Vincenzo Pavone*
Girolamo Gaeta
Angela Lombardi
Flavia Nastri
Ornella Maglio
Centro di Studio di Biocristallografia—CNR
Department of Chemistry
via Mezzocannone, 4
University of Napoli
80134 Napoli, Italy

Carla Isernia
Seconda Università di Napoli
Facoltà di Scienze Ambientali
via Arena 22
81100 Caserta, Italy

Michele Saviano
Centro di Studio di Biocristallografia—CNR
Department of Chemistry
University of Napoli
Napoli, Italy

Discovering Protein Secondary Structures: Classification and Description of Isolated α-Turns

Irregular protein secondary structures are believed to be important structural domains involved in molecular recognition processes between proteins, in interactions between peptide substrates and receptors, and in protein folding. In these respects tight turns are being studied in detail. They also represent template structures for the design of new molecules such as drugs, pesticides, or antigens. Isolated α-turns, not participating in α-helical structures, have received little attention due to the overwhelming presence of other types of tight turns in peptide and protein structures. The growing number of protein X-ray structures allowed us to undertake a systematic search into the Protein Data Bank of this uncharacterized protein secondary structure. A classification of isolated α-turns into different types, based on conformational similarity, is reported here. A preliminary analysis on the occurrence of some particular amino acids in certain positions of the turned structure is also presented. © *1996 John Wiley & Sons, Inc.*

INTRODUCTION

The complex macromolecular architecture of protein three-dimensional structure is characterized by well-defined elements such as α-helices and β-sheets.[1-5] These repetitive motifs have been recognized at the early stage of protein structural characterization and successively 3_{10}-helices,[5] β-

Received June 6, 1995; accepted September 26, 1995.
* To whom correspondence should be addressed.
Biopolymers, Vol. 38, 705–721 (1996)

CCC 0006-3525/96/060705-17

hairpins,[6–8] and tight turns[9–11] have been identified. In particular, hairpins and turns, which are generally considered as irregular secondary structures,[12,13] are claimed to be relevant structural determinants for protein–protein and protein–substrate interactions, for protein folding processes,[14,15] and for the formation of active site of enzymes.[16,17] Among these irregular structures, tight turns have received much attention and in particular γ- and β-turns have been studied in detail and precisely classified.[18–21] They occur frequently (especially β-turns) in proteins,[5] and in small linear and cyclic peptides.[11,22,23]

It is well known that the γ-turn is the smallest of these structures and it involves three amino acid residues. It is often stabilized by an intraturn hydrogen bond between the C′O group of the first residue and the NH group of the third residue. Two possible γ-turns may exist. They differ in the position of the C^{β} atom of the central residue and they display ϕ,ψ torsion angles around the central residue of approximately opposite sign.

β-Turns correspond to a chain reversal of larger size, compared to the γ-turn, which involves four amino acid residues. This folded structure may also be stabilized by a hydrogen bond between the C′O group of the first residue and the NH group of the fourth. β-Turns containing a *trans* peptide bond between the two central residues are mainly classified in two groups. Type I (I′) and type II (II′) (I′ and II′ refer to the mirror images of the backbone conformation of type I and II, respectively) differ mainly in the orientation of the peptide bond between the second and third residue, being rotated about 180°. Type III (III′) β-turn is a distortion of type I (I′) and corresponds to the basic repetitive unit of the 3_{10}-helix.

Isolated 5-[24] and α-turns[22] are little investigated especially for their lower occurrence in proteins and peptides. The 5-turn corresponds to a chain reversal involving six amino acid residues and it has been found at the C-terminal helical part of proteins[25] and peptides[26] and at the loop side of β-hairpins.[27] The typically considered α-turn is the basic repetitive unit of right-handed α-helix. This turn may also exist with opposite handedness and is characteristic of the basic repetitive unit of left-handed α-helix. It corresponds to a chain reversal involving five amino acid residues and may be stabilized by a hydrogen bond between the C′O group of the first residue and the NH group of the fifth. The α_L α-turn type was recently identified as a motif of α-helices C-capping.[28] No other type of α-turn has been identified or classified so far, to the best of our knowledge. Both these larger turns (5- and α-turns) are difficult to be recognized because they are often stabilized by internal hydrogen bonds typical of smaller tight turns (such as β-turns). In this respect their classification may be somehow ambiguous.

We have hypothesized, using simple molecular models, that isolated α-turns, with all *trans* peptide bonds, may also exist in four basic motifs, together with the corresponding backbone mirror images. They would differ, in analogy with β-turns, in a rotation of about 180° of either one of the two peptide bonds between the three central residues of the turn.

In order to verify this hypothesis, we have undertaken a systematic search of isolated α-turns into the Protein Data Bank and compared their structures, using a clustering procedure.

METHODS

The search for isolated α-turns was performed using an ANSI C-language program, written for this purpose, and running on a Personal Iris workstation 4D35 GT Turbo from Silicon Graphics.

Protein Selection

All protein x-ray structures[29,30] refined with an R factor ≤ 0.23 at a resolution ≤ 2.5 Å were initially selected.

The identification and classification of isolated α-turns was carried out using the following steps.

1. *All α-Turn Selection:* The criterion adopted for α-turn selection is based solely on the electrostatic interaction energy between dipoles as previously described[24]: those protein pentapeptide segments with a calculated energy ≤ -0.5 kcal/mol between C'_iO_i and HN_{i+4} dipoles were selected for further calculations (hydrogen atoms were fixed in their stereochemical expected positions and segments containing proline at position $i + 4$ were not considered). The large interval for the identification of α-turns is justified by the errors in protein structure positional parameters and by the unnecessary strong hydrogen bond to define this secondary structure. α-Turns in α-helices are included in the data set when using the above criteria.
   Preliminary Clustering: The α-turns containing all *trans* peptide bonds were grouped in eight possible different clusters on the basis of the ϕ_{i+1}, ϕ_{i+2}, and ϕ_{i+3} sign, corresponding to eight possible permutations; this criterion was chosen because no cluster was expected to be found around the borderline of $\phi = 0$, where α-amino acids possesses locally a considerable high potential energy. These clusters, formed by three pairs of consecutive ϕ,ψ angles, were named, according to the ϕ_{i+1}, ϕ_{i+2},

Table I Naming Scheme of α-Turns According to the Sign of the ϕ_{i+1}, ϕ_{i+2}, and ϕ_{i+3} Angles

Names	ϕ_{i+1}	ϕ_{i+2}	ϕ_{i+3}
I-α_{RS}	−	−	−
I-α_{LS}	+	+	+
II-α_{RS}	−	+	−
II-α_{LS}	+	−	+
I-α_{LU}	−	+	+
I-α_{RU}	+	−	−
II-α_{LU}	−	−	+
II-α_{RU}	+	+	−

and ϕ_{i+3} sign, as "X-α_{YZ} turns", where X = I,II depends on residue $i + 2$ and $i + 3$ ϕ angles ["I" if sign(ϕ_{i+2}) = sign(ϕ_{i+3}); "I" if sign(ϕ_{i+2}) ≠ sign(ϕ_{i+3})]; Y = R,L depends on the sign of residue $i + 3$ ϕ angle ("R" if $\phi_{i+3} < 0$; "L" if $\phi_{i+3} > 0$); Z = S,U depends on residue $i + 1$ and $i + 3$ ϕ angles ["S" if sign(ϕ_{i+1}) = sign(ϕ_{i+3}); "U" if sign(ϕ_{i+1}) ≠ sign(ϕ_{i+3})]. The rules for this naming scheme are summarized in Table I. Figure 1 evidences the rational for this naming scheme. In fact, "R" or "L" represent a clockwise (right-handed) or a counterclockwise (left-handed) chain reversal, when viewing the structure down the C'_iO_i bond, "U" or "S" represents a U-shaped or screw-like-shaped turn,[31] and "I" or "II" discriminate structures with similar topology but different conformations.

Furthermore, a few *cis* peptide bond containing sequences were detected and therefore they were not analyzed automatically. Those containing a *cis* peptide bond were assigned to a cluster arbitrarily named I-α_C.

2. *Isolated α-Turn Selection:* Cluster I-α_{RS} and I-α_{LS} are characterized by a triplet of consecutive ϕ angles (ϕ_{i+1}, ϕ_{i+2}, ϕ_{i+3}) either all negative or positive. Therefore these clusters also contain α-turns included in either right- or left-handed α-helices. In order to avoid overlapping with α-turns belonging to α-helices, isolated turns were selected within cluster I-α_{RS} and I-α_{LS}, if two consecutive hydrogen bonds (defined here) $C'_iO_i \leftarrow HN_{i+4}$ and $C'_{i+1}O_{i+1} \leftarrow HN_{i+5}$ would not exist. Despite this reduction, several turns appeared to be located within distorted helices. Some of these turns were proline containing helices. These structures were judged as α-helical because the screw-like motif was clearly visible with ribbon drawings and they were considered distorted because the hydrogen-bond pattern of α-helices was lacking. These turns were excluded from further calculations.
3. *Reduction for Data Base Redundancy:* In order to avoid redundancy in the data base (deriving for instance from different x-ray structures of the same protein, different mutants, etc.), identical α-turns were taken once, keeping those with the strongest $i \rightarrow i + 4$ dipole interaction. Identity was ascertained by the pentapeptide sequence overlapping with the same residue numbers. A few sequences with different sequence numbers, belonging to structurally related proteins, were finally removed manually.
4. *Analysis of the Final Clusters and Distorted α-Turns:* The mean values of the ϕ_j,ψ_j ($j = i + n$; $n = 1, 2, 3$) torsion angles were evaluated for each reduced cluster. An arbitrary circular area in the ϕ,ψ plane with a radius of ±45° around these mean values was chosen. Sequences displaying any ϕ_j,ψ_j angle coordinates outside this area were considered distorted turns and not included in further calculations, except those cases in which only one of the three pairs of ϕ,ψ angles displayed values not exceeding ±90° from the mean. The average ϕ,ψ angles were then iteratively recalculated.

