Protein Science (1994), *3*:876–882. Cambridge University Press. Printed in the USA.

Structural characteristics and stabilizing principles of bent β-strands in protein tertiary architectures

CAROLA DAFFNER, GARETH CHELVANAYAGAM,[1] AND PATRICK ARGOS
European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 10.2209, 69012 Heidelberg, Germany

(RECEIVED February 7, 1994; ACCEPTED March 23, 1994)

Abstract

β-Strands as constituents of β-pleated sheets in protein tertiary structures often display considerable distortion from a purely extended conformation. The dislocation types are often characterized as "bulging," "twisting," and "bending." The former 2 properties have been extensively studied and classified. In this work an investigation of bent β-structures is undertaken. The structural characteristics examined included the bending angles within and out of the principal strand plane, their distribution among various strand types such as parallel and antiparallel, the amino acid preferences at bend sites, and the usage of charged and polar residues for stabilization through interactive anchoring with other atoms of the β-sheet within which the bent strand lies.

Keywords: β-sheets; β-strands; extended conformation; protein design; protein folding; protein structure

β-Pleated sheets in protein tertiary structures have been the subject of considerable study, especially given their major status as 1 of the 2 important secondary structural types along with α-helices. Sheets are composed of β-strands, which are individual oligopeptide segments that align in a side-by-side fashion and adjacently hydrogen bond through interaction of their main-chain NH and CO groups. Though strands are typically in extended conformation with backbone (ϕ, ψ) dihedral angles near (−122°, 143°) (Richardson et al., 1978), they often display considerable structural distortion characterized by "bulging," "twisting," and "bending" (Fig. 1).

β-Bulges were first extensively studied by Richardson et al. (1978) and more recently by Chan et al. (1993). Bulges are those peptide segments (generally 2 residues in length) that pucker out from an extended substructure between 2 consecutive β-type hydrogen bonds joining adjacent strands. The residues that splay out are opposite a single residue on the adjacent nonbulged strand (Fig. 1). Twisting is concerned with distortions from a classical flat β-sheet of hydrogen bonds such that the local extended strand chain dislocates (or twists) relative to the direction of its adjacent strand neighbor (Fig. 1). Salemme (1983) has described in detail such sheet properties. Both Chan et al. (1993) and Salemme (1983), through their thorough and systematic studies and classification schemes, have also aimed to explain the etiology of the distortions. Chothia and Janin (1981, 1982) have carefully examined the relative orientations of interacting β-pleated sheets.

In this work, similar efforts will be undertaken for bends where the extended strand conformation does not follow a line but rather curves (or bends) at discrete sites resulting in a uniformly bent or multiply contorted strand (Fig. 2). The significance and pervasiveness of bent strand segments in protein architecture is amply exemplified in a recent minireview of Chothia and Murzin (1993), who discuss many interesting and surprising all-β-stranded topologies. A particularly salient example relying heavily on bent β-substructures is the parallel β-helix of pectate lyases (Yoder et al., 1993). All the illustrations of newly discovered protein folds shown by Chothia and Murzin (1993) contain at least 1 strongly curved strand and generally many more.

The bent substructures in the present work are characterized by their bending angles within and out of the overall strand plane, their distribution among various strand types, the compositional preference for particular amino acid types at the point of bend, preferred positions of bends within strands, and the use of charged and polar residues at the bend for stabilization. The latter salt bridges and hydrogen bonds may well be responsible for the bend during protein folding. The results presented here should prove useful in designing and engineering protein 3-dimensional structures.

Results and discussion

Bend angle distributions

A total of 946 strands with length 4 residues or more was examined from the 247 protein tertiary structure data set. The

Reprint requests to: Patrick Argos, European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 10.2209, 69012 Heidelberg, Germany; e-mail: argos@embl-heidelberg.de.
[1] Present address: Department of Organic Chemistry, ETH-Zentrum, Zürich, Switzerland.

