Protein Engineering vol.12 no.10 pp.811–814, 1999

SHORT COMMUNICATION

Novel protein structural motifs containing two-turn and longer 3_{10}-helices

Lipika Pal and Gautam Basu[1]

Department of Biophysics, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700 054, India

[1]To whom correspondence should be addressed; email: gautam@boseinst.ernet.in

The 3_{10}-helix constitutes a small but significant fraction of secondary structural elements in proteins. Protein data base surveys have shown these helices to be present as α-helical extensions, in loops and as connectors between β-strands. The present work focuses on two-turn and longer 3_{10}-helices where we establish that two-turn and longer 3_{10} helices, unlike the more abundant single-turn 3_{10}-helices, frequently occur independent of any other contiguous secondary structural elements. More importantly, a large fraction of these independent two-turn and longer 3_{10}-helices, along with α-helices and β-strands, are found to form novel super-secondary structural motifs in several proteins with possible implications for protein folding, local conformational relaxation and biological functions.

Keywords: 3_{10}-helix/α-helix/secondary structure/structural motif

Introduction

Although the α-helix is the most ubiquitous helical secondary structure in proteins, several recent studies have demonstrated that amino acid residues in proteins can also significantly populate the non-canonical 3_{10}-helical conformation (Baker and Hubbard, 1984; Barlow and Thornton, 1988; Karpen *et al.*, 1992; Blundell and Zhu, 1995; Doig *et al.*, 1997). The 3_{10}-helix, suggested to be the intermediate between a nascent helix and an α-helix (Millhauser, 1995) and observed in simulation studies of α-helix melting (Fan *et al.*, 1991; Soman *et al.*, 1991), has also been demonstrated to have functional roles in several proteins (McPhalen *et al.*, 1992; Kavavaugh *et al.*, 1993, Kostrikis *et al.*, 1994; De Guzman *et al.*, 1998; Hashimoto *et al.*, 1998). The general consensus from analysis of protein helices present in the protein data bank (PDB; Bernstein *et al.*, 1977) is that 3_{10}-helices comprise of only 3–4% of all residues and are short. They have also been reported to occur frequently at the termini of α-helices (Baker and Hubbard, 1984; Barlow and Thornton, 1988) and as connectors between two β-strands (Barlow and Thornton, 1988).

Since the majority of 3_{10}-helices in proteins are very short, comprising of three residues (or one-turn) only, the results of PDB analyses of 3_{10}-helices have almost always been determined by these single-turn helices. However, with increasing number of structures in PDB, the number of longer (two-turn or more) 3_{10}-helices have grown to a small but a finite size. Therefore, as the number of known protein structures grow, there is a need to re-examine the role of 3_{10}-helices in proteins, especially the longer ones, in the context of their immediate structural environment. The present analysis, where we focus exclusively on two-turn and longer 3_{10}-helices in proteins, establishes that unlike the single-turn 3_{10}-helices, two-turn and longer 3_{10}-helices in proteins mostly occur independent of any contiguous α-helix or other secondary structural element (SSE), and often as part of a super-secondary structural motif (SSSM), defined as a set of sequence-contiguous SSEs that pack into well-defined three-dimensional super-secondary structural or folding units (Efimov, 1984).

Materials and methods

The December 1997 pdb_select data set (Hobohm and Sander, 1994), with less than 25% sequence identity, was used in this analysis; this set was further reduced to the final set of 267 protein chains from the PDB by the following screening criteria: (a) resolution $\leqslant$ 2.0 Å and (b) R-factor $\leqslant$ 20%. Although these stringent criteria result in a small data set, it improves the quality of the helices studied, as required in this study. Secondary structure assignments were made using the DSSPprogram (Kabsch and Sander, 1983). This procedure yielded a total of 267 low sequence-identity and high resolution protein (chain) structures which contained 40 two-turn or longer 3_{10}-helices. The organization of structural motifs containing the 3_{10}-helices were identified by visual inspection along with the DSSP classification of residues. To confirm the results on the small data set and to calculate amino acid composition of 3_{10}-helices, a larger data set, the March 1999 culledpdb data set (Hobohm *et al.*, 1993), with less stringent screening criteria ($<$30% sequence identity and $\leqslant$2.5 Å resolution) was also used. Over-representation of amino acids in the helices were estimated (5% significance level) as described elsewhere (Karpen *et al.*, 1992).