Analysis of Amino Acids Preferences

The analysis of amino acid compositions in the isolated α-turn identified was made, accounting for the different frequency of each residue [$F_{(r)}$] in the final data base of selected protein structures. The frequency for each amino acid residue in either one of the $i + 1$, $i + 2$, and $i + 3$ positions for all the α-turns [$F^t_{(r)}$] was evaluated. The function $R = [(F^t_{(r)} - F_{(r)})/F_{(r)}]$ was then used to better

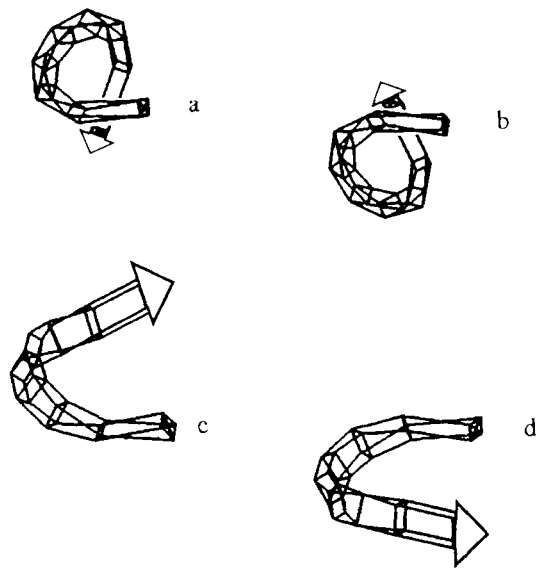

FIGURE 1 Ribbon structures of pentapeptidic chain reversals view down the $C'_i{=}O_i$ bond: (a) right-handed (clockwise) chain reversal (R) with a screw-like shape (S), (b) left-handed (counterclockwise) chain reversal (L) with a screw-like shape (S), (c) right-handed chain reversal (R) with a U-like shape (U), and (d) left-handed chain reversal (L) with a U-like shape (U). Arrows indicate the direction from residue i to $i + 4$.

Table II Summary of α-Turn Selection from the Protein Data Bank: (1) all α-Turn Including Those Belonging to α-Helices; (2) Nonhelical α-Turns; (3) Unique α-Turns Obtained After Reduction for Protein Data Base Redundancy; (4) Final Selection

Cluster	Step 1	Step 2	Step 3	Step 4
I-α_{RS}	28772	554	251	238
I-α_{LS}	33	11	5	5
II-α_{RS}	108	108	41	39
II-α_{LS}	91	91	8	8
I-α_{LU}	37	37	17	14
I-α_{RU}	151	151	30	28
II-α_{LU}	23	23	19	8
II-α_{RU}	26	26	9	9
I-α_c	31	31	9	7
Total	29272	1032	389	356
No. of proteins	439	439	199	193

visualize on bar diagrams the different propensities of each amino acid residue to be accommodated in the α-turns. Values of R close to zero are indicative of frequencies in the turn similar to those observed in any position of protein sequences, while positive or negative values represent higher or lower frequencies, respectively.

RESULTS

Four hundred sixty protein x-ray structures were selected from the entire protein data base (PDB) according to an adequate resolution and refinement. Protein structures numbering 21 are lacking of α-turns. These include 8 hydrolases (1ADA, 1HSC, 1PO5, 1TGL, 2BLM, 2SC2, 3CBH, 3HTC; PDB codes will be used throughout the text),[32-39] 2 erabutoxine (3EBX, 5EBX),[40,41] 2 oncogene proteins (3P21, 2P21),[42] 2 cytokines (31BI, 4I1B),[43] 2 oxytocin derivatives (1XY1, 1XY2),[44] γ-δ resolvase (1RSL),[45] glycosidase inhibitor (1HOE),[46] FKBP (1FKB),[47] alamethicin (1AMT),[48] and gramicidin A (1GMA).[49]

The selection of the various α-turns is summarized in Table II. Four hundred thirty-nine proteins contain 29,272 α-turns, which were identified using an energy cutoff for the interactions between C'_iO_i and HN_{i+4} dipoles. α-Turns numbering 29,241, corresponding to those pentapeptidic sequences with all-*trans* peptide bonds, were then grouped according to the signs of the ϕ_{i+1}, ϕ_{i+2}, and ϕ_{i+3} angles in eight different clusters. The remaining 31 pentapeptide sequences containing a *cis* peptide bond between residues $i + 1$ and $i + 2$ were assigned to a separate cluster. No other type of *cis* containing pentapeptide sequences with an α-turn structure was detected. Among all the α-turns, 28,218 corresponded to α-turns included in right-handed α-helices and 22 in small stretches of left-handed helices. These turns were removed from the corresponding clusters I-α_{RS} and I-α_{LS}. It is worth mentioning here that the N- and C-terminal helix ending by typical helical α-turns have been removed (they were considered as part of the helix and not as capping motif) and therefore they were not analyzed, while helices capping by nonhelical α-turns (belonging to clusters different from clusters I-α_{RS} and I-α_{LS}) were kept. The remaining 1032 isolated α-turns were reduced for redundancy deriving from the presence of several structurally related proteins in the data base, giving 389 α-turns within 199 protein structures.

Finally, a total of 356 isolated α-turns was detected in 193 protein structures and they are listed in Table III. Seven of these contain a *cis* peptide bond. 33 out of 389 α-turns were rejected from the data set because they display larger differences from the average values of ϕ,ψ angles for each cluster. These turns were considered distorted turns and therefore unclassified (see Table III). They include also two turns in the homologous proteins 1CSE E209–E213[50] and 3TEC E213–E217.[51] They contain a *cis* peptide bond between residues $i + 1$ and $i + 2$, but they show ϕ,ψ torsion angles different from those observed for cluster I-α_C. Thus, they were not assigned to a specific cluster because of the very low occurrence among the analyzed protein structures.

Figure 2a–h shows the Ramachandran plots of the ϕ_j,ψ_j angles ($j = i + n$; $n = 1, 2, 3$) for the clusters containing all-*trans* peptide bonds, while Figure 2i shows the plot of the cluster containing one *cis* peptide bond. The ϕ_j,ψ_j mean values with the estimated standard deviations are reported in Table IV for each cluster. The averaged conformations are also shown as stereo view in Figure 3a–i. The superimpositions of all the backbone atoms (from C_i^α to C_{i+4}^α) with these average structures are reported in Figure 4a–i. The averages of the superimposition root mean square displacements (RMSD) of all backbone atoms (from C_i^α to C_{i+4}^α) are reported along the diagonal of Table V for each cluster. Intercluster RMSDs including all backbone atoms from C_i^α to C_{i+4}^α and based on average conformations are also reported in Table V. Furthermore, the average structures calculated on orthogonal coordinates display bond geometry for all the residues (data not shown) in agreement with

Table III Protein Structures Containing Isolated α-Turns (PDB Codes Are Used), Grouped According to the Different α-Turn Type (the Corresponding Sequences Are Indicated as Single Letter Codes)

Protein	Turn		Sequence	Protein	Turn		Sequence	Protein	Turn		Sequence
					Cluster I-α_{RS}						
1aap B	24	28	DVTEG	1ova D	277	281	SSNVM	2er6 E	171	175	DTTAY
1alc	56	60	SNALW	1ova D	318	322	LSGIS	2fb4 H	205	209	HKPSN
1ald	79	83	FHETL	1p02 A	140	156	GRTTG	2fb4 H	61	65	ADSVK
1bbp B	171	175	SEAAC	1p06 A	55	59	AGHCG	2fb4 H	73	77	NDSKN
1bbp D	122	126	DEDKK	1p07 A	221B	224	PASQR	2fb4 L	27A	30	NIGSS
1bbp D	19	23	DWSNY	1paz	78	82	CTPHY	2fb4 L	91	95	DVSLN
1ca2	8	12	GKHNG	1pfk B	249	253	HIQRG	2fbj H	190	194	PAVEC
1ccr	2	6	SFSEA	1phh	285	289	GDAAH	2fcr	148	152	DMVND
1ccr	22	26	CAQCH	1ppd	82	86	YRNTY	2fcr	94	98	DAEGY
1cla	97	101	HQETE	1ppd	96	100	RSREK	2fd2	56	60	SEDEV
1cms	127	131	YPSLA	1prc C	161	165	THVER	2gbp	200	204	NANKI
1cms	311	315	DRANN	1prc C	321	325	RLKDY	2gcr	153	157	RYLDW
1cox	115	119	GSLVN	1prc C	82	86	SPQEG	2hhb C	113	117	LPAEF
1cox	391	395	NAGSG	1prc C	87	91	CTYCH	2ifb	95	99	RVDNG
1dhf B	152	156	DLEKY	1prc H	136	140	PLRVA	2lbp	103	107	APELT
1dtx	26	30	NQKKK	1prc H	174	178	DRSEH	2lbp	225	229	PEGVG
1er8 E	46	50	TASEV	1prc L	19	23	GDLFD	2ltn A	110	114	DKTTQ
1er8 E	57	61	TPSKS	1prc L	60	64	DPFAI	2ltn A	125	129	NAAWD
1fcb A	469	473	SIAEL	1prc M	3	7	YQTIY	2ltn A	167	171	NAATN
1fcb A	76	80	APEKK	1psg	34P	38P	ASKYF	2ltn A	54	58	DRETG
1fcb B	400	404	LKDKL	1rbp	16	20	DKARF	21ym	121	125	QAWIR
1fd2	35	39	HPDEC	1rbp	5	9	RVSSF	2mba	43	47	FADFK
1fnr	233	237	VSREQ	1rdg	14	18	DPAKG	2pfk B	54	58	RYSVS
1fxi B	75	79	LTCVA	1sdh B	47	51	TIGYF	2pfk D	53	57	DRYSV
1fxi C	68	72	QIQAG	1sgc	140	156	GSTTG	2prk	46	50	HPEFE
1gcr	25	29	LQPYF	1sgt	177	180	VANEE	2prk	79	83	SRTYG
1gd1 O	47	51	DSVHG	1sgt	185	186	YPDTG	2rhe	28	32	DIGSN
1gd1 P	209	213	GAAKA	1snc	46	50	HPKKG	2rhe	93	97	NDSLD
1gd1 P	83	87	AWGEI	1tec E	45	49	HPDLA	2rsp B	46	50	SEEDW
1gd1 Q	138	141	DPKAH	1tec I	57	61	NPGTN	2rus A	291	295	HGAVT
1gd1 Q	192	196	DLRRA	1tgs Z	23	27	GANTV	2rus B	75	79	DEARE
1gd1 R	198	202	AAAES	1tlp E	229	233	GVHIN	2sar B	7	11	CLSAL
1gd1 R	279	283	VSRDY	1tmb H	185	186C	KPDEG	2sdh B	51	55	FKRLG
1gox	345	349	SLKEI	1tmb H	95	98	NWREN	2tec E	57	61	DFVDN
1gox	81	85	QKMAH	1trm A	95	99	DRKTL	2tgd	95	99	NSNTL
1hgt I	60	64	PEEYL	1ubq	55	59	TLSDY	2tgp I	3	7	DFCLE
1hgt L	7	11	FEKKS	1ycc	14	18	CLQCH	2tim A	155	159	KKADW
1hne E	62A	65	NVRAV	1ypi B	213	217	NGSNA	2trx A	59	63	NIDQN
1hne E	95	99B	DPVNL	2act	85	89	TEENY	2trx A	8	12	TDDSF
1hrh A	447	451	NRETK	2aza B	117	121	HWAMM	2tsc A	134	138	NVGEL
1iht H	203	205	SPFNN	2ca2	34	38	DTHTA	2tsc A	37	41	NLQDG
1iht H	56	60	AHCLL	2cab	124	128	NSAKY	2tsc B	80	84	WDEWA
1ldm	200	204	VWSGM	2cab	181	185	PSTLL	2wgc B	27	31	GGDYC
1lhm	60	64	NSRYW	2cab	34	38	KTSET	3app	239	243	DSNAG
1lz1	46	50	NAGDR	2cab	8	12	DDKNG	3apr E	272	276	SPDSL
1lzt	79	83	PSCAL	2ccp	288	292	TLEEQ	3b5c	81	85	PDDRS
1ntp	95	99	DSNTL	2ccy B	72	76	PEAFG	3blm	106	110	SPILE
1omd	2	6	ITDIL	2cd4	51	55	LNDRA	3blm	50	54	DTKSG

