)	1.05 (2)	0.53 (1)
Met	0.39 (1)	0.80 (2)	0.39 (1)	1.19 (3)
Ala	0.99 (11)	0.82 (9)	0.09* (1)	0.64 (7)
Gly	0.48 (5)	**1.92* (20)**	**8.66* (91)**	0.38* (4)
Cys	1.27 (3)	0.43 (1)	0.00 (0)	0.00 (0)
Tyr	**2.54* (12)**	0.43 (2)	1.27 (6)	1.28 (6)
Pro	0.50 (3)	0.00* (0)	0.00* (0)	0.00* (0)
Thr	1.06 (8)	0.00* (0)	0.13* (1)	0.67 (5)
Ser	1.10 (9)	0.99 (8)	0.37 (3)	0.62 (5)
His	2.06 (6)	**4.16* (12)**	0.69 (2)	0.00 (0)
Glu	0.40 (3)	0.67 (5)	0.40 (3)	1.61 (12)
Asn	0.35 (2)	**5.26* (30)**	0.70 (4)	1.75 (10)
Gln	0.45 (2)	0.68 (3)	0.22 (1)	1.58 (7)
Asp	0.94 (7)	**2.58* (19)**	0.54 (4)	0.82 (6)
Lys	1.23 (9)	0.55 (4)	0.41 (3)	**3.44* (25)**
Arg	0.86 (5)	0.70 (4)	0.17* (1)	1.74 (10)

(*continued*)

Table 3. *Continued*

Residue	*i*	*i* + 1	*i* + 2	*i* + 3
E. Type II′ turns				
Ile	0.42 (2)	0.21 (1)	0.62 (3)	0.00* (0)
Phe	1.40 (5)	0.28 (1)	0.84 (3)	1.38 (5)
Val	1.64 (10)	0.00* (0)	0.33 (2)	0.81 (5)
Leu	0.55 (4)	0.00* (0)	0.55 (4)	0.54 (4)
Trp	0.00 (0)	0.00 (0)	0.75 (1)	0.74 (1)
Met	0.00 (0)	0.00 (0)	0.56 (1)	0.56 (1)
Ala	0.77 (6)	0.51 (4)	1.16 (9)	1.15 (9)
Gly	1.09 (8)	**9.24* (68)**	1.09 (8)	**1.75* (13)**
Cys	1.20 (2)	0.60 (1)	0.00 (0)	1.19 (2)
Tyr	**2.42* (8)**	0.00 (0)	0.30 (1)	1.20 (4)
Pro	0.48 (2)	0.00* (0)	1.19 (5)	0.00* (0)
Thr	0.95 (5)	0.00* (0)	0.38 (2)	**2.25* (12)**
Ser	1.57 (9)	0.70 (4)	**2.10* (12)**	1.21 (7)
His	1.47 (3)	0.00 (0)	1.47 (3)	0.00 (0)
Glu	1.33 (7)	0.19 (1)	0.95 (5)	0.56 (3)
Asn	0.99 (4)	0.25 (1)	**2.48* (10)**	1.47 (6)
Gln	1.28 (4)	0.32 (1)	0.64 (2)	0.95 (3)
Asp	1.16 (6)	0.77 (4)	**1.93* (10)**	0.38 (2)
Lys	0.19 (1)	0.39 (2)	1.17 (6)	1.35 (7)
Arg	0.74 (3)	0.25 (1)	0.49 (2)	1.46 (6)

	Types VIa and VIa2				Type VIb			
Residue	*i*	*i* + 1	*i* + 2	*i* + 3	*i*	*i* + 1	*i* + 2	*i* + 3
F. Type VI turns[b]								
Ile	0 (0)	0 (2)	0 (0)	0 (0)	1	2	0	2
Phe	0 (1)	1 (0)	0 (0)	1 (0)	5	1	0	3
Val	1 (1)	0 (0)	0 (0)	0 (0)	2	1	0	1
Leu	1 (0)	1 (0)	0 (0)	2 (0)	0	4	0	2
Trp	0 (0)	1 (0)	0 (0)	1 (0)	0	0	0	0
Met	0 (0)	0 (0)	0 (0)	0 (0)	0	0	0	1
Ala	1 (0)	3 (0)	0 (0)	1 (0)	2	1	0	6
Gly	3 (0)	0 (0)	0 (0)	0 (1)	9	0	0	2
Cys	0 (0)	1 (0)	0 (0)	0 (0)	0	0	0	0
Tyr	0 (0)	4 (0)	0 (0)	1 (0)	1	6	0	3
Pro	0 (0)	3 (0)	15 (5)	0 (0)	2	0	35	1
Thr	1 (1)	0 (0)	0 (0)	1 (3)	4	4	0	3
Ser	0 (0)	0 (1)	0 (0)	0 (0)	3	3	0	2
His	0 (1)	0 (0)	0 (0)	0 (1)	1	2	0	1
Glu	2 (0)	1 (0)	0 (0)	2 (0)	1	5	0	0
Asn	0 (1)	0 (2)	0 (0)	5 (0)	1	3	0	2
Gln	0 (0)	0 (0)	0 (0)	0 (0)	1	2	0	0
Asp	3 (0)	0 (0)	0 (0)	1 (0)	2	0	0	0
Lys	3 (0)	0 (0)	0 (0)	0 (0)	0	1	0	1
Arg	0 (0)	0 (0)	0 (0)	0 (0)	0	0	0	5

[a] The number of examples is given in parentheses after each potential. Values labeled with * are statistically significant. Significantly high values are indicated by bold typeface.