Results and discussion

Quality of the helices

The average backbone dihedral angles of the 40 two-turn and longer 3_{10}-helices were found to be ϕ = –69.3 ($\pm$20.9) and ψ = –18.1 ($\pm$19.7). This compares well with a recent analysis with a larger dataset (Smith *et al.*, 1996) where the corresponding values were ϕ = –62.8 ($\pm$38.0) and ψ = –16.5 ($\pm$34.7). The average backbone dihedral angles for the N-cap and the C-cap residues were [ϕ = –90.1 ($\pm$60.3), ψ = 82.3 ($\pm$73.5)] and [ϕ = –68.5 ($\pm$63.6), ψ = 67.5 ($\pm$87.2)] respectively. The average pitch of the helices were found to be 5.88($\pm$0.29) Å. These average values as well as an inspection of individual backbone dihedral angles for the helices show that these are indeed well formed 3_{10}-helices and not just a distorted α-helix.

Amino acid composition of 3_{10}-helices

The larger culledpdb data set, consisting of 1085 protein chains, was used to determine the amino acid composition of 3_{10}-helices. The number of three, four, five, six, seven, eight, nine and 10 residue 3_{10}-helices present in this set are 2365, 396, 108, 94, 21, 8, 4 and 5 respectively. Here we summarize the results for amino acid over-representation where the helices

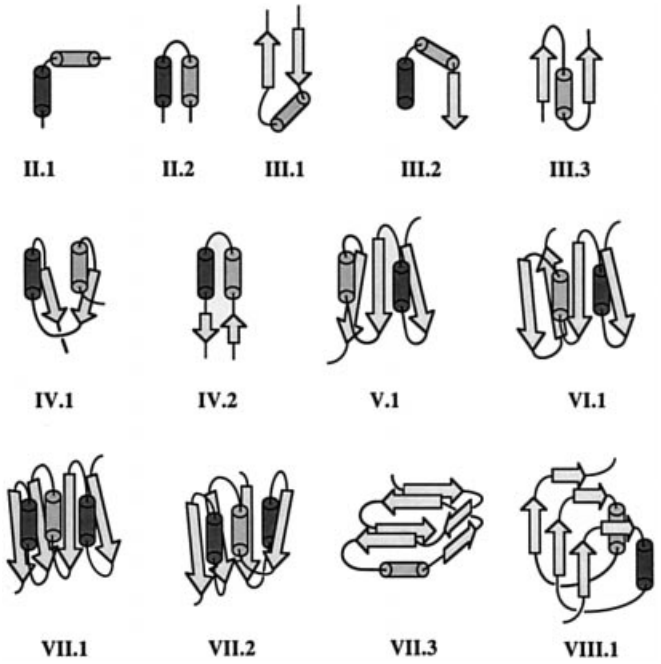

Fig. 1. A schematic representation of all the SSSMs containing a 3_{10}-helix found in this work. α- and 3_{10}-helices are depicted by dark and light barrels respectively.