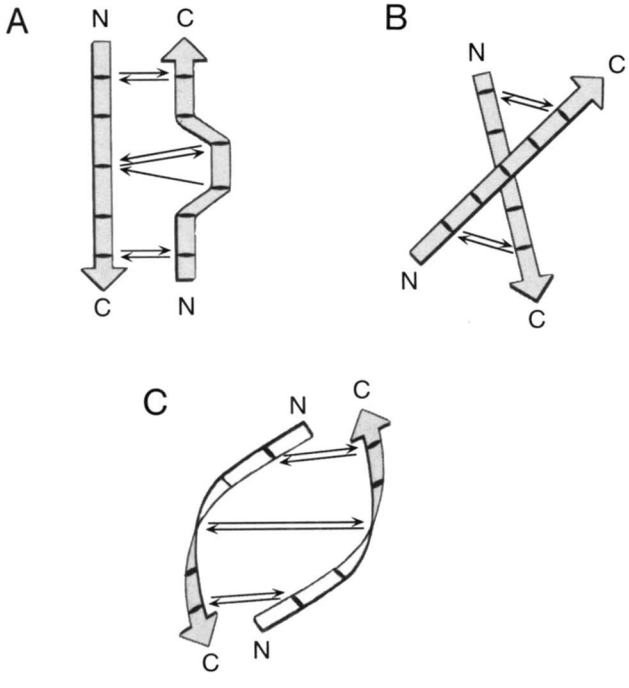

Fig. 1. Structural distortion of β-strands. β-Strands are indicated by thick arrows from the N- to C-terminus along the polypeptide chain sequence. Blackened ellipses illustrate successive C_α positions along the backbone, whereas thin arrows refer to an NH to CO hydrogen bond between main-chain atoms. **A:** A C+ 2-residue β-bulge in 1 of 2 antiparallel strands, which represents the most observed class (Chan et al., 1993). **B:** Two antiparallel strands twisted relative to each other (Salemme, 1983). **C:** Two bent antiparallel strands.

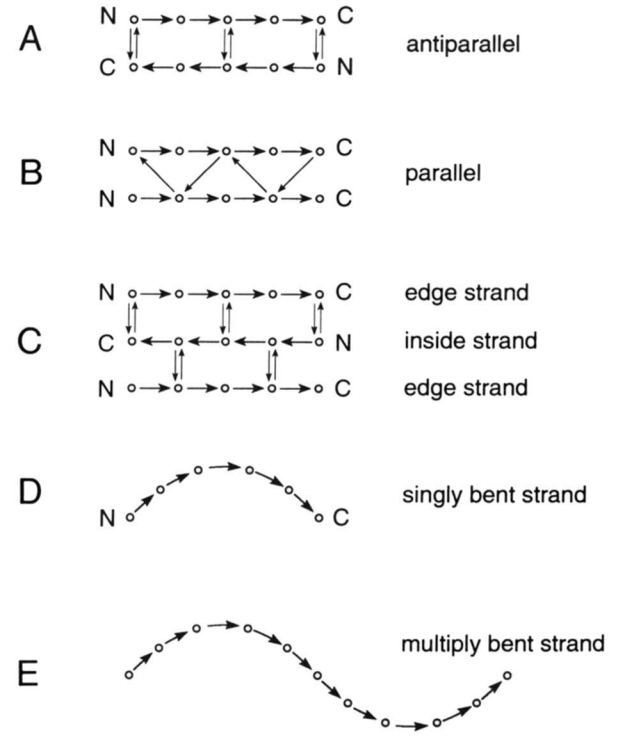

Fig. 2. Different strand types considered in this work. **A:** Antiparallel strand pair. **B:** Parallel strand pair. **C:** Inside and edge β-sheet strands. **D:** Singly bent strands. **E:** Multiply bent strands. In all cases, C_α atoms are indicated by open circles with arrows connecting successive atoms along the N- to C-terminus of the hypothetical polypeptide chain. Arrows across strands indicate hydrogen bonds (NH to CO).

strands were constituted by 4,466 amino acids. Table 1 lists the distribution of the extended segments among the various strand structural types. Antiparallel strands at 71% of the total dominate, as do strands that are without bulges (88%), and are uniformly bent (85%) and appear in sheets composed of more than 3 extended segments (73%). Strands with a total bend angle greater than 25° make up 31% of the sample, whereas only 10% have total bend angle greater than 50°. Strands with at least 1 local bend angle greater than 20° make up about 19% of the total strand count; if the angle is reduced to 15°, the fraction becomes 38%.

Table 1. *Distribution of strands in the data set composed of 946 members according to structural type*

Strand type	Fraction of sample (%)
Antiparallel	71
Parallel	29
Short (less than 5 residues)	53
Long (more than 5)	47
Inside	50
Edge	50
With bulges	12
Without bulges	88
Singly bent	85
Multiply bent	15
In large sheet (more than 3 strands)	73
In small sheet (3 or fewer)	27
Total bend greater than 25°	31
Total bend greater than 50°	10

The frequency distribution of total bend angles within the strand planes is shown in Figure 3A, and Figure 3C shows similar statistics for the total bend component out of the major strand plane. The 2 distributions correlate at the level of 0.65 according to Pearson statistics (Press et al., 1988:484–488). Thus, strands that bend strongly in the plane also tend to bend more out of the plane. Clearly, strands bend more in their planes, with the most frequent angle at 11°, rather than out of their planes where the most observed angle is 2°. Figure 3B and D relate similar distributions for local bend angles; however, there is much less correlation between strong local bend angles in and out of the strand plane (0.15).

