RESEARCH ARTICLE

α-Turns in protein structures

D. V. Nataraj, N. Srinivasan*, R. Sowdhamini* and C. Ramakrishnan†

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India
*Present address: ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, London, WC1E 7HX, UK

The occurrence of 5 → 1 type of hydrogen bonds (α-turn) in proteins has been studied using a data set comprising of 107 proteins with resolution ≤ 2.0 Å. A very large majority of such α-turn segments (96%) form part of regular α-helices. The examples (84) which do not form part of an α-helix are termed 'isolated α-turns' and are grouped into two major families and seven minor groups along with two isolated examples based on the similarity of conformational angles. The family with large number of examples (50) have (φ, ψ) angles close to an α-helix and hence belong to the class of the shortest α-helices. The 'end to end' distances of these α-turns vary between 4.7 and 6.7 Å, the range being nearly the same as that of α-helices. The propensity calculations show that some amino acids such as Glu, Ser and Thr have statistically significant higher preferences to occur in α-turns than in α-helices. In addition to the 5 → 1 type, the residues in the α-turn are involved in hydrogen bonds with other parts of the chain. The residues are in general more hydrophilic compared to those in α-helices. In many cases (70%) the α-turn occurs at the ends of extended strands, and whenever it occurs at the loop regions connecting two extended strands, it brings about a hairpin bend.

POLYPEPTIDE chain reversal is an essential feature in bringing about the globular shape of proteins. Turns are also associated with various aspects of protein function[1-3]. The direction of the main chain can be reversed by characteristic conformations at a single amino acid residue (γ-turns[4,5]) or at two residues (β-turns[2,6,7]) or by longer loops[8-10]. Both γ- and β-turns are locally stabilized by hydrogen bonds in the main chain. The 3 → 1 hydrogen bond in the γ-turn is formed between carboxyl oxygen at position i and amide at position $i+2$, with the turn forming residue at the position $i+1$. Most β-turn types are stabilized by a 4 → 1 hydrogen bond between carbonyl oxygen at position i and amide at position $i+3$.

These two kinds of turns occur frequently in globular proteins, with β-turn being one of the more common substructures[11-14]. In this communication, attention is drawn to 'three-residue segments' locally stabilized by a 5 → 1 hydrogen bond between carbonyl oxygen at position i and amide at position $i+4$, reversing the course of the polypeptide direction. The analysis of a large number of known three-dimensional structures of proteins reveals that such a local chain reversal is a common occurrence. This class of turns is referred to as α-turns since the hydrogen bonding pattern is the same as in the α-helix. A schematic diagram of an α-turn is shown in Figure 1. The present analysis shows that α-turns can have different characteristic main chain conformational angles and amino acid preferences. They connect regular secondary structures and quite often act as linkers of two β-strands.

Materials and methods

Data set used

The data set used in the present study consists of a set of 107 proteins whose coordinates are available in Protein Data Bank[15,16]. The choice of the set has been guided by the condition that their resolution is 2 Å or better, and the cut-off percentage of the sequence homology between any two proteins in the set is not more than 40. Also in the case of proteins having multiple chains with the same sequence, only one chain has been considered.

Identification of α-turns in proteins

Since the 5 → 1 hydrogen bond is the crux of the

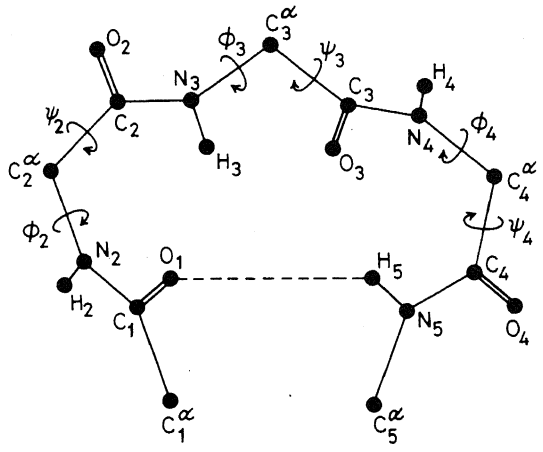

Figure 1. Schematic diagram of an α-turn, occurring in a system of four linked peptide units. The designations of atoms and torsion angles involved are also marked.

†For correspondence.

