Protein Science (1993), *2*, 1574–1590. Cambridge University Press. Printed in the USA.

Identification, classification, and analysis of beta-bulges in proteins

A.W. EDITH CHAN,[1,2] E. GAIL HUTCHINSON,[1] DANIEL HARRIS,[1] AND JANET M. THORNTON[1]

[1] Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, Gower Street, London WC1E 6BT, United Kingdom
[2] Italfarmaco-Research Center, via dei Lavoratori 54, Milan 22092, Italy

(RECEIVED February 24, 1993; ACCEPTED July 21, 1993)

Abstract

A β-bulge is a region of irregularity in a β-sheet involving two β-strands. It usually involves two or more residues in the bulged strand opposite to a single residue on the adjacent strand. These irregularities in β-sheets were identified and classified automatically, extending the definition of β-bulges given by Richardson et al. (Richardson, J.S., Getzoff, E.D., & Richardson, D.C., 1978, *Proc. Natl. Acad. Sci. USA 75*, 2574–2578). A set of 182 protein chains (170 proteins) was used, and a total of 362 bulges were extracted. Five types of β-bulges were found: classic, G1, wide, bent, and special. Their characteristic amino acid preferences were found for most classes of bulges. Basically, bulges occur frequently in proteins; on average there are more than two bulges per protein. In general, β-bulges produce two main changes in the structure of a β-sheet: (1) disrupt the normal alternation of side-chain direction; (2) accentuate the twist of the sheet, altering the direction of the surrounding strands.

Keywords: β-bulges; classification; proteins; protein structure

Beta-pleated sheets, being one of the two major structural elements found in globular proteins, are formed from two or more β-strands aligned side-by-side and hydrogen bonded. In each strand, the polypeptide chain is in an almost fully extended conformation, with typical ϕ, ψ torsion angles of $-122°$ and 143°, respectively (Richardson et al., 1978). The hydrogen bonds between adjacent strands are formed between the main-chain CO and NH groups.

The regular arrangements of the strands in β-sheets can be distorted by "β-bulges," which have been classified as follows: classic, G1, wide, parallel, and some pseudobulges (Richardson et al., 1978; Richardson, 1981). This classification was based largely on an examination of protein backbone drawings in the *Atlas of Molecular Structure on Microfiche* (Feldmann, 1976) and the hydrogen-bonding diagrams in published reports of solved X-ray structures. A β-bulge is defined as the region between two consecutive β-type hydrogen bonds, which include two residues on one strand opposite a single residue on the other strand.

Reprint requests to: Janet M. Thornton, Biomolecular Structure and Modelling Unit, Biochemistry and Molecular Biology Department, University College, Gower Street, London WC1E 6BT, United Kingdom.

β-Bulges, like β-turns, affect the directionality of β-strands, but in much less drastic a manner. The bulge is caused by the extra residues on the bulged strand, which increase the backbone length, thus causing the strand to bulge out of the plane of the sheet. At the same time, the twist of the sheet is slightly accentuated.

In this paper we present a systematic study and classification of β-bulges and aim to explain how and why the bulges are formed. All bulges are identified automatically using a computer program. Many more X-ray crystal structures have become available since the original classification was made. As a result, some new classes are introduced to add to the original bulge definitions.

Analysis

Before defining β-bulges, let us describe β-sheets in more detail. Adjacent β-strands can either run in the same direction (parallel β-sheet) or in opposite directions (antiparallel β-sheet). Their side chains extend above and below the sheet, with the C_α-C_β bond of each residue being approximately perpendicular to the plane of the sheet. The direction of these side chains alternates as one goes along

a strand but is in register on adjacent strands. The strands in a β-sheet are connected by hydrogen bonds.

Kabsch and Sander (1983) defined residues i and j, which form the appropriate hydrogen bonds between adjacent strands (or are covalently bonded to residues that form the required hydrogen bonds), and have termed this a bridge. There are two types of bridges: parallel and antiparallel. Their hydrogen-bonding patterns are depicted in Figure 1.

In antiparallel β-sheet, the hydrogen-bonding pattern between a given pair of strands has alternately wide and narrow spacing. This is because the hydrogen bonds from a given strand alternate between the two strands either side of it as one moves from one residue position to the next. Thus, for example, all the even-numbered residues will have hydrogen bonds from their CO and NH groups to one adjacent strand, while the odd-numbered residues will bond in the opposite direction to the other adjacent strand. In parallel β-sheet, on the other hand, the hydrogen bonds between the bridges are evenly spaced, going across the sheet at an angle to the strands.

For each protein, the main-chain hydrogen-bonding strengths and torsion angles were calculated using a program called SSTRUC (D.K. Smith, unpubl.). The patterns of hydrogen bonds were used to assign secondary structures using a modified implementation of the Kabsch and Sander (1983) algorithm, whereby a residue that lies at either end of a secondary structure is included if it has one "correct" hydrogen bond. In a β-bulge, the two residues on the bulged side are labeled "1" and "2", whereas the residue on the opposite strand is labeled "X" (Richardson et al., 1978). Here, the original definition of β-bulge has been extended to any irregularity in the hydrogen-bonding pattern of a β-sheet, where the regular pattern is disrupted by at most one extra residue on one strand and at most four extra residues on the other strand. The irregularities were classified into: classic, wide, bent, and special bulge types in parallel and antiparallel β-sheets. However, even this extended definition does not include cases where a bulge occurs outside the definition of β-sheet. Nevertheless, we have chosen to restrict the definition to cases within the β-sheets only.

The bulges were first classified by the number of extra residues in each strand. Then we determined whether they belonged to antiparallel or parallel β-sheets. Finally, the hydrogen-bonding pattern in the bulge residues was compared with the patterns illustrated in Figures 2–4. In addition, for classic bulges; the ϕ, ψ angles of residue 1 must fall into an α_R conformation.

The following examples illustrate the classification process for bulges in antiparallel β-sheets. If there is one extra residue on one strand between the two (perfect) bridges, but none on the adjacent one, then the bulge can be defined as classic or wide by the specific hydrogen-bonding patterns illustrated in Figure 2. If there is one extra residue on each strand (at the same location), then it is a bent bulge. Special bulges are defined to have between two and four extra residues on one strand.

Because residue 1 in a G1 bulge does not form part of a bridge, a different algorithm has been developed to search for the G1 bulges. First, we have excluded all the classic bulges because both G1 and classic bulges have identical hydrogen-bond patterns. Then we searched for all examples that match the hydrogen patterns illustrated for G1G

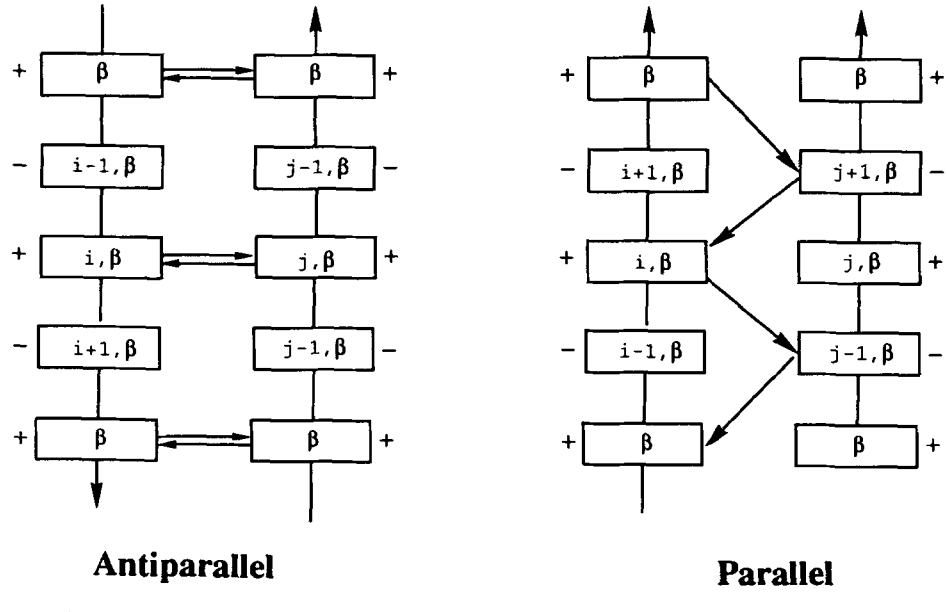

Fig. 1. Description of antiparallel and parallel β-sheet, with illustration of Kabsch and Sander's (1983) definition for bridge.