(*Continued*)

Table III *(Continued from the previous page.)*

Protein	Turn		Sequence	Protein	Turn		Sequence	Protein	Turn		Sequence
1ova C	232	236	MASEK	2cga A	146	150	YTNAN	3blm	98	102	NKDDI
1ova C	376	380	HIATN	2csc	59	63	DPDEG	3bp2	67	71	NPYTN
1ova D	199	203	KDEDT	2cyp	58	62	DKHDN	3c1a	194	198	HHAVC
3cpp	77	81	DYRHF	4gr1	164	168	HESQI	5rub A	263	267	DGYVA
3dfr	148	152	NPALT	4gr1	300	304	LNKLG	5rub A	93	97	NITDG
3er5 E	240	244	SSSVG	4gr1	330	334	GDVCG	5rxn	29	33	DFKDI
3fgf	67	71	GVCAN	4gr1	471	475	SEELV	5sga E	219	223	NCRTG
3gbp	303	307	LSEFT	4htc H	60I	64	TENDL	6dfr	129	133	EPDDW
3gch	55	59	AAHCG	4mba	116	120	VASVA	6ldh	181	185	HSCSC
3lhm	122	126	RQYVQ	4pcy	87	91	HQGAG	6rxn	6	10	CNVCG
3lym	59	63	NSRWW	4pep	47	51	SLACS	7cpp	328	332	DEREN
3lzm	158	162	WDAYK	4rxn	14	18	DPEDG	7enl	346	350	VNQIG
3mcg 2	186	190	PEQWK	4rxn	39	43	CPLCG	7enl	432	436	HGDKL
3p2p A	17	21	HPLMD	4rxn	6	10	CTVCG	7est E	23	27	QRNSW
3p2p A	67	71	YPYTE	4sgb E	220	222A	NCSSG	7est E	55	59	AAHCV
3p2p B	112	116	NKEHK	4sgb I	5	9	NCCAG	7est E	98	100	DVAAG
3pal	78	82	TDAET	4tln	189	193	GEDVY	7gch	168	172	CKKYW
3pep	57	61	NPDDS	5acn	422	426	GPCIG	7gch	95	99	NSLTI
3pfk	221	225	AEGVG	5acn	451	455	GRNDA	7icd	230	234	KGNIM
3pfk	249	253	HVQRG	5acn	481	485	NPETD	7icd	268	272	NPNTG
3pfk	53	57	EVGDV	5acn	49	53	DPANQ	7icd	319	323	GIGIA
3rp2 A	173	180	EYKFQ	5acn	596	600	NIENR	7icd	342	346	APKYA
3rp2 B	165	169	EKACV	5acn	607	611	NAVTQ	7pti	24	28	NAKAG
3tec E	23	27	AWDIA	5acn	687	691	DYNKI	7rxn	14	18	DPAEG
451c	12	16	CVACH	5cpp	407	411	DPATT	7rxn	39	43	CPVCG
4cms	251	255	NLSYM	5enl	320	324	DDLTV	8adh	165	169	PLEKV
4enl	102	106	KSKLG	5i1b	86	90	DPKNY	8adh	283	287	QEAYG
4enl	202	206	GASAG	5mbn	43	47	FDRFK	8adh	342	346	LDPLI
4er1 E	270	274	GDYID	5p21	145	149	SAKTR	8dfr	152	156	DYKDF
3er2 E	294	298	SAGIG	5pcy	52	56	ASKIS	9pap	6	10	DWRQK
4er4 E	269	273	PGDYI	5pep	110	114	SFLYY				
4fd1	16	20	CVEVC	5pep	252	256	SIASL				
							Cluster I-α_{LS}				
1fcb B	372	376	NHGGR	3enl	55	59	KWMGK	4ccp	36	40	YDNYI
1gox	253	257	NHGAR	3fxn	56	60	MGDEV				
							Cluster II-α_{RS}				
1cox	453	457	LLNSA	1sgt	23	27	AQGEF	2mcg 1	28	32	DVGGY
1cse I	57	61	NPGTN	1snm	46	50	HPKKG	2mhr	114	118	YKGKL
1fkf	87	91	HPGII	1tec E	261	265	GTGTY	2pfk A	136	140	TIGFF
1gdl O	129	133	VMGVN	1trm B	172	176	YPGKI	2tec E	48	52	LAGKV
1gd1 O	267	271	LKGIL	1trm B	191	195	CQGDS	2tga	172	176	YPGQI
1gd1 R	221	225	LKGKL	1ypi A	26	30	RLNTA	2tgt	190	194	SCQGD
1ihs H	77	80	ERNIE	2aza A	114	118	FPGHW	2trx B	49	53	YQGKL
1mba	123	127	PAGAD	2cyp	82	86	NAGLQ	3pfk	136	140	TIGFD
1ova A	84	87A	LPGFG	2fb4 H	193	197	SLGTQ	3rp2 B	191	195	FMGDS
1p07 A	192A	195	GRGDS	2fb4 H	64	68	VKGRF	4htc H	191	195	CEGDS
1phh	49	53	EQGMV	2fd2	8	12	CIKCK	5acn	56	60	ERGKT
1prc C	211	215	LVGVS	2hmz B	109	113	YRGKI	7enl	237	241	HDGKV
1prc M	284	288	LTGTF	2lbp	289	293	TALER	8gch	191	195	CMGDS

(Continued)

Table III (*Continued from the previous page.*)

Protein	Turn		Sequence	Protein	Turn		Sequence	Protein	Turn		Sequence
					Cluster II-α_{LS}						
1176	27	31	IGIGH	1tlp E	106	110	YSQGY	2lbp	37	41	GGIKG
1ppd	181	185	WGENG	2act	188	192	WGEEG	4fxn	56	60	MGDEV
1tgs Z	145	149	KSSGT	2apr	11	15	YGNDI				
					Cluster I-α_{LU}						
1gcr	136	140	MPSYR	2fbj H	100	104	HYYGY	4ccp	35	39	EYDNY
1gcr	47	51	RPNYQ	2gcr	136	140	MPNYR	4pal	1	6	SFAGL
1gcr	8	12	DRGFQ	2gcr	95	99	REDYR	4tln	224	228	TQDNG
1s01	209	213	LPGNK	2pka X	93	95B	GFNLS	5dfr	9	13	AVDRV
1thr I	59	62	DIEQQ	2prk	212	216	WIGGS				
					Cluster I-α_{RU}						
1bbp C	142	146	TGEAK	1ova B	69	73	KDSTR	3grs	55	59	GGTCV
1cox	289	293	AGSVG	1pfk B	39	43	DGYLG	3mcg 2	185	189	TPEQW
1fcb B	207	211	EGEKD	1prc C	125	129	VAQTG	3pfk	39	43	HGYAG
1fcb B	442	446	YGRNG	1prc M	51	55	LGASG	4grl	424	428	ANKEE
1fkf	82	86	YGATG	1tmn E	158	162	QNESG	4ins B	19	23	CGERG
1gox	318	322	EGEAG	1ton	173	177	KDNVT	5p2p A	25	29	YGCYC
1gox	87	91	EGEYA	1ypi B	196	200	LGDKA	7tln	91	95	LSYDG
1l59	106	110	MGENG	2cts	238	242	HEGGN	7xia	345	349	DGLQA
1l62	106	110	MGEDG	2wgc B	46	50	CGSQA				
1l75	106	110	MGETG	3bc1	334	338	GGAYA				
					Cluster II-α_{LU}						
1gp1 B	75	79	NQFGH	2prk	149	153	QSSGV	3gch	35	39	DKTGF
1sgt	33	42	LSMGC	2wgc A	148	152	SKWGS	4htc L	14I	14M	SYIDG
1tec E	239	243	ASQGR	3c2c	109	112	KTLKK				
					Cluster II-α_{RU}						
1cms	11	15	YLDSQ	2fb4 L	48	52	YRDAM	2rhe	50	54	YYNDL
1gd1	300	304	IDGKM	2fbj L	48	52	YEISK	5pep	9	13	YLDTE
2alp	35	40	INNAS	2ltn C	100	104	YLGVF	8adh	140	144	FLGTS
					Cluster I-α_{C}						
1s01	166	170	GYPAK	2sni E	166	170	GYPGK	6cpp	98	102	FIPTS
1tec E	170	174	NYPAY	3blm	165	169	YEIEL				
2prk	169	173	YSPAS	6cpp	104	108	DPPEQ				
					Unclassified						
1fcb B	480	484	LSTLK	2gcr	112	116	LQDRF	1ihs H	145	149	KETWT
1ldm	217	221	TNKDK	1ova A	147	150	FQTAA	1lzt	98	102	IVSDG
1phh	164	168	ISRQS	1lzt	99	103	VSDGN	1rbp	63	67	LLNNW
2cab	137	141	KADGL	2rus A	441	445	PGDAD	1ypi A	25	29	ERLNT
2dhf A	61	65	PEKNR	1fcb A	21	25	KPDDC	2rus A	440	444	FPGDA
2fbj H	132	136	PPALS	2fxb	28	32	DEDGI	4fxn	40	44	IDELL
2fbj L	166	170	DSKDS	2tgd	184A	188	YLEGG	5pep	241	245	NSDGE
2gn5	23	27	GKPYS	2gn5	70	74	VGQFG	4fxn	73	77	ISTKI
3p2p A	119	123	DTKKY	2rus A	450	454	GWRKA	2fgf	25	29	CKNGG
4cpv	2	6	FAGVL	1cd4	30	34	NSNQI	1cse E	209	213	YPTNT
5rub B	393	397	GGGAF	2mcg 1	25	29	TSSDV	3tec E	213	217	YPTST