Preferred structural characteristics of bent strands

Tables 2 and 3 list, respectively, the composition of strand structural types among extended segments with in-the-plane total bend angle greater than 25° and with at least 1 in-the-plane local bend angle greater than 20°. It is clear that strands that are highly distorted globally and locally are antiparallel, long (more than 5 residues), not bulged, uniformly bent, and in large sheet

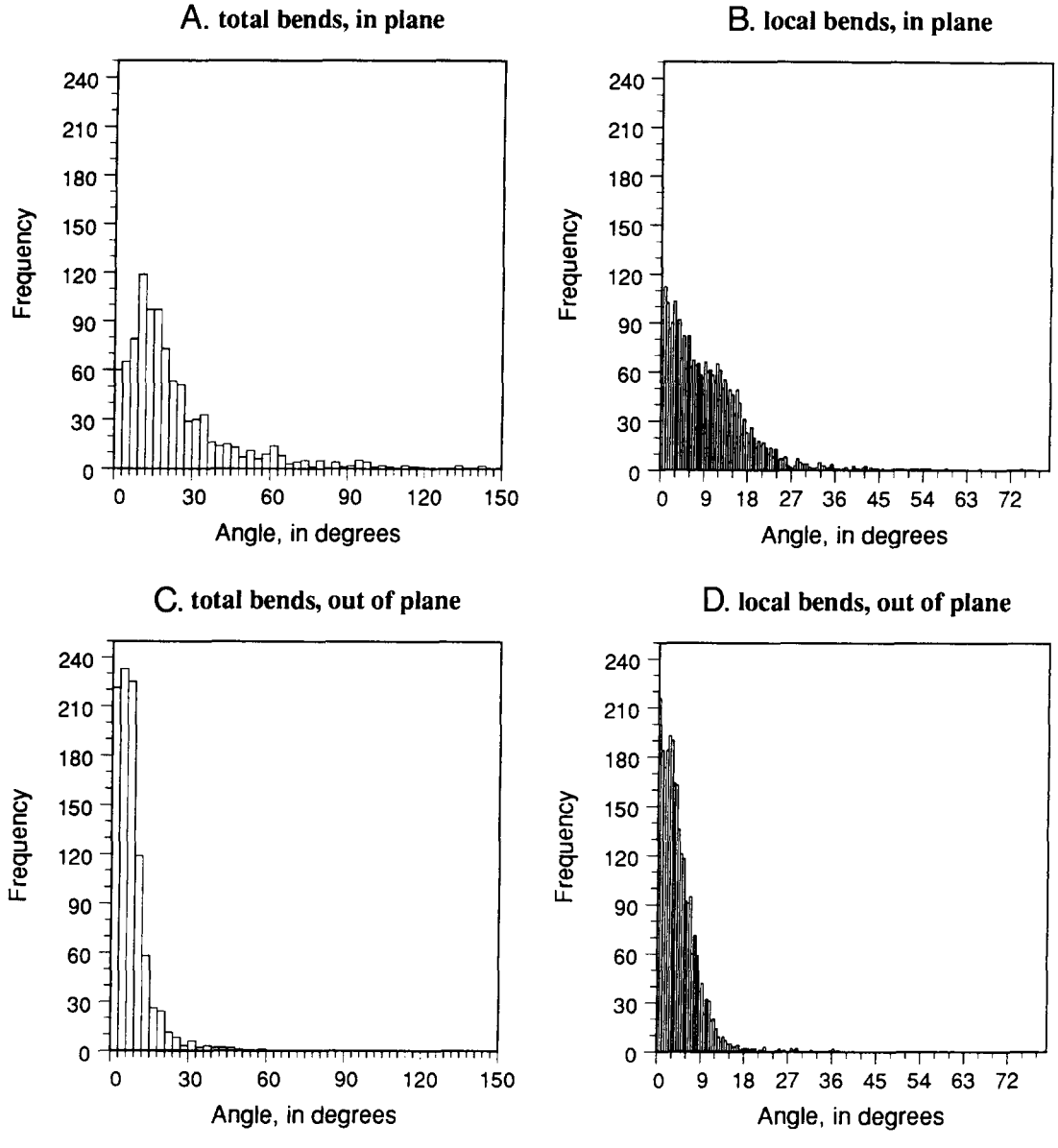

Fig. 3. Frequency distribution for various bend angles. **A:** Total bend angle component within major plane defined by the strand C_α positions. **B:** Local bend angle component within the major strand plane. **C:** Total bend component out of the major plane delineated by the strand C_α atom sites. **D:** Local bend component out of the major strand plane.

structures, composed of more than 3 strands. Even though about 45% of strands displaying at least 1 local bend greater than 20° are bulged, only 20% of these local bends occur directly across the bulge.

Preferred strand position for bends

Each strand constituted by 6 or more residues was sectioned according to the position of local bend: N-terminus, middle, and C-terminus. If the N-terminal residue is numbered 2 and the C-terminal as c, then the N- or C-terminal bends are, respectively, between strand sites (2, 3) and ($c-2$, $c-1$). All other possible local bend sites are designated "middle." Of all local bends greater than 20°, 31%, 60%, and 9% were observed to be at N-terminal, middle, or C-terminal sites, whereas the respective expected percentages were 17.5%, 65%, and 17.5%, resulting in respective preferences of 1.8, 0.9, and 1.1. Though most strong bends occur in the middle of extended structures, clearly there is a skewed preference for the N-terminus. This phenomenon may correlate with the process of protein folding and suggests their early involvement.