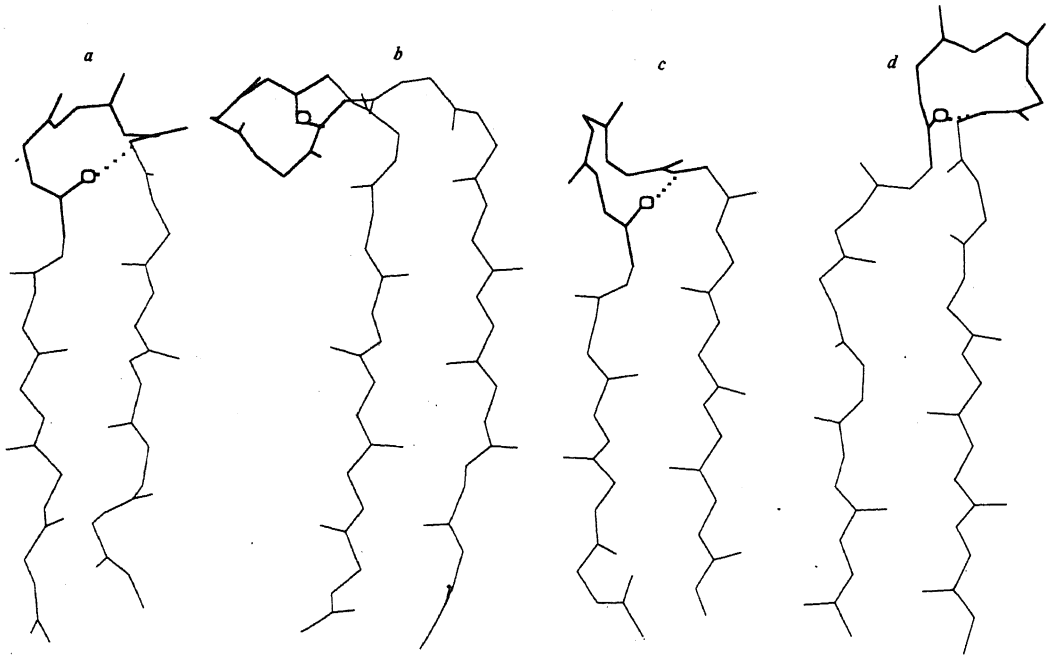

Figure 8. Two antiparallel β-strands connected by an α-turn. The α-turn segment is shown as thick lines and the $5 \rightarrow 1$ hydrogen bond by a dotted line. The oxygen atom taking part in the hydrogen bond is explicitly shown. ***a***, segment 388–399 of the protein 1cox (α-turn 391–395: family F1); ***b***, segment 41–58 of the protein 1gcr (α-turn 47–51: group g1); ***c***, segment E5–E21 of the protein 3apr (α-turn E11–E15: group g4); ***d***, segment L160–L172 of the protein 2fbj (α-turn L166–L170: isolated example).

Conclusion

The present study shows that the proteins do contain $5 \rightarrow 1$ hydrogen-bonded α-turns which do not necessarily form part of a larger secondary structure such as α-helix. The fact that the number of such examples of isolated α-turns is small as compared to β-turns, points to the fact that in these examples, α-turns have, for some reason, been preferred over the more populous β-turns. The small but significant number of examples of α-helical α-turns indicate that, in these cases, (i) either the α-helix which has been initiated has not progressed or, alternatively, (ii) the α-helix thus formed by the α-helical α-turn has not been able to propagate itself in either direction.

Since these examples occur in different types of proteins, the reason for these short α-helical α-turns may not lie so much in the functional/folding aspects of the protein as in the combination of residues in the α-turns as well as those which immediately precede and succeed these turns. In any case the isolated α-turns deserve to be considered as a secondary structural feature, though of a minor nature.