[b] For these turns only the number of examples of each residue is included as there are insufficient examples for meaningful potentials to be calculated. For the type VIa turns, the data for the type VIa1 turns are shown followed by the data for the type VIa2 turns in parentheses.

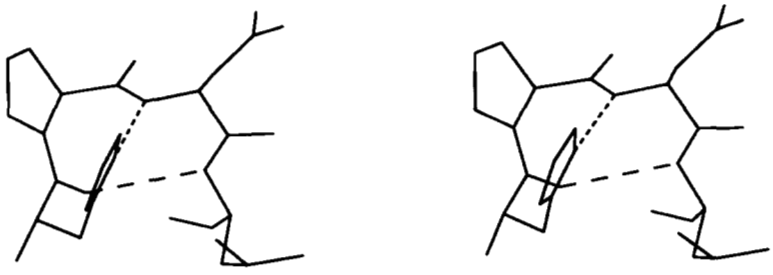

Fig. 1. Stereo plots illustrating an example of a type I β-turn with histidine at position *i*. The histidine side-chain $N\delta_1$ accepts a hydrogen bond from the N-H of the central peptide (small dotted line). The larger dashed line represents the main-chain hydrogen bond between residues *i* and $i + 3$. The residues involved in the turn are 39–42 (His-Pro-Asp-Leu) in chain E of subtilisin Carlsberg (Brookhaven code 1CSE [Bode et al., 1987]). The plots were generated using the program MOLSCRIPT (Kraulis, 1991).

bond with the main-chain nitrogen of the third residue in the turn. In this case, the side chain and main chain of residue *i*, together with residues $i + 1$ and $i + 2$, form another turn-like structure known as the "Asx turn" (Rees et al., 1983). The new data also show a preference for histidine at this position, and this residue can fulfill the same role. Seven of the 45 examples of histidine at position *i* hydrogen bond to the NH of the central peptide (see Fig. 1). A further 8 examples are stabilized by the histidine side chain forming hydrogen bonds to other residues in the turn. Proline is also preferred at position *i*, and this may be due to the high occurrence of multiple turns (see later). Proline is by far the most favored residue at the second position because of the restriction on ϕ to about $-60°$. The other preferred residues at this position (glutamic acid and serine) can stabilize the ϕ angle by forming a hydrogen bond between the oxygen atoms on their side chains and the main-chain amide (see Fig. 2).

At position $i + 2$, aspartic acid, asparagine, serine, and threonine are favored. Analysis of ϕ, ψ distributions of residues in high-resolution, good-quality protein structures (Laskowski et al., 1993) shows that these residues are more likely than most residues to adopt the correct conformation for this position ($\phi = -90°$, $\psi = 0°$) in the turn. This is at the edge of the "bridge" region between α-helical and β-sheet regions of the Ramachandran plot.

The new data confirm glycine as the only residue to show a preference for being at position $i + 3$. Although the ϕ, ψ angles at position $i + 3$ do not cluster around a single value as they do at positions $i + 1$ and $i + 2$, the α_L conformation is unusually common at this position in type I turns (data not shown). Indeed, the majority of glycines at this position do adopt the α_L or ϵ conformations. This contrasts with the conformations of glycine at position *i*, for example, which are more scattered. An α_L conformation at position $i + 3$ does facilitate the return of the polypeptide chain to run antiparallel to its original direction after completion of the turn, and this may be the reason for the preference for glycine at this position.

Type II turns

No strong preferences were previously identified for position *i*. In this analysis, proline and tyrosine residues are favored at this position, but there is no obvious explanation for this. As for type I turns, proline is strongly favored at position $i + 1$ because the ϕ angle at this position ($-60°$) corresponds to the preferred value for proline. In type II turns, lysine is also common. The reason for this is unclear; the lysine residues do not appear to be involved in any local interactions; rather they form hydrogen bonds with residues elsewhere in the structure or with solvent molecules. As observed previously, position $i + 2$ is dominated by glycine and to a lesser extent asparagine because these residues most readily adopt the α_L conformation. At the $i + 3$ position, cysteine, serine, and lysine are favored. Turns involving serine can be stabilized by hydrogen bonding of the serine side chains to the main chain of the first residue in the turn. The lysine side chain is only involved in local hydrogen bonding in about 25% of cases, and most of these involve hydrogen bonds to residue $i - 1$ (i.e., the residue before the first residue in the turn). There is no obvious reason for the preference for cysteine, but the potential calculated for this residue is artificially high because 4 of the 13 examples observed belong to similar regions of wheatgerm agglutinin.