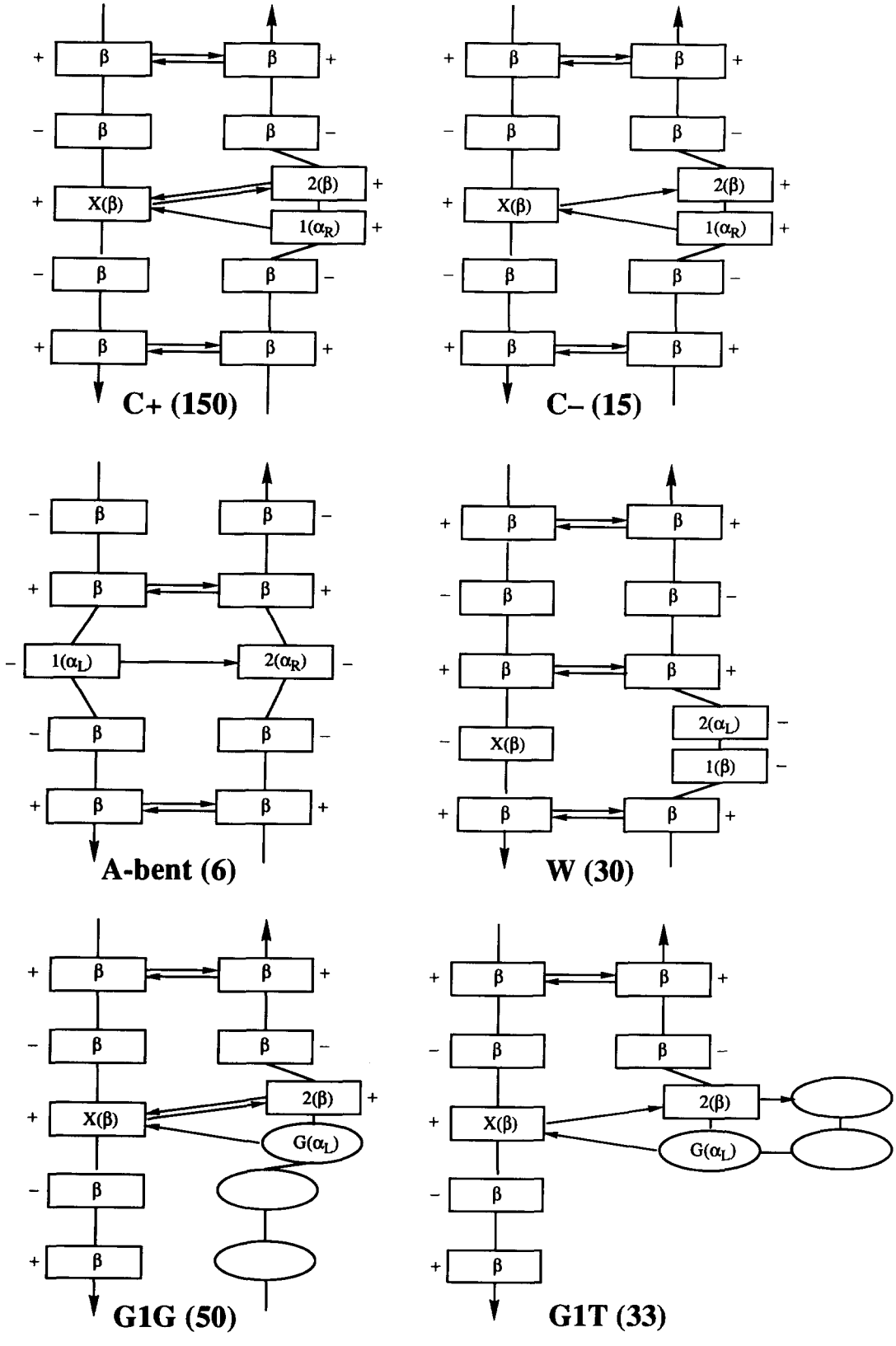

Fig. 2. Hydrogen-bonding diagrams for β-bulges in antiparallel β-sheets. The name of each subclass is listed underneath each figure. The number in parentheses denotes the total number of bulges found in that class. Arrows indicate an NH to CO hydrogen bond. Rectangles represent residues in β-strands, and ellipsoids correspond to any residue not in a β-strand. The conformation of each residue is indicated inside each box. Residues involved in the bulge are labeled as X, 1, and 2. The "+" and "−" signs indicate the orientations of the side chains into and out of the paper. G is glycine.

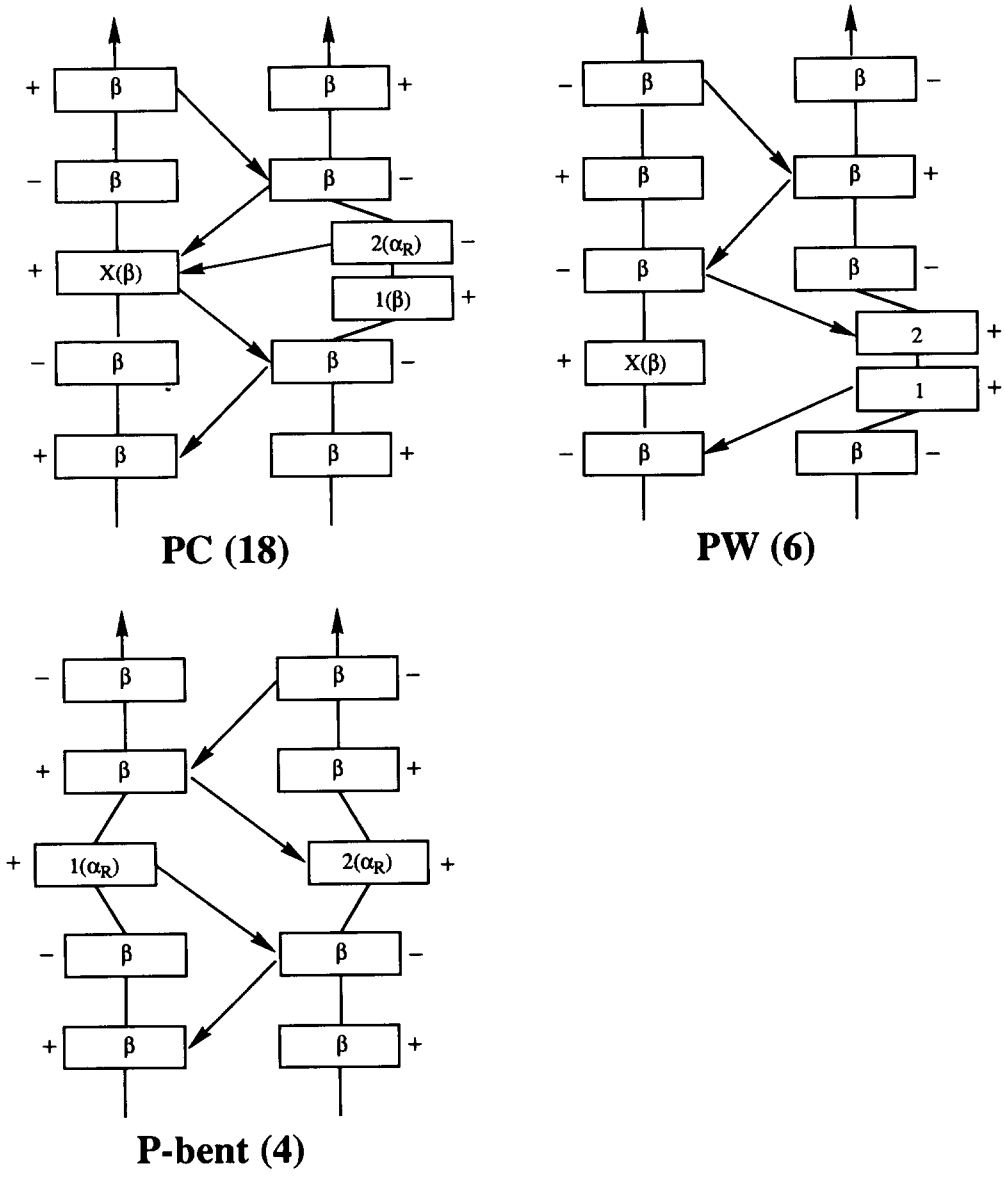

Fig. 3. Hydrogen-bond diagrams for β-bulges in parallel β-sheets.

and G1T in Figure 2, with the additional condition that residues X and 2 must form part of a bridge, but allowing residue 1 to have any conformation.

The hydrogen-bonding patterns were visualized schematically using diagrams produced by the HERA program (Hutchinson & Thornton, 1990), and the proteins themselves were examined visually using the QUANTA™ computer graphics program.

Results and discussion

Classification of β-bulges

Table 1 lists the numbers of β-bulges in each of the five main classes and their subclasses. Our data set included 362 bulges from 170 proteins (Table 2). This represents an average of greater than two bulges per protein in the data set. Most examples involve antiparallel strands, with fewer than 10% of β-bulges being between parallel strands even though the ratio of antiparallel to parallel β-sheets is 4.5 to 1 in our data set. Hence, β-bulges are more common in antiparallel sheets. All classes except the G1 were found in both parallel and antiparallel β-sheets, whereas the G1 class was found only at the edges of antiparallel β-sheets.

Almost all bulges occur at the edge of a sheet, with the bulged strand being the outermost one. The distortion produced by the bulge is most pronounced in the special, bent, and G1 bulges, and least pronounced in the classic and wide bulges. These distortions are demonstrated in Figure 5 and kinemages. The common structural effect of

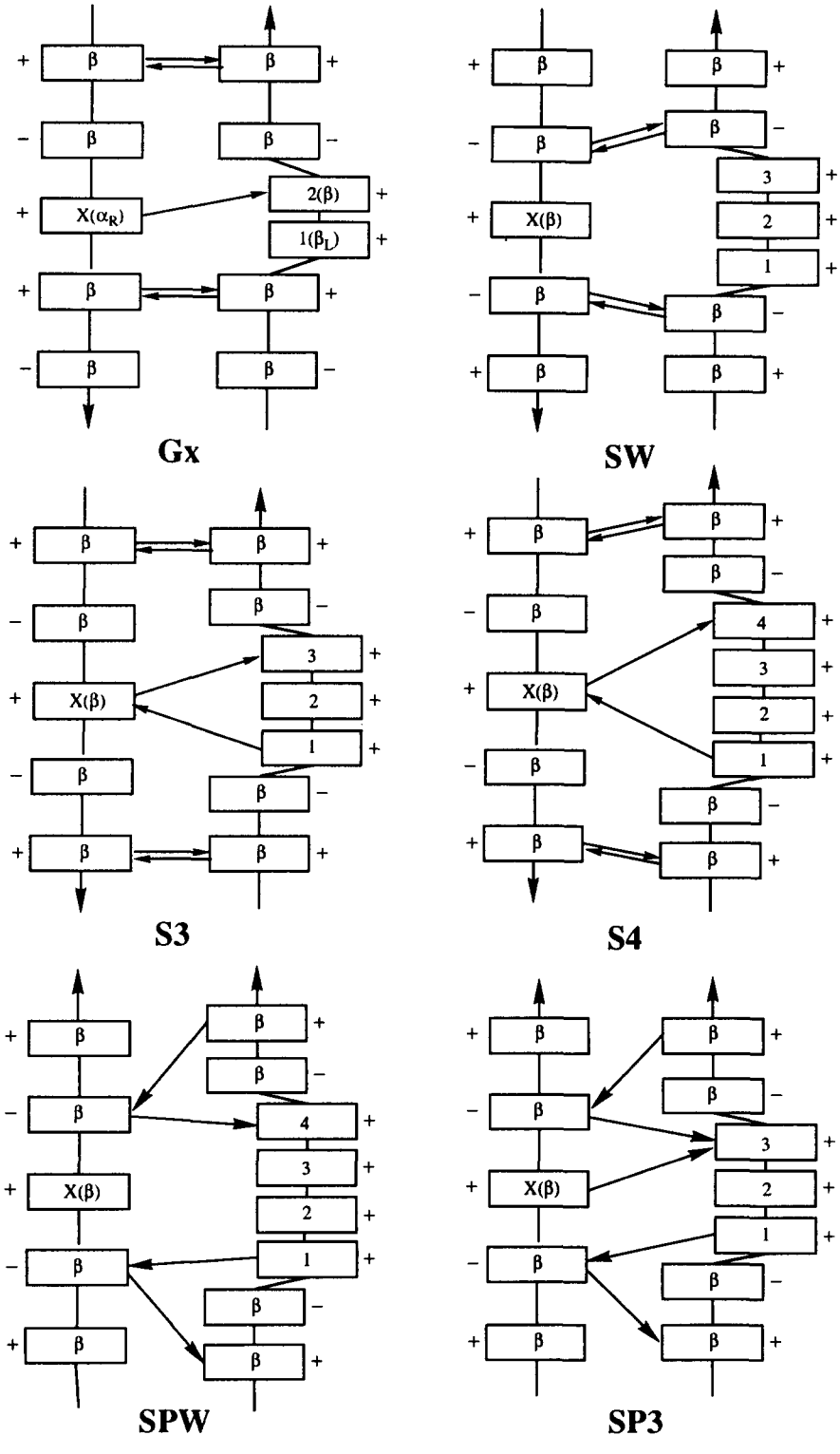

Fig. 4. Hydrogen-bond diagrams for special β-bulges in β-sheets. Residues involved in the bulge are labeled X, 1, 2, 3, and 4.

a β-bulge is to introduce a twist into the β-strand. The local twist is around 35° and 45° for a classic and a bent bulge, respectively. In a regular β-strand, the twist angle is around 10°.