1. Richardson, J. S., *Adv. Protein Chem.*, 1981, **34**, 167–339.
2. Rose, G. D., Gierasch, L. M. and Smith, J. A., *Adv. Protein Chem.*, 1985, **37**, 1–109.
3. Richardson, J. S. and Richardson, D. C., *Trends Biochem. Sci.*, 1989, **14**, 304–309.
4. Nemethy, G. and Printz, M. P., *Macromolecules*, 1972, **5**, 755–758.
5. Matthews, B. W., *Macromolecules*, 1972, **5**, 818–819.
6. Venkatachalam, C. M., *Biopolymers*, 1968, **6**, 1425–1436.
7. Chou, P. Y. and Faman, G. D., *J. Mol. Biol.*, 1977, **115**, 135–175.
8. Efimov, A. V., *Molecular Biology (USSR)*, 1986, **20**, 258–266.
9. Thornton, J. M., Sibanda, B. L., Edwards, M. S. and Barlow, D. J., *BioEssays*, 1988, **8**, 63–69.
10. Srinivasan, N., Sowdhamini, R., Ramakrishnan, C. and Balaram, P., in *Molecular Conformation and Biological Interactions* (eds Balaram, P. and Ramaseshan, S.), Indian Academy of Sciences, Bangalore, 1991.
11. Wilmot, C. M. and Thornton, J. M., *J. Mol. Biol.*, 1988, **203**, 221–232.
12. Wilmot, C. M. and Thornton, J. M., *Prot. Eng.*, 1990, **3**, 479–493.
13. Milner-White, E. J., Ross, B. M., Ismail, R., Bethadj-Mostefa, K. and Poet, R., *J. Mol. Biol.*, 1988, **204**, 777–782.
14. Milner-White, E. J., *J. Mol. Biol.*, 1990, **216**, 385–397.
15. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, Jr, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M., *J. Mol. Biol.*, 1977, **112**, 535–542.
16. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. and Weng, J., in *Crystallographic Databases – Information Content, Software Systems, Scientific Applications*, (eds Allen, F. H., Bergerhoff, G. and Sievers, R.), Data Commission of International Union of Crystallography, Bonn, 1987, pp. 107–132.
17. Mitra, J. and Ramakrishnan, C., *Int. J. Pept. Protein Res.*, 1981, **17**, 401–411.
18. Corey, R. B. and Pauling, L., *Proc. R. Soc. London*, 1953, **B141**, 10–20.
19. Stein, P. E., Leslie, A. G. W., Finch, J. T. and Carrell, R. W., *J. Mol. Biol.*, 1991, **221**, 941–959.
20. Ramachandran, G. N. and Sasisekharan, V., *Adv. Protein Chem.*, 1968, **23**, 283–437.
21. Lebioda, L. and Stec, B., *J. Am. Chem. Soc.*, 1989, **111**, 8511–8513.

'extended conformational region' of Ramachandran map and (ii) no distinction has been made between β_i and β_s strands and U (uncharacterized) for other segments.

It can be seen from the table that there are examples for all the possible combinations. The maximum number of examples (19) occur for EE linkers (E–α_T–E). In fact, in about 70% of the examples the α-turn either precedes or succeeds the E-strand. This points to the tendency of α-turns to occur at the ends of extended strands.

The EE linkers are further examined from the point of view of immediate local chain reversal. Very interestingly, in 17 of these 19 examples, the α-turn occurs at the loop regions of β-hairpin. Even in the two cases 1sn3 (α-turn segment 7–11) and 2rsp (α-turn segment A46–50), the E-strands adjoining the α-turn though not forming a sheet are antiparallel in orientation and hence the full stretch can be considered as a hairpin bend without intermolecular hydrogen bonds. These are listed in Table 6, along with the α-turn families/groups to which they belong. Nine of these belong to family F1 and the remaining are distributed among the various groups. The β-hairpins occurring in proteins have been grouped into four classes by different workers[29–31]. Following the definition given by Pavone[30] the classes to which these linkers belong to have also been identified and these are also given in Table 6. Examples are found belonging to all the four classes (I, II, III and IV), with class IV having the maximum number (11).

Four examples, one in each from the families F1 (1cox)[32], g1 (1gcr)[33], g4 (3apr)[34] and one of the isolated examples (2fbj)[35] are shown in Figure 8, which brings out this aspect pictorially.

The α-turns which form linkers between other secondary structures (viz. (i) helix and helix, (ii) helix and strand and (iii) strand and helix) do not exhibit any conspicuous pattern or uniformity.