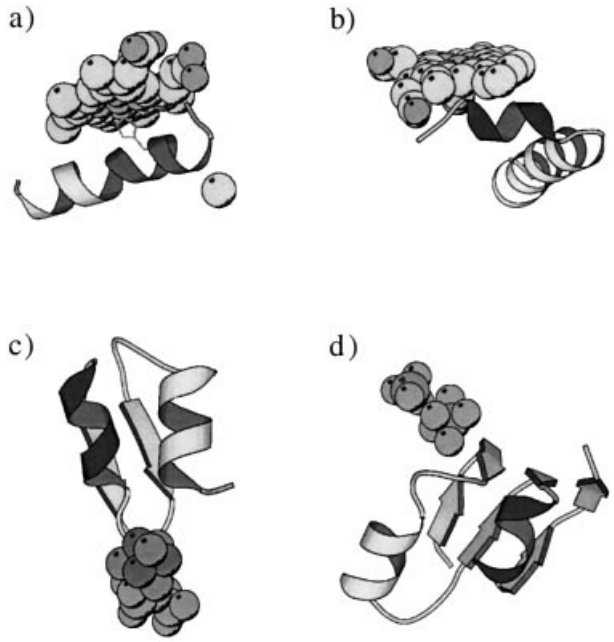

Fig. 2. Cartoon representation of a select set of two turn 3_{10}-helices in the context of their structural environment generated by MOLSCRIPT (Kraulis, 1994). 3_{10}-helices are darker than the α-helices and all hetero atoms are shown by cpk model: (**a**) unusual 3_{10}-helical termini of an α-helix in peroxidase coordinating the heme Fe atom; the lone hetero atom is calcium and the side chain of the distal His coordinated to heme is shown in wireframe model; (**b**) co-planar heme–3_{10}-helix (present as motif II.1) interaction in myoglobin (ferric); (**c**) the interaction of motif IV.1 in aconitase with the Fe4-S4 cluster and nitroisocitrate; and (**d**) motif VIII.1 in xanthine-guanine phosphoribosyltransferase with a sulphate and a magnesium ion (with five co-ordinated water molecules).

are classified as short (3–5 residues) and long (6 and more residues). N-cap (N0): Asp, Asn, His, Pro (short) and Asp, Ser (long). N1: Pro, Ala, Gly, Trp (short), Pro, Met (long). N2: Glu, Asp, Ser, Asn, His, Trp (short), Asp, Glu, Gly, Pro (long). N3: Asp, Glu, Phe, His, Leu, Lys, Asn, Gln, Tyr, Ser, Trp, Ala (short), Trp, Tyr, Asp, Glu, Met (long). C-cap (C0): Leu, Ile, Phe, Val, Cys, Gly (short), Phe, Leu, Ile (long). The results on the short 3_{10}-helices are almost in accordance with a previous study (Karpen *et al.*, 1992) performed exclusively on 77 three-residue 3_{10}-helices. While an amino acid composition study is important, because of its statistical nature, its application to a limited number of 3_{10}-helices is prone to errors. However, our results confirm the reported amino acid composition trend of the more abundant short 3_{10}-helices (Karpen *et al.*, 1992) and points to the interesting variation of amino acid composition as a function of chain length, which should be looked at more carefully only when a larger dataset becomes available before any definite conclusion can be made.

Two-turn and longer 3_{10}-helices present in SSSM

When the 3_{10}-helices from the small dataset were examined carefully, in more than 50% of cases (25 structures) they showed an unprecedented tendency to occur as part of a SSSM. A summary of the observed SSSMs are shown schematically in Figure 1 and a few representative 3_{10}-helices are shown in Figure 2 in the context of their immediate structural environments. The remaining 15 helices were found to occur as termini of α-helices (6), termini of the protein chain (1) or part of a long loop (8). Table I summarizes all the 3_{10}-helices and proteins studied in this work.

3_{10}-Helices contiguous with a SSE and present in long loops

The 3_{10}-helix was found contiguous with an α-helix in five proteins. In peroxidase, the contiguous α-helical and 3_{10}-helical stretches in the entire helix are of equal length. The 3_{10}-helix contributes the distal histidine coordinating the heme, suggesting an important indirect structural role related to protein function as shown in Figure 2a. In acetohydroxy acid isomeroreductase only one turn of the α-helix was contiguous to the two turn 3_{10}-helix. This is unusual as more often the α-helix is longer than the 3_{10}-helix. Three other cases of α-helical termini were more typical of a long α-helix with a short 3_{10}-helical termini. Unlike the above case, in phosphoribosyl anthranilate isomerase, the two turn 3_{10}-helix is contiguous with one more turn of 3_{10}-helix. In only one protein was a C-terminal 3_{10}-helix found, that in 2Fe-2S ferredoxin.