Type VIII turns

These are the most common nonclassic β-turns. The ϕ, ψ requirements at position $i + 1$ are identical to type I turns, whereas residue $i + 2$ is in the β_E region of the Ramachandran plot, rather than α_R as in the type I turn. At position *i*, the only significant preferences shown are for proline and glycine. As in type I turns, proline is favored at position $i + 1$ due to the ϕ, ψ

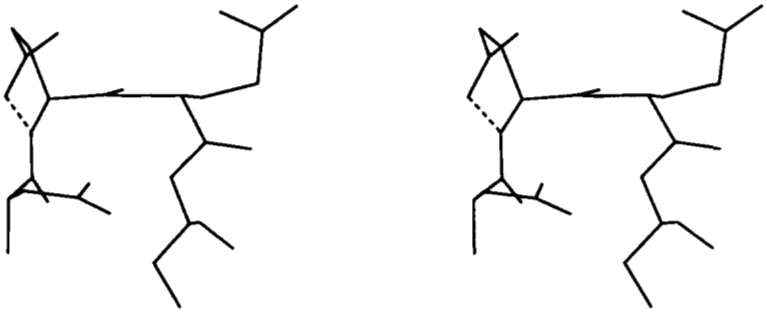

Fig. 2. Type I β-turn with glutamic acid at position $i + 1$. The side-chain $O\epsilon_1$ of the glutamic acid residue hydrogen bonds back to its own main chain, stabilizing the turn. The residues involved are 129–132 (Asn-Glu-Glu-Ser) in chain B of superoxide dismutase (Brookhaven code 2SOD [Tainer et al., 1982]). The plots were generated using MOLSCRIPT (Kraulis, 1991).

constraints at this position. Aspartic acid is also favored at this position because it can stabilize the ϕ angle by the Oδ atom of the side chain interacting with the main-chain amide. Unlike serine and glutamic acid at position $i + 1$ of type I turns, no hydrogen bonds are formed in this case. At position $i + 2$, significant preferences are observed for asparagine and aspartic acid. The side-chain Oδ atoms of both these residues can hydrogen bond to the main-chain nitrogen of the first residue beyond the turn (residue $i + 4$) (Fig. 3). This forms a classic "Asx turn" (Rees et al., 1983) analogous to that discussed for position i of type I turns. An alternative stabilizing hydrogen bond can be formed with the main chain of residue $i + 3$. More surprisingly, there is also a preference for hydrophobic residues: valine and, less significantly, phenylalanine and isoleucine. These residues may be favored because they prefer to adopt the β conformation, which is required at this position (Chou & Fasman, 1974). As observed previously, proline is the dominant residue at position $i + 3$; it encourages the formation of a type VIII rather than a type I turn by restricting the conformation of the previous residue to β. Threonine is also common at this position. There is no obvious reason why this residue in particular should be preferred. However, most of the examples observed are associated with the beginning or end of a β-strand, consistent with the β_E conformation of residue $i + 2$, and threonine does show a preference to be in β-sheets.

Sequence preferences for type I′ and II′ turns were not previously analyzed due to insufficient data. Now, although the numbers involved are still smaller than the other turn types, some significant preferences are observed.

Type I′ turns

It has already been observed that short β-hairpins frequently incorporate type I′ turns (Sibanda & Thornton, 1985). Of the 127 examples of these turns in the current data set, 57 (45%) are involved in short 2:2 hairpins (Kinemage 1), and a further 18 (14%) are involved in longer hairpins. This may explain the preference for β-sheet-preferring residues to be in the first position of type I′ turns. Tyrosine is the only residue to show a significant preference to be in this position, but the potentials for valine and isoleucine are also high. The majority of these residues (Tyr 10/12; Ile 8/11; Val 8/13) do occur in short β-hairpins, and would represent the final residue of a β-strand. At the 2 central positions, the residue preferences are dominated by the restrictions on the ϕ, ψ angles to the α_L (residue $i + 1$) and γ_L (residue $i + 2$) regions of the Ramachandran plot. At position $i + 1$, asparagine, aspartic acid, and glycine are favored, whereas glycine is the dominant residue at position $i + 2$. Histidine is also strongly preferred at position $i + 1$, although the number of examples is quite small. However, this residue shows no preference for the α_L conformation, and there are no distinctive hydrogen bonding patterns involved. Charged and polar residues are favored at position $i + 3$, with lysine the most popular. The reason for this is not clear because these residues do not appear to make any hydrogen bonding or electrostatic interactions with other residues in the turn. The side chains of the arginine residues in this position do form hydrogen bonds to residues around and within the turn, but there is no definite pattern.

Type II′ turns

These turns are also involved in β-hairpins. Of the 90 examples of type II′ turns, 34 (38%) are involved in 2:2 hairpins and a further 18 (20%) are part of longer β-hairpins. As for type I′ turns, tyrosine shows a significant preference to be in the first position, and valine is more weakly preferred. At position $i + 1$, the ϕ, ψ angles are constrained to be around $+60°$, $-120°$, and hence the vast majority of examples have glycine in this position. Aspartic acid, asparagine, and serine are all favored in position $i + 2$. These residues show a preference for adopting the correct ϕ, ψ conformation ($\phi = -80$, $\psi = 0°$) as was discussed for position $i + 2$ in type I turns. Threonine and glycine are the only preferred residues at position $i + 3$ of type II turns. Five of the 12 examples involving threonine are stabilized by forming a hydrogen bond from the $O_{\gamma 1}$ atom of threonine to the oxygen of the main chain at position i (Fig. 4). There is no obvious reason for the preference for glycine; unlike the corresponding position in type I turns, there is no preference for the α_L conformation at position $i + 3$ of type II′ turns.

