J. Mol. Biol. (1992) **223**, 845–851

Orthogonal ββ Motifs in Proteins

R. Sowdhamini, N. Srinivasan, C. Ramakrishnan and P. Balaram†

Molecular Biophysics Unit, Indian Institute of Science
Bangalore 560012, India

(Received 22 May 1991; accepted 25 October 1991)

A super-secondary structural motif comprising two orthogonally oriented β-strands connected by short linking segments of ≤5 residues has been identified from a data set of 65 independent protein crystal structures. Of the 42 examples from 14 proteins, a vast majority have only a single residue as the linking element. Analysis of the conformational angles at the junction reveals that the recently described type VIII β-turn occurs frequently at the connecting hinge, while the type II β-turn is also fairly common.

Keywords: protein folding; protein super-secondary structures; β-strands; protein structural motifs; type VIII β-turns.

Super-secondary structures formed by linking contiguous secondary structure modules by connecting loops, with varied backbone conformations, are important elements determining the overall folding of globular proteins (Rao & Rossmann, 1973; Efimov, 1984*a*,*b*, 1986*a*,*bc*; Thornton *et al.*, 1988; Richardson & Richardson, 1989). The α,α-corner, a unit comprising two approximately orthogonal helices, is a relatively widespread structural feature in proteins (Efimov, 1984*a*,*b*, 1986*a*). Less attention has been paid to isolated β-strands (extended regions of the polypeptide chain), although an incisive analysis of β-hairpins has served to identify the conformational features favoured by tight turns that facilitate chain reversal in these structures (Sibanda & Thornton, 1985; Sibanda *et al.*, 1989). β-Pleated sheets have been the subject of several analyses with specific emphasis on chain geometry, coiling and packing (Salemme, 1983; Chothia & Janin, 1982; Chothia, 1983, 1984). In this communication we draw attention to a structural feature, consisting of two β-strand elements (defined as a segment in an extended conformation) at right angles to each other generating an "L" structure, and examine details of the connecting segments for the case of short loops (5 residues or less). The type VIII β-turn, a relatively new addition to the family of turns observed in proteins (Wilmot & Thornton, 1988, 1990), is frequently found at the connecting junction in L structures. Using a data set of 65 non-homologous protein structures determined at resolutions ≤2·0 Å (Srinivasan *et al.*, 1991; 1 Å = 0·1 nm), a sub-data base of β-strands was generated using the following criterion for strand identification: at least four successive residues in the protein sequence should have their ϕ values in the range $-180°$ to $-60°$ and ψ in the range $+60°$ to $+180°$. Structures identified using this criterion correspond to extended polypeptide chain conformations and include segments that do not form a part of intramolecular hydrogen-bonded hairpins and sheets. A total of 470 extended polypeptide segments were identified in the data set; 175 ββ motifs linked by five residues or less were then chosen for further analysis. β-Strand axes were computed by the procedure of Chou *et al.* (1984). The majority (59·4%) of motifs had interstrand angles less than 60°. Several ββ motifs with an interstrand angle (ρ) of $90(\pm 20)°$ were identified, corresponding to an approximately orthogonal orientation of the two strands. Table 1 lists the 42 examples found in 14 of the 65 proteins examined. Figure 1 shows representative examples from unrelated proteins with differing linker segment size. The observed motifs give rise to an L-shaped orientation of the polypeptide backbone (Efimov, 1984*a*,*b*). The angular distribution in the ββ motifs does not show a distinct peak at a value of 90°. This is in contrast to the distribution observed in αα, αβ and βα motifs (Srinivasan *et al.*, 1991). Nevertheless, ββ motifs with an approximately orthogonal orientation have been chosen for analysis because these represent a major distortion from an isolated β-hairpin or a continuous extended strand. Furthermore, there are a significant number of examples of such an orientation in the data set.

† Author to whom all correspondence should be addressed.