The ϕ, ψ torsion angles for residues 1 and 2 in the bulge are always irregular; the other residues in the strands remain in the β conformation. Table 3 summarizes the average ϕ, ψ torsion angles observed at positions 1, 2, and X

Table 1. *Total numbers of bulges extracted*

	Antiparallel		Parallel		Total
Classic					187
	C+	150	PC	18	
	C−	15			
	Irregular	4			
G1					114
	G1G	50			
	G1T	33			
	G1A	25			
	G1AT	6			
Wide					36
	β-α_L	19	PW	6	
	α_L-β	11			
Bent					
	A-bent	6	P-bent	4	10
Special					15
	Gx	3			
	S3	4	SP3	1	
	S4	4			
	SW	1	SPW	2	
Total		331		31	362

for each of the different β-bulge types. Figure 6 shows the Ramachandran plots for these three positions in the most commonly occurring bulge types in antiparallel β-sheets: classic, wide, and G1. From these plots, the characteristic ϕ, ψ angles of each bulge could be easily identified. The different types of β-bulges will now be described in detail.

Classic bulges (antiparallel)

The most common bulge is the classic-type, which occurs both in parallel and antiparallel β-sheet. In antiparallel β-sheet, they are found between narrowly spaced pairs of hydrogen bonds. The 187 classic bulges identified, including parallel and antiparallel examples, represent 51% of the total number of bulges. In antiparallel β-sheet, there are two subclasses: C+ and C−, which differ only by one hydrogen bond. The C+ subclass appears more frequently (150 examples) and has NH(1) and NH(2) both hydrogen bonded to CO(X), and NH(X) bonded to CO(2) (Fig. 2; Kinemage 1). Thus the CO(X) accepts two hydrogen bonds. However, the strength of the NH(2) to CO(X) hydrogen bond (Kabsch & Sander, 1983) (-1.4 ± 0.6 kcal/mol) is weaker than that of the other two bonds, which have approx. equal strengths (around -2.4 ± 0.8 kcal/mol). These hydrogen bonds are just as strong as the narrow hydrogen bonds in undistorted antiparallel β-sheets, which have an average energy of -2.3 ± 0.8 kcal/mol. The second subclass, the C− (15 examples), lacks the weak NH(2) to CO(X) bond. The absence of this bond is compensated by one of the other two bonds being strengthened.

In general, both C+ and C− types behave similarly. Residue 1 adopts an approximate α_R helical conformation with ϕ around −96° and ψ around −31°. Residues 2 and X are in β conformation. The net effect of this is that the side chain of residue 1 is brought onto the same side of the β-sheet as both residues 2 and X. Hence, all three residues have their side chains pointing in the same direction.

Table 4 shows the amino acid preferences for each of the three positions. Position 2 favors small amino acids such as glycine, alanine, and serine, whereas position 1 favors large hydrophobic amino acids such as isoleucine, valine, and leucine. Position X has a preference for tryptophan, valine, arginine, and isoleucine. These preferences seem to indicate that a small residue is required at position 2 in order to minimize repulsive interactions between its side chain and those of the large hydrophobic residues at positions 1 and X. The preferred residues at position 1 and X are classical amino acids required for β-conformation, with the exception of tryptophan and arginine, which occur frequently at position X. There are eight cases where tryptophan is at position X, and of these, four occur in the immunoglobulin family (Brookhaven codes for the entries are 2RHE, 2FB4, 1CD4, 1CD8). There are nine examples of arginine in position X. In five of these cases, arginine uses its side chain to form a salt-bridge with another bulged residue (either position 1 or 2).

Four examples fitted into neither the C+ nor the C− subclasses and have been classified as irregular. One was in Rubisco (9RUB; Lundqvist & Schneider, 1991), where the hydrogen bonds are from NH(X) to CO(2), NH(X) to CO(1), and NH(1) to CO(X). The other three examples (in human class 1 histocompatibility antigen [3HLA], immunoglobulin FC [1FC2], and barnase [1RNB]) all have glycine at position 1. Their hydrogen-bonding pattern is exactly that of a classic bulge, but their ϕ, ψ angles are all positive, in the α_L conformation.

G1 class (antiparallel)

The next most common class is the G1 type (117 examples; 33% of total, see Table 1), which occurs only in antiparallel β-sheets. Position 1 adopts α_L conformation. Residues 2 and X are both in β conformation but with ϕ centered at −90° (compared to the more usual −122°) and ψ centered at 150° (compared to the more usual 143°). They have greater distortions in their ϕ and ψ torsion angles than the classic bulges (see Table 3). In our study, the G1 class contains four subclasses, as follows: G1G, glycine at position 1; G1T, glycine at position 1 and a Type I′ or Type II β-turn between position 2 and another residue, with glycine at position $i + 2$ of the turn; G1A, any amino acid (excluding glycine) at position 1; G1AT, any amino acid (excluding glycine) at position 1, plus turn as in G1T.

Table 2. *Bulges extracted from the data set (these data also appear on the Diskette Appendix)*

PDB	1	2	X	PDB	1	2	X	PDB	1	2	X
Classic – C+											
1BBP(A)	L106	S107	I115	1PII	S432	Q433	I436	2RHE	L48	I49	W36
1BBP(A)	V30	A31	V134	1PRC(H)	A165	G166	V157	2SGA	G45	F46	L53
1BMV(1)	A106	C107	L95	1PRC(H)	L6	A7	L10	2SGA	L210	G211	L199
1BMV(1)	M11	A12	V162	1PRC(H)	R181	Y182	W172	2SGA	S214	G215	F227
1C2R(A)	I27	V28	I19	1PRC(H)	T169	D170	E184	2SGA	T163	G164	Q182
1CD4	I36	L37	W28	1PSG	D195	S196	T262	2SGA	T89	G90	I105
1CD8	L49	L50	W35	1PSG	I204	A205	I197	2SNS	I15	K16	K24
1COB(A)	L142	A143	V117	1RBP	A130	D131	L122	2STV	L128	K129	W89
1COX	T251	E252	T264	1RBP	M27	A28	V136	2STV	V68	S69	V191
1COX	V275	A276	Q267	1RBP	Q38	D39	R60	2TRX(A)	A87	A88	L80
1CTF	K59	A60	E116	1RBP	R121	L122	D131	2TS1	N14	Q15	T219
1CTF	L94	K95	V56	1RBP	V107	D108	V116	2TSC(A)	A148	F149	Y164
1F3G	K147	E148	K167	1RNB(A)[a]	G53	D54	E73	3B5C	I76	G77	V29
1F3G	V119	I120	M59	1RNB(A)	I109	R110	I96	3BCL	A272	G273	B276
1F3G	V163	I164	I22	1RVE(A)	C21	G22	I24	3BCL	B200	S201	R214
1FBP(A)	V160	A161	I138	1SNV	L231	G232	S243	3BCL	S316	Y317	L355
1FBP(A)	V196	D197	C183	1SNV	V227	A228	I219	3BCL	S323	S324	K331
1FC2(D)[a]	G371	F372	F404	1TGS(I)	Q51	K52	C24	3DFR	I13	G14	T126
1FC2(D)	G385	Q386	S383	1TNF(A)	L93	L94	I80	3DFR	V139	S140	V157
1FKF	F36	D37	G28	1TRB	E208	E209	R221	3FGF	E58	E59	V62
1FKF	I7	S8	R71	1TRB	N85	K86	N97	3GAP(A)	L61	S62	V49
1FKF	L104	K105	V23	1TRB	T218	G219	T211	3GRS	K252	E253	S264
1FNR	L43	L44	V61	1VSG(A)	E117	S118	S121	3GRS	V431	G432	V422
1FXI(A)	V52	S53	V86	1VSG(A)	Y214	V215	M210	3HLA(B)[a]	G29	F30	F62
1GCR	D38	S39	K2	2AZA(A)	I81	A82	L50	3LZM	L33	T34	Y25
1GCR	G165	S166	Y134	2CA2	I91	Q92	V121	4DFR(A)	V136	F137	I155
1GCR	H122	S123	Y93	2CA2	L57	R58	E69	4ENL	S3	K4	T24
1GCR	L127	E128	R89	2ER7(E)	F275	G276	F284	4ENL	V153	L154	F169
1GCR	N33	S34	Y6	2ER7(E)	K105	K106	S81	4PTP	F41	C42	L33
1GCR	R76	S77	Y45	2FB4(H)	V48	A49	W36	4PTP	I47	N48	W51
1GD1(O)	V168	R169	E245	2GLS(A)	S145	G146	S143	4PTP	Q210	G211	V199
1GP1(A)	V172	R173	L164	2HIP(A)	V41	Q42	W45	4PTP	S214	W215	V227
1HGE(A)	N248	G249	R201	2LTN(A)	L33	T34	H35	4PTP	S86	K87	K107
1HOE	R68	Y69	V36	2LTN(A)	V147	N148	I139	4SGB(I)	A45	Y46	Y16
1HOE	V56	G57	S21	2LTN(B)	L38	S39	T68	4SGB(I)	I23	C24	Y15
1LDM	S298	N299	V291	2MEV(1)	I43	G44	V238	5CPA	I38	G39	I47
1LFG	E648	C649	N414	2MEV(3)	L192	T193	K123	5TMN(E)	L202	R203	I188
1LFG	I305	G306	V81	2MEV(3)	L77	A78	V189	7API(B)	L383	F384	M374
1LFG	L247	A248	A94	2OVO	S51	H52	C24	7RSA	D121	A122	I107
1LFG	L591	A592	A437	2PAB(A)	E89	H90	V94	7RSA	V118	H119	A109
1LFG	V410	L411	V599	2PAB(A)	F44	A45	V32	8ADH	E74	S75	R37
1LFG	V77	A78	V255	2PLV(1)	V87	T88	V259	8ADH	V41	A42	A70
1NSB(A)	A94	L95	I444	2PLV(2)	L234	A235	A121	8DFR	I16	G17	T146
1NSB(A)	G246	I247	S243	2PLV(3)	I82	L83	V192	9API(A)	S330	K331	E354
1NSB(A)	H172	M173	L155	2PMG(B)	V533	M534	I500	9API(A)	V181	F182	A355
1NSB(A)	I261	K262	K254	2POR	I257	D258	I254	9PAP	D108	G109	V210
1NSB(A)	M399	V400	M376	2POR	L271	G272	A275	9PAP	D158	H159	L134
1NSB(A)	T210	D211	I203	2REB	I228	G229	E241	9PAP	V164	G165	L172
1NSB(A)	W176	S177	V192	2REB	I298	G299	Y291	9RNT	A87	G88	V79
1OVA(A)	L383	F384	I374	2REB	V238	G239	V231	9RUB[a]	A580	Y579	K490
1OVA(A)	V210	T211	E214	2REB	V247	K248	R222	9RUB(B)	Y71	E72	K81
1PHH	R214	S215	R218								
Classic – C–											
1ACX	V78	G79	G70	1NSB(A)	W407	Y408	E427	2FB4(H)	D106	Y107	R98
1ALD	K110	G111	Q125	1PYP	L139	G140	I155	2FB4(H)	K43	G44	A40
1COB(A)	L82	G83	F43	1RVE(A)	L107	G108	I189	2HIP(A)	G38	E39	R47
1F3G	G68	K69	S78	2CA2	K45	P46	G82	2MEV(1)	F222	G223	L71
1NSB(A)	S396	G397	L378	2ER7	T205	S206	T195	9RUB(B)	H123	D124	A26
G1G											
1ACX	G75	T76	T72	1PRC(H)	G73	G74	L70	3FGF	G38	R39	H35
1AKE(A)	G130	R131	H126	1PSG	G244	E245	N241	3FGF	G80	R81	K77
1BMV(1)	G37	G38	D33	1RBP	G127	T128	N124	3LZM	G23	Y24	D20
1C2R(A)	G24	T25	A21	1RBP	G51	Q52	D48	4ENL	G15	N16	D12
1COX	G272	N273	D269	1SNV	G224	R225	D221	4ICD	G272	K273	N268
1COX	G395	K396	N391	1TGS(I)	G28	I29	G25	4SGB(I)	G20	A21	S17
1F3G	G84	V85	S81	1UBQ	G10	K11	T7	5CPA	G44	R45	S41
1FBP(A)	G191	E192	D187	2CYP	G228	Y229	S225	5PTI	G28	L29	N24
1FC2(D)	G402	S403	D399	2ER7(E)	G78	S79	Y75	5TMN(E)	G15	D16	G12
1FKF	G33	K34	L30	2FB4(H)	G179	L180	Q176	8ACN	G492	K493	G489
1FNR	G125	E126	N122	2FXB	G31	I32	D28	8ACN	G720	T721	H717
1GLY	G407	D408	D403	2LIV	G338	T339	H335	8RXN(A)	G10	Y11	C6
1GP1(A)	G169	V170	G166	2PAB(A)	G22	S23	D18	9WGA(A)	G108	F109	S105
1HRH(A)	G462	R463	T459	2POR	G38	L39	T35	9WGA(A)	G151	S152	S148
1LAP	G323	K324	A320	2TSC(A)	G106	R107	T103	9WGA(A)	G22	Y23	S19
1LFG	G578	K579	C575	3BLM	G54	K55	D50	9WGA(A)	G65	H66	S62
1PRC(H)	G162	V163	A159	3FGF	G29	F30	C25				