As an independent SSE, the occurrence of 3_{10}-helices in long loops have been mentioned earlier (Martin *et al.*, 1995). We found eight cases where 3_{10}-helices were present in long loops. In half of these, the 3_{10}-helix acts as a connector between two SSSMs. For example, in glutathione synthetase, a 3_{10}-helix connects two SSSMs. When not a connector, the loops are often part of a SSSM and exhibit a functional role, like the metal binding site in leucine aminopeptidase.

SSSMs containing a 3_{10}-helix

The simplest SSSM containing a 3_{10}-helix is the helix–helix motif comprising of one α- and one 3_{10}-helix. These occur either as a $3_{10}/\alpha$ corner (motif II.1) or a $3_{10}/\alpha$ hairpin (motif II.2). The II.1 motif is mostly present in heme proteins with a globin-like fold exhibiting strong co-planar interaction with the heme. A typical case, that of myoglobin, is shown in Figure 2b. The only non-heme protein containing the II.1 motif is glycogen phosphorylase where, unlike in the globin counterparts, the 3_{10}-helix is attached to the N-terminal end of the α-helix. Of the four cases where the motif II.2 was found, no SSEs were found close to the motif (sequence or spatial) suggesting inherent stability of the motif, as an

Table I. Structural environment around two-turn and longer 3_{10}-helices

Protein name	Code[a]	Residue number[b]	SSSM[c,d]	Connectivity[e,f]
3-Ketoacetyl-CoA thiolase	1afwA	312–317	α-t	-NC-α
Peroxidase	1arv	181–186	α-t	α-NC-
Methane monooxygenase hydroxylase	1mtyB	336–341	α-t	α-NC-
P-Hydroxybezoate hydroxylase	1pbe	322–327	α-t	-NC-α
Acetohydroxyacid isomeroreductase	1yveI	112–117	α-t	-NC-α
Phosphoribosyl anthranilate isomerase	1nsj	145–150	3_{10}-t	3_{10}-NC-
2Fe-2S Ferredoxin	1doi	121–126	ch-t	ch-NC-
Glutathione synthetase	1gsa	126–131	loop	α-NC-α
Aconitase	8acn	685–690	loop	β-NC-β
Transducin-alpha	1tag	201–208	loop	β-NC-β
Oligo-peptide binding protein	2olbA	155–160	loop	β-NC-α
L-Fuculose-1-phosphate aldolase	1fua	113–118	loop	α-NC-β
Arabinose operon regulatory protein	2arcA	102–107	loop	β-NC-β
Thymidylate synthase	1tys	133–140	loop	β-NC-β
Leucine aminopeptidase	1lcpA	267–272	loop	β-NC-α
Hemoglobin thionville	1babB	36–41	II.1	α-NC-
Hemoglobin I (carbon monoxy)	3sdhA	45–54	II.1	α-NC-
Leghemoglobin(oxy)	2gdm	38–43	II.1	α-NC-
Myoglobin (ferric)	2fal	37–42	II.1	α-NC-
Glycogen phosphorylase	1gpb	515–524	II.1	-NC-α
Chitobiase	1qba	703–708	II.2	α-NC-
Amylase	1smd	274–279	II.2	α-NC-
Endoglucanase V	2eng	153–158	II.2	-NC-α
Aconitase	8acn	574–579	II.2	-NC-α
T-Cell surface glycoprotein CD4	1cdy	59–64	III.1	β-NC-β
Streptavidin	1sriA	116–121	III.1	β-NC-β
Pseudoazurin	1zia	79–84	III.1	β-NC-β
Triosephosphate isomerase	5timA	156–161	III.2	α-NC-β
Rubisco	5rubA	252–257	III.2	α-NC-β
Endo-β-*N*-acetylglucosaminidase F1	2ebn	23–28	III.3	β-NC-β
Old yellow enzyme	1oyc	309–314	IV.1	-NC-β
Aconitase	8acn	168–174	IV.1	β-NC-β
Cellulase CelC	1ceo	240–248	IV.2	α-NC-β
Cellulase CelC	1ceo	179–184	V.1	β-NC-β
Klebsiella aerogenes urease	2kauC	284–289	V.1	β-NC-β
Deoxyribonuclease I	3dni	178–183	VI.1	β-NC-α
Dienelactone hydrolase	1din	149–158	VII.1	β-NC-β
Ascorbate oxidase	1aoZA	466–471	VII.3	β-NC-β
Flavodoxin	2fcr	40–45	VII.2	β-NC-β
Xanthine-guanine phospho-ribosyltransferase	1nulA	116–121	VIII.1	β-NC-β