0022-2836/92/040845-07 $03.00/0

Table 1

Orthogonal or L-shaped ββ motifs

Protein code[a]	Strand preceding linker	Strand following linker	ρ (°)	β-Turn type[b]		MRA[c]
A. *Single-residue linkers*[d,e]						
2ALP	15B–18	31–35	78·8	II	(G18–G19)	BB
2ALP	112–119	120A–120D	103·7			BB
2ALP	180–183	185–191	70·4			BB
3GRS	123–127	129–133	85·2			BB
3GRS	145–149	151–154	75·8	VIII	(P150–H151)	BE
2APP	63–67	69–75	88·0	II	(S67–G68)	PE
2APP	95–98	100–105	97·4	II	(H98–G99)	BE
2APP	190–195	197–200	98·5	VIII[f]	(D196–S197)	BE
2APP	203–207	209–213	87·9			BB
3RP2	A 80–A 85	A 87–A 90	107·1	VIII	(E86–K87)	PE
3RP2	A102–A108	A110–A115	87·8	VIII	(E109–K110)	BE
5RSA	72–75	77–85	70·1	VIII	(Y76–S77)	BP
1TPP	80–85	87–90	105·9	VIII	(S86–K87)	PE
1TPP	118–124	127–130	94·7	VIII	(T125–Q127)	PE
1ACX	62–65	67–70	92·2	VIII	(K66–S67)	BP
2CAB	207–212	214–219	102·9	VIII	(K213–E214)	BP
2PKA	A 80–A 85	A 87–A 91	101·3	VIII	(T86–A87)	BE
2PKA	B102–B108	B110–B114	86·3	VIII	(Q109–S110)	BE
2PKA	B118–B124	B127–B132	100·1	VIII[f]	(T125–Q127)	BE
2PKA	B127–B132	B134–B139	72·9	II	(L132–G133)	BP
1GD1	O115–O118	O120–O122A	70·3			BB
1GD1	O271–O274	O276–O279	74·7			BE
1TON	81–85	87–91	106·7	VIII	(R86–Q87)	PE
1TON	118–124	128–131	94·9			PE
1TON	148–152	154–159	70·2	VIII	(H153–D154)	BP
B. *Two-residue linkers*[d,e]						
3RP2	A110–A115	A118–A121	84·7	I	(P116–A117)	PE
3RP2	A128–A131	A134–A140	78·2	II	(P132–G133)	BP
1TPP	127–130	134–140	80·6			PE
2CAB	38–41	44–52	95·2	I	(T42–S43)	PE
1TON	87–91	94–95B	80·7	I	(D92–Y93)	PE
1TON	128–131	134–139	79·0	II	(V132–G133)	PE
C. *Three-residue linkers*[d,e]						
2ALP	120I–121	125–131	76·3			PP
3GRS	284–287	291–294	109·5			BP
3RP2	A148–A152	A156–A164	72·7	VIII	(Y153–T154)	BB
2CAB	31–34	38–41	97·8	III[g]	(T35–S36)	PE
				I	(S36–E37)	
2PKA	B110–B114	B118–B124	86·7	III	(D116–A117)	PP
2CPP	380–383	387–393	107·2	II	(P385–G386)	PE
2CYP	130–134	138–141	104·2	I[g]	(E135–D136)	PE
				I[f]	(D136–T137)	
D. *Four-residue linkers*[d,e]						
1FB4	H112–H117	H122–H132	85·6			BP
E. *Five-residue linkers*[d,e]						
1ACX	40–44	50–53	94·4			PP
2CAB	44–52	58–61	78·7			BB
2CAB	66–71	77–80	86·8	I[g]	(N73–Q74)	BP
				I	(Q74–D75)	
				VIII[f]	(D75–R76)	