Preferred amino acids at bend sites

Table 4 lists amino acid preferences for residues appearing at either side of a large local bend (greater than 20°). Because residue sets (i to $i+2$) and ($i+1$ to $i+3$) form the plane containing major axes used to determine the local bend angle, the residue types residing at positions ($i+1$) and ($i+2$) were counted and referred to as "inside" the bend, whereas (i) and ($i+3$) are

Table 2. *Observed counts among strand structural types for extended segments with total in-the-plane bend greater than 25°*

Structural type	Count
Antiparallel	247
Parallel	38
Short	28
Long	257
Inside	154
Edge	131
Bulged	85
Not bulged	200
Singly bent	236
Multiply bent	49
In large sheet	215
In small sheet	70

called "next-to." Amino acid type preferences are given in Table 4 for inside, next-to, inside plus next-to, loop (taken from Palau et al., 1982), and strand (calculated from the present data set) conformations. Actual observed and expected counts are listed for the inside plus next-to case.

The strong local bend amino acid preferences, quite consistent between the "inside" and "next-to" categories and measured relative to residue type composition found in β-strands, focus on the turn residues Gly and Pro, the exceptional residues Cys and Trp, and the charged and polar types Asp, Arg, Lys, His, Ser, and Thr. Only 2 of the latter group are also preferred by loop- or turn-configured residues in proteins; namely, Asp and Ser (Table 4). The β-strand preferences are for hydrophobic amino acids and share in preference only Thr and Trp with bends. Clearly, residue types preferred near strand bends are mostly not those inclined to be in turn conformations, are rarely preferred in extended substructures, and are concentrated on charged (Asp, Lys, Arg, His) and polar (Ser, Thr) amino acids as well as the important loop residues Gly and Pro. The unique status of bend residues within β-strand structures is emphasized by the fact that the average bend preference for charged and polar residues is 1.14, whereas for the same amino acid types in extended structures, the mean preference is only 0.75. The residue types included His, Lys, Arg, Asp, Glu, Gln, Ser, and Thr. Of the local strand bends with angles greater than 20°, 62% had at least 1 polar residue at the inside or next-to sites, 53% similarly had at least 1 charged amino acid, and 89% had at least 1 charged or polar type in 1 of the 4 positions. Obviously such residues are important for the structural stabilization of the bend through not only their presence but also through interaction with other groups, the most significant likely to be salt bridging and hydrogen bonding. It is also possible during protein folding that such interactions help establish the bend and are thus essential in achieving the final protein topology.

The interactions of the polar and charged side chains at strong local bends with other protein atoms were investigated in detail for those sheets that contained at least 3 strands and at least 1 bend angle greater than 20° in each of 3 or more strands. All polar or charged atoms within 4.00 Å of similar atoms in side groups at flanking bend sites were listed and classified as main chain or side chain and within the strand containing the flanking side group, in an adjacent strand, elsewhere within the β-pleated sheet containing the bend, or in flanking loops at the

Table 3. *Observed counts among strand structural types with extended segments having at least 1 local bend angle greater than 20°*

Structural type	Count
Antiparallel	155
Parallel	28
Short	39
Long	144
Inside	81
Edge	102
Bulged	84
Not bulged	99
Singly bent	147
Multiply bent	36
In large sheet	123
In small sheet	60

Table 4. *Amino acid type preferences for various conformations*[a]

			Preference				
Amino acid	Observed count	Expected count	Inside	Next-to	Inside plus next-to	Turn	Strand
Ala	47	61	0.67	0.87	0.77	0.84	0.83
Cys	10	8	1.00	1.50	1.25	0.69	0.95
Asp	28	17	2.00	1.50	1.65	1.28	0.33
Glu	27	30	1.00	0.80	0.90	0.78	0.64
Phe	44	51	0.76	0.96	0.86	0.88	1.57
Gly	75	42	2.38	1.14	1.79	1.76	0.59
His	23	15	1.28	1.75	1.53	0.53	0.84
Ile	61	78	0.55	1.00	0.78	0.55	1.72
Lys	43	42	1.05	1.05	1.02	0.95	0.80
Leu	63	85	0.59	0.91	0.74	0.49	1.27
Met	20	22	1.00	0.82	0.91	0.52	1.36
Asn	12	15	1.14	0.50	0.80	1.48	0.41
Pro	29	11	3.67	1.17	2.60	1.47	0.29
Gln	21	26	1.00	0.62	0.81	1.00	0.89
Arg	34	26	1.23	1.38	1.31	0.91	0.79
Ser	56	50	0.92	1.31	1.12	1.29	0.86
Thr	74	65	1.44	0.84	1.14	1.05	1.20
Val	104	123	0.77	0.92	0.85	0.51	2.05
Trp	19	18	0.67	1.44	1.06	0.88	1.47
Tyr	40	46	0.72	1.04	0.87	1.28	1.53