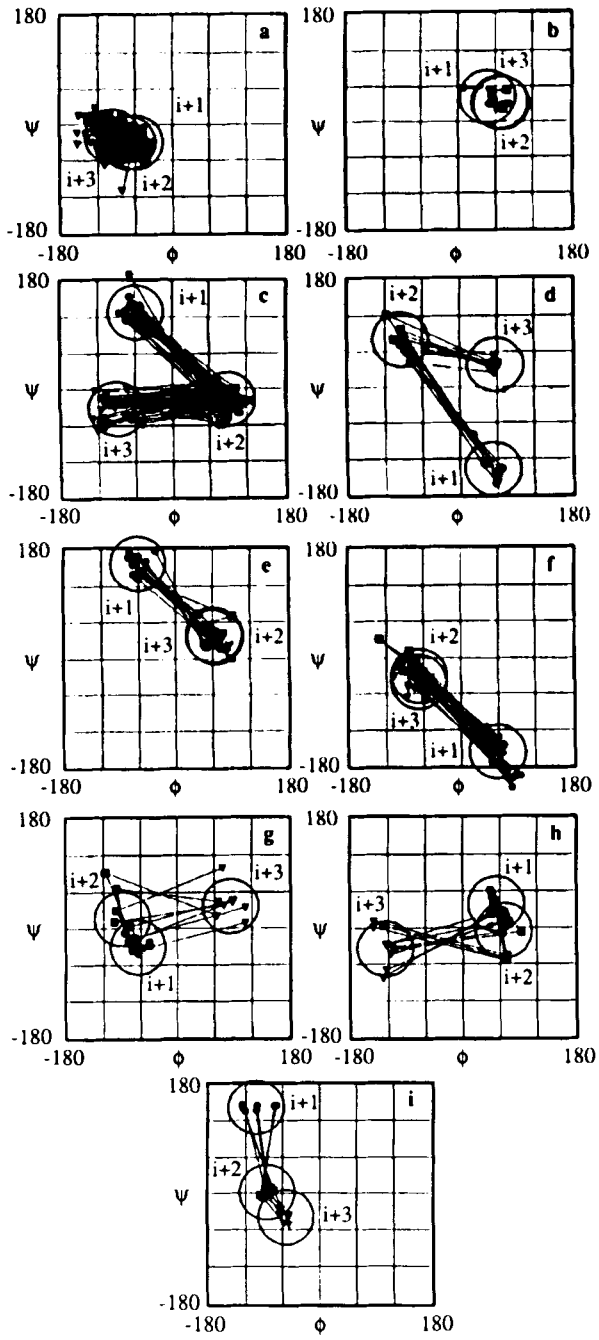

FIGURE 2 (a)–(i): Ramachandran plots of the ϕ_j,ψ_j angles ($j = i + n$; $n = 1, 2, 3$) for all the clusters. Circles with a radius of 45° centered on the average values of ϕ,ψ angles are also reported.

the expected values,[52] thus confirming the consistency of each family.

Table VI reports the hydrogen bond geometrical parameters between residues i and $i + 4$ for each type of α-turn (calculated on average structures), together with the electrostatic contributions between other dipoles. The hydrogen bond distance $O_i \cdots N_{i+4}$ is comprised between 2.8 and 3.6 Å, with an angle $N_{i+4} \cdots O_i = C'_i$ between 123° and 161°. The strongest interactions occur in type I-α_C and II-α_{LS} turns and the weakest correspond to type II-α_{LU} and I-α_{LS} turns. The analysis of dipole interactions indicates that generally these secondary structures are further stabilized by other hydrogen bond interactions. In particular, the C'_iO_i group is also a hydrogen bond acceptor of $N_{i+3}H$ group, except for type II-α_{LS} and I-α_C turn. In the case of type II-α_{RS} and II-α_{LU} turns, this interaction is even stronger than the corresponding interaction between residues i and $i + 4$. Type II-α_{LS} turn is instead further stabilized by a hydrogen bond between residues $i + 1$ and $i + 3$. The type I-α_C turn does not contain any other favorable hydrogen-bond interaction. A small stabilizing contribution also occurs between dipoles $C'_{i+1}O_{i+1}$ and HN_{i+4} in type I-α_{YZ} turns (Y = R,L; Z = S, U).

Figure 5a reports the frequency $F_{(r)}$ of each amino acid in the selected 193 protein structures. The various amino acids were ordered according to their hydrophobicity score.[53] Figure 5b shows R (defined above), representing the variation from the mean frequency, in either one of the three central positions of the α-turns. Figure 5c shows the deconvolution of R for positions $i + 1$, $i + 2$, and $i + 3$ of the α-turns.

DISCUSSION

The identification and classification of isolated α-turns clearly indicates the relevance of this irregular secondary structure. In fact, on average about two isolated α-turns were identified in each protein structure analyzed (356/193 turn/protein). Only 7 structurally unrelated protein structures (including three peptides) do not contain isolated α-turns.

The naming scheme used to distinguish structurally unrelated isolated α-turns is based solely on the sign of the ϕ_{i+1}, ϕ_{i+2}, and ϕ_{i+3} angles and it provides a complete description of the global topology of the turned structures, as far as it regards the right-handed or left-handed screw-like or U-like mode of chain reversal. The intrinsic chirality of protein chain folding by irregular secondary structures was rationalized: by the use of an α-turn, protein structures have preferentially a right-handed fold (90% of the observed α-turns) and they rarely fold in a left-handed manner. Screw-like motifs are also preferred. Type I-α_{RS} is the most frequently observed (about 70% of the detected α-turns).

The structural homology within each cluster, as

Table IV The ϕ_j,ψ_j Mean Values with the Estimated Standard Deviations for Each Cluster

Cluster	111 ϕ_{i+1} (°)	ψ_{i+1} (°)	ϕ_{i+2} (°)	ψ_{i+2} (°)	ϕ_{i+3} (°)	ψ_{i+3} (°)
I-α_{RS}	−60 ± 11	−29 ± 13	−72 ± 14	−29 ± 15	−96 ± 20	−20 ± 17
I-α_{LS}	48 ± 22	42 ± 14	67 ± 9	33 ± 14	70 ± 11	32 ± 12
II-α_{RS}	−59 ± 10	129 ± 15	88 ± 15	−16 ± 19	−91 ± 22	−32 ± 18
II-α_{LS}	53 ± 15	−137 ± 25	−95 ± 12	81 ± 23	57 ± 5	38 ± 8
I-α_{LU}	−61 ± 12	158 ± 15	64 ± 17	37 ± 21	62 ± 12	39 ± 8
I-α_{RU}	59 ± 18	−157 ± 31	−67 ± 17	−29 ± 20	−68 ± 12	−39 ± 12
II-α_{LU}	−65 ± 15	−20 ± 15	−90 ± 17	16 ± 44	86 ± 18	37 ± 27
II-α_{RU}	54 ± 8	39 ± 15	67 ± 13	−5 ± 31	−125 ± 11	−34 ± 32
I-α_C	−103 ± 23	143 ± 4	−85 ± 8	2 ± 6	−54 ± 6	−39 ± 9

can be seen qualitatively by inspection of Figures 2 and 4, and quantitatively from Table V, was ascertained by using several parameters. On average, the intracluster RMSD of all backbone atoms superimposition from C_i^α to C_{i+4}^α is about 0.28 Å, while the intercluster RMSD for the average structures is 1.1 Å. Smaller structural differences between X-α_{YU} and X-α_{YS} turns (X = I, II; Y = R, L) are present (intercluster RMSD for these average structures is about 0.55 Å).

The I-α_{RS} turn and I-α_{RU} turn are characterized by a right-handed chain reversal with a screw-like and U-like shape, respectively. A comparison, by the backbone atoms' superimposition from residue $i + 2$ to residue $i + 3$, of the type I-α_{RS} and I-α_{RU} turns average structures is reported as a stereo view in Figure 6a. The type I-α_{LS} turn and type I-α_{LU} turn correspond to about the mirror images of I-α_{RS} and I-α_{RU} turns, respectively, as far as the backbone conformation it regards. I-α_{LS} and I-α_{LU} turns are characterized by a left-handed chain reversal with a screw-like and U-like shape, respectively. A comparison of the type I-α_{LS} and I-α_{LU} turns is also reported as a stereo view in Figure 6b. The type I-α_{RS} and I-α_{LS} turns correspond approximately to the basic repetitive units of the right- and left-handed α-helices, respectively. Type I-α_{RU} and I-α_{LU} turns are partly similar to type I-α_{RS} and I-α_{LS} (intercluster RMSDs are 0.55 and 0.58 Å, respectively), but they differ in a rotation of about 90° of the peptide bond between residues i and $i + 1$. This different peptide bond orientation determines the screw-like or U shape of the turn.

II-α_{RS} and II-α_{RU} turns are characterized by a right-handed chain reversal with a screw-like and U-like shape, respectively. A comparison of type II-α_{RS} and II-α_{RU} turns is reported as a stereo view in Figure 6c. Type II-α_{LS} and II-α_{LU} turns correspond to about the mirror images, as far as the backbone conformation is concerned, of II-α_{RS} and II-α_{RU} turns, respectively. A comparison of type II-α_{LS} and II-α_{LU} turns is reported as stereo view in Figure 6d. Type II-α_{RS} and II-α_{LS} turns differ from type II-α_{RU} and II-α_{LU} (intercluster RMSDs are 0.61 and 0.58 Å, respectively) in the relative orientation of the peptide bond between residues i and $i + 1$ (they are rotated about 90°), thus determining the screw-like or U shape of the turn, in a manner similar to type I-α_{YZ} turns (Y = R, L; Z = S, U).