[a] Protein code as in the Brookhaven Protein Data Bank: 2ALP, α-lytic protease (Fujinaga *et al.*, 1985); 3GRS, glutathione reductase (Karplus & Schulz, 1987); 2APP, penicillopepsin (James & Sielecki, 1983); 3RP2, rat mast cell protease (Reynolds *et al.*, 1985); 5RSA, ribonuclease A (Wlodawer *et al.*, 1986); 1TPP, β-trypsin (Marquart *et al.*, 1983); 1ACX, actinoxanthin (Pletnev *et al.*, 1982), 2CAB, carbonic anhydrase B (Kannan *et al.*, 1984); 2PKA, kallikrein A (Bode *et al.*, 1983); 1GD1, hologlyceraldehyde 3-phosphate dehydrogenase (Skarzynski *et al.*, 1987); 2CYP, cytochrome c peroxidase (Finzel *et al.*, 1984); 3BCL, bacteriochlorophyll A protein (Tronrud *et al.*, 1986); 1FB4, immunoglobulin FAB (Marquart *et al.*, 1980); 2CPP, cytochrome *P*450 CAM (Poulos *et al.*, 1987); 1TON, tonin (Fujinaga & James, 1987). The present data set consists of 5 serine proteases which might bias the data slightly. They have been considered nevertheless, since the primary sequences and detailed stereochemical features vary significantly.

[b] β-turn-type involving linker residues are classified. Deviations of 30° from ideal (ϕ,ψ) values at the $i+1$ and $i+2$ positions of the β-turn are allowed. The ideal values are: type I β-turn: $(-60°, -30°)$,

The vast majority of linking segments correspond to only a single residue at the junction with a non-β-conformation. If the connecting residue adopts a right-handed α-helical (α_R) conformation, then a type VIII β-turn as defined by Wilmot & Thornton (1988) is generated by the linker residue and the succeeding residue in the sequence. Figure 2(a) shows the distribution of backbone conformations at the single linking residue of L structures represented on a Ramachandran map. Interestingly, of the five examples of Gly as the linking residue *all* have positive ϕ values with four in an approximately α_L conformation. A majority of the other residues adopt the α_R conformation. Figure 2(b) shows the distribution of backbone conformational angles for the α_R and α_L linking residues, together with the conformations of the succeeding and preceding residues, respectively. The α_R conformation followed by a β-conformation results in a type VIII β-turn, whereas β followed by α_L generates a type II β-turn. Thus, centrally placed type II and type VIII β-turns result in a right-angled bend of the polypeptide backbone, rather than a sharp chain reversal. Indeed, earlier analyses of β-hairpins have suggested that type I′ or III′ β-turns are found more frequently at the hairpin corner. In the present data set, out of 43 two-residue linkers, 12 are in type I′ (III′) β-turns, ten in type II β-turns and six in type I β-turns. Most of the β-strands involved in the L-structures are short (about 5 residues long). Accessibility calculations were performed for the 43 L structures by the method of Lee & Richards (1971). The accessible surface area of all the residues, as well as those in the linker region alone, in the protein L motifs have been compared with the corresponding model system, Ala-X-Ala, where X assumes a silk conformation ($\phi = -140°$, $\psi = +135°$) (see Table 1 for details). Many of the L structures are buried (60% of the observed examples) with the remaining being partially buried. When the linker residues of L-shaped $\beta\beta$ motifs are considered alone, half of them are exposed, with about one-third being partially buried. A preliminary analysis on the occurrence of the L structures in homologous proteins indicates that the motif corresponding to H112–H132 of immunoglobulin FAB (1FB4) is conserved well in a family of immunoglobulins. The L structures present in families of acid and serine proteases are also reasonably well conserved.