[a] "Inside" refers to the residues flanking a large in-the-plane local bend greater than 20°, whereas "next-to" are the 2 residues neighboring the inside ones. Counts are given for the inside plus next-to case. Preferences are also given for residues in turn conformations taken from Palau et al. (1982) and in extended configurations calculated from data used in the present work. A preference value less than 1.00 indicates relative avoidance of the residue type, 1.00 neutrality, and greater than 1.00, preference. The preferences are normalized relative to the entire data sample for turn and strand and relative to strands for the remaining preferences.

Table 5. *Counts of interactions (salt bridges and hydrogen bonds) between side groups in bent strands and atoms in various regions of the encompassing β-pleated sheet*

Interacting side group	Interaction with			
	Same strand	Opposite strand(s)	Other residues, mostly in sheet	Flanking turns of bent strand
Asp	13	12	2	12
Glu	1	9	5	6
Arg	2	4	5	2
His	4	10	3	7
Lys	1	8	9	0
Asn	12	21	2	6
Glu	4	2	4	1
Ser	17	8	8	1
Thr	15	14	5	0
Tyr	1	6	8	1
Total	70	94	51	36
Interacting main-chain:side-chain atoms	55:15	36:58	21:30	28:8

termini of the bent strand. Table 5 lists the counts for each side-group type. Most of the interactions (salt bridges or hydrogen bonds) between the bend participating side group and other atoms involve those in adjacent strands of the sheet (38%), and most of these (62%) are with other side chains. Nonetheless, 28% of the interactions between the bend side chain and other atoms are within the same bent strand where main-chain associations with the bend side group predominate (79%). The remaining 34% of the interactions are with remaining sheet regions (mostly involving other side groups) or flanking turns of the bent strand segment (most involving main-chain atoms of loops). The side-chain/side-chain associating pairs are not unexpected and those with counts greater than 5 included Glu-Lys, Asp-Arg, Asp-His, Asn-Tyr, Asp-Lys, Tyr-Asp, and Thr-Thr. Though interactions between hydrophobic residues must also be significant for bend stabilization, the strong preferences for charged and polar residues at bend sites point to their special importance to bend stability. For purposes of protein design and engineering, side-group/side-group interactions between the bent and opposite strand side chains are recommended to promote and stabilize any intended bend of an extended substructure. Two structural examples of bent strands and their stabilizing side groups and respective interactions are shown as stereo drawings in Figures 4 and 5 and Kinemages 1 and 2.

Data and methods

Protein structures studied

A data set of 247 known 3-dimensional protein structures was examined. Their atomic coordinates were taken from the Brookhaven Protein Data Bank (Bernstein et al., 1977), and they were selected by the procedure of Heringa et al. (1992) such that the largest possible subset of structures from those known is found under the constraint that the amino acid sequences of no protein pair be no greater than 50% in residue identity over all matched amino acids after alignment by an appropriate method. The 4-letter Brookhaven codes associated with each protein file containing the coordinate records are subsequently listed, with chain identity indicated in parentheses if there was information given for more than 1 polypeptide such as for oligomeric proteins:

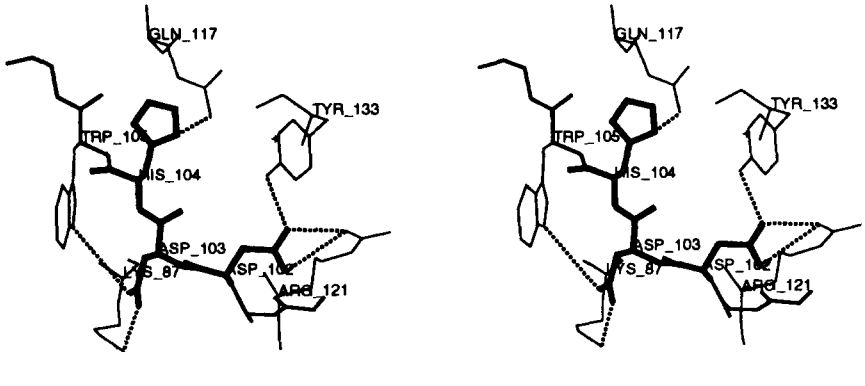

Fig. 4. Stereo illustration of an exemplary local bend of more than 20° taken from retinol binding protein (Cowan et al., 1990). The central strand shown encompasses residues 100–106 with single-letter-coded sequence GNDDHWI. Asp 102 and Asp 103 constitute the residues closest to the bend region shown as thick lines, whereas His 104 is in the "next-to" site (see text). The bend is stabilized by interactions of His 104 with Gln 117 appearing in a strand adjacent to the bent strand, Asp 103 with both Trp 105 of the same strand and Lys 87 of an adjacent strand, and Asp 102 with Tyr 133 appearing in a strand of the sheet containing the bent segment. The main- and side-chain atom sites are indicated for each residue noted.