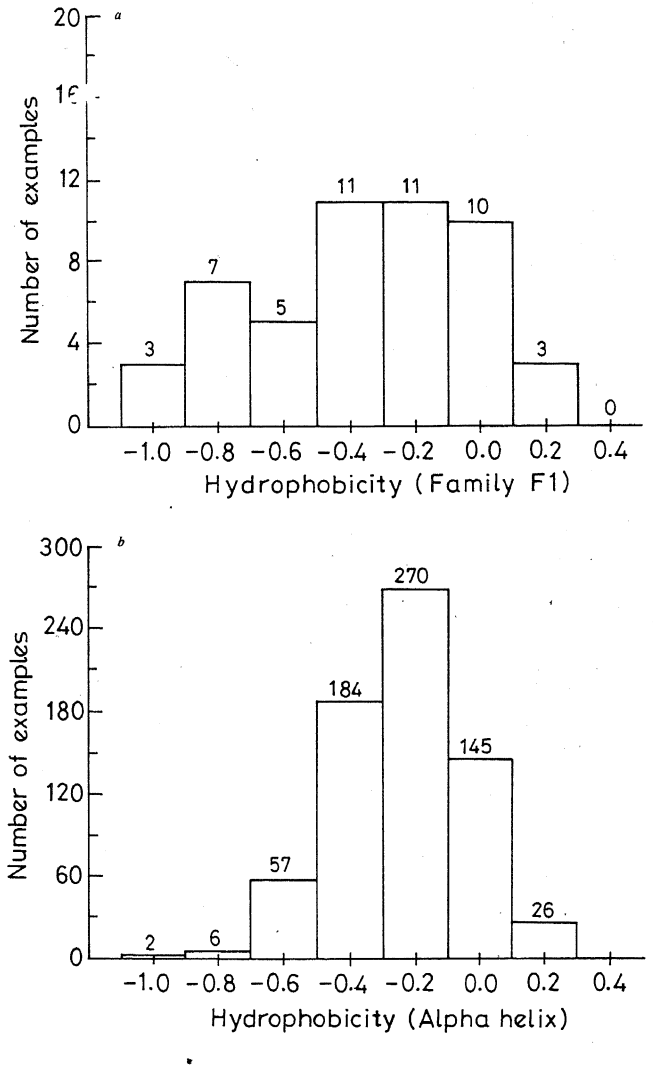

Figure 7. The distribution of average hydrophobicity: *a*, of α-turn segments in family F1; *b*, of α-helical segments in the data set.

Table 5. Number of examples of α-turns connecting different secondary structures (H–helix; E – extended strand and U – uncharacterized)

Sec. struc. preceding α-turn	Sec. struc. succeeding α-turn			
	H	E	U	Total
H	8	7	6	21
E	4	19	17	40
U	4	10	9	23
Total	16	36	32	84

Table 6. EE linker examples of α-turns which fall in different β-hairpin classes

Protein code	α-turn segment	Family/ group	β-hairpin class
1aap	A24–28	F1	IV
1cox	391–395	F1	IV
2er7	E240–244	F1	IV
2ltn	A54–58	F1	IV
2ltn	A167–171	F1	IV
2rhe	93–97	F1	IV
3blm	50–54	F1	IV
3cla	97–101	F1	IV
3fgf	67–71	F1	IV
3apr	E11–15	g4	IV
1ak3	A129–133	g7	IV
1gdl	O300–304	g6	III
4pep	9–13	g6	III
1rbp	63–67	#	III
1gcr	47–51	g1	II
1gcr	136–140	g1	II
2fbj	L166–170	#	I

– Isolated examples.

Since all the 20 amino acids have occurred in α-turns, it would be more appropriate to look at the product propensity of the triplet of amino acids that occur at C_2^α, C_3^α and C_4^α (the product propensity being the product of the individual propensities) and select those triplets that may have a better preference to occur in α-turns as compared to α-helices. This is done by evaluating a quantity ΔP, which is the difference between the product propensity for an α-turn and the product propensity for α-helix. The distribution of ΔP is shown as a histogram in Figure 6. It can be seen that, though the peak occurs around 0, there are few triplets that occur in α-turns which have a large ΔP. Those with $\Delta P > 4.0$ are listed in Table 4. In the six distinct triplets listed in Table 4, Thr, Cys and Ser contribute 78%, which is to be expected since these are the predominantly high-propensity amino acids for α-turns.

Other hydrogen bonds of residues involved in α-turn segment

Since the isolated α-turns do not have consecutive hydrogen bonds to stabilize the segment, it may be interesting to find out whether these α-turn segments are involved in hydrogen bonds with other parts of the chain. On doing this, it was found that except in two examples, the segment is invariably involved with other hydrogen bonds (in addition to the $5 \rightarrow 1$ hydrogen bond). In fact, of the 252 residues in the α-turns (84 examples with three residues in positions $i+1$, $i+2$ and $i+3$), 178 (71%) act either as a donor or an acceptor for hydrogen bond. It is interesting to note that 79 out of 84 (94%) residues in position $i+3$ take part in hydrogen bonding. (The corresponding values for residues in positions $i+1$ and $i+2$ are 61% and 57%, respectively).

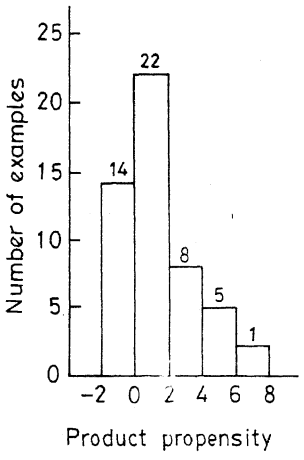

Figure 6. Histogram showing the distribution of the difference product propensity (ΔP) in the examples of α-turns in family F1 (see text for the definition of ΔP).

Table 4. Triplets having large preference to occur in α-turns arranged in decreasing order

Triplet	No. of examples	ΔP*
SER–THR–THR	1	7.027
CYS–ALA–THR	1	5.620
THR–HIS–THR	1	5.284
SER–CYS–SER	1	4.898
THR–VAL–CYS	1	4.769
CYS–ARG–THR	1	4.231

*Cut-off is 4.0.