In contrast to β-hairpins or segments forming β-sheet structures, isolated L structures are not stabilized by the simultaneous formation of several interstrand hydrogen bonds. A preliminary analysis using the method described by Soman & Ramakrishnan (1986) reveals that, of the examples in Table 1, eight are isolated motifs. Of the remaining motifs, in 21, one of the arms is in register with a proximal β-strand; while in 13 cases both the arms are in register with adjacent strands. Indeed, examples of orthogonal β-sheet packing involving a 90° twist in both strands have been described. A specific example is illustrated by the occurrence of two "in-register" L structures, namely, residues 190 to 200 and 203 to 213 in penicillopepsin, 2APP (Chothia & Janin, 1982). The sheet-forming tendency of L structures can be viewed in light of the fact that about 91% of all the extended strands (470) in the data set form components of β-sheets. It is conceivable that isolated L motifs may not contribute significantly to the stabilization of the overall fold, but, may in fact be an artifact, a kind of molecular fossil of the folding process. The distortion of an antiparallel hairpin (interstrand angle $\leq 20°$) to an L structure is an extreme situation with intermediate possibilities as visualized in Figure 3(a) to (c). Structural examples from protein observations of the parent hairpin, and the distorted offspring, are shown in Figure 3(d) to (f). The relevance of structural observations on native protein crystals to the mechanism of protein folding (Baldwin, 1989; Dill, 1990; Richards, 1991) remains to be established.

This research was supported by a grant from the Department of Science and Technology. R.S. was the recipient of a senior research fellowship from the Council of Scientific and Industrial Research.

($-90°$,0°); type II β-turn: ($-60°$,120°), (80°,0°); type III β-turn: ($-60°$, $-30°$), ($-60°$, $-30°$) and type VIII β-turn: ($-60°$, $-30°$), ($-120°$,120°). The single letter codes and residue numbers of the amino acids at the $i+1$ and $i+2$ positions of the β-turn are indicated within parentheses. Of the 33 β-turns found in the 42 L-structures, 31 have their $C^{\alpha}_{i} \ldots C^{\alpha}_{i+3}$ distances $\leq 7{\cdot}5$ Å. For single-residue cases, the linking residue occupies either the $i+1$ or $i+2$ position of the turn. For multiple residue linkers the turn may be constituted by any 2 residues in the linker region.

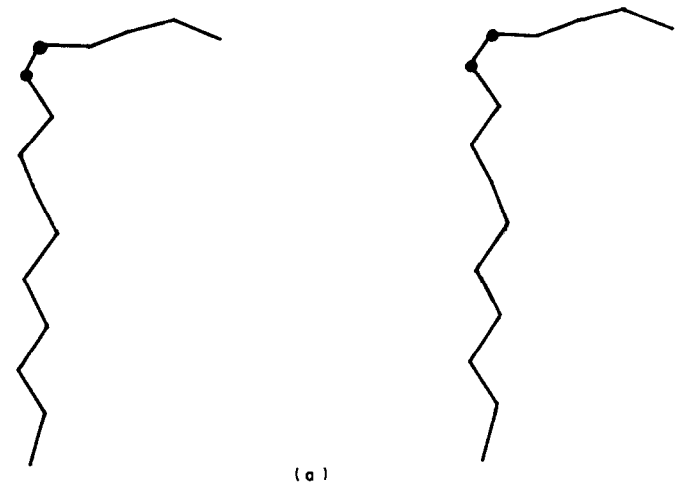
[c] Mean residue accessibility (MRA) expressed in terms of single letter codes: B, buried (0 to 30%); P, partially buried (30 to 50%); E, exposed (>50%). The first code in the column corresponds to MRA of L structures, while the second code refers to MRA of the linker residues of L structures. The method of residue accessibility calculation has been adopted from that described by Thornton (1981) where a particular amino acid in the protein environment is compared with that of a standard (see the text for details). MRA refers to the average residue accessibility of all the residues involved.

[d] In some cases the residue numbering is not continuous in the Protein Data Bank file, so as to permit sequence alignment. Difference between the residue numbers on either side of the linker may therefore not correspond to linker size. In some cases, a prefix is attached to the residue number which denotes the chain identifier in the Protein Data Bank file.