The type VI turns occur very infrequently, and thus it is not possible to draw conclusions about the sequence preferences for these turns. Positional potentials were not calculated in these cases, but Table 3F shows the frequency of occurrence of each residue in the different positions of type VIa and VIb turns (see Kinemages 4 and 5).

Multiple turns

Only 42% of the 3,899 β-turns identified occur in isolation (file Single.trn in the SUPLEMNT directory); the remainder are involved in multiple turns (file Multiple.trn) in that they have at least 1 residue in common with another turn (Table 4; Kinemage 2). Double turns are the most common multiple turns. Of these, the (I, I + 1) double turns, which consist of 5 residues such that residues 1–4 and residues 2–5 form turns, occur most frequently. An analysis of the classes of turn involved in the (I, I + 1) dou-

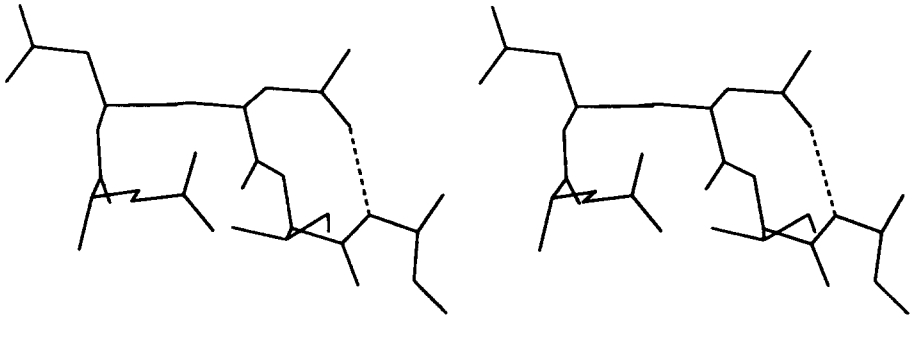

Fig. 3. Type VIII β-turn with asparagine at position $i + 2$. The side-chain Oδ_1 of Asn hydrogen bonds to the main chain of the first residue after the end of the turn (residue $i + 4$). The residues involved are 86–90 (Gln-Asp-Asn-Ile) from human lysozyme (Brookhaven code 1LZ1 [Artymiuk & Blake, 1981]). The plots were generated using MOLSCRIPT (Kraulis, 1991).

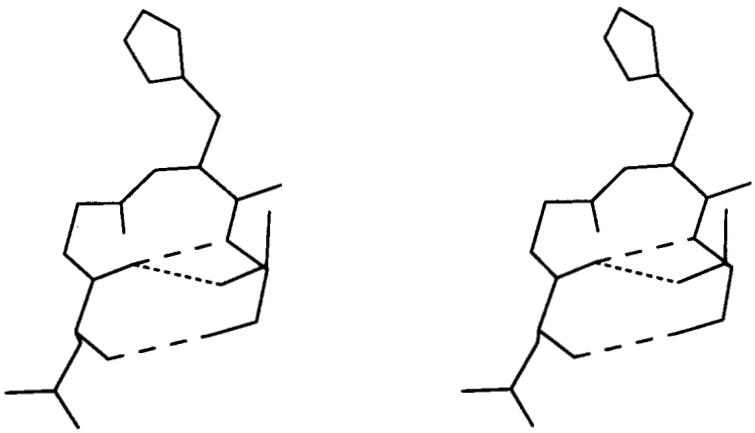

Fig. 4. Type II′ β-turn with threonine at position $i + 3$, illustrating the hydrogen bonding between the $O\gamma_1$ atom of the threonine residue and the main-chain oxygen of residue i (small dotted line). The hydrogen bond at the main-chain oxygen is bifurcated, as it also accepts a hydrogen bond from the main-chain N-H of the threonine (larger dashed line). The example is taken from residues 19 to 22 (Asn-Glu-His-Thr) from chain 2 of foot and mouth disease virus (Brookhaven code 1BBT [Acharya et al., 1989]). The plots were generated using MOLSCRIPT (Kraulis, 1991).

ble turns is shown in Table 5. Many combinations are disallowed because of the incompatibility of the ϕ,ψ angles of residue $i + 2$ of the first turn and residue $i + 1$ of the second turn. These are indicated by dashes in Table 5. For the other combinations, the expected number of examples, based on the total frequency of the turn classes (Table 1), is shown in parentheses after the number of observed examples. Three turn combinations occur with an unexpectedly high frequency (indicated by bold typeface in Table 5).

Double turns consisting of type IV overlapping with type I′ turns occur more than 10 times as frequently as expected. In the majority of these examples (59 of 63), the $i + 1$ residue of the type IV turn is in the β_E conformation, and these turns correspond to the distorted type II turn ($\beta_E\gamma$) previously described (Wilmot & Thornton, 1990). The vast majority of these distorted type II–I′ double turns occur in 2:2 β-hairpins (Fig. 5A; Kinemage 1); indeed most of the I′ turns observed to be part of a β-hairpin do not occur in isolation but form part of these double turns.