Hydrophobicity of residues in α-turns

Another aspect that can be looked into is the nature (hydrophobic or otherwise) of the residues involved in the α-turn examples of family F1. Using the consensus values of Eisenberg[27], the average hydrophobicity of the α-turn segments (H_α^t) are evaluated (i.e. average of the hydrohobicity values of the residues at positions $i+1$, $i+2$ and $i+3$). This distribution is shown in Figure 7 *a*. Figure 7 *b* gives the corresponding distribution of hydrophobicity using the residues occurring in α-helices in the data set. It is clear that the residues involved in the isolated α-turns are more polar (hydrophilic) compared to those in α-helices. This would mean that these α-turns are more likely to occur on the surface, accessible to the solvent. This is in agreement with the now well-known concept[2] that the turns in protein usually tend to be on the surface as compared to the other longer secondary structures such as helices and sheets.

Location of α-turns in proteins

The α-turn, being a small segment with a well-defined geometry, can act as a linker between two secondary structural segments in proteins. This aspect has been looked into by examining the proximal secondary structure that (giving an allowance of maximum of four residues on either side) occurs on either side of the α-turn. The secondary structures were identified by examining the contiguous occurrence of representative (ϕ, ψ) values[26]. The results are tabulated in Table 5, in which H stands for helix, E for extended strand. (The commonly used notation is β-strands.) In an earlier communication[28] it has been shown that some of the β-strands are not involved in the formation of the β-sheet. The notation β_s is used to designate those which are involved in forming the β-sheet and β_i those which did not. We have preferred to use the notation E-strands since (i) these have been identified with the criteria that the (ϕ, ψ) values in the segment all lie in the

Table 2. Distribution of end-to-end distances (C_1^α–C_5^α) in the examples of α-turns in different ranges

Family	Range (Å) 4.7–5.2	5.2–5.7	5.7–6.2	6.2–7.0	Total
F1	10	15	16	8	49
F2	2	6	1	—	9
g1	4	—	—	—	4
g2	1	1	—	—	2
g3	4	—	—	—	4
g4	1	3	—	—	4
g5	—	—	1	2	3
g6	2	1	—	—	3
g7	—	1	1	—	2
#	1	—	—	—	1
#	1	—	—	—	1
Total	26	27	19	10	82

– Isolated examples.

may be interesting to compare propensity values of amino acids involved in such turns. The propensity of occurrence of an amino acid at the specified position of the α-turn is given by

$$p_j^k = \frac{f_j^k}{F_j},$$

where f_j^k is the fraction of amino acid residue of type j at position k and F_j is the fraction of the same amino acid in the complete data set.

Using the above formula, the propensity of these amino acids to occur in the positions $i+1$, $i+2$, $i+3$ of the α-helical α-turn have been computed. These along with average propensity are given in columns 2–5 of Table 3. It is evident that four amino acids Cys, Glu, Ser and Thr show higher propensity (> 1.2) to occur in α-turns. (These are shown in bold in the table.)

Statistical significance of the propensity values has been determined using the d-test[11]. The statistic d is given by

$$d = \frac{(n \times f_j^k) - (n \times F_j)}{\sqrt{n \times F_j \times (1 - F_j)}},$$

where n is the number of examples in F1.

The amino acids with statistically significant propensity values for 5% tail ($d > 1.97$) are shown with an asterisk in Table 3. Of the four amino acids with high propensity values, only three (Glu, Ser and Thr) are seen to be statistically significant. The amino acid Cys, in spite of showing a high propensity value, is not statistically significant (probably due to the small sample size).

Since the α-turn when repeated, leads to an α-helix, it will be interesting to compare these propensity values with the corresponding ones for α-helical segments in proteins. The propensity values for amino acids to occur in an α-helix as computed using the data set used are

Table 3. Propensity of the 20 amino acids to occur in α-helical α-turn (family F1) and in α-helix. Those whose average propensity values are > 1.2 are in bold

Amino acid	Propensity to occur in α-turn in position $i+1$	$i+2$	$i+3$	Ave	α-helix
ALA	1.38*	1.38*	0.69	1.15	**1.48**
ARG	1.05	1.58	0.00	0.88	**1.23**
ASN	0.00	1.30	0.87	0.73	0.80
ASP	0.67	1.67*	1.00	1.11	0.94
CYS	**2.50**	**2.50**	**2.50**	**2.50**	0.78
GLN	0.57	0.57	1.71	0.95	**1.28**
GLU	1.54*	**2.69**	**2.31***	**2.18**	**1.46**
GLY	0.00	0.53	0.26	0.26	0.41
HIS	0.69	1.38	1.38	1.15	1.03
ILE	0.40	0.00	0.80	0.40	1.02
LEU	0.55	0.55	0.82	0.64	**1.31**
LYS	1.23*	1.54*	0.62	1.13	**1.27**
MET	0.80	0.00	0.00	0.27	**1.33**
PHE	0.00	0.00	1.11	0.37	1.06
PRO	**3.26***	0.00	0.00	1.09	0.56
SER	**2.19***	1.37*	0.82	**1.46**	0.75
THR	1.93*	0.97	**3.87***	**2.26**	0.75
TRP	0.00	0.00	0.83	0.28	1.14
TYR	0.59	0.00	0.00	0.20	0.85
VAL	0.63	1.25*	0.94	0.94	0.90