(*continued*)

Table 2. *Continued*

PDB	1	2	X	PDB	1	2	X	PDB	1	2	X
G1T											
1COB(A)	G112	R113	I147	1PAZ	G28	D29	V68	2PMG(B)	G300	F301	G297
1F3G	G110	Q111	I67	1PCY	G24	E25	L74	2SAR(A)	G34	V35	E54
1F3G	G116	D117	A61	1PCY	G67	E68	N31	2SGA	G133	Q134	V162
1FKF	G19	Q20	L50	1PRC(H)	G154	L155	V168	2SGA	G19	E29	L44
1FKF	G69	Q70	L103	1SNV	G216	R217	V230	2SGA	G196	G197	T213
1FNR	G142	A143	C42	1TNF(A)	G129	D130	V50	2STV	G169	A170	D92
1HGE(A)	G240	D241	N170	2AZA(A)	G90	E91	H35	3BLM	G114	K115	I97
1HOE	G51	Q52	N25	2CA2	G171	K172	N61	3GAP(A)	G67	D68	I42
1LAP	G315	D316	V328	2CPP	G315	D316	L301	4PTP	G133	T134	I162
1LAP	G49	K50	G68	2LIV	G243	L244	W334	4PTP	G196	G197	V213
1MSB(A)	G202	L203	V199	2PMG(B)	G297	K298	F301	8ADH	G86	D87	V73
G1A											
1CSE(I)	N61	V62	N57	1TNF(A)	N46	Q47	R44	2RHE	D97	E98	N93
1HRH(A)	K451	L452	N447	2CYP	N219	N220	N216	3FGF	N102	Y103	E99
1LAP	N81	W82	D77	2I1B	N129	M130	A28	3FGF	N71	R72	G67
1LDM	C265	R266	L292	2LTN(A)	N171	V172	N167	3HLA(B)	W60	S61	S57
1LFG	N234	T235	C231	2OVO	N28	K29	G25	7API(B)	K380	S381	E376
1LZ1	R50	S51	N46	2PAB(A)	D39	T40	A36	8ACN	Q611	E612	N607
1OVA(A)	N380	A381	H376	2PAB(A)	N27	V28	T49	8ATC(B)	E142	K143	C138
1PSG	D281	S282	D279	2REB	F255	K256	V246	9RNT	N84	Q85	N81
1RNB(A)	W94	L95	S91								
G1AT											
1CD4	Q165	K166	Q163	1VSG(A)	I140	G141	V143	2SGA	S174	G175	Y171
1GP1(A)	F160	E161	N40	2MEV(3)	D184	G185	V82	8DFR	D145	T146	G17
Wide (antiparallel)											
1ACE	L31	G32	N98	1PSG	T97	N98	T88	2PLV(2)	C257	C258	S388
1BBP(A)	C43	G44	H61	1RBP	N40	I41	G59	2PLV(2)	P236	L237	G120
1CTF	E96	G97	D55	1RNH	T40	R41	P19	2SGA	L165	N166	I181
1FBP(A)	P119	L120	T134	1RVE(A)	G165	V166	Y137	2SGA	Y178	G179	T168
1FNR	G149	P150	V76	2ER7(E)	T205	S206	T195	2STV	S34	G35	D185
1HGE(A)	Q132	N133	L154	2ER7(E)	T97	G98	T88	3GRS	P280	D281	E262
1LAP	G56	L57	P62	2FB4(H)	K148	D149	S182	5HVP(A)	D60	Q61	T74
1NSB(A)	E116	P117	A55	2MEV(1)	N244	M245	C125	8ACN	E553	D554	T698
1NSB(A)	E275	C276	C215	2MEV(3)	P194	L195	G122	8ATC(B)	D19	H20	L58
1PHH	E126	V127	F141	2PLV(1)	H265	I266	F130	8ATC(B)	L46	N47	D57
PC											
1ALD	I73	G74	G28	2LIV	V169	V170	I141	4MDH(A)	L64	K65	I35
1CSE(E)	V44	V45	L90	2TS1	C186	R187	T32	5CPA	F189	K190	A61
1GD1(O)	V27	V28	V3	2YHX	I204	K205	H387	8ACN	G656	R657	W630
1LAP	I252	T253	C303	2YHX	S183	V184	L133	8ACN	I350	R351	T440
1RVE(A)	Y128	I129	K86	4DFR(A)	V93	I94	I41	8ACN	I557	L558	V631
2GBP	Q175	L176	F143	4ENL	I291	V292	I243	8ADH	I346	T347	R369
PW											
1HGE(A)	T276	C277	N53	2PLV(2)	A29	N30	A490	8ATC(A)	A127	G128	R105
2GLS(A)	L332	A333	S342	2TSC(A)	G197	D198	S160	8DFR	D110	M111	N48
A-bent											
1PAZ	M16	N9		2PLV(1)	T126	C270		2PLV(3)	T108	L224	
2MEV(3)	A102	M221		2PLV(2)	Y100	G262		2TSC(A)	G31	G204	
P-bent											
1COX	N283	P13		1TRB	D106	K7		2TSC(A)	G204	R166	
1OVA(A)	R345	E195									
S2											
1GCR	G13	H14	E7	1GCR	G52	H53	E46	1GCR	G141	R142	E135

		X	1	2	3	4
Special						
1ACE	SPW	T110	D190	P191	K192	T193
1FBP(A)	SP3	C92	G111	K112	Y113	
1HGE(A)	S4	S145	N137	A138	C139	K140
1MSB(A)	S3	A215	N115	H116	E117	
1NSB(A)	S3	I113	H133	Y134	A135	
1NSB(A)	S4	G300	S401	M402	K403	E404
1PSG	S3	L84	T67	S68	Q69	
1RVE(A)	S3	G190	G178	D179	L180	
1VSG(A)	S4	G5	D182	N183	D184	A185
2ER7(E)	S4	V84	L66	S67	G68	A69
2PMG(B)	SPW	G59	G88	Q89	N90	G91
3LZM	SW	I27	R14	L15	K16	I17

[a] Irregular bulges.