[a]Three letter pdb code is followed by the chain ID where relevent.
[b]Residues forming the 3_{10}-helix.
[c]SSSMs are shown in Figure 1 and other conformations are discussed in the text.
[d]ch and t represent chain and termini respectively.
[e]Relevent sequence contiguous SSEs at the C- or N-termini of the 3_{10}-helix, represented by the '-NC-' unit, are indicated by α (helix) or β (strand).

independent folding unit. Only in one case, that of endoglucanase V, did a disulphide bond covalently connected the two helices.

A single 3_{10}-helix also forms a SSSM with two other SSEs. The 3_{10}-helix occurs as an anti-parallel β-strand connector (motif III.1) in three proteins. In streptavidin, the anti-parallel β-strands are part of a β-barrel with Trp120 from the 3_{10}-helix being part of the binding site. In T-cell surface glycoprotein, the anti-parallel β-strands are part of a larger β-sandwich super structure, and in pseudoazurin, His81, part of the 3_{10}-helix, and Cys78 and Met86, flanking the helix, bind the copper ion. Two proteins contain the 3_{10}-helix as the 'turn' in an α-β-hairpin motif (motif III.2)–rubisco and triosephosphate isomerase, both sharing the β/α (TIM) barrel fold. The other motif, 3_{10}-helix as a parallel β-strand connector (motif III.3), occurs in endo-β-*N*-acetylglucosaminidase where, strictly speaking, the 3_{10}-helix is part of the long loop described earlier.

A 3_{10}-helix was also found to occur in SSSMs containing more than three SSEs as described here. In old yellow enzyme and aconitase, a pair of 3_{10}- and α-helices pack against a pair of parallel β-strands (motif IV.1). In both cases, a loop connecting the SSEs in the motif exhibits strong binding activity or is part of the active site. Figure 2c shows the interaction of motif IV.1 of aconitase with the Fe4-S4 cluster. In cellulase CelC, an α-helix–3_{10}-helix hairpin motif forms a single layer SSSM with a pair of short antiparallel β-strands (motif IV.2). Cellulase CelC exhibits yet another two layer motif (motif V.1) containing three parallel β-strands and a pair of 3_{10}- and α-helices which recurs in *Klebsiella aerogenes* urease as well. In both cases several residues that are part of the motif's loops and turns participate in the active site.

Deoxyribonuclease I exhibits a two layer motif (motif VI.1) containing two pairs of anti-parallel β-strands and a pair of 3_{10}- and α-helices where several loop residues and a β-strand residue are involved in oligopeptide binding. Four parallel β-strands along with three helices, two α-helices and one 3_{10}-helix, form a two layer motif (motif VII.1) in dienelactone hydrolase, where, like the previous motifs, two residues in the loop of the motif form part of the active site. The same set of SSEs, with different connectivity, form a three layer motif

(motif VII.2) in flavodoxin where a part of the motif, away from the 3_{10}-helix, is involved in FMN binding. Three pairs of anti-parallel β-strands form a pseudo barrel SSSM (motif VII.3) along with a 3_{10}-helix in ascorbate oxidase, where three His residues, contributed by a β-strand and a turn, are part of the active trinuclear copper site. In xanthine-guanine phosphoribosyltransferase, three parallel β-strands, each 90° bent via a turn, form a motif (motif VIII.1) along with a pair of 3_{10}- and α-helices, where Asp89 in the turn is the magnesium ion interaction site, as shown in Figure 2d.