Type I-α_{YZ} and type II-α_{YZ} (Y = R, L; Z = S, U) are structurally different, since the peptide bond between residues $i + 1$ and $i + 2$ is rotated about 180° (intercluster RMSDs are about 1.1 Å). This different peptide bond orientation is ineffective in determining the global shape of the turned structure.

The type I-α_C turn corresponds to a group of structures containing one *cis* peptide bond between residues $i + 1$ and $i + 2$. The chain reversal is clockwise with a screw-like shape. There is no structural similarity with all the other α-turns.

The analysis of the hydrogen-bond interactions inside the different types of α-turns thus far described also indicates these secondary structures may be fused with tighter turns, as expected. Table VII reports a summary of smaller tight turns included in different α-turn types. All α-turns include a favorable dipole interaction similar to a β-turn between residues i and $i + 3$, except type I-α_C and II-α_{LS}. The ϕ,ψ torsion angles around residues $i + 1$ and $i + 2$ correspond to about the expected values for a β-turn (type I, I', II, II', III, or III', depending on the particular α-turn type). Type I-α_{LS} and I-α_{LU} turns may also contain a dipole interaction similar to a β-turn between residues $i + 1$ and $i + 4$. Therefore, type I-α_{LS} turns can be seen as formed by two consecutive type III' β-turns, while type I-α_{LU} turns as two consecutive, slightly dis-

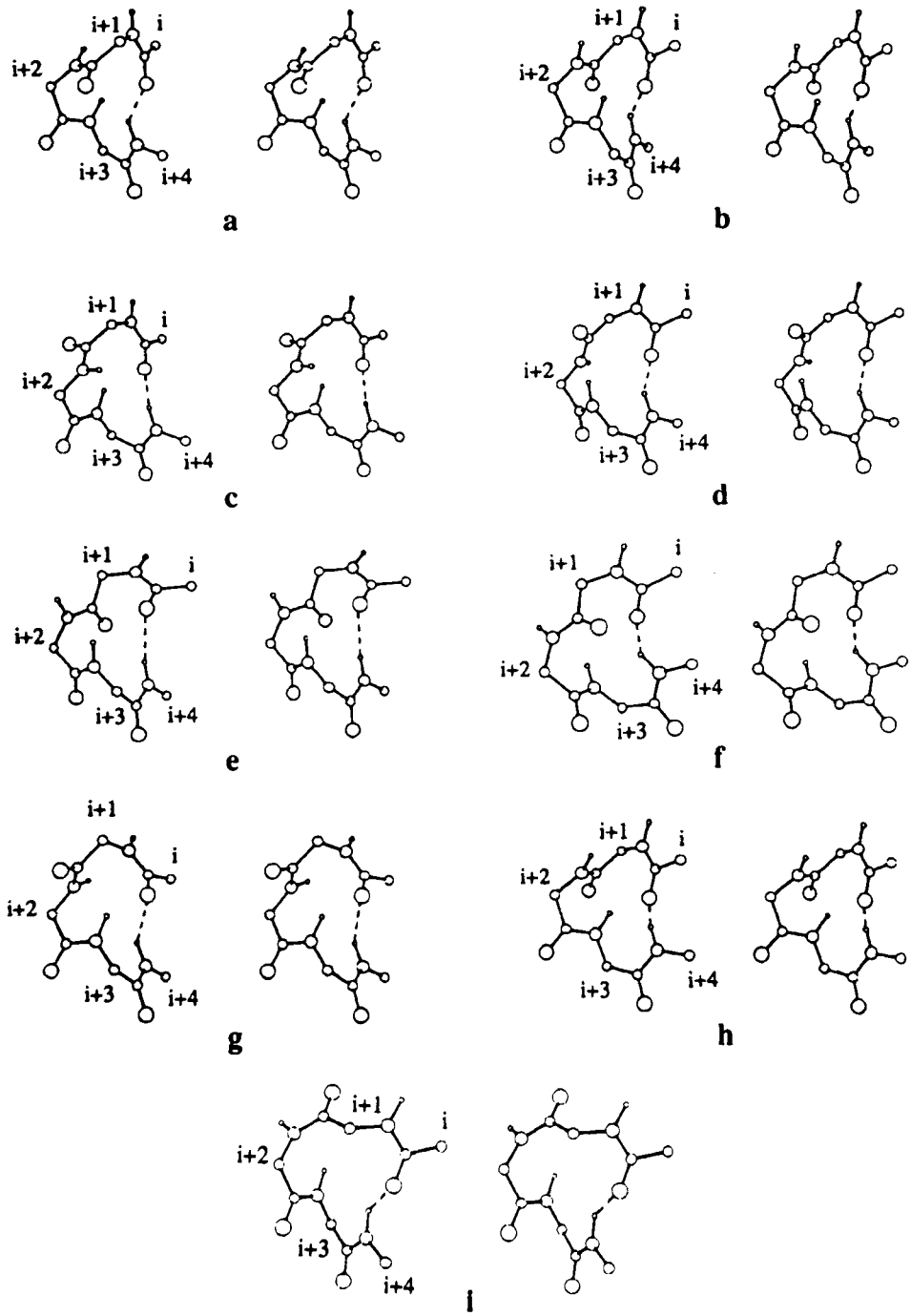

FIGURE 3 The averaged conformations of different types of α-turn are reported as stereo view: (a) type I-α_{RS}, (b) type I-α_{LS}, (c) type I-α_{RU}, (d) type I-α_{LU}, (e) type II-α_{RS}, (f) type II-α_{RU}, (g) type II-α_{LS}, (h) type II-α_{LU}, and (i) type I-α_{C}.

torted type II and III' β-turns. The ϕ,ψ torsion angles around residues $i + 2$ and $i + 3$ correspond to about the expected values for a β-turn.

Furthermore, type II-α_{LS} turns are unique because they include a γ^{i}-turn between residues $i + 1$ and $i + 3$, being the conformation ($\phi = -95°$, $\psi = +81°$) around the residue at position $i + 2$ of the turn in the typical expected range of this tighter turn. The type II-α_{RS} turn, about corresponding to the mirror image of the type II-α_{LS} turn, does not contain a γ-turn between residues $i + 1$ and $i + 3$, but it is slightly distorted (compared to the mirror image of type II-α_{RS} turn), preferring to incorporate a type II β-turn between residues i and $i + 3$.

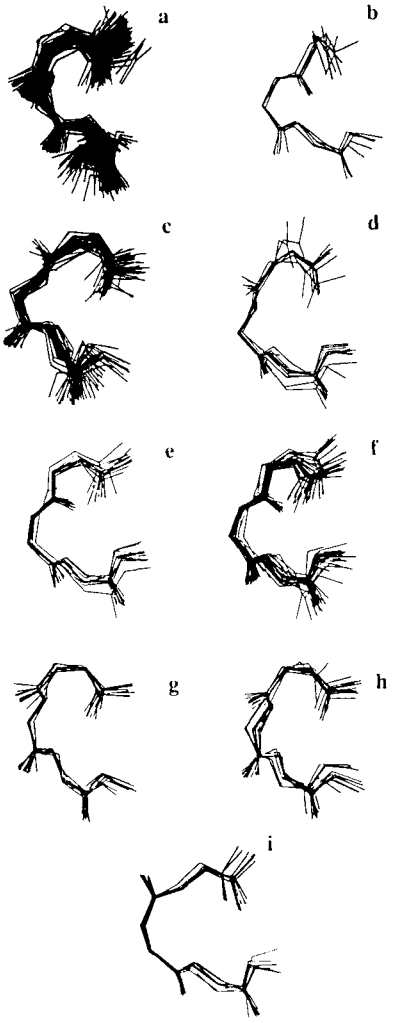

FIGURE 4 All backbone atom superimposition (from C_i^α to C_{i+4}^α) with the mean structures for each cluster: (a) type I-α_{RS}, (b) type I-α_{LS}, (c) type I-α_{RU}, (d) type I-α_{LU}, (e) type II-α_{RS}, (f) type II-α_{RU}, (g) type II-α_{LS}, (h) type II-α_{LU}, and (i) I-α_C.

These observations on the presence of inner hydrogen bonds indicate a possible ambiguity in determining whether the dominating motif corresponds to β- or α-turns. This ambiguity is particularly severe when the electrostatic interaction $i \leftarrow i + 4$ is smaller than $i \leftarrow i + 3$, as it is, for example, for the average structures of type II-α_{RS} and II-α_{LU} turns. This ambiguity could not be solved because the compactness and the size of the α-turned structure may give a variety of different patterns of internal favorable electrostatic interactions of variable strength. Anyhow, 172 out of 219 isolated α-turns, possessing an $i \leftarrow i + 4$ electrostatic interaction ≤ -1.0 kcal/mol, are characterized by weaker favorable interactions between the other dipoles.