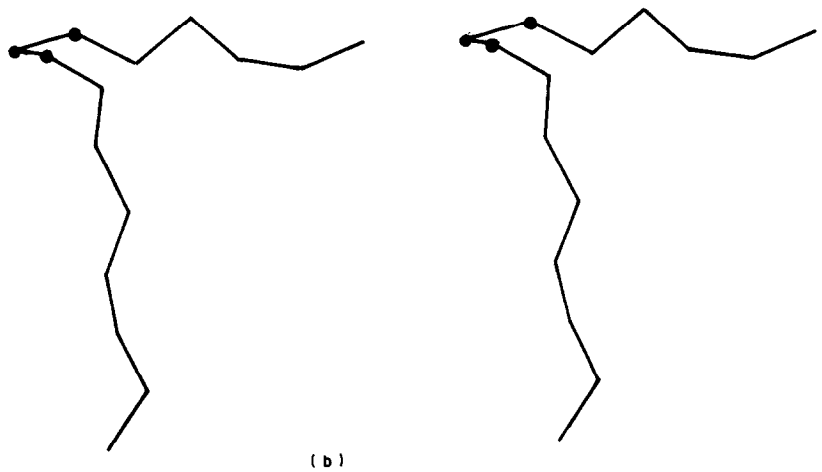
[e] In the 175 $\beta\beta$ motifs identified in the data set, the number of examples with 1, 2, 3, 4 and 5 residues in the linker region are 48, 43, 30, 31 and 23, respectively.

[f] Distorted β-turns where only one torsion angle deviates as much as 45° from ideal values.

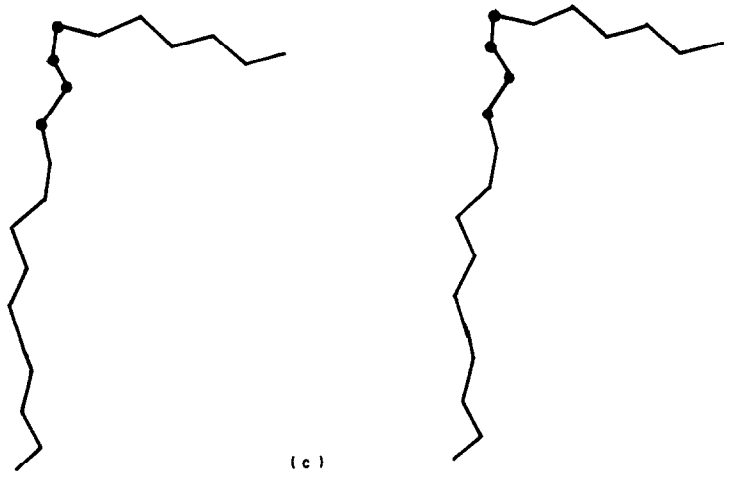
[g] Constitutes multiple β-turn linkers, with each row referring to a β-turn.



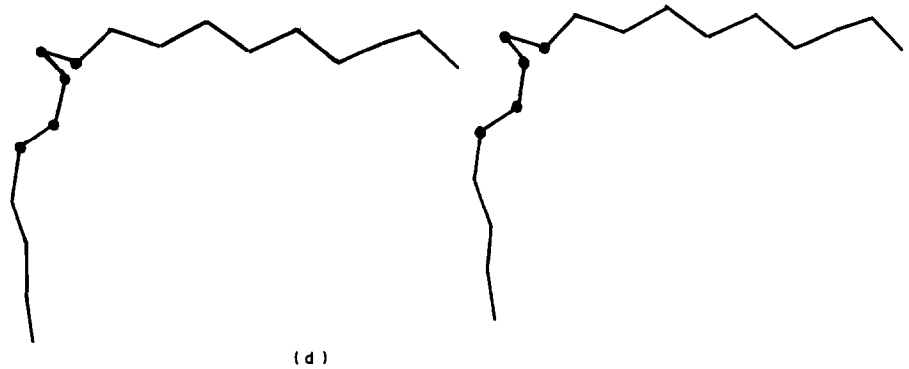
(a)



(b)



(c)



(d)