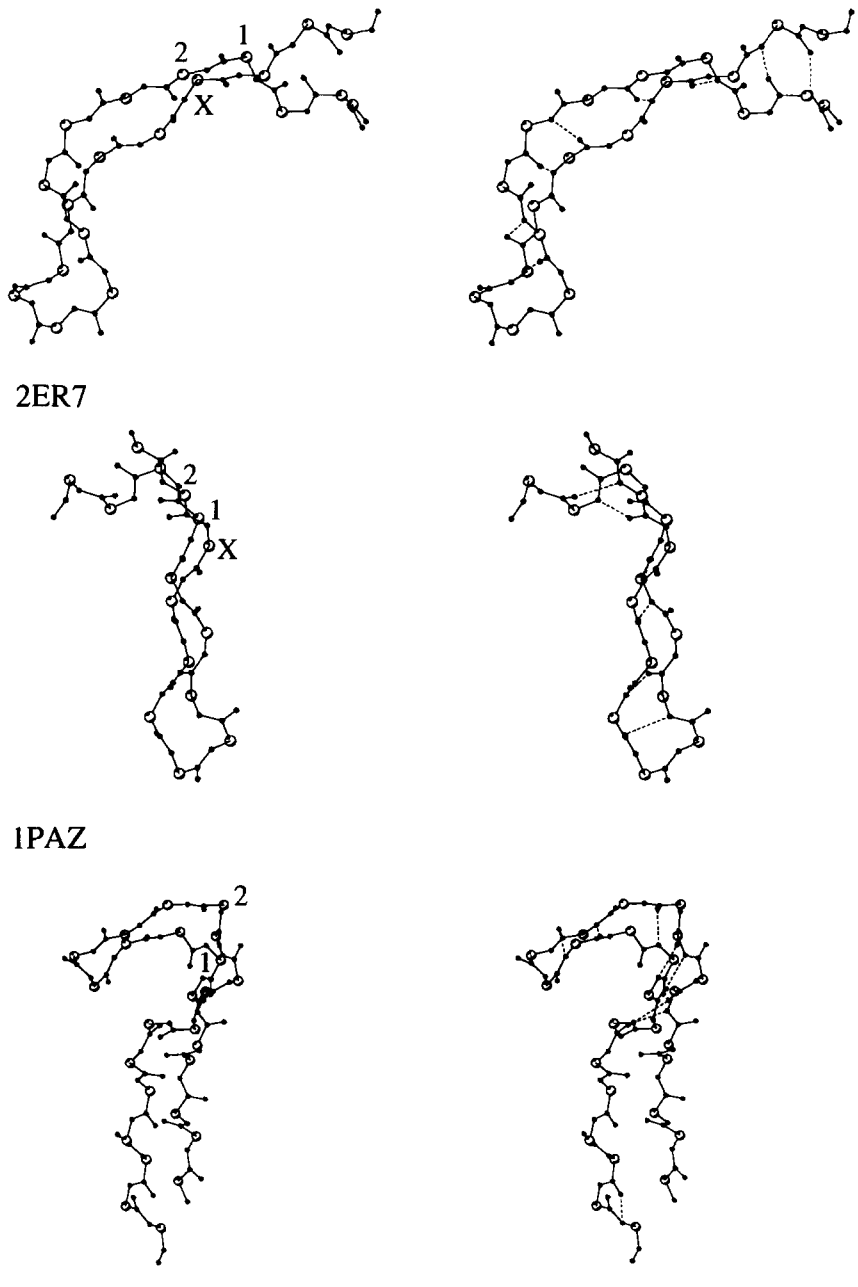

Fig. 5. Stereo pictures of (**A**) classic (2SGA; Sielecki et al., 1979), (**B**) wide (2ER7; Veerapandian et al., 1991), and (**C**) bent (1PAZ; Petratos et al., 1987) bulges. Only main-chain atoms are shown, with Cα positions given by the white spheres. Residues 1, 2, and X are marked in each case at the Cα position. Hydrogen bonds are indicated by dotted lines in the right-hand stereo picture. The illustrations were produced using the MOLSCRIPT program (Kraulis, 1991).

The only six examples of the G1AT subclass differ quite considerably from one another in conformation; therefore, no attempt was made to classify them further.

The hydrogen-bonding pattern for G1 bulges is similar to that of the classic bulges, except that residue 1 is always at the beginning of the β-strand (or at the end of a loop). The subclasses G1T and G1AT differ from the other G1 subclasses by virtue of the Type I′ or Type II turn involving the residue at position 2. The turn uses up one of that residue's available hydrogen bonds, namely that from the NH(2) group; hence, the NH(2) to CO(X) hydrogen bond is absent in these bulges. The plane of the turn and its hydrogen bond is almost perpendicular to the plane of the G1 bulge. A wide range of G1 bulges, along with β-turns, has already been studied by Milner-White (1987) because G1 bulges are usually associated with regions containing a β-hairpin turn (Sibanda & Thornton, 1985; Fujinaga & James, 1987; Sibanda et al., 1989). Milner-White proposed that G1 bulges be treated as a new sort of loop-turn. This combined structure has a consistent handedness that is constrained by the requirements of the three hydrogen bonds (see Kinemage 5). Also, ev-

Table 3. *ϕ, ψ angles for different classes of bulges*[a,b]

	1		2		X	
	ϕ	ψ	ϕ	ψ	ϕ	ψ
Antiparallel β-sheet						
Classic						
C+	−96.0 (32.4)	−31.3 (19.8)	−159.6 (25.5)	163.9 (78.7)	−114.0 (20.0)	128.3 (12.1)
C−	−110.5 (8.4)	−22.1 (47.0)	−200.5 (73.4)	160.4 (18.4)	−100.8 (29.9)	134.1 (23.8)
G1						
G1G	82.0 (39.0)	13.8 (26.7)	−96.3 (31.6)	146.3 (15.0)	−85.3 (16.7)	154.3 (27.5)
G1T	85.1 (37.3)	−10.6 (32.0)	−77.5 (18.6)	149.7 (17.0)	−82.2 (18.5)	127.9 (23.4)
G1A	64.9 (10.6)	26.6 (17.3)	−109.9 (29.4)	150.0 (11.6)	−94.0 (16.8)	145.4 (39.9)
Wide						
β-α_L	−106.9 (15.5)	154.7 (11.7)	58.2 (9.5)	44.6 (12.1)	−110.1 (23.1)	136.9 (10.0)
α_L-β	44.3 (18.9)	47.7 (18.2)	−73.7 (14.1)	138.0 (9.5)	−156.5 (19.8)	159.9 (15.0)
Parallel β-sheet						
PC	−84.3 (13.6)	117.4 (16.6)	−95.0 (21.0)	−23.0 (16.2)	−120.7 (20.5)	148.4 (19.3)
PW	−97.8 (8.2)	8.2 (96.1)	181.4 (75.6)	163.8 (21.3)	−102.9 (25.4)	99.7 (72.4)

[a] Each number represents the average value of the angles in a set of β-bulges.
[b] Number in parentheses denotes standard deviation.

Table 4. *Frequency of occurrence for the classic antiparallel bulges*

Amino acid	hd[a]	P_β[b]	1		2		X	
			$F_{aai/bulge}$	N_{aai}	$F_{aai/bulge}$	N_{aai}	$F_{aai/bulge}$	N_{aai}
I	0.73	1.60	1.60	21	0.23	3	1.60	21
F	0.61	1.28	0.50	4	0.75	6	0.38	3
V	0.54	1.65	1.72	31	0.17	3	1.89	34
L	0.53	1.22	1.96	27	0.65	9	0.94	13
W	0.37	1.19	0.41	1	0.41	1	3.29	8
M	0.26	1.67	0.88	3	0.58	2	1.46	5
A	0.25	0.97	0.89	9	2.26	23	0.79	8
G	0.16	0.81	0.20	2	2.74	28	0.10	1
C	0.04	1.30	0.28	1	1.13	4	0.85	3
Y	0.02	1.29	0.44	3	0.59	4	1.32	9
P	−0.07	0.62	0.00	0	0.00	0	0.00	0
T	−0.18	1.20	0.56	6	0.46	5	0.74	8
S	−0.26	0.72	1.17	11	2.02	19	0.64	6
H	−0.40	0.71	0.64	2	1.28	4	0.32	1
E	−0.62	0.26	1.15	7	1.15	7	1.15	7
N	−0.64	0.65	0.88	4	0.66	3	0.44	2
Q	−0.69	1.23	0.74	3	0.98	4	0.49	2
D	−0.72	0.80	0.87	5	1.39	8	0.17	1
K	−1.10	0.74	0.50	4	1.50	12	1.00	9
R	−1.80	0.90	0.90	5	0.90	5	1.61	9

[a] The scale of hydrophobicity was taken from Eisenberg et al. (1982).
[b] P_β is the propensity of a residue forming a β-structure (Chou & Fasman, 1974). The normalized frequency of occurrence for a particular amino acid inside a bulge ($F_{aai/bulge}$) was calculated by the following formula:

$$F_{aai/bulge} = (N_{aai}/N_{tot})/F_{aai/\beta},$$

where N_{aai} = total number of amino acid i found in the bulge region (i.e., position 1, 2, or X) from a particular class of β-bulge, N_{tot} = total number of a particular class of β-bulge, and $F_{aai/\beta}$ = percentage frequency of occurrence of amino acid i in β-sheet in this data set.

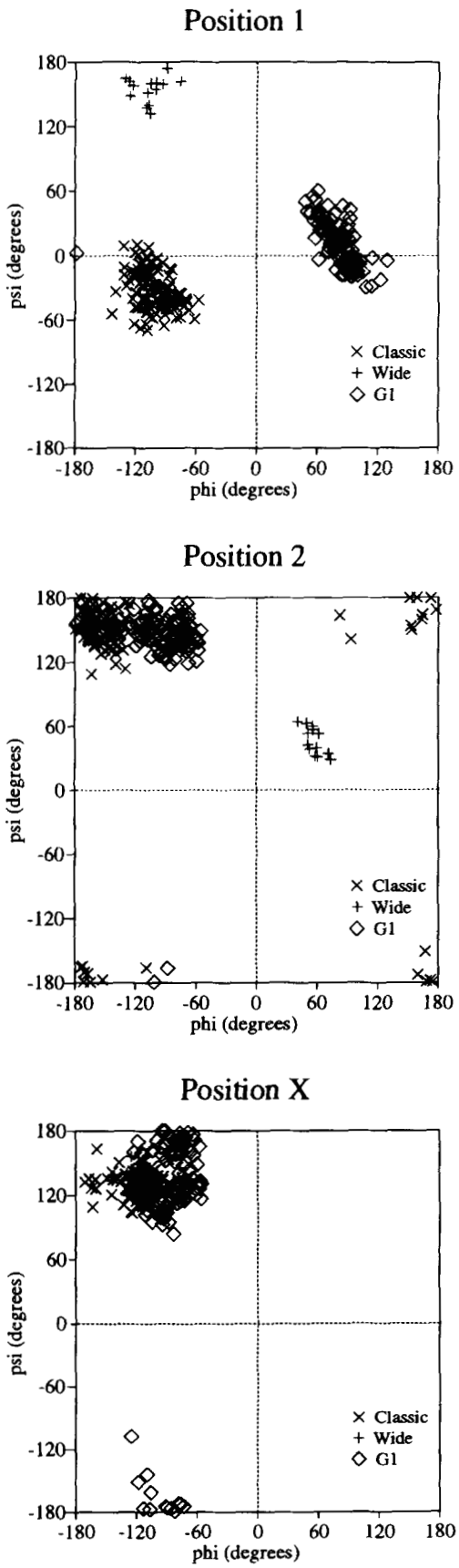

Fig. 6. Ramachandran plot of residues 1, 2, and X for classic, wide (α_L-β) and G1 bulges in antiparallel β-sheets.

ery one of the G1G examples is a bulged hairpin, with position 1 being located at either X + 3 (38 cases) or X + 4 (12 cases) along the chain (see Kinemage 4). This hairpin loop structure is very common (Sibanda & Thornton, 1985) and presumably represents a particularly stable conformation.

The first thing to note is that proline is entirely absent from G1 bulges. By far the most favored amino acid at position 1 is glycine (see Table 5). This is easily explained: in all G1 bulges, the residue at position 1 adopts an α_L conformation and this is best achieved by having a glycine residue at this position. Thus, the examples in the G1G and G1T subclasses (which are defined as those having glycine at position 1) outnumber the G1A and G1AT examples.