Double type I turns are quite common; these are similar to 2 successive turns of a 3_{10}-helix (Fig. 5B; Kinemage 3). As discussed earlier, proline is quite common at position i of a type I turn (Table 3A); 9 of the 76 examples with proline at this position occur as the second half of a double type I turn. The proline then acts additionally as the second residue ($i + 1$) of a type I turn, where it is preferred. Thus, the high frequency of double type I turns provides an explanation for the high occurrence of proline at position i of type I turns.

The third common class of double turns consists of the rare type VIa followed by type IV (Kinemage 4). This occurs 6 times as frequently as expected; in fact, 40% of the type VIa turns in the data set are involved in such a double turn. Unlike the other common classes of double turn, in this case there appears to be

Table 4. *Division of the data set into single and multiple turns*

Turn type	No. of examples	No. of single turns involved[a]	% of total no. of turns[b]
Single turns	1,632	1,632	41.9
Double turns			
(I, I + 1) type	436		
(I, I + 2) type	94		
(I, I + 3) type	112		
Total	642	1,284	32.9
Triple turns[c]			
(I, I + 1) type	60		
Others	127		
Total	187	561	14.4
Quadruple turns	53	212	5.4
5-Turns	27	135	3.5
6-Turns	9	54	1.4
7-Turns	3	21	0.5
Total multiple turns	921	2,267	58.1

[a] The equivalent number of single turns involved in the multiple turn (e.g., double turn is equivalent to 2 single turns).

[b] The percentage of the total number of turns that are involved in that type of multiple turn.

[c] For triple turns only the (I, I + 1) type is listed separately as the numbers of the other types are much smaller. For larger orders of multiple turns there are a large number of combinations, none of which occurs with very high frequency.

Table 5. *Occurrence of turn types in (I, I + 1) double turns*[a]

	I	I′	II	II′	VIa1	VIa2	VIb	VIII	IV
I	**82(43)**	–	–	–	–	–	–	7(11)	63(59)
I′	–	0(0)	–	–	–	–	–	–	2(6)
II	–	0(1)	–	–	–	–	–	–	22(19)
II′	2(3)	–	–	–	–	–	–	0(0)	4(4)
VIa1	1(1)	–	–	–	–	–	–	0(0)	**6(1)**
VIa2	0(0)	–	–	–	–	–	–	2(0)	1(0)
VIb	–	–	0(0)	–	0(0)	0(0)	0(0)	–	0(2)
VIII	–	–	–	–	–	0(0)	0(0)	–	4(15)
IV	44(59)	**63(6)**	0(19)	12(4)	0(1)	0(0)	0(2)	12(15)	109(80)

[a] The row indicates the turn type of the first 4 residues in the turn; the column indicates the turn type of the last 4 residues in the turn. Dashes indicate turn combinations that are disallowed because of conflicting ϕ, ψ criteria. The expected number of each allowed combination is indicated in parentheses. Values that are considerably higher than expected are indicated by bold typeface.