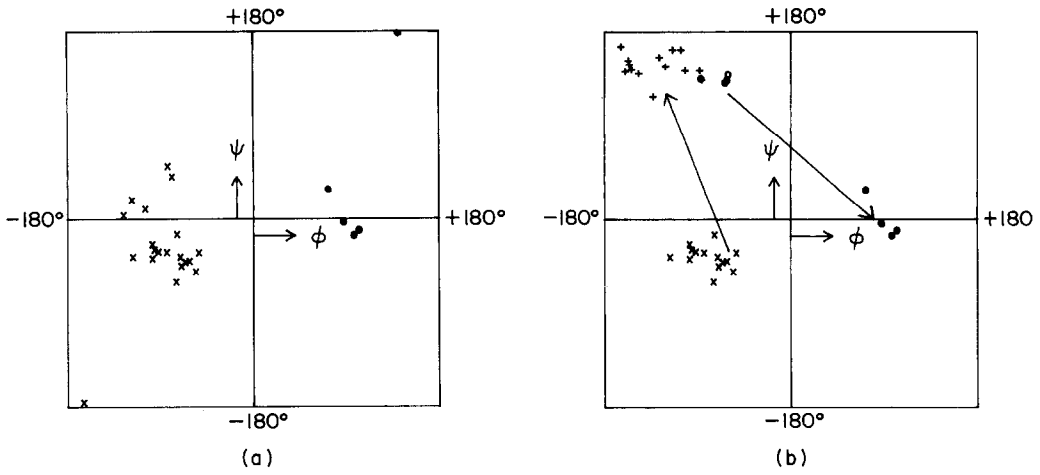


Figure 2. (ϕ,ψ) plot of: (a) linkers of single residue L-shaped $\beta\beta$ motifs (× denotes non-Gly residues and ● indicates Gly residues) and (b) residues at the $i+1$ and $i+2$ positions of single-residue L structure with either type II or type VIII β-turns at the centre (○ for $i+1$ position of type II β-turn; ● for $i+2$ position of type II β-turn; × for $i+1$ position of type VIII β-turn; and + for $i+2$ position of type VIII β-turn). Ideal II and VIII β-turns are shown as arrows with the arrowhead corresponding to position ($i+2$) of the turns.