Bulges that do not have glycine at position 1 (i.e., subclasses G1A and G1AT) tend to favor asparagine instead. Asparagine has less main-chain flexibility than glycine, which is reflected in the statistics for the ϕ torsion angles at position 1 for these subclasses. For the G1G bulges, the mean value is around 82 ± 39°, whereas in the G1A bulges it is 65 ± 11°.

Position 2 favors charged groups, whereas position X favors asparagine or aspartic acid (see Table 5). We have looked at the side-chain hydrogen bonding in those bulges with a charged group at position 2. Fifty percent of them have their side chains hydrogen-bonded to the bulge residues at position X (which favors Asp or Asn). Forty percent have their side chains hydrogen-bonded to a residue around the bulged region, and the remaining 10% are hydrogen-bonded to some other residue in the protein. In G1A and G1AT bulges, the side chain of residue 1 is on the opposite side of the β-sheet from the side chains of positions 2 and X.

Wide class (antiparallel)

Wide bulges in antiparallel β-sheets are defined as those occurring between the widely spaced pairs of hydrogen bonds. This class can be subdivided into the β-α_L and α_L-β subclasses, describing the conformations of the residues at position 1 and 2. Because these residues are not involved in any main-chain bonding with other residues, they can adopt either the β or α_L conformations. The β-α_L-bulge is the more common of the two subclasses.

Because the residue at X is already involved in hydrogen bonding to another strand, it is not available to hydrogen bond with one of the reversed residues. The orientation of the side chains depends on the subclass (see Kinemage 2).

Given that there are only 30 examples of wide bulges, it is difficult to draw conclusions about the preferences of amino acids for positions 1, 2, and X. There are, however, some general rules. At position 2 (in the α_L conformation), no hydrophobic group is observed, and glycine,

Table 5. *Frequency of occurrence for the G1 bulges (G1G + G1A + G1T)*

Amino acid	hd[a]	P_β[b]	1		2		X	
			$F_{aai/bulge}$	N_{aai}	$F_{aai/bulge}$	N_{aai}	$F_{aai/bulge}$	N_{aai}
I	0.73	1.60	0.00	0	0.21	2	0.53	5
F	0.61	1.28	0.17	1	0.52	3	0.17	1
V	0.54	1.65	0.00	0	0.54	7	0.76	10
L	0.53	1.22	0.00	0	0.80	8	0.80	8
W	0.37	1.19	1.13	2	0.57	1	0.57	1
M	0.26	1.67	0.00	0	0.40	1	0.00	0
A	0.25	0.97	0.00	0	0.54	4	0.81	6
G	0.16	0.81	11.30	84	0.54	4	1.08	8
C	0.04	1.30	0.39	1	0.00	0	2.34	6
Y	0.02	1.29	0.00	0	1.01	5	0.20	1
P	−0.07	0.62	0.00	0	0.00	0	0.00	0
T	−0.18	1.20	0.00	0	1.02	8	0.90	7
S	−0.26	0.72	0.00	0	1.17	8	1.61	11
H	−0.40	0.71	0.00	0	0.44	1	2.65	6
E	−0.62	0.26	0.23	1	2.03	9	0.68	3
N	−0.64	0.65	3.95	13	0.91	3	5.47	18
Q	−0.69	1.23	0.34	1	3.04	9	0.34	1
D	−0.72	0.80	0.71	3	2.38	10	3.57	15
K	−1.10	0.74	0.34	2	2.59	15	0.17	1
R	−1.80	0.90	0.25	1	2.71	11	0.25	1

[a] The scale of hydrophobicity was taken from Eisenberg et al. (1982).
[b] P_β is the propensity of a residue forming a β-structure (Chou & Fasman, 1974).

asparagine, and aspartic acid occur most frequently. This is due to the ϕ, ψ angle preferences for these amino acids. Position 1 favors proline, aspartic acid, and glutamic acid, whereas position X has a high occurrence of threonine.

Parallel bulges: PC and PW

Bulges are less common in parallel β-sheets than in antiparallel ones. Two subclasses are defined for them: parallel classic (PC) and parallel wide (PW).

In the PC subclass, the residue at position 2 is hydrogen bonded to residue X, with NH(2) bonded to CO(X), and is in the α_R conformation. Both residues 1 and X remain in the β conformation, but with the ϕ angles clustered at approx. −84°. The distortion of residue 2 into the α_R conformation disrupts the usual side-chain alternation and results in residues 1 and 2 pointing in opposite directions. It also causes a shift in the bonding pattern of the residues preceding the one at position 2 and places two residues opposite one (see Fig. 3). In this subclass, all positions favor hydrophobic residues, but because there are only 18 examples, it is not possible to draw general conclusions about the amino acid preferences for these bulges. The PW subclass is similar to the antiparallel wide class in that it involves no hydrogen bonding between the residues of the bulge (see position 2 of PW in Fig. 3). Residues 2 and X are in the β conformation, whereas the ϕ torsion angle of residue 1 is around −98 ± 8°, and its ψ torsion angle varies between −120° and 120°. The net result is that if ψ has a negative value, then all three side chains are on the same side. When ψ has a positive value, side chains of 1 and X are on the same side opposite to that of 2. Figure 3 illustrates the first case. Only six examples of the PW subclass were found in the data set.

Bent-type

"Bent" bulges do not occur frequently in β-sheets. However, the distortion affects the β-sheet in quite a drastic manner. These bulges have an extra residue in each strand of the β-ladder (Figs. 2, 3). We have called these two residues 1 and 2. There is no X position in this class. In the antiparallel examples (Fig. 2; Kinemage 3), residue 1 donates a hydrogen bond to residue 2 of the opposite strand (NH(1) to CO(2)). There are three cases where NH(1) is also hydrogen-bonded to the residue preceding residue 2. The conformations of residues 1 and 2 are either α_L-α_R or α_R-α_L. In the cases of parallel sheets (Fig. 3), the normal hydrogen-bonding pattern is interrupted and the main-chain conformation of both "inserted" residues is α_R. This special combination of angles gives a bend or a curve to the β-ladder. Instead of one residue bulging out as in classic bulges, the whole ladder forms a hump. The twist angle at the bulged residue is 45° and is slightly larger than that observed in a classic bulge.

Special bulges

A total of six different kinds of special bulges were found in antiparallel and parallel sheets, although all of these are rare. One of them, labeled as "Gx", is the same as those found by Richardson et al. (1978). The others are labeled by a prefix "S", and their hydrogen-bonding diagrams are shown in Figure 4. Basically these bulges are very similar to the classic ones in both kinds of β-sheets, except that there are more extra residues in the bulged strand. The S3- and S4-types in the antiparallel sheets have three and four residues involved in the bulged region, respectively. The SW-type is similar to the antiparallel wide-type, but with two more residues involved at the bulge region. In these cases, the bulge region is very large and all these special bulges occur at the edges of β-sheets. The side chains of the bulge residues are more or less on the same side of the β-sheet. The main-chain conformations of the bulged residues are very irregular. For example, in mannose-binding protein (1MSB; Weis, 1991), the ϕ, ψ angles for position 1 are $-120°$ and 10°, respectively, and residue 2 is in the α_L conformation. In another example, influenza virus neuraminidase (1NSB; Burmeister, 1992), the ϕ, ψ angles for position 1 are $-90°$ and 13°, respectively, and position 2 is again α_L. Another type Gx bulge was only found in γ-crystallin (1GCR; Wistow et al., 1983). It can be viewed as two extra residues opposite one extra on the other strand (see Fig. 4), with a specific hydrogen bond from NH(X) to CO(2).

In parallel sheets, SP3 is similar to PC and SPW is similar to PW. As in the cases of antiparallel sheets, their ϕ, ψ angles are very irregular. For example, in acetylcholinesterase (1ACE; Sussman et al., 1991), the angles are as follows: X = (−51°, 135°), 1 = (−84°, 99°), 2 = (−59°, −12°), 3 = (−92°, 0°), 4 = (−148°, 75°).

Interconversion of classic and wide bulges

In principle, a classic bulge can interchange with a wide one by changing the conformation of one residue from α_R to α_L. The local twist produced by both types is very similar, and they superimpose well (Fig. 7; Kinemage 6). Residue 1 in the classic bulge is equivalent to residue 2 in the wide-type (Fig. 8). The classic bulge has an additional hydrogen bond between the residue marked A in Figure 8 and the shaded residue. If this hydrogen bond exists, and the shaded residue is in α_R conformation, a classic bulge is formed. Alternatively, if the hydrogen bond is not formed, and the shaded residue has an α_L conformation, then a wide bulge is formed. Because most amino acids prefer an α_R to an α_L conformation, classic bulges are more likely to be formed, as can be seen from the relative numbers of the two bulge types in our data set. However, as discussed earlier, residues that can adopt the α_L conformation (e.g., glycine, asparagine, and aspartic acid) are favored at position 2 of a wide bulge, and so the sequence at this position may determine the type of bulge formed.

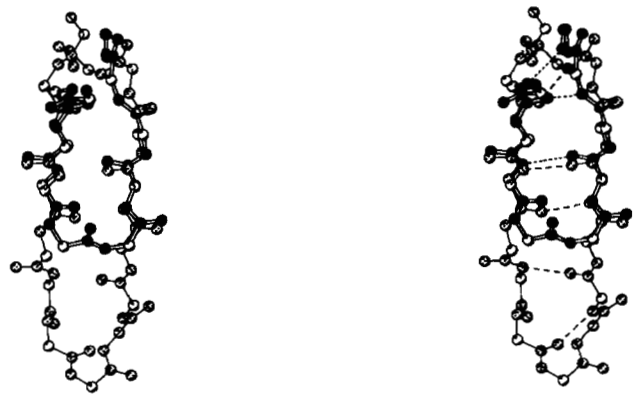

Fig. 7. Stereo pictures showing superimposed a classic (thick bonds and black non-Cα main-chain atoms) and a wide (thin bonds and gray non-Cα main-chain atoms) bulges. Both bulges are taken from 2SGA (Sielecki et al., 1979). Hydrogen bonds for the wide bulge are indicated by dashed lines in the right-hand picture, whereas those for the classic bulge are indicated by dotted lines.

Proline in β-bulges

It has been speculated that proline may be involved in the formation of β-bulges because it has no free NH group and therefore cannot be a hydrogen-bond donor. It therefore may distort a β-ladder. However, proline is only occasionally observed in β-bulges. In this data set, only 10 proteins have a proline residue in the bulge region, with eight of them found in the wide bulges. Even in wide bulges, proline can only occur in the position where it does not need to adopt the α_L conformation (see Fig. 9). The inability of proline to donate a hydrogen bond prevents its occurrence at positions 1, 2, or X in a C+ bulge. However, we have found eight proteins where proline immediately precedes position 1 in these bulges, which could be interpreted as one factor in bulge formation. The inability of proline to take part in the formation of a narrow hydrogen-bonding residue pair in antiparallel β-strands may cause subsequent residues in the strand to form a bulge.