Implications for SSSMs containing a 3_{10}-helix

Although we report only a small number of SSSMs containing two-turn and longer 3_{10}-helices in this work, upon analyzing the larger culledpdb dataset with less stringent structural quality, the number of SSSMs containing 3_{10}-helices grew. The larger dataset contained 132 two-turn and longer 3_{10}-helices where 63% 3_{10}-helices occur as part of a SSSM. The remaining 3_{10}-helices were found to occur as termini of α-helices (18%), termini of the protein chain (4%) or part of a long loop (15%). This clearly establishes the high propensity of two-turn and longer 3_{10}-helices to be present in SSSMs. It should be noted that these SSSMs, especially III–VIII, are not meant to represent typical and frequently observed motifs containing a 3_{10}-helix. Rather, they represent structural motifs within which the independent 3_{10}-helices occur, and therefore, even for single occurrences, they demonstrate the diversity of structural environment in which two-turn and longer 3_{10}-helices may occur. On the other hand, SSSM II can be considered as a 'conserved' SSSM that occurs again and again in unrelated proteins. A simple sequence and structural analysis of these helices did not yield any clear statistically significant trend dictating their occurrences or stabilities, probably due to the small dataset used. Specific interactions, from within the protein and from solvent and ligand molecules, possibly account for the stability of these 3_{10}-helices. Only a protein-specific and case-by-case detailed atomic level examination would reveal if the presence of the 3_{10}-helices imposes any crucial structural constraints that ultimately translates into a functional role.

These motifs are novel in that they contain a 3_{10}-helix in place of the more expected α-helix. It is remarkable that in the majority of the representative SSSMs illustrated here, either the 3_{10}-helix itself (direct) or some part of the motif (indirect) was found to have some functional role. A natural question that arises at this point is: 'Why did a particular motif substitute a 3_{10}-helix for an α-helix?' As demonstrated from computer simulations (Tirado-Rives *et al.*, 1993; Basu *et al.*, 1994; Huston and Marshall, 1994), energetically the two helical forms are very close, both in terms of inter-helical free energy difference $\Delta G°$ and free energy of activation $\Delta G^{‡}$. Experimental data also suggest that they are inter-convertible under suitable conditions (Li *et al.*, 1997; Rigby *et al.*, 1997) or that the 3_{10}-helix may be an intermediate in α-helix formation (Sundaralingam and Sekharudu, 1989). At the same time, structural motifs are thought to be independent folding units (Efimov, 1984, 1997). This is because identical structural motifs are often found to occur in unrelated proteins. Recent NMR studies of protein folding intermediates also lend support to such hypothesis (Barber *et al.*, 1996). In the context of folding and stability, these motifs seem to have stabilized some conformational intermediate containing a 3_{10}-helix during their folding pathway. Further, since the motifs often exhibit direct or indirect functional roles, they could also be special in that they can undergo local conformational relaxation (3_{10}-helix ⇋ α-helix) under structural constraints arising from their environment like binding a ligand. It will be interesting to see, perhaps from computer simulations, the conformational pathways of folding of these motifs and if these motifs indeed show a 3_{10}-helix ⇋ α-helix transition under suitable conditions.

Acknowledgements

We would like to thank D.Pal, P.Chakrabarti, K.Kinoshita and K.Mizuguchi for their critical comments. K.Kinoshita kindly prepared Figure 2. Part of this work was done at the Bioinformatics Centre, Bose Institute, with financial support from CSIR, India.