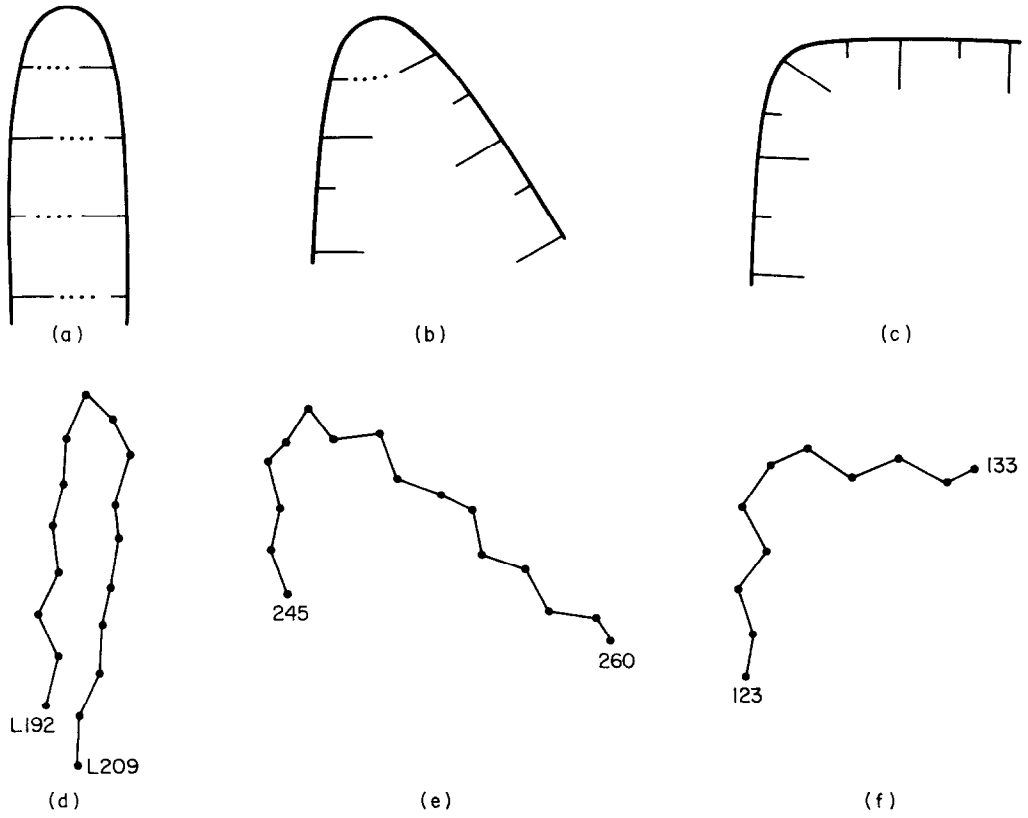


Figure 3: Schematic diagram and observed examples of the distortion of (a) regular antiparallel β-hairpin to (c) an L structure, through (b) a possible intermediate state. (d) to (f) Observations from proteins: (d) 1FB4 (L192–L209); (e) 2APP (245–260); (f) 3GRS (123–133).

Figure 1. Stereo pairs of representative examples of L structures from protein observations: (a) carbonic anhydrase (2CAB), 38–52 (2-residue linker); (b) kallikrein A (2PKA), 110–124 (3-residue linker); (c) immunoglobulin FAB (1FB4), H112–H132 (4-residue linker); (d) carbonic anhydrase (2CAB), 44–61 (5 residue linker). The C^{α} tracing with linker residues marked as filled circles is shown for each of the examples. The structure appears as an inverted "L", the first residue of the strand preceding the linker occupying the top right corner and the last residue following the linker occupying the bottom left corner.

References

Baldwin, R. L. (1989). How does protein folding get started? *Trends Biochem. Sci.* **14**, 291–294.

Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G. & Bartunik, H. (1983). Refined 2 Å X-ray structure of porcine pancreatic kallikrein A, a specific trypsin-like serine protease. Crystallisation, structure determination, crystallographic refinement, structure and its comparison with bovine trypsin. *J. Mol. Biol.* **164**, 237–282.

Chothia, C. (1983). Coiling of β-pleated sheets. *J. Mol. Biol.* **163**, 107–117.

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

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

Chou, K.-C., Nemethy, G. & Scheraga, H. A. (1984). Energetic approach to the packing of α-helices. II Treatment of non-equivalent and non-regular helices. *J. Amer. Chem. Soc.* **106**, 3161–3170.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.

Efimov, A. V. (1984*a*). New supersecondary protein structure: the αα corner. *Mol. Biol. (U.S.S.R.)*, **20**, 1239–1251.

Efimov, A. V. (1984*b*). A novel super-secondary structure of proteins and the relation between the structure and the amino acid sequence. *FEBS Letters*, **166**, 33–38.

Efimov, A. V. (1986*a*). Standard conformations of polypeptide chain in irregular regions of proteins. *Mol Biol. (U.S.S.R.)*, **20**, 208–216.

Efimov, A. V. (1986*b*). Standard structures in protein molecules. α–β Hairpins. *Mol. Biol. (U.S.S.R.)*, **20**, 258–266.

Efimov. A. V. (1986*c*). Standard structures in protein molecules. β–α Hairpins. *Mol. Biol. (U.S.S.R.)*, **20**, 267–271.

Finzel, B. C., Poulos, T. L. & Kraut, J. (1984). Crystal structure of yeast cytochrome *c* peroxidase refined at 1·7 Å resolution. *J. Biol. Chem.* **259**, 13027–13036.

Füjinaga, 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.

Füjinaga, M., Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1985). Refined structure of alpha-lytic protease at 1·7 Å resolution. Analysis of hydrogen bonding and solvent structure. *J. Mol. Biol.* **195**, 479–502.

James, M. N. G. & Sielecki, A. R. (1983). Structure and refinement of penicillopepsin at 1·8 Å resolution. *J. Mol. Biol.* **163**, 299–361.