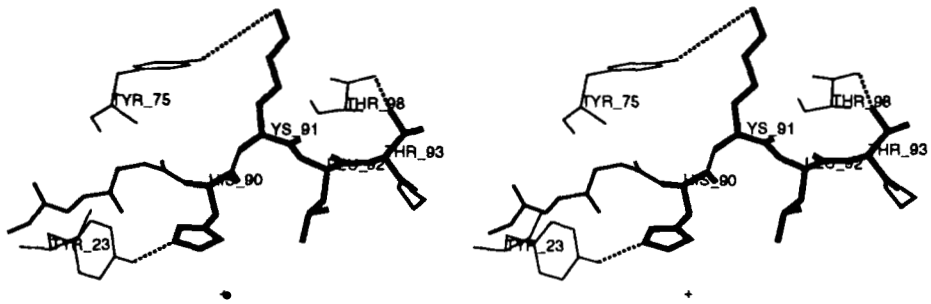

Fig. 5. Stereo illustration of a strand segment (residues 87–94 with single-letter-coded sequence KIYHKLTY) containing 2 local bends greater than 20° from bilin binding protein (Huber et al., 1987). The bends are on either side of Leu 92, and the corresponding bent region is shown in thick lines. The bends are stabilized by interactions between His 90 and Tyr 23 contained in a flanking loop of a strand in the sheet containing the bend strand, Lys 91 and Tyr 75 of an adjacent strand, and Thr 93 and Thr 90 of an adjacent strand.

ACT1(A), ACT1(D), ROM7, SOD0(Y), ZIF1(C), 155C, 1AAP(A), 1AAT, 1ACE, 1ACX, 1ADA, 1AK3(B), 1ALC, 1ALD, 1ATX, 1BBP(C), 1BDS, 1BN2(1), 1BP2, 1CBP, 1CC5, 1CDT(A), 1COX, 1CPB, 1CPK(E), 1CPK(I), 1CRN, 1CRO(A), 1CTF, 1CTX, 1CY3, 1DPI, 1DRF, 1DTX, 1ECD, 1EFM, 1F19(H), 1FC1(A), 1FC2(C), 1FCB(A), 1FDX, 1FKF, 1FNR, 1FX1, 1FXI(A), 1GCR, 1GD1(R), 1GOX, 1GPD(R), 1GSG(P), 1HCC, 1HCO(A), 1HCO(B), 1HDD(D), 1HIP, 1HRH(A), 1HSC, 1HSD(A), 1IFB, 1LAP, 1LDB, 1LDM, 1LRD(3), 1LYM(B), 1MAD(H), 1MAD(L), 1MCA(B), 1MLE, 1MON(E), 1MON(H), 1MRT, 1MSB(A), 1NRD, 1OMD, 1OVA(A), 1PAD, 1PAL, 1PCD(A), 1PCD(B), 1PFK(A), 1PHV(I), 1PHY, 1PP2(L), 1PPT, 1PRC(C), 1PRC(L), 1PRC(M), 1PRC(H), 1PTE, 1R09(1), 1R09(3), 1R1E(E), 1RBP, 1RHD, 1RIG, 1RNH, 1RNS(S), 1RNS, 1RNT, 1RSL(C), 1SDG, 1SGC, 1SGT, 1SH1, 1SN3, 1TAB(I), 1TEC(I), 1TGC, 1TGL, 1TGS(I), 1TNF(A), 1TPT, 1VSG(A), 1WSY(A), 1WSY(B), 1ZNF, 256B(A), 2ABX(A), 2ACT, 2APR, 2BJL(1), 2BLM(A), 2BUS, 2CA2, 2CBH, 2CCP, 2CCY(A), 2CD4, 2CDV, 2CHY, 2CNA, 2ETI, 2FBJ(H), 2FCR, 2FXB, 2GN5, 2HFL(L), 2HLA(B), 2HMZ(D), 2IL8(B), 2LBP, 2LH1, 2LHB, 2LTN(A), 2LTN(D), 2MB5, 2MHR, 2MLT(A), 2MRT, 2PAB(B), 2PHH, 2PKA(B), 2PKA(X), 2PLV(1), 2PLV(2), 2PLV(3), 2PLV(4), 2PRK, 2RSP(A), 2RUS(A), 2SAR(B), 2SC2(A), 2SC2(B), 2SDH(B), 2SNI(E), 2SNI(I), 2SNS, 2SOD(3), 2STV, 2TBV(C), 2TEC(E), 2TIM(B), 2TMV(B), 2TRX(B), 2TSC(A), 2UTG(B), 2YHX, 31BI, 351C, 3ADK, 3AIT, 3APP, 3AT1(A), 3AT1(D), 3B5C, 3BCL, 3BLM, 3C2C, 3CBH, 3CLN, 3CRO(R), 3CSC, 3CYT(O), 3DFR, 3DPA, 3ER3(E), 3FBP(B), 3FGF, 3GAP(B), 3GBP(O), 3HLA(A), 3HMG(B), 3HMG(C), 3ICB, 3INS(D), 3LZM, 3MBA, 3PEP, 3PGK, 3PGM, 3PHV, 3RP2(A), 3TRX, 3WRP, 3ZNF, 4CHA(B), 4CPA(I), 4DFR(A), 4FD1, 4FXN, 4GR1, 4INS(C), 4MDH(A), 4PCY, 4SBV(C), 4SGB(I), 4TGF, 4TNC, 4TS1(A), 5EBX, 5ENL, 5ICD, 5P21, 5RXN, 6ACN, 6ADH(B), 6CPA, 6CPP, 6EST, 6HIR, 6PTI, 6RXN, 7ABP, 7TLN, 7WGA(A), 8CAT(B), 9XIA.