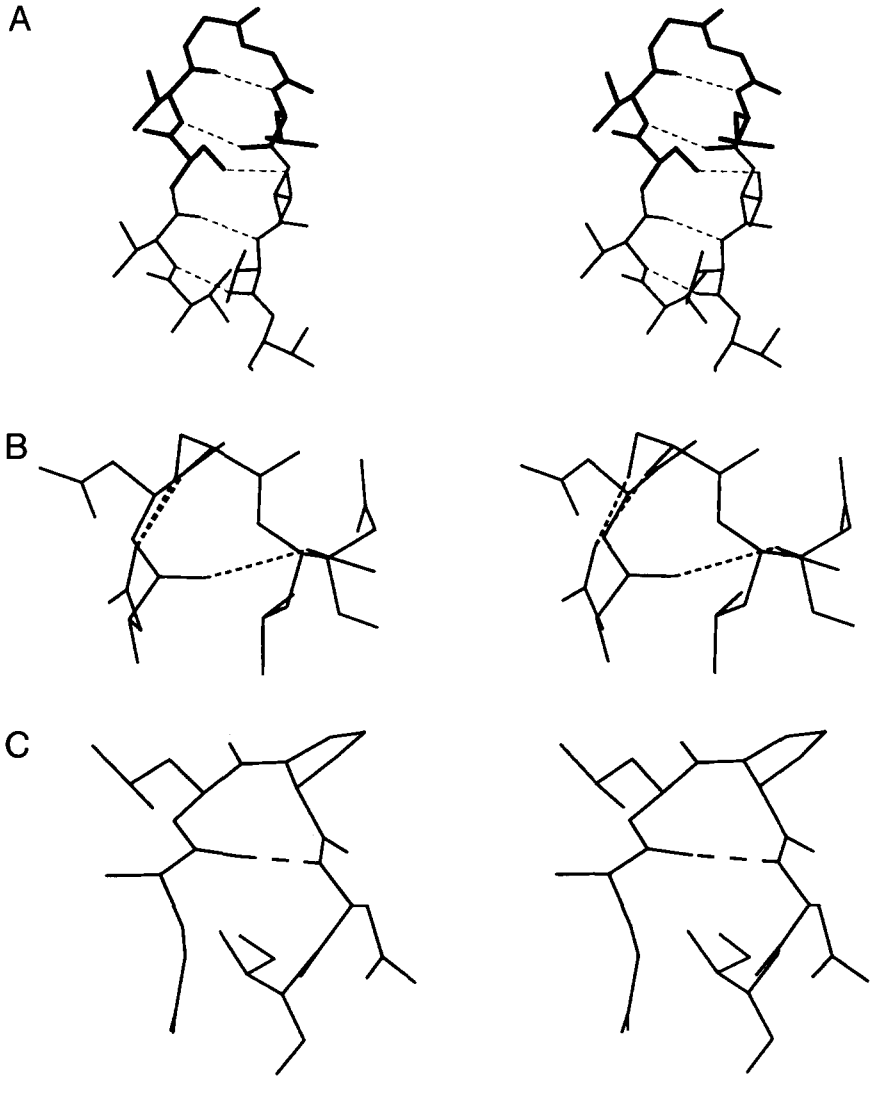

Fig. 5. Stereo pictures illustrating examples of the common types of double β-turns. The plots were generated using MOLSCRIPT (Kraulis, 1991). **A:** Type IV-I′ turn found in a β-hairpin. Residues 88–94 of endothiapepsin (Brookhaven code 2ER7 [Veerapandian et al., 1991]) are shown with the residues involved in the turn (residues 90–94) indicated by thicker lines. **B:** Double type I turn forming a "distorted 3_{10} helix." The example is taken from Bence Jones protein (Brookhaven code 2RHE [Furey Junior et al., 1983]) and involves residues 93–97. **C:** Type VIa–IV double turn from *Chironomus thummi thummi* erythrocruorin (Brookhaven code 1ECA [Steigemann & Weber, 1979]) residues 72–76.

no preferred conformation for the type IV turn, and these double turns are not involved in any single structural motif. Rather these double turns have variable conformation. All the type VIa turns have proline in the third position, which is the least favored position for a proline in turns. However, the second overlapping turn will have proline in the second position, which is the most favored position. Thus, formation of a double turn may stabilize the relatively unfavorable type VIa turn. An example of one of these double turns is shown in Figure 5C.

For the remaining classes of double turns, the degree of overlap is lower and there is no restriction on class imposed by the ϕ,ψ angles alone. Thus, it is not surprising that the combinations of turn class involved in these double turns show few departures from what would be expected based on the overall frequencies of turn classes in Table 1.

For triple and higher orders of multiple turn, the number of examples is much lower. The majority of these multiple turns are heterogeneous, consisting of various combinations of the (I, I + 1), (I, I + 2), and (I, I + 3) double turns; Kinemages 6 and 7 show two quite different cases. Only triple turns consisting of 2 overlapping (I, I + 1) double turns occur reasonably frequently.

Conclusions

The conformations and sequence preferences for each class of β-turn are summarized in Figure 6. Most of the preferences described can be rationalized in terms of preferred conformations of particular residues and the formation of hydrogen bonds or other interactions that stabilize the turn. For type I′ and II′ turns, some of the preferences for the end residues can be explained by the involvement of these turn types in β-hairpins. A number of preferences cannot be explained using these mainly local factors. This is not surprising because the formation of a particular turn type is also affected by long-range interactions within the protein and with the solvent (e.g., Mattos et al., 1994). The high frequency of multiple turns in the data set provides at least 1 example of turns being involved with more than the basic 4 residues. The turn propensities described here will be useful in turn prediction and in design of β-turn mimetics and novel