Kannan, K. K., Ramanadham, M. & Jones, T. A. (1984). Structure, refinement and function of human carbonic anhydrase I. *Ann. N.Y. Acad. Sci.* **429**, 49–55.

Karplus, P. A. & Schulz, G. E. (1987). Refined structure of glutathione reductase at 1·54 Å resolution. *J. Mol. Biol.* **195**, 701–729.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980). Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3·0 Å and 1·9 Å resolution. *J. Mol. Biol.* **141**, 369–391.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. sect. B*, **39**, 480–490.

Pletnev, V. Z., Kuzin, A. P., Trakharov, S., Popovich, V. & Tsigannik, I. (1982). X-ray investigation of three-dimensional structure of actinoxanthin. In *Chemistry of Peptides and Proteins* (Voelter, W., Wuensch, E., Ovchinnikov, Y. & Ivanov, I., eds), vol. 1, pp. 429–444, Walter de Gruyter, Berlin.

Poulos, T. L., Finzel, B. C. & Howard, A. J. (1987). High-resolution crystal structure of cytochrome P450 CAM. *J. Mol. Biol.* **195**, 687–700.

Rao, S. T. & Rossmann, M. G. (1973). Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76**, 241–256.

Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L. & Matthews, B. W. (1985). Structure of a serine protease from rat mast cells determined from twinned crystals by isomorphous and molecular replacement. *Acta Crystallogr. sect. B*, **41**, 139–147.

Richards, F. M. (1991). The protein folding problem. *Sci. Amer.* **264**, 34–39.

Richardson, J. S. & Richardson, D. C. (1989). Principles and patterns of protein conformation. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D. ed.), pp. 1–98, Plenum Press, New York and London.

Salemme, F. R. (1983). Structural properties of protein β-sheets *Progr. Biophys. Mol. Biol.* **42**, 95–133.

Sibanda, B. L. & Thornton, J. M. (1985). β-Hairpin families in globular proteins. *Nature (London)*, **316**, 170–174.

Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989). Conformation of β-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.

Skarzynski, T., Moody, P. E. C. & Wonacott, A. J. (1987). Structure of holo-glyceraldehyde 3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1·8 Å resolution. *J. Mol. Biol.* **193**, 171–187.

Soman, K. V. & Ramakrishnan, C. (1986). Identification and analysis of extended strands and β-sheets in globular proteins. *Int. J. Biol. Macromol.* **8**, 89–96.

Srinivasan, N., Sowdhamini, R., Ramakrishnan, C. & Balaram, P. (1991). Analysis of short loops connecting secondary structure elements in proteins. In *Molecular Conformation and Biological Interactions* (Balaram, P. & Ramaseshan, S., eds), Indian Academy of Sciences, Bangalore, India, in the press.

Thornton, J. M. (1981). Disulphide bridges in proteins. *J. Mol. Biol.* **151**, 261–287.

Thornton, J. M., Sibanda, B. L., Edwards, M. S. & Barlow, D. J. (1988). Analysis, design and modification of loop regions in proteins. *BioEssays*, **8**, 63–69.

Tronrud, D. E., Schmid, M. F. & Matthews, B. W. (1986). Structure and X-ray amino acid sequence of a bacteriophyll A protein from *Prosthecochloris aestuarii* refined at 1·9 Å resolution. *J. Mol. Biol.* **188**, 443–454.

Wilmot, C. M. & Thornton, J. M. (1988). Analysis and prediction of the different types of β-turn in proteins. *J. Mol. Biol.* **203**, 221–232.

Wilmot, C. M. & Thornton, J. M. (1990). β-Turns and their distortions: a proposed new nomenclature. *Protein Eng.* **3**, 479–493.

Woldawer, A., Borkakoti, N., Moss, D. S. & Howlin, B. (1986). Comparison of two independently refined models of ribonuclease-A. *Acta Crystallogr. sect. B*, **42**, 372–387.

Edited by A. Klug