*Statistically significant as per d-test.

given in column 6 of Table 3. The values are in good agreement with those reported earlier by other workers[24,25]. The amino acids that exhibit high propensity (> 1.2) are shown in bold. (Glu, Met, Gln, Lys and Arg). It is interesting to note that, except for Glu, none of the other amino acids show large propensity both for α-helix and α-turn. This shows that the *amino acids that have larger propensity to occur in α-turn have no greater tendency to occur in α-helices and vice versa.* This qualitatively lends an explanation to the isolated nature of the α-turns. The above analysis shows that among the 20 amino acids, Ser and Thr have formed α-turns in spite of their low propensity for α-helices.

Another way of looking at the α-turns is to consider them as an overlap of two successive β-turns ($i \rightarrow i+3$ and $i+1 \rightarrow i+4$), with the β-turns being of type I and/or type III. Identifying β-turns by the criteria of Ramakrishnan and Srinivasan[26] (and not necessarily by the existence of $4 \rightarrow 1$ type hydrogen bond), 42 out of 50 examples of family F1 fall under this category (in 27 examples two type I β-turns overlap and in 15 examples type III and type I β-turns successively overlap). The propensities of these amino acids to occur in various types of β-turns have been reported by Wilmot and Thornton[11]. The interesting aspect is that the propensities of the amino acids Cys, Ser and Thr (which show greater propensity for α-turns) are also greater than 1.0 for type I β-turns. A further analysis shows that less than one-third of the overlapping β-turns (42 out of 153) alone form the α-turn.