It is noteworthy that the type I-α_C turn, contain-

Table V Intercluster Superimposition RMSD (Å) of Each Average Conformations for All Backbone Atoms from C_i^α to C_{i+4}^α (Diagonal Terms Indicate the Mean Intracluster Superimposition RMSD (Å) for All Backbone Atoms from C_i^α to C_{i+4}^α)

Cluster	I-α_{RS}	I-α_{LS}	II-α_{RS}	II-α_{LS}	I-α_{LU}	I-α_{RU}	II-α_{LU}	II-α_{RU}	I-α_C
I-α_{RS}	0.26 ± 0.12								
I-α_{LS}	1.53	0.21 ± 0.09							
II-α_{RS}	1.11	0.99	0.32 ± 0.24						
II-α_{LS}	1.10	1.09	1.23	0.22 ± 0.15					
I-α_{LU}	1.41	0.51	0.90	0.94	0.24 ± 0.09				
I-α_{RU}	0.55	1.38	1.04	1.07	1.34	0.28 ± 0.14			
II-α_{LU}	0.87	1.26	1.16	0.58	1.09	1.06	0.41 ± 0.12		
II-α_{RU}	1.28	0.74	0.61	1.34	0.94	1.08	1.06	0.36 ± 0.12	
I-α_C	1.27	1.24	0.77	1.43	1.25	1.08	1.38	0.87	0.19 ± 0.05

Table VI Geometrical Parameters of the $C^{\alpha}_{i}O_i \leftarrow HN_{i+4}$ Hydrogen Bonds of the Average Structures for Each Type of α-Turn Are Reported, Together with the Electrostatic Contributions Between C'O and HN Dipoles

Type	$N_i - O_{i+4}$ (Å)	$N_i - O_{i+4}$ C (°)	$i \leftarrow i+4$ (kcal/mol)	$i \leftarrow i+3$ (kcal/mol)	$i+1 \leftarrow i+4$ (kcal/mol)	$i+1 \leftarrow i+3$ (kcal/mol)
I-α_{RS}	3.2 ± 0.5	156 ± 10	−1.6	−1.1	−0.1	1.0
I-α_{LS}	3.5 ± 0.4	146 ± 15	−1.1	−0.8	−0.5	1.0
II-α_{RS}	3.6 ± 0.7	158 ± 13	−1.4	−1.9	0.3	0.5
II-α_{LS}	3.0 ± 0.6	161 ± 12	−2.4	0.1	0.2	−1.3
I-α_{LU}	3.4 ± 0.3	126 ± 17	−1.3	−0.5	−0.5	1.1
I-α_{RU}	3.5 ± 0.5	126 ± 19	−1.2	−0.9	−0.2	1.0
II-α_{LU}	3.6 ± 0.2	123 ± 14	−0.8	−1.7	0.3	0.5
II-α_{RU}	3.1 ± 0.2	149 ± 16	−1.9	−1.3	0.1	0.3
I-α_{C}	2.8 ± 0.1	152 ± 13	−2.6	0.2	0.3	0.5

ing a *cis* peptide bond, is characterize by the strongest $i \leftarrow i + 4$ electrostatic interaction. It is reasonable to believe that the energy loss associated with the *cis* isomer is partly compensated by a strong hydrogen-bond interaction. Furthermore, the type I-α_C turn is characterized by a Pro residue at position $i + 2$ (see Table III), except for 3BLM 165–169.[54] This observation is in line with the well-known higher propensity of Xaa-Pro peptide bond to be accommodated in a *cis* configuration.

The analysis of the frequency of the 20 naturally occurring amino acids in α-turned structures is reported in Figure 5b and reveals that these structures are mainly characterized by hydrophilic amino acids. The hydrophobic residues have a negative frequency variation of about 30–50% from the natural frequency in the analyzed proteins. Exceptions are proline and cysteine, which show a positive frequency variation in the α-turns, probably because these residues can provide some local constraints and thus stabilize the α-turn conformation. Within the hydrophilic amino acids there is an overwhelming presence of glutamic acid and, to a less extent, of aspartic acid. Figure 5c describes the occurrence of the various amino acids in the central positions of the α-turns. The major difference, with respect to the natural occurrence of the various amino acids in any position of protein structures, regards the high frequency of proline residues at position $i + 1$ of the turn, glutamic acid at position $i + 2$, and cysteine, threonine, and glutamic acid at position $i + 3$. Proline residues have never been observed at position $i + 3$ of the α-turn. It noteworthy that when combining all the α-turn types to derive the amino acid preferences, residue type preferences for the different positions in the individual α-turn type could be masked. On the other hand, due to the small number of structures in each type of α-turn, it was not possible to have a reliable statistical analysis of amino acid distribution for each type of α-turn. However, the low occurrence of Gly residues in the type I-α_{RS} turn and, conversely, the high frequency of the Gly residue in type II-α_{RS} and I-α_{RU} turns, is noteworthy (see Table III). Furthermore, the type II-α_{RS} turn is also characterized by a high occurrence of Lys residues.

It also must be mentioned that the amino acid frequency within the α-turns here described may only slightly be affected (from a quantitative point of view) by the data base reduction criterion adopted to avoid redundancy from homologous proteins. In fact, some identical sequences, or pentapeptide segments differing by only one amino acid, belonging to structurally related proteins, are still present in the final data set (see Table III). By performing the α-turn selection first and then the data base reduction, we were able to describe the α-turn conformation better, even though the accuracy of the amino acid frequency was somehow sacrificed. We observed that even identical α-turn sequences may adopt different α-turn types. Examples are as follows: (HPKKG) 1SNC 46–50 type I-α_{RS}[55] and 1SNM 46–50 type II-α_{RS}[56]; (NPGTN) 1TEC I57–I61 type I-α_{RS}[57] and 1CSE I57–I61 type II-α_{RS}[50]; (MGDEV) 3FXN 56–60 type I-α_{LS}[58] and 4FXN 56–60 type II-α_{LS}.[58] This detailed information would have been lost by using the reversed reduction-selection procedure.

The previous observation that α-turns are mainly characterized by hydrophilic amino acids in the central positions of the turn is in agreement with the observation that 353 out of 356 isolated α-turns so far identified are exposed to the solvent. Only three cases (1COX 115–119,[59] 1PHH 285–289,[60] and 8ADH 140–144[61]) correspond to structures buried in the interior of the protein. Therefore, the relevance of different types of α-turns described here in molecular recognition processes can

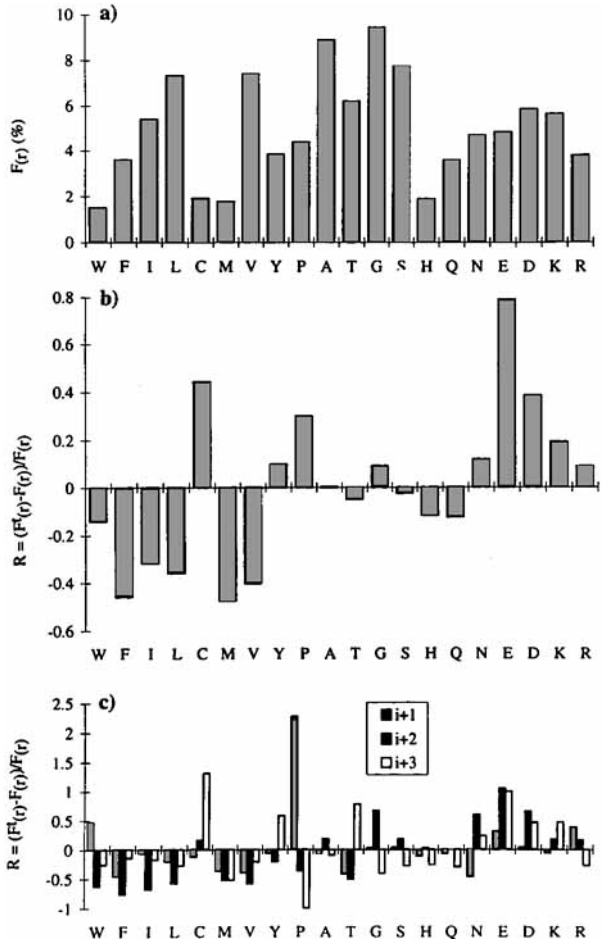

FIGURE 5 (a) Diagram of the natural amino acid occurrence $F_{(r)}$ within the 193 protein structures analyzed. (b) Diagram of $R = [F^t_{(r)} - F_{(r)}]/F_{(r)}$ representing the percent variation from the mean frequency in either one of the three central positions of the α-turns. (c) Diagram of the deconvoluted function R over the $i + 1$, $i + 2$, and $i + 3$ positions of the α-turns.

be ascribed to their occurrence on the protein surface. It should be mentioned that a detailed analysis of the α-turn position in the protein structure revealed that in most cases this particular secondary structure is not only exposed to solvent, but surprisingly it protrudes outward from the protein surface with a hook-like shape and therefore these structural motifs can function as keys in a lock–key interaction mechanism between proteins. Figure 7 shows a stereo view of 1HRH[62] and 1SNM[56] ribbon structure.

The α-turned structures mainly identified on the protein surface correspond to the most difficult to determine, experimentally, parts of the proteins. We have examined the temperature B factors of the backbone atoms in the α-turns in an attempt to identify surface loops that are characterized by higher temperature factors and thus may be poorly defined in the electron density map. The mean B values of the backbone atoms from N_i to $C'_{i+4}O_{i+4}$ in the α-turns, when compared to the B factors of the backbone atoms in the whole protein, indicate that most of the α-turns are sufficiently well defined. In particular, 33 out of 389 α-turns show average backbone atom B factors greater than 40 Å.^{2} Fourteen of these turns corresponds to unclassified (distorted) α-turn and another 14 turns are of type I-α_{RS}. The conformation of these fragments might thus be affected by the use of geometrical constraints in the refinement procedure. This observation is particularly intriguing for rarely occurring α-turn types which may not be statistically significant. Nevertheless, there is enough convincing evidence on their existence. The I-α_{LS} α-turn type is only 1.5% of the total number of α-turns, but its con-

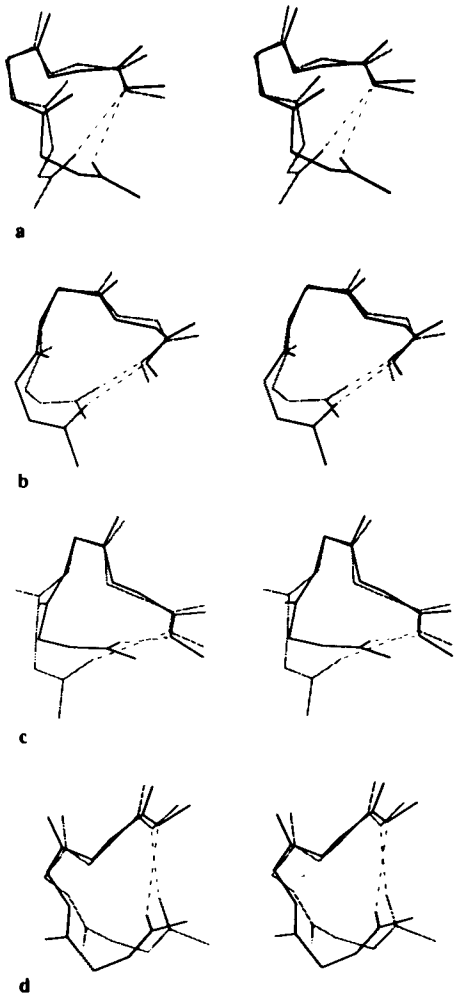

FIGURE 6 Stereo view of all backbone atom superimposition from C^α_{i+2} to C^α_{i+4}: (a) between α-turns type I-α_{RS} and I-α_{RU}, (b) type I-α_{LS} and I-α_{LU}, (c) type II-α_{RS} and II-α_{RU}, and (d) type II-α_{LS} and II-α_{LU}. Thick line correspond to U-shaped turns.