Delineation of secondary structure and hydrogen bonds

Residues participating in extended β-strand conformations were detected by the computer program DSSP of Kabsch and Sander (1983), which generally relies on recognition of hydrogen bonds through threshold properties (such as distance) of main-chain NH and CO interactions. DSSP also lists hydrogen bond partners, allowing delineation of β-pleated sheet structures. Bulges were detected by deviation in typical hydrogen bond patterns of adjacent strands (see Fig. 1 for an example and Chan et al. [1993] for a discussion). Local bend angles were measured across bulges by simply excluding the puckered residues; for example, if residues 4 and 5 in an 8-residue strand were bulged, then the local bend angles across the bulge were defined by the 2 local major axes (see below) of principal planes defined by the main-chain atoms of the residue triplets (1, 2, 3) and (2, 3, 6); (2, 3, 6) and (3, 6, 7); and (3, 6, 7) and (6, 7, 8). The local bend angle directly across the bulge is defined with the sets (2, 3, 6) and (3, 6, 7).

Definition of strand bends

For purposes of this work, each residue was assigned 3 main-chain atoms: namely, C′, C_α, and N, where C′ is the carbonyl carbon, C_α is the carbon atom to which the residue side group is covalently bonded, and N is the peptide nitrogen. Several bend angle characterizations were considered: local bend angle within a strand, total bend angle over an entire strand, and bend angle components within and out of the principal strand plane.

The principal plane through given strand atoms (3 per residue) is determined by the method of Chelvanayagam et al. (1992), who applied principal component analysis (Press et al., 1988:353–397) to the positions of atoms intended to lie in the plane. This plane will define a major axis representing the best line through the main-chain strand atoms and with direction from the N- to C-terminus along the amino acid sequence, a minor axis, which is orthogonal to the major axis and within the principal plane, and the normal perpendicular to the latter 2 axes and the principal plane (Chelvanayagam et al., 1992). The local bend angle is defined as that angle made by the intersection of 2 major axes associated with the planes defined by 2 consecutive sets of 3 residues (9 atoms) along the strand; for instance, the angle between major axes of strand residues (i to $i + 2$) and residues ($i + 1$ to $i + 3$) (see Fig. 6 for an illustration). This local bend angle has 2 components: namely, that within the major plane defined by all backbone atoms in the entire strand and that component outside or deviant from the overall strand plane.

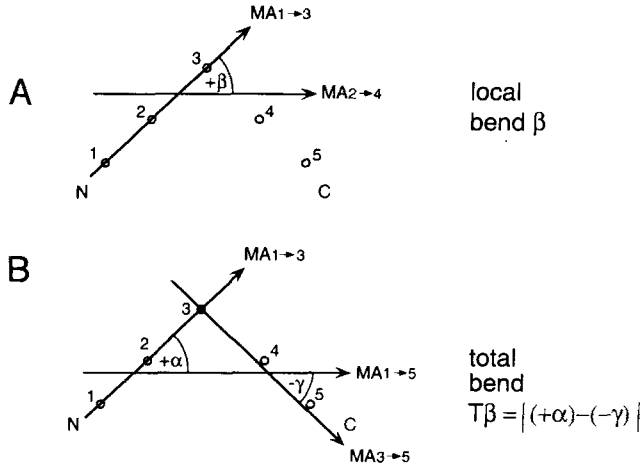

Fig. 6. Hypothetical local and total bend angles. The open circles refer to successive C_α atoms along the strand counted from the N- to C-termini of the polypeptide chain. $MA_{i\to j}$ indicates the major axis of the plane containing backbone atoms (3 per C_α) encompassing consecutive residues i to j. **A:** A local bend angle, β, which is always defined as a positive angle less than 180°. **B:** The total strand bend angle, TB, which is taken as a positive value; angles above the major strand axis are taken as positive $(+\alpha)$ and those below are negative $(-\gamma)$. See text for further details.