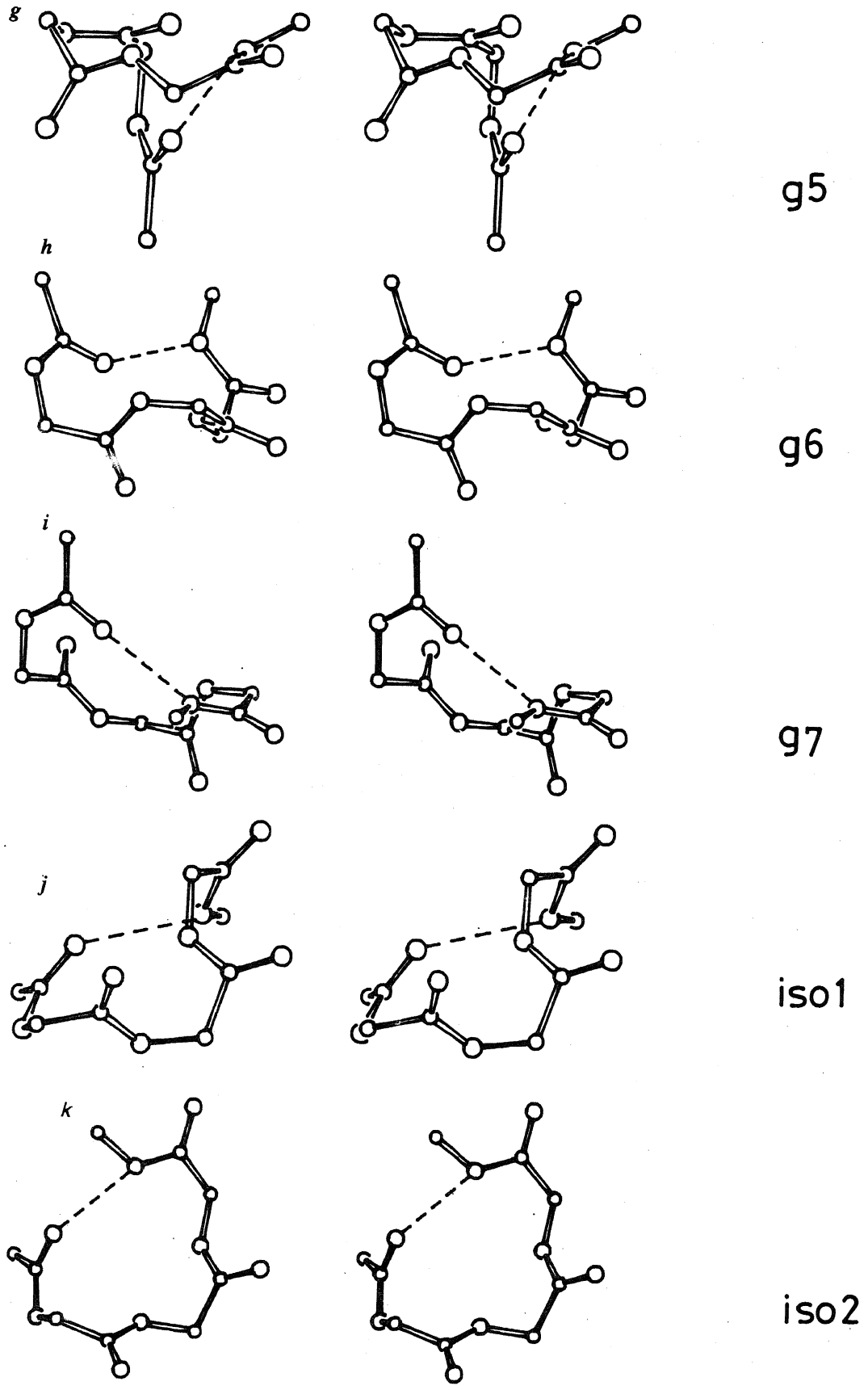

Figure 5 *g–k*.

Figure 5. *a–k*, ORTEP stereo diagrams of one representative example from different categories of α-turns. The 5 → 1 hydrogen bond characterizing the α-turn is shown as dotted lines and the category is marked. The diagrams have been taken using a version of ORTEP modified to suit PC.