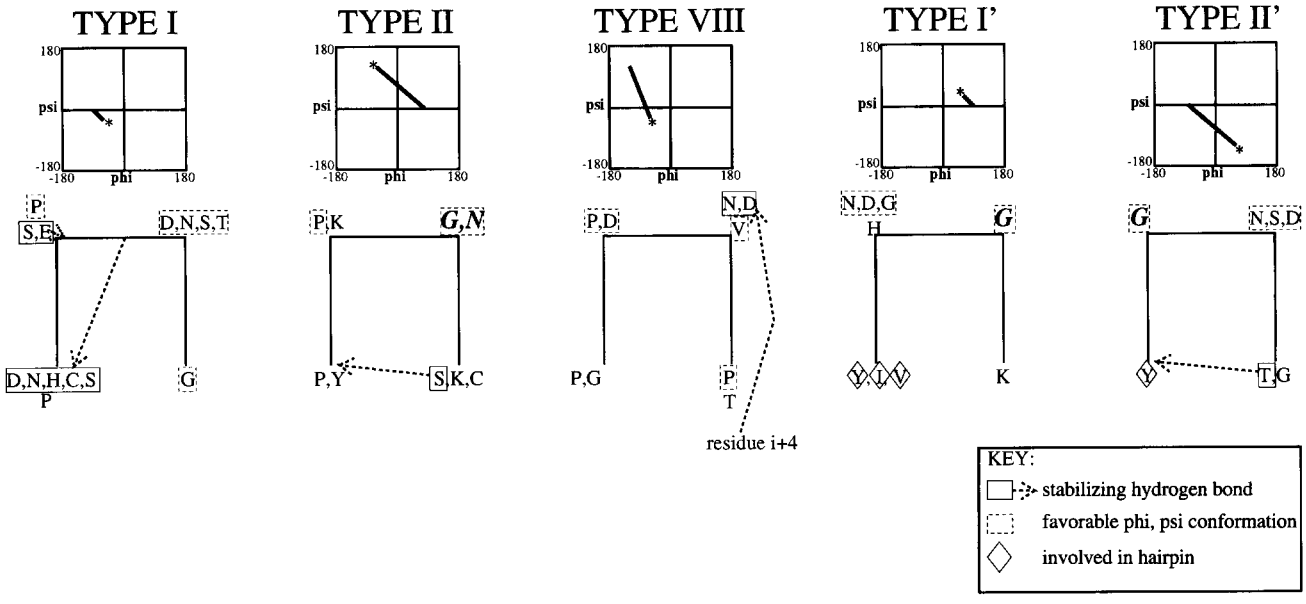

Fig. 6. Schematic diagram showing the ϕ, ψ conformations (Ramachandran plots in usual format with ϕ along the x-axis and ψ along the y-axis) and residue preferences for the main classes of β-turn. In the Ramachandran plots (top) the line connects the conformations of residues $i + 1$ (indicated by asterisks) and $i + 2$. Below each of the Ramachandran plots is a schematic β-turn, indicating sequence preferences. Where reasons for the preferences have been found, these are indicated by boxes around the corresponding residues as coded in the key. The stabilizing hydrogen bonds are indicated by dotted arrows linking the hydrogen bonded residues.

proteins. These are available by anonymous ftp from the /pub directory at IP address 128.40.46.39.

Methods

A nonhomologous data set of 205 protein chains from the April 1993 release of the Brookhaven Protein Data Bank (Bernstein et al., 1977) was used in the analysis; they are listed in file TurnPBB.lst in the SUPLEMNT directory on the Diskette Appendix. The protein chains were selected so that no 2 chains had more than 35% sequence identity. Structural analogues were eliminated from this data set using the structural alignment program SSAP (Orengo et al., 1992).

Main-chain dihedral angles and hydrogen bonds were calculated using a slightly modified version (D.K. Smith, unpubl. data) of the DSSP algorithm (Kabsch & Sander, 1983). β-Turns were defined and classified as in Wilmot and Thornton (1990). Turns were defined as 4 consecutive residues (i to $i + 3$) with the distance between C_i^α and $C_{i+3}^\alpha < 7$ Å, and where the central residues were not helical. Because some distorted helices may not be classified as helix by DSSP, turns whose central residues were assigned as helical by the crystallographers were also excluded. We subsequently eliminated "turns" that satisfied the above criteria, but where all 4 residues were assigned as β-sheet (by DSSP). As noted previously (Wilmot & Thornton, 1988), many of these form β-bulges and may have different sequence preferences to β-turns. The remaining turns were grouped into 7 classes using the ϕ and ψ angles of residues $i + 1$ and $i + 2$. These angles were allowed to vary by ±30° from the ideal values for each class, with one angle being allowed to deviate by as much as 45° (Lewis et al., 1973). The definitions used for the ϕ, ψ angles of the type VI turns were taken from Richardson et al. (1981). The two subclasses of type VIa turns identified by Richardson have been called VIa1 and VIa2 turns in this paper. In addition, for the type VI turns the $i + 2$ residue had to be a *cis*-proline. The miscellaneous category, type IV, contained all the remaining turns that were not assigned to any of the other categories. A second classification was done using less stringent tolerances of ±40 and ±50.

Average values were calculated for the ϕ and ψ values at the $i + 1$ and $i + 2$ positions in each of the turn types. The conformational potential for each residue (j) to be in a β-turn was calculated as

$$P_t(j) = \frac{f_t(j)}{\langle f_t \rangle},$$

where

$$f_t(j) - \frac{\text{number of residue } j \text{ in turns}}{\text{number of residue } j \text{ in proteins}}$$

and

$$\langle f_t \rangle - \frac{\text{total number of residues in turns}}{\text{total number of residues in proteins}}.$$

The positional potential for residue j at position i in turns was calculated as

$$P_{ti}(j) = \frac{f_i(j)}{\langle f_i \rangle},$$

where

$$f_i(j) = \frac{\text{number of residue } j \text{ at position } i \text{ of turns}}{\text{number of residue } j \text{ in proteins}}$$

and

$$\langle f_i \rangle = \frac{\text{total number of residues at position } i \text{ of turns}}{\text{total number of residues in proteins}}.$$

For each turn type (k) type-dependent positional potentials were calculated as

$$P_{tik}(j) = \frac{f_{ik}(j)}{\langle f_{ik} \rangle},$$

where

$$f_{ik}(j) = \frac{\text{number of residue } j \text{ at position } i \text{ of turn type } k}{\text{number of residue } j \text{ in proteins}}$$

and

$$\langle f_{ik} \rangle = \frac{\text{total number of residues at position } i \text{ of turn type } k}{\text{total number of residues in proteins}}$$

(adapted from Chou & Fasman, 1977). Type-dependent potentials were not calculated for turn types VIa and VIb because there were not sufficient data to do so. The statistical significance of the individual values was assessed using a d-test as in Wilmot and Thornton (1988). Significant preferences were investigated further by displaying individual examples on a computer graphics screen using Quanta™ and by calculating hydrogen bonds using the program HBPLUS (McDonald & Thornton, 1994).