Chelvanayagam et al. (1992) described in detail and diagrammatically how these components were determined.

The total bend angle of a strand is defined as the absolute difference of the 2 bend angles defined by the major axes of the N- and C-terminal 3 residues, respectively, and the major axis of the entire strand (Fig. 6). Positive (negative) angles were defined by counterclockwise (clockwise) rotations relative to the strand major axis. Consecutive local bend angles for 3 residues measured relative to the strand major axis should monotonically increase or decrease for a curved, uniformly bending (singly bent) strand; when this was not the case, creating a zig-zag (multiply bent) type strand, the total bend angle of the strand was taken as the difference between the maximum and minimum angles achieved at the intersection of the major strand axis and the local 3-residue line (Fig. 2). Components of the total bend angle were taken to be within and without the major strand plane as also determined for local bends.

Strand types

In this work, strands with various characteristics are considered. Extended conformations in β-sheets can be parallel or antiparallel, edge or inside strands, and singly or multiply bent. These types are illustrated in Figure 2. Strands were also considered with and without bulges and as long (more than 5 residues in length) or short (less than or equal to 5 residues); 5 provided the demarcation because it was near the average strand length (4.7) of the sample used here. Extended substructures in large (more than 3-stranded) and small (2- or 3-stranded) sheets were also distinguished.

Amino acid preferences

Preferences were always based on an observed count divided by the expected number such that resulting ratios greater than 1.00 indicated an affinity for the residue type, equal to 1.00, neutrality, and less than 1.00, avoidance. Preferences were calculated for amino acid types to be in β-strands and to be at specific sites surrounding large bends in strands. The expected number of a particular amino acid type to be in a strand structure was defined as the product of the number of residues of that type in the entire database of structures and the fraction of residues in the entire database that appear in β-strands. The expected number for particular residue types to be at sites such as those flanking a strand bend (inside sites) was taken as the product of the number of residues of that type in all β-strands examined and the fraction of residues in all strands that appear at the appropriate bend sites.

Acknowledgments

We are grateful for useful discussions with Simon Hubbard whose computer programs that determine atomic interactions were also kindly provided. Jaap Heringa's automated routine to select known protein structures for analysis was also essential in performing this study. Fiona Berrie, Petra Riedinger, and Simon Hubbard provided expert technical help in the preparation of the manuscript.

References

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.

Chan AWE, Hutchinson EG, Harris D, Thornton JM. 1993. Identification, classification, and analysis of beta-bulges in proteins. *Protein Sci 2*:1574–1590.

Chelvanayagam G, Heringa J, Argos P. 1992. The anatomy and evolution of proteins with virus capsid jellyroll topologies. *J Mol Biol 228*:220–242.

Chothia C, Janin J. 1981. Relative orientation of close packed β-pleated sheets in proteins. *Proc Natl Acad Sci USA 78*:4146–4150.

Chothia C, Janin J. 1982. Orthogonal packing of β-pleated sheets in proteins. *Biochemistry 21*:3955–3965.

Chothia C, Murzin AG. 1993. New folds for all-β proteins. *Structure 1*:217–222.

Cowan SW, Newcomer ME, Jones TA. 1990. Crystallographic refinement of human serum retinol binding protein at 2 Ångstroms resolution. *Proteins Struct Funct Genet 8*:44–61.

Heringa J, Sommerfeldt H, Higgins D, Argos P. 1992. OBSTRUCT: A program to obtain largest cliques from a protein sequence set according to structural resolution and sequence similarity. *Comput Appl Biosci (CABIOS) 8*:599–600.

Huber R, Schneider M, Mayr I, Mueller R, Deutzmann R, Suter F, Zuber H, Falk H, Kayser H. 1987. Molecular structure of the bilin binding protein from *Pieris brassicae* after refinement at 2.0 Ångstroms resolution. *J Mol Biol 198*:499–513.

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

Palau J, Argos P, Puigdomenech P. 1982. Protein secondary structure: Studies on the limits of prediction accuracy. *Int J Protein Pept Res 91*:394–401.

Press WH, Flannery BP, Teukolsky SK, Vetterling WT. 1988. *Numerical recipes in C*. New York: Cambridge University Press.

Richardson JS, Getzoff ED, Richardson DC. 1978. The β-bulge: A common small unit of nonrepetitive protein structure. *Proc Natl Acad Sci USA 75*:2574–2578.

Salemme FR. 1983. Structural properties of protein β-sheets. *Progr Biophys Mol Biol 42*:95–133.

Yoder MD, Keen NT, Jurnak F. 1993. New domain motif: The structure of pectate lyase C, a secreted plant virulence factor. *Science 260*:1503–1507.