Role of β-bulges

The effect of bulges on the directions of β-strands and the orientation of side chains can affect the positioning of crucial residues in a protein. Thus, bulges are thought to play an important biological role in proteins. In superoxide dismutase, for example, the strategic placement of two bulges at the end of each loop allows the two loops to peel away from the β-barrel and enclose the active site (Getzoff et al., 1989). The bulges in human lactoferrin were thought to be a common mode of entry to and exit

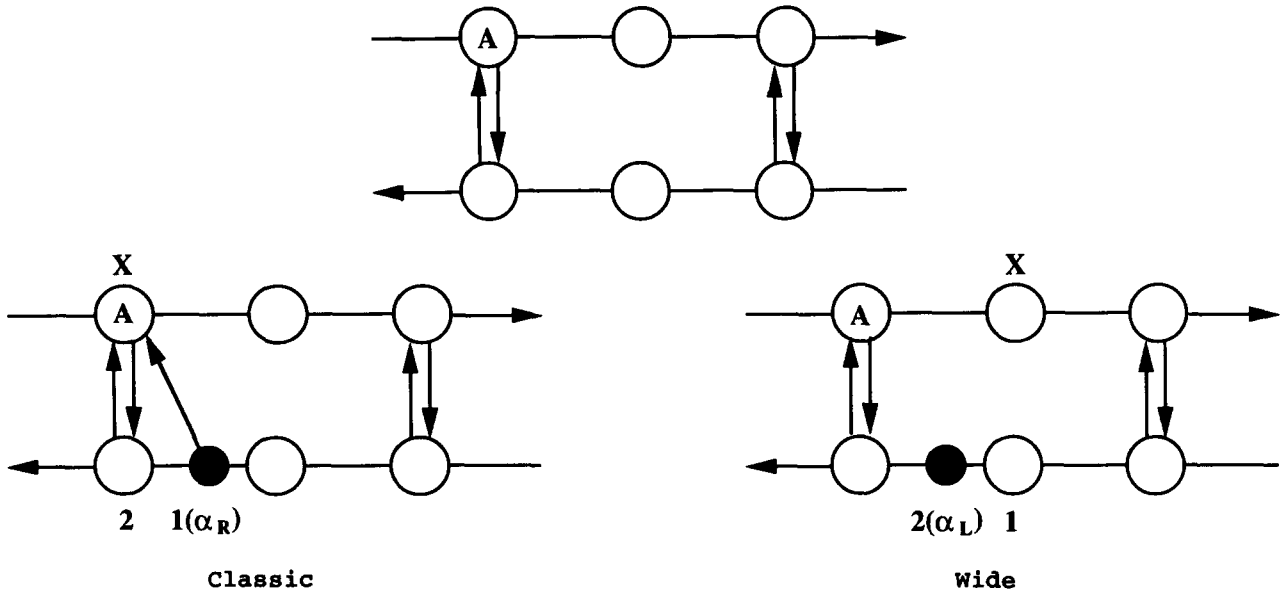

Fig. 8. Relationship between the classic and wide bulges in antiparallel sheets. Each circle denotes a residue.

from a β-sheet (Anderson et al., 1989), whereas in β- and α-crystallin, bulges at the ends of the *d* strands accentuate the local twist of the sheets and allow the strands to turn abruptly (Lapatto et al., 1991).

Because of these important functional roles, it is thought that β-bulges must be well conserved (Bajaj & Blundell, 1984; Chothia, 1984). This has been found in dihydrofolate reductase, where two bulges involved in positioning active-site residues are conserved across several species (Howell et al., 1990). The wide β-bulge found in the common cold virus, human rhinovirus 14, is conserved in amino acid composition from one viral capsid protein to another (Arnold & Rossmann, 1990). An alternative possible role for β-bulges is to accommodate insertions or deletions in β-strands, while preserving the hydrogen-bonding pattern in the remainder of the strand. In order to understand more about the function of β-bulges, we have looked at the immunoglobulins and aspartic proteinase[1] families,

[1] Immunoglobulins studied: 2RHE, 4FAB, 2HFL, 2IGF, 2MCP, 1REI, 2FB4, 1CD4, 1CD8; aspartic proteinases studied: 3APP, 2ER7, 4CMS, 5PEP.

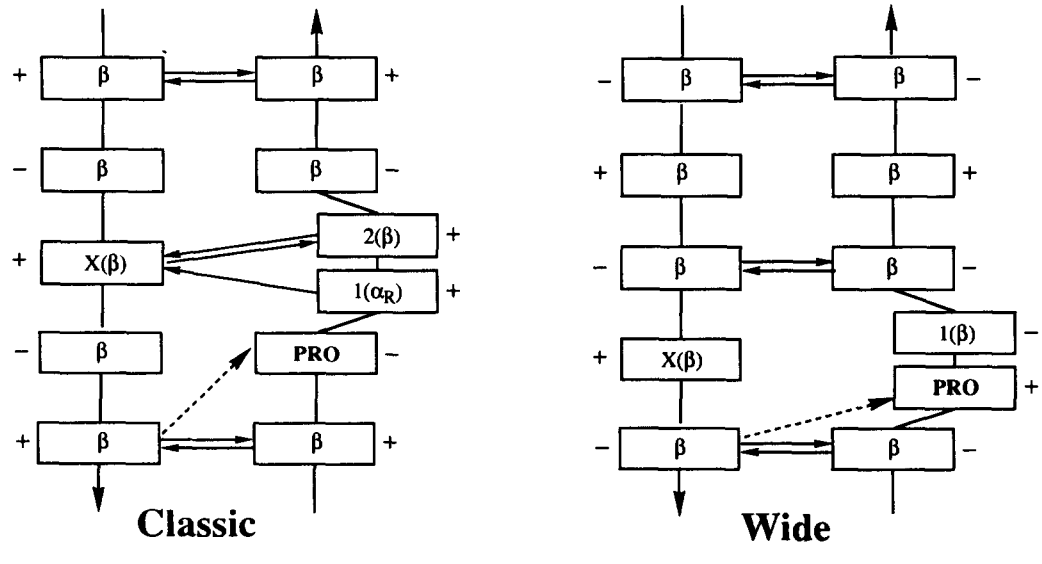

Fig. 9. Proline as a favorable element in wide bulges but not in classic bulges. The dashed arrow denotes a possible hydrogen bond that could form (for a non-Pro residue) in the absence of a β-bulge.

for which a reasonable number of crystal structures are known.

β-Bulges in the immunoglobulin family are quite well conserved. A C+ bulge was found in all nine proteins studied (Fig. 10). The sequence of the bulge is also conserved: residue X is always tryptophan, whereas residues 1 and 2 are hydrophobic groups (isoleucine or leucine). A wide bulge is also found in all proteins except CD4, and the sequence of this bulge is also conserved. These two bulges are believed to be involved in the dimerization of the two domains of the immunoglobulin family because the bulges give a twist to the corresponding β-strand (Chothia et al., 1985; Jones et al., 1992). All these proteins dimerize except CD4, and the absence of the wide bulge in CD4 may be responsible for its inability to do so.

Figure 11 shows the alignment of the four aspartic proteinases. A special β-bulge is found in all these proteins, but the subclass is not the same in each case. Pepsin (5PEP; Cooper et al., 1990) and chymosin (4CMS; Newman et al., 1991) have an SP3 bulge, whereas endothiapepsin (2ER7; Veerapandian et al., 1991) and penicillopepsin (3APP; James & Sielecki, 1983) have an SP4 bulge. The change in bulge type seems to accommodate the extra residue (glycine) in 2ER7 and 3APP. A wide bulge is conserved in another region of each of these proteins, though the residues involved are different in each case.

```
C+ Bulge:    X           12
2RHE 37 NSVIWYQQVPGKAPKLLIYYN
4FAB 36 TYLRWYLQKPGQSPKVLIYKV
2HFL 30 NYMYWYQQKSGTSPKRWIYDT
2IGF 31 TYLEWYLQKPGQSPKLLIYKV
2MCP 37 NFLAWYQQKPGQPPKLLIYGA
1REI 31 KYLNWYQQTPGKAPKLLIYEA
2FB4 30 STVNWYQQLPGMAPKLLIYRD
1CD4 24 IQFHWKNSN    QIKILGNQ
1CD8 33   CSWLFQPRGASPTFLLYLS

Wide Bulge: X              12
2RHE 84 EADYYCAAWNDSLDEPGFGGGTKLTVLG
4FAB 88 LGVYFCSQSTH  VPWTFGGGTKLEIKR
2HFL 82 AAEYYCQQWG   RNPTFGGGTKLEIKR
2IGF 83 LGVYYCFQGSH  VPPTFGGGTKLEIKR
2MCP 80 LAVYYCQNDHS  YPLTFGAGTKLEIKR
1REI 83 IATYYCQQYQS  LPYTFGQGTKLQIT
2FB4 82 ETDYYCAAWDVSLNAYVFGTGTKVTVLG
1CD4 79 SDTYICEVED        QKEEVQLLVFG
1CD8 88 NEGYYFCSALSNSIMYFSHFVPVFLPA
```

Fig. 10. Alignment of regions of the nine members of the immunoglobulin family showing classic and wide bulges. All sequences were from the light chain. Only the first residue is numbered. Bold type denotes bulges.

```
Special bulge:
position              1234          bulge-type
3APP    58 NPS  ATGKELSGYTWSISY        SP4
2ER7    57 TPSKSTTAKLLSGATWSISY        SP4
4CMS    57 DPRKSSTFQNLG KPLSIHY        SP3
5PEP    57 NPDDSSTFEATS QELSITY        SP3

Wide bulge:
position                X        12
3APP    76 GDGSSASGNVFTDSVTVGGVTAHGQAVQAA
2ER7    76 GDGSSSSGDVYTDTVSVGGLTVTGQAVESA
4CMS    76 GT GSMQGILGYDTVTVSNIVDIQQTVGLS
5PEP    76 GT GSMTGILGYDTVQVGGISDTNQIFGLS
```

Fig. 11. Alignment of regions of the four members of the aspartic proteinase family, showing special and classic bulges. Bold type represents the β-bulge. Only the first residue is numbered.

This brief analysis of two protein families shows examples of at least two possible roles for β-bulges in proteins (Richardson et al., 1978). Some bulges, which are important for the function of the protein, are very well conserved. Other bulges seem to play the role of accommodating insertions of a small number of residues within β-strands.

Conclusions

We have developed a program that can automatically extract and classify β-bulges according to their conformation and hydrogen-bonding patterns. Bulges occur frequently in proteins; on average there are more than two bulges per protein. This, together with the observation that hydrogen-bond energies in classic bulges and undistorted antiparallel β-sheets are very similar, suggests that a bulge must be a relatively low-energy conformation. This may be supported by a recent experimental study in which isolated tripeptides were found to adopt backbone torsion angles consistent with a classic β-bulge (Parthasarathy et al., 1993). The amino acid preferences for the β-bulges, summarized in Table 6, should be useful in assigning residues to electron-density maps and in modeling by homology.