Table VII Summary of Smaller Tight Turns Included in the Different Types of α-Turns

$i \leftarrow i+4$ α-Turn Type	$i \leftarrow i+3$ β-Turn Type	$i+1 \leftarrow i+4$ β-Turn Type	$i+1 \leftarrow i+3$ γ-Turn Type
I-α_{RS}	III	—	—
I-α_{LS}	III′	III′	—
II-α_{RS}	II	—	—
II-α_{LS}	—	—	γ^i
I-α_{LU}	II	III′	—
I-α_{RU}	II′	—	—
II-α_{LU}	I	—	—
II-α_{RU}	I′	—	—
I-α_C	—	—	—

formation corresponds to that typical of the well-established left-handed α-helix (which is also rare). The II-α_{LS} α-turn type represents 2.2% of the total number of α-turns, but its significance derives from its much higher occurrence (8.8%) before reduction of the protein data base for redundancy. II-α_{LU} and II-α_{RU} are actually rare α-turn types (2.2 and 2.5%, respectively). We believe that the statistical significance of these turns should be taken with caution, before further evidence will be reported. Finally, the significance of I-α_c turns is supported by the occurrence in small cyclic peptides (see below). In addition to this, rarely occurring turns are mirror images (as far as the backbone torsion angles it regards) of more frequently observed α-turns. This is the case for the rare I-α_{LS}, II-α_{LS}, and II-α_{LU} α-turn types, which are mirror images of the more frequently observed α-turns I-α_{RS}, II-α_{RS}, and II-α_{RU}. Therefore, the rarity of some of the α-turns seems to be related to a higher content of residues that adopt a conformation on the right side of the Ramachandran map. Generally, these conformations correspond to higher energy minima.

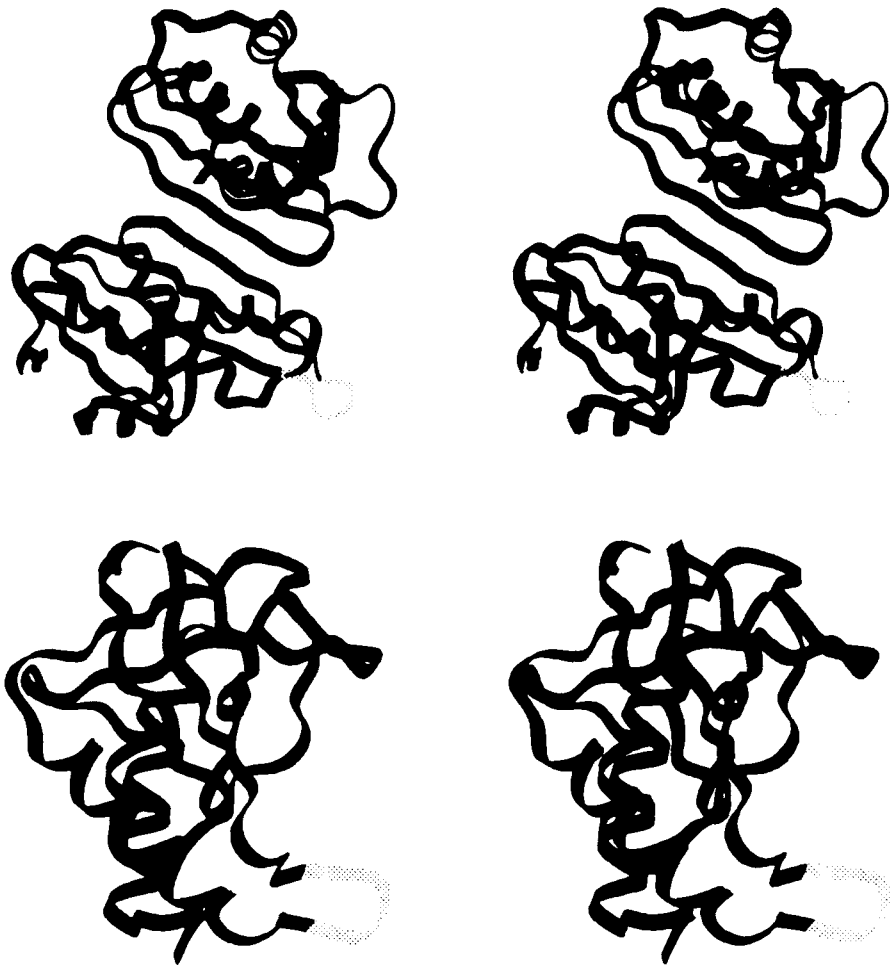

FIGURE 7 Ribbon structure stereo view of 1HRH (top)[62] and 1SNM (bottom).[56] Shadowed ribbon stretches indicate the 1HRH 447–451 and 1SNM 46–50 α-turn region.

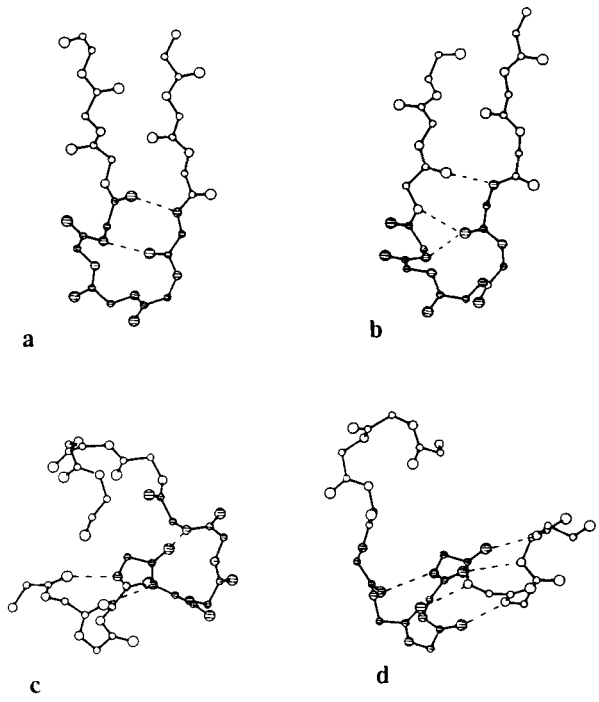

FIGURE 8 Examples of α-turn positions in protein structures; dashed atoms are in α-turned conformation: (a) on the loop side of a type 3 β-hairpin from 1SGC 140–156,[74] (b) on the loop side of a type 4 β-hairpin from 1CMS 315–319,[75] (c) at the end (C-terminal) of an α-helix from 1TEC E239–243,[57] and (d) at the beginning (N-terminal) of an α-helix from 1OVA B69–73.[76] Residues involved in the α-turned structure are dashed.

In addition, we wish to speculate that rarely occurring motifs, whenever present on the protein surface, might contain specific information about unique possible types of recognition precesses.

Furthermore, an examination of the structural domains adjacent to the isolated α-turns here identified reveals that these turns are generally connecting stretches of extended peptide chains. In several cases these turns are located at the tip of type 3 and 4 β-hairpins[7]; type II-α_{LU} and I-α_{RU} turns may instead be observed at the end (C-terminal) or at the beginning (N-terminal) of α-helices, respectively. These turns correspond to the α-helical Ccap or Ncap previously described.[63] It is worth mentioning that type II-α_{LU} turn corresponds to the α_L motif recently proposed as helix C-capping signal.[28] In a few cases the α-turns are located within highly irregular conformations. Figure 8 shows some of these examples. Detailed analysis is presently in progress.

In conclusion, we report here for the first time an identification and classification of isolated α-turns in proteins. These turns have been classified into four different types (beside their mirror images of the backbone atoms). Type I-α_{RS} (I-α_{LS}) corresponds to the basic repetitive units of α-helices and it is also frequently observed in nonhelical structures, while type I-α_{RU} (I-α_{LU}), II-α_{RS} (II-α_{LS}), and II-α_{RU} (II-α_{LU}) are classified presently as new irregular secondary protein structures. The α-turns containing a $i + 1$–$i + 2$ *cis* peptide bond also have been identified, whereas α-turns with a $i + 2$–$i + 3$ *cis* peptide bond never have been observed.

These different types of α-turns found in proteins partly confirms our initial hypothesis that isolated α-turns, containing all-*trans* peptide bonds, may exist in four distinct basic motifs (together with the corresponding backbone mirror images). They would differ in a rotation of about 180° of either one of the two peptide bonds between the three central residues of the turn. Surprisingly, only two of these four basic motifs were identified in protein structures. They differ in the relative orientation of the peptide plane between residues $i + 1$ and $i + 2$. These two basic motifs could be divided further into two more groups because α-turns containing a 90° rotation of the peptide plane between residue i and $i + 1$ were observed. Those containing an inversion of the peptide bond between residues $i + 2$ and $i + 3$ of the turn have not been observed so far.

Conformational energy calculations are presently in progress to understand the relative stability of these newly characterized protein secondary structures better.

We also have attempted to understand the role of the α-turned structure in protein active site, molecular recognition, and protein folding with a preliminary search in the wide literature available. For instance, it was found[64] that the residues in the α-turn 1LZ1 46–50 participate in a cluster of hydrogen bonds and they are located in the active site cleft, suggesting the possibility of a functional role. Cys^{39} in 1FD2 35–39 α-turn[65] and His^{117} and Met^{121} in 2AZA 117–121 α-turn[66] are involved in the metal ion coordination. The 1COX 115–119 α-turn[59] contains residues in close contact with the flavin adenine dinucleotide. 2CD4 51–55 α-turned residues[67] are located in the putative binding region of HIV gp120 protein and at site where changes affect the binding of OKT4d. 1P02 140–156 residues in the α-turn[68] are located next to the active site triad of the serine protease.

α-Turns are also relevant structural domains in small peptides, particularly in cyclo-peptides containing 7–9 residues in their sequence. Ilamicin b1,[69] cyclolinopeptide A,[70] and its analogues[71] are examples of cyclo-epta- and cyclo-nona-peptides

containing the type I-α_C turned structure. They have with important biological functions: Ilamicin b1 is an antibiotic, inhibiting the growth of mycobacteria, while cyclolinopeptide A shows cytoprotective properties against toxic agents, such as phalloidine. Amatoxins[72,73] a family of toxic peptides isolated from *Amanita phalloides*, are examples of cyclo-octa-peptides containing both type I-α_{RS} and II-α_{RS} turns.