References

Baker,E.N. and Hubbard,R.E. (1984) *Prog. Biophys. Mol. Biol.*, **44**, 97–179.
Basu,G., Kitao,A., Hirata,F. and Go,N. (1994) *J. Am. Chem. Soc.*, **116**, 6307–6315.
Barber,E., Barany,G. and Woodward,C. (1996) *Folding Des.*, **1**, 65–76.
Barlow,D.J. and Thornton,J.M. (1988) *J. Mol. Biol.*, **201**, 601–619.
Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,N. (1977) *J. Mol. Biol.*, **112**, 535–542.
Blundell,T.L. and Zhu,Z.-Y. (1995) *Biophysical Chem.*, **55**, 167–184.
De Guzman,R.N., Wu,Z.R., Stalling,C.C., Pappalardo,L., Borer,P.N. and Summers,M.F. (1998) *Science*, **279**, 384–388.
Doig,A.J., Macarthur,M.W., Stapley,B.J. and Thornton,J.M. (1997) *Protein Sci.*, **6**, 147–155.
Efimov,A.V. (1984) *FEBS Lett.*, **166**, 33–38.
Efimov,A.V. (1997) *Proteins Struct. Funct. Genet.*, **28**, 241–260.
Fan,P., Kominos,D., Kitchen,D.B. and Levy,R.M. (1991) *Chem. Phys.*, **158**, 295–301.
Hashimoto,Y., Kohri,K., Kaneko,Y., Morisaki,H., Kato,T., Ikeda,K. and Nakanishi,M. (1998) *J. Biol. Chem.*, **26**, 16544–16550.
Hobohm,U., Scharf,M. and Schneider,R. (1993) *Protein Sci.*, **1**, 409–417.
Hobohm,U. and Sander,C. (1994) *Protein Sci.*, **3**, 522–524.
Huston,S.E. and Marshall,G.R. (1994) *Biopolymers*, **34**, 75–90.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Kavavaugh,J.S., Moo-Penn,W.F. and Arnone,A. (1993) *Biochemistry*, **32**, 2509–2513.
Karpen,M.E., DeHaseth,P.L. and Neet,K.E. (1992) *Protein Sci.*, **1**, 1333–1342.
Kostrikis,L.G., Liu,D.J. and Day,L.A. (1994) *Biochemistry*, **33**, 1694–1703.
Kraulis,P.J.J. (1994) *Appl. Crystallogr.*, **24**, 946–950.
Li,T., Horan,T., Osslund,T., Stearns,G. and Arakawa,T. (1997) *Biochemistry*, **36**, 8849–8857.
Martin,A.C.R., Toda,K., Stirk,H.J. and Thornton,J.M. (1995) *Protein Engng*, **8**, 1093–1101.
McPhalen,C.A., Vincent,M.G., Picot,D., Jansonius,J.N., Lesk,A.M. and Chothia,C. (1992) *J. Mol. Biol.*, **227**, 197–213.
Millhauser,G.L. (1995) *Biochemistry*, **34**, 3873–3877.
Rigby,A.C., Baleja,J.D., Li,L., Pedersen,L.G., Furie,B.C. and Furie,B. (1997) *Biochemistry*, **16**, 15677–15684.
Smith,L.J., Bolin,K.A., Schwalbe,H., MacArthur,M.W., Thornton,J.A. and Dobson,C.M. (1996) *J. Mol. Biol.*, **255**, 494–506.
Soman,K.V., Karimi,A. and Case,D.A. (1991) *Biopolymers*, **31**, 1351–1361.
Sundaralingam,M. and Sekharudu,Y.C. (1989) *Science*, **244**, 1333–1337.
Tirado-Rives,J., Maxwell,D.S. and Jorgensen,W.J. (1993) *J. Am. Chem. Soc.*, **115**, 11590–11593.

Received April 23, 1999; revised July 2, 1999; accepted July 2, 1999