Table 6. *Amino acid preferences for β-bulges*

Class	1	2	X
C+	Ile, Val, Leu	Gly, Ala, Ser	Trp, Val, Arg, Ile
Gl	Gly, Asn	Arg, Lys, Asp, Glu, Gln	Asn, Asp, His, Cys, Ser
Wide (anti)	Pro, Asp, Glu	Gln, Asn, Asp	Thr

In general, β-bulges produce two main changes in the structure of a β-sheet: (1) disrupt the normal alternation of side-chain direction; (2) accentuate the twist of the sheet, altering the direction of the surrounding strands. In some cases, these effects are clearly important for the function of the protein, and the bulges are conserved within a protein family. Other bulges occur when additional residues are inserted in a β-strand. However, there clearly is no simple reason for the occurrence of a bulge; rather, formation of bulges is the result of a complex interplay of sequential and environmental factors.

Data

For this study, a set of 182 protein chains (170 proteins) was extracted from the April 1992 release (including the prereleases) of the Brookhaven Protein Data Bank (Bernstein et al., 1977). The protein chains were selected such that no two have a sequence homology greater than 35%. The method used was that of Orengo et al. (1993), in which the proteins were first sorted by resolution and clustered into families using a pairwise homology matrix. The four-letter Brookhaven codes for the 182 protein chains in the data set are listed below, with the chain identity in parentheses (Brookhaven codes and protein names appear on the Diskette Appendix):

1ABP, 1ACE, 1ACX, 1AKE(A), 1ALD, 1AMT(A), 1BBP(A), 1BMV(1), 1CD4, 1C2R(A), 1CD8, 1COB(A), 1COL(A), 1COX, 1CRN, 1CSE(E,I), 1CTF, 1CY3, 1ECD, 1F3G, 1FBP(A), 1FC2(D), 1FKF, 1FIA(A), 1FNR, 1FXI(A), 1GCR, 1GD1(O), 1GLY, 1GMF(A), 1GOX, 1GP1(A), 1GST(A), 1HDD(C), 1HGE(A,B), 1HOE, 1HRH(A), 1ITH(A), 1LAP, 1LDM, 1LFG, 1LMB(A), 1LZ1, 1MBA, 1MBD, 1MSB(A), 1NSB(A), 1OVA(A), 1PAZ, 1PCY, 1PFK(B), 1PGD, 1PHH, 1PII, 1PPT, 1PRC(C,H,L), 1PSG, 1PYP, 1R69, 1RBP, 1RHD, 1RNB(A), 1RNH, 1ROP(A), 1RVE(A), 1SDH(A), 1SNV, 1TAB(I), 1TGL, 1TGS(I), 1THB(A), 1TNF(A), 1TPK(A), 1TRB, 1UBQ, 1UTG, 1VSG(A), 1WSY(A), 256B(A), 2AZA(A), 2CA2, 2CCY(A), 2CDV, 2CI2(I), 2CPP, 2CSC, 2CYP, 2ER7(E), 2FB4(H), 2FCR, 2FXB, 2GBP, 2GLS(A), 2GN5, 2HIP(A), 2HMZ, 2I1B, 2LH3, 2LIV, 2LTN(A,B), 2MEV(1,3,4), 2OVO, 2PAB(A), 2PLV(1,2,3,4), 2PMG(B), 2POR, 2REB, 2RHE, 2RSP(A), 2SAR(A), 2SCP(A), 2SGA, 2SIC(I), 2SNS, 2STV, 2TIM(A), 2TMV(P), 2TRX(A), 2TS1, 2TSC(A), 2WRP(R), 2YHX, 2ZTA(B), 3B5C, 3BCL, 3BLM, 3CHY, 3DFR, 3EBX, 3FGF, 3GAP(A), 3GRS, 3HLA(B), 3LZM, 3PGK, 3PGM, 3SDP(A), 451C, 4BP2, 4CPV, 4DFR(A), 4ENL, 4FD1, 4FXN, 4ICB, 4ICD, 4INS(B,C), 4MDH(A), 4PTP, 4SGB(I), 5CPA, 5CYT(R), 5HVP(A), 5PTI, 5TMN(E), 5TNC, 6RLX(A), 6XIA, 7API(B), 7RSA, 8ACN, 8ADH, 8ATC(A,B), 8DFR, 8RXN(A), 9API(A), 9PAP, 9RNT, 9RUB(B), 9WGA(A).

Acknowledgments

Italfarmaco partially supported this work and financed A.W.E.C. during her stay in London. We thank R.A. Laskowski for help and comments. We acknowledge financial support from SERC (E.G.H.).

References

Anderson, B.F., Baker, H.M., Norris, G.E., Rice, D.W., & Baker, E.N. (1989). Structure of human lactoferrin: Crystallographic structure analysis and refinement. *J. Mol. Biol. 209*, 711–734.

Arnold, E. & Rossmann, M.G. (1990). Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 Å. *J. Mol. Biol. 211*, 763–801.

Bajaj, M. & Blundell, T.L. (1984). Evolution and the tertiary structure of proteins. *Annu. Rev. Biophys. Bioeng. 13*, 453–492.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., 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.

Burmeister, W.P., Ruigrok, R.W.H., & Cusack, S. (1992). The 2.2 Å resolution crystal structure of influenza B neuraminidase and its complex with sialic acid. *EMBO J. 11*, 49–56.

Chothia, C. (1984). Principles that determine the structure of proteins. *Annu. Rev. Biochem. 53*, 537–572.

Chothia, C., Novotny, J., Bruccoleri, R., & Karplus, M. (1985). Domain association on immunoglobulin molecules—The packing of variable domains. *J. Mol. Biol. 186*, 651–663.

Chou, P.Y. & Fasman, G.D. (1974). Conformational parameters for amino acids in helical, β-sheet and random coil regions calculated from proteins. *Biochemistry 13*, 211–222.

Cooper, J.B., Khan, G., Taylor, G., Tickle, I.J., & Blundell, T.L. (1990). X-ray analyses of aspartic proteinases. 2. 3-Dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 Å resolution. *J. Mol. Biol. 214*, 199–222.

Eisenberg, D., Weiss, R.M., Terwilliger, T.C., & Wilcox, W. (1982). Hydrophobic moments and protein structure. *Faraday Symp. Chem. Soc. 17*, 109–120.

Feldmann, R.J. (1976). *Atlas of Macromolecular Structure on Microfiche*. Tracor-Jitco, Rockville, Maryland.

Fujinaga, M. & James, M.N.G. (1987). Rat submaxillary gland serine protease, tonin structure solution and refinement at 1.8 Å resolution. *J. Mol. Biol. 195*, 373–396.

Getzoff, E.D., Tainer, J.A., Stempien, M.M., Bell, G.L., & Hallewell, R.A. (1989). Evolution of CuZn superoxide dismutase and the Greek key β-barrel structure motif. *Proteins Struct. Funct. Genet. 5*, 322–336.

Howell, E.E., Booth, C., Farnum, M., Kraut, J., & Warren, M.S. (1990). A second-site mutation at phenylalanine-137 that increases catalytic efficiency in the mutant aspartate-27, refined at a resolution of 3.0 Å. *Biochemistry 29*, 8561–8569.

Hutchinson, E.G. & Thornton, J.M. (1990). HERA—A program to draw schematic diagrams of protein secondary structure. *Proteins Struct. Funct. Genet. 8*, 203–212.

James, M.N.G. & Sielecki, A.R. (1983). The structure and refinement of penicillopepsin at 1.8 Å resolution. *J. Mol. Biol. 163*, 299–361.

Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K., & Stuart, D.I. (1992). Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature 360*, 232–239.

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

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

Lapatto, R., Bax, N.B., Driessen, H., Lindley, P.F., Blundell, T.L., & Slingsby, C. (1991). High resolution structure of an oligometric eye lens β-crystallin. *J. Mol. Biol. 222*, 1067–1083.

Lundqvist, T. & Schneider, G. (1991). Crystal-structure of activated ribulose-1,5-bisphosphate carboxylase complexed with its substrate, ribulose-1,5-bisphosphate. *J. Biol. Chem. 266*, 12604–12611.

Milner-White, E.J. (1987). Beta-bulges within loops as recurring features of protein structures. *Biochim. Biophys. Acta 911*, 261–265.

Newman, M., Safro, M., Frazao, C., Kahn, G., Zdanov, A., Tickle, I.J., Blundell, T.L., & Andreeva, N. (1991). X-ray analyses of aspartic proteinases 4. Structure and refinement at 2.2 Å resolution of bovine chymosin. *J. Mol. Biol. 221*, 1295–1309.

Orengo, C.A., Flores, T.P., Taylor, W.R., & Thornton, J.M. (1993). Identification and classification of protein fold families. *Protein Eng. 6*, 485–500.

Parthasarathy, R., Go, K., & Chaturvedi, S. (1993). Helix-forming tendencies of amino acids depend on their sequence contexts: Tripeptides AFG and FAG show incipient β-bulge formation in their crystal structures. *Biopolymers 33*, 163–171.

Petratos, K., Banner, D.W., Beppu, T.A., Wilson, K.S., & Tsernoglou, D. (1987). The crystal structure of pseudoazurin from *Alcaligenes faecalis* S-6 determined at 2.9 Å resolution. *FEBS Lett. 218*, 209–214.

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

Richardson, J.S., Getzoff, E.D., & Richardson, D.C. (1978). The β-bulge: A common small unit of nonrepetitive protein structure. *Proc. Natl. Acad. Sci. USA 75*, 2574–2578.

Sibanda, B.L., Blundell, T.L., & 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. & Thornton, J.M. (1985). Beta-hairpin families in globular proteins. *Nature 316*, 170–174.

Sielecki, A.R., Hendrickson, W.A., Broughton, C.G., Delbaere, L.T.J., Brayer, G.D., & James, M.N.G. (1979). Protein structure refinement: *Streptomyces griseus* serine protease A at 1.8 Å resolution. *J. Mol. Biol. 134*, 781–804.

Sussman, J.L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., & Silman, I. (1991). Atomic-structure of acetylcholinesterase from *Torpedo californica*—A prototypic acetylcholine-binding protein. *Science 253*, 872–879.

Veerapandian, B., Cooper, J.B., Sali, A., & Blundell, T.L. (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.

Weis, W.I., Kahn, R., Fourme, R., Drickamer, K., & Hendrickson, W.A. (1991). Structure of the calcium dependent lectin domain from a rat mannose binding protein determined by MAD phasing. *Science 254*, 1608–1615.

Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P., & Blundell, T. (1983). X-ray analysis of the eye lens protein γ2 crystallin at 1.9 Å resolution. *J. Mol. Biol. 170*, 175–202.