The β-turns were also classified automatically according to whether they occurred as single turns or as part of a multiple (i.e., overlapping) turn. A turn was defined to be part of a multiple turn if it had at least 1 residue in common with another β-turn. The original classification of multiple turns (Isogai et al., 1980) was extended and used to classify double turns. (I, I + 1) double turns consist of 2 turns beginning at consecutive residues, i.e., beginning at residues i and $i + 1$. Correspondingly, the (I, I + 2) double turns consist of 4-residue turns beginning at residues i and $i + 2$. These were termed "successive turns" by Isogai et al. (1980). The (I, I + 3) double turns, not previously analyzed, consist of pairs of turns beginning at residues i and $i + 3$. Multiple turns of higher order, viz. triple, quadruple, etc. turns, were analyzed in terms of combinations of these basic types.

Acknowledgments

We thank Dr. Jane Richardson for useful discussions, particularly concerning type VI turns, and Christine Orengo for providing the list of non-homologous proteins. We acknowledge financial support from the SERC (E.G.H.).

References

Acharya R, Fry E, Stuart D, Fox G, Rowlands D, Brown F. 1989. The 3-dimensional structure of foot and mouth disease virus at 2.9 Å resolution. *Nature 337*:709–716.

Artymiuk PJ, Blake CCF. 1981. Refinement of human lysozyme at 1.5 Å resolution. Analysis of non-bonded and hydrogen bond interactions. *J Mol Biol 152*:737–762.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer based archival file for macromolecular structure. *J Mol Biol 112*:535–542.

Bode W, Papamokus E, Musil D. 1987. The high resolution X-ray crystal structure of the complex formed between subtilisin carlsberg and eglin c, an elastase inhibitor from the leech hirudo-medicinalis—Structural analysis, subtilisin structure and interface geometry. 2. *Eur J Biochem 166*:673–692.

Chou PY, Fasman GD. 1974. Conformational parameters for amino acids in helical, β-sheet and random coil regions calculated from proteins. *Biochemistry 13*:211–223.

Chou PY, Fasman GD. 1977. β-Turns in proteins. *J Mol Biol 115*:135–175.

Furey Junior W, Wang BC, Yoo CS, Sax M. 1983. Structure of a novel Bence Jones protein (rhe) fragment at 1.6 Å resolution. *J Mol Biol 167*:661–692.

Isogai Y, Leach SJ, Scheraga HA. 1980. Characterization of multiple bends in proteins. *Biopolymers 19*:1183–1210.

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

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Laskowski RA, Macarthur MW, Moss DS, Thornton JM. 1993. Procheck—A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Lewis PN, Momany FA, Scheraga HA. 1971. Folding of polypeptide chains in proteins. A proposed mechanism for folding. *Proc Natl Acad Sci USA 68*:2293–2297.

Lewis PN, Momany FA, Scheraga HA. 1973. Chain reversals in proteins. *Biochim Biophys Acta 303*:211–229.

Mattos C, Petsko GA, Karplus M. 1994. Analysis of 2 residue turns in proteins. *J Mol Biol 238*:733–747.

McDonald I, Thornton JM. 1994. Satisfying hydrogen bonding potential in proteins. *J Mol Biol 238*:777–793.

Orengo CA, Brown NP, Taylor WR. 1992. Fast structure alignment for Protein Data Bank searching. *Proteins Struct Funct Genet 14*:139–167.

Rees DC, Lewis M, Lipscomb WN. 1983. Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J Mol Biol 168*:367–387.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:167–339.

Rose GD, Gierasch LM, Smith JA. 1985. Turns in peptides and proteins. *Adv Protein Chem 37*:1–109.

Sibanda BL, Thornton JM. 1985. Beta-hairpin families in globular proteins. *Nature 316*:170–174.

Steigemann W, Weber E. 1979. Structure of erythrocruorin in different ligand states refined at 1.4 Å resolution. *J Mol Biol 127*:309–338.

Tainer JA, Getzoff ED, Beem KM, Richardson JS, Richardson DC. 1982. Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *J Mol Biol 160*:181–217.

Veerapandian B, Cooper JB, Sali MA, Blundell TL. 1991. X-ray analyses of aspartic proteinases. 3. 3-Dimensional structure of endothiapepsin complexed with a transition-state isostere inhibitor of renin at 1.6 Å resolution. *J Mol Biol 216*:1017–1029.

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

Wilmot CM, Thornton JM. 1988. Analysis and prediction of the different types of β-turn in proteins. *J Mol Biol 203*:221–232.

Wilmot CM, Thornton JM. 1990. β-Turns and their distortions: A proposed new nomenclature. *Protein Eng 3*:479–493.