The growing number of detailed protein and peptide x-ray structures will probably clarify the biological relevance and the occurrence of this peculiar and irregular protein secondary structure.

REFERENCES

1. Pauling, L. & Corey, R. (1951). *Proc. Natl. Acad. Sci. USA* **37,** 729–740.
2. Pauling, L., Corey, R. & Branson, H. (1951) *Proc. Natl. Acad. Sci. USA* **37,** 205–211.
3. Levitt, M. & Chothia, C. (1976) *Nature* **261,** 552–557.
4. Richardson, J. S. (1977) *Nature* **268,** 495–500.
5. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
6. Milner-White, E. J. & Poet, R. (1987) *TIBS* **12,** 189–192.
7. Pavone, V. (1988) *Int. J. Biol. Macromol.* **10,** 238–240.
8. Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989) *J. Mol. Biol.* **206,** 759–777.
9. Venkatachalam, C. M. (1968) *Biopolymers* **6,** 1425–1436.
10. Crawford, J. L., Lipscomb, W. N. & Schellman, C. G. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 538–542.
11. Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Protein Chem.* **37,** 1–109.
12. Thornton, J. M., Sibanda, B. L., Edwards, M. S. & Barlow, D. J. (1988) *Bio. Essays* **8,** 63–69.
13. Lesczcynski, J. F. & Rose, G. D. (1986) *Science* **234,** 849–855.
14. Thomas, D. J. (1991) *J. Mol. Biol.* **222,** 805–817.
15. Ptitsyn, O. B. (1981) *FEBS Lett.* **131,** 197–202.
16. Novotny, J., Handschumacher, M. & Haber, E. (1986) *J. Mol. Biol.* **189,** 715–721.
17. Joseph, D., Petsko, G. A. & Karplus, M. (1990) *Science* **249,** 1425–1428.
18. Matthews, B. W. (1972) *Macrmolecules* **5,** 818–819.
19. Chou, P. Y. & Fasman, G. D. (1977) *J. Mol. Biol.* **115,** 135–175.
20. Wilmot, C. M. & Thornton, J. M. (1988) *J. Mol. Biol.* **203,** 221–232.
21. Milner-White, E. J., Ross, B. M., Belhadj-Mostefa, K. & Poet, R. (1988) *J. Mol. Biol.* **204,** 777–782.
22. Toniolo, C. (1980) *CRC Crit. Rev. Biochem.* 1–44.
23. DeGrado, W. F. (1988) *Adv. Protein Chem.* **40,** 51–124.
24. Kabsch, W. & Sander, C. (1983) *Biopolymer* **22,** 2577–2637.
25. Schellman, C. (1980) in *Protein Folding, Proceedings of 28th Conference of the German Biochemical Society,* Rainer, J., Ed., Elsevier/North-Holland Biomedical Press Amsterdam, New York, pp. 53–61.
26. Karle, I. L., Flippen-Anderson, J. L., Uma, K. & Balaram, P. (1993) *Int. J. Peptide Protein Res.* **42,** 401–410.
27. Milner-White, E. J. & Poet, R. (1986) *Biochem. J.* **240,** 289–292.
28. Aurora, R., Srinivasan, R. & Rose, G. D. (1994) *Science* **264,** 1126–1130.
29. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1987) *J. Mol. Biol.* **112,** 535–542.
30. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987) in *Crystallographic Databases-Information Content, Software Systems, Scientific Applications,* Allen, F. H., Bergerhoff, G. & Sievers, R., Eds., Data Commission of the International Union of Crystallography, Bonn-Cambridge-Chester, pp. 107–132.
31. Calascibetta, F. G., Den Santis, P., Morosetti, S., Muñiz-Miranda, M., & Forni, E. (1976) *Gazz. Chim. Ital.* **106,** 491–497.
32. Wilson, D. K., Rudolph, F. B. & Quiocho, F. A. (1991) *Science* **252,** 1278–1284.
33. Flaherty, K. M., De Luca-Flaherty, C. & McKay, D. B. (1990) *Nature* **346,** 623–628.
34. Bone, R., Frenk, D., Kettner, C. A. & Agard, D. A. (1989) *Biochemistry* **28,** 7600–7609.
35. Brady, L., Brzozowski, A. N., Derewenda, Z. S., Dodson, E. J., Dodson, S., Tolley, S., Turkenburs, J. P., Kristiansen, L., Huge-Jensen, B., Norslov, L., Thim, L. & Menge, U. (1990) *Nature* **343,** 767–770.
36. Moews, P. C., Knox, J. R., Dideberg, O., Charlier, P. & Frere, J.-M. (1990) *Protein Struct. Funct. Genet.* **7,** 156–171.
37. Liao, D. I. & Remington, S. (1990) *J. Biol. Chem.* **265,** 6528–6531.
38. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C. & Jones, T. A. (1990) *Science* **249,** 380–386.
39. Rydel, T., Ravichandran, K. G., Tulinski, A., Bode, W., Huber, R., Roitsch, C. & Fenton, J. W., II (1990) *Science* **249,** 277–280.
40. Smith, J. L., Corfield, P. W. R., Hendrickson, W. A. & Low, B. W. (1988) *Acta Crystallogr.* **A44,** 357–368.
41. Corfield, P. W. R., Lee, T. J. & Low, B. W. (1989) *J. Biol. Chem.* **264,** 9239–9242.
42. Tong, L., Milburn, M. V., de Vos, A. M. & Kim, S.-H. (1989) *Science* **245,** 244–244.
43. Veerapandian, B. (1992) *Biophys. J.* **62,** 112–115.
44. Wood, S. P., Tickle, I. J., Treharne, A. M., Pitts,

J. E., Mascarenhas, Y., Li, J. Y., Husain, J., Cooper, S., Blundell, T. L., Hruby, V. J., Buku, A., Fischman, A. J. & Wyssbrod, H. R. (1986) *Science* **232,** 633–636.
45. Sanderson, M. R., Freemont, P. S., Ryce, P. A., Goldmann, A., Atfull, G. F., Grindley, N. D. F. & Steitz, T. A. (1990) *Cell* **63,** 1323–1329.
46. Pflugrath, J. W., Wiegand, G., Huber, R. & Vertesy, L. (1986) *J. Mol. Biol.* **189,** 383–386.
47. Van Duyne, G. D., Standaert, R. F., Schreiber, S. L. & Clardy, J. (1991) *J. Am. Chem. Soc.* **113,** 7433–7434.
48. Fox, R. O. & Richards, F. M. (1982) *Nature* **300,** 325–330.
49. Langs, D. A. (1988) *Science* **241,** 188–191.
50. Bode, W., Papamokos, E. & Musil, D. (1987) *Eur. J. Biochem.* **166,** 673–692.
51. Gros, P., Betzel, C., Dauter, Z., Wilson, K. S. & Hol, W. G. J. (1989) *J. Mol. Biol.* **210,** 347–367.
52. Engh, R. A. & Huber, R. (1991) *Acta Cryst.* **A47,** 392–400.
53. Furet, P., Sele, A. & Cohen, N. C. (1988) *J. Mol. Graphics* **6,** 182–189.
54. Herzberg, O. (1991) *J. Mol. Biol.* **217,** 701–719.
55. Loll, P. J. & Lattman, E. E. (1989) *Protein Struct. Funct. Genet.* **5,** 185–201.
56. Loll, P. J. & Lattman, E. E. (1990) *Biochemistry* **29,** 6866–6873.
57. Gros, P., Fujinaga, M., Dijkstra, B. W., Kalk, K. H. & Hol, W. G. J. (1989) *Acta Crystallogr. Sect. B Struct. Sci.* **B45,** 488–499.
58. Smith, W. W., Burnett, R. M., Darling, G. D. & Ludwig, M. L. (1977) *J. Mol. Biol.* **117,** 195–225.
59. Vrielink, A., Lloyd, L. F. & Blow, D. M. (1991) *J. Mol. Biol.* **219,** 533–554.
60. Schreuder, H. A., Van Der Laan, J. M., Hol, W. G. J. & Drenth, J. (1988) *J. Mol. Biol.* **199,** 637–648.
61. Eklund, H., Samama, J. I. & Jones, T. A. (1984) *Biochemistry* **23,** 5982–5996.
62. Davies, J. F., Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. (1991) *Science* **252,** 88–95.
63. Richardson, J. S. & Richardson, D. C. (1988) *Science* **240,** 1648–1652.
64. Artymiuk, P. J. & Blake, C. C. F. (1981) *J. Mol. Biol.* **152,** 737–762.
65. Stout, C. D. (1989) *J. Mol. Biol.* **205,** 545–555.
66. Baker, E. N. (1988) *J. Mol. Biol.* **203,** 1071–1095.
67. Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, J. E., Hosain, Y., Reinherz, E. L. & Harrison, S. C. (1990) *Nature* **348,** 411–418.
68. Fujinaga, M., Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1985) *J. Mol. Biol.* **183,** 479–502.
69. Iitaka, Y., Nakamura, H., Takada, K. & Takida, T. (1974) *Acta Cryst.* **B30,** 2817–2825.
70. Di Blasio, B., Rossi, F., Benedetti, E., Pavone, V., Pedone, C., Temussi, P. A., Zanotti, G. & Tancredi, T. (1989) *J. Am. Chem. Soc.* **111,** 9089–9098.
71. Di Blasio, B., Rossi, F., Benedetti, E., Pavone, V., Saviano, M., Pedone, C., Zanotti, G. & Tancredi, T. (1992) *J. Am. Chem. Soc.* **114,** 8277–8283.
72. Kostansek, E. C., Lipscomb, W. N., Yocum, R. R. & Thiessen, W. E. (1978) *Biochemistry* **17,** 3790–3795.
73. Zanotti, G., Wieland, T., Benedetti, E., Di Blasio, B., Pavone, V. & Pedone, C. (1989) *Int. J. Peptide Protein Res.* **34,** 222–228.
74. Delbaere, L. T. J. & Brayer, G. D. (1985) *J. Mol. Biol.* **183,** 89–103.
75. Gilliland, G. L., Winborne, E. L., Nachman, J. & Wlodawer, A. (1990) *Proteins Struct. Funct. Genet.* **8,** 82–101.
76. Stein, P. E., Leslie, A. G. W., Finch, J. T. & Carrell, R. W. (1991) *J. Mol. Biol.* **221,** 941–959.