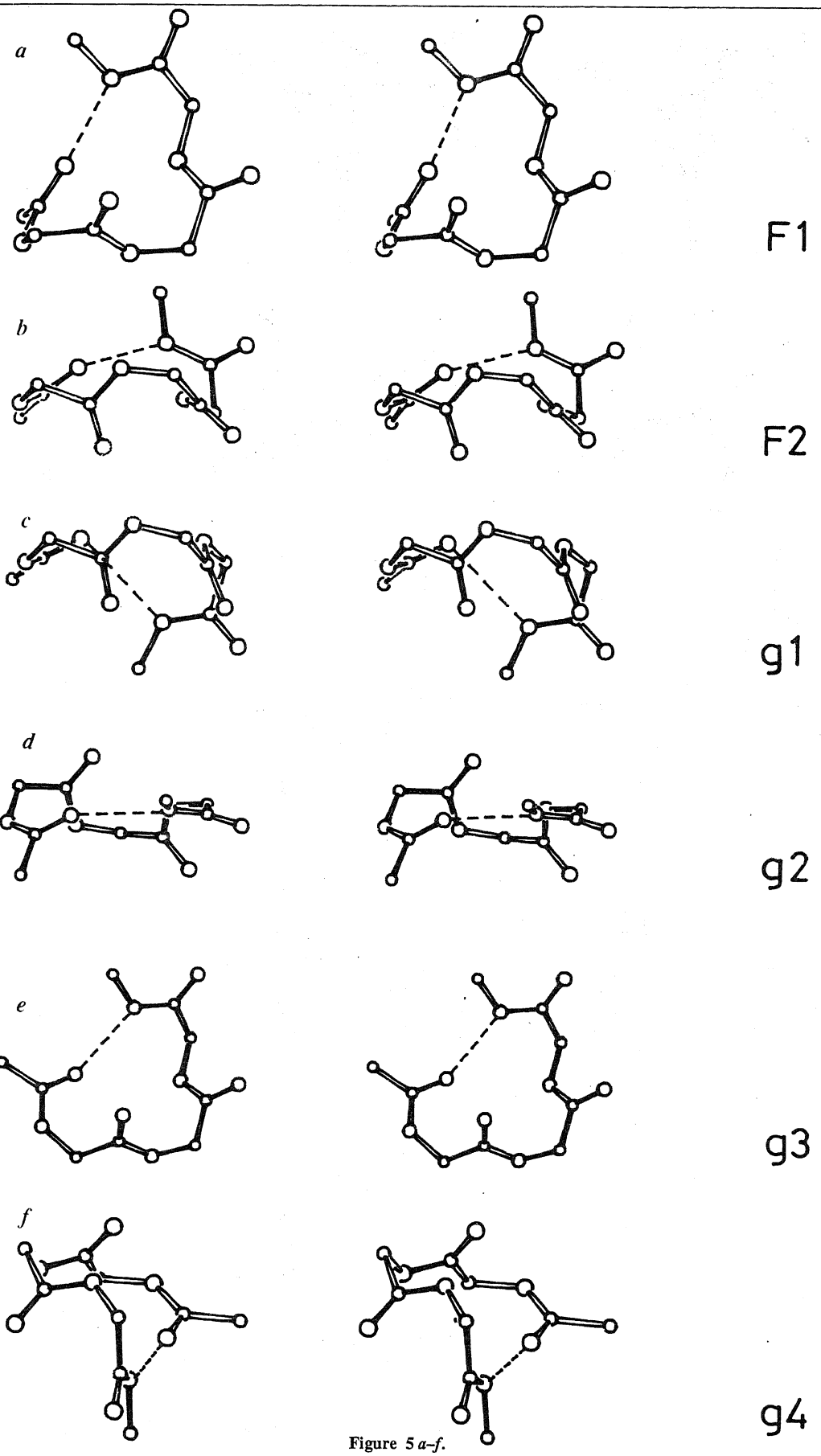

Figure 5 *a–f.*

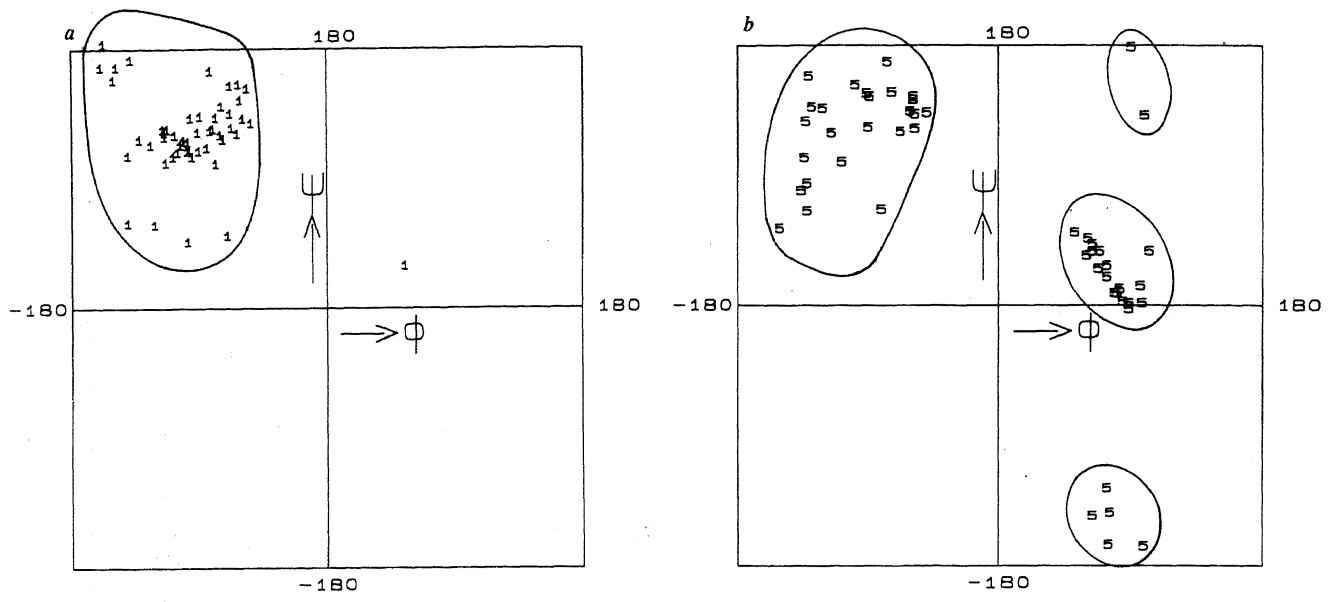

Figure 4. Plot of the (ϕ, ψ) values at the two α-carbon atoms immediately preceding and succeeding the α-turn in the examples of family F1. The clustering is also marked: *a*, at C_1^α; *b*, at C_5^α.

examples). It is very interesting to note that all the points except one in Figure 4 *a*, occur in the same quadrant of the (ϕ, ψ) plane, with a cluster near the extended conformation. There are two major and one minor clusters observable in Figure 4 *b*. One of the two major clusters occurs close to the extended region and the other in the α_L region. The minor cluster occurs in the inverse extended region (similar to the one designated as E′ in Figure 2 *a* and all these are glycyl residues). Thus, it is possible to have three motifs (while the term family is used to denote the segment with similar values at (ϕ, ψ) C_2^α, C_3^α and C_4^α, the term motif is used to denote the segment with similar (ϕ, ψ) values at all the five positions C_1^α to C_5^α) for this family, which can be represented by ($\beta - \alpha_R - \alpha_R - \alpha_R - \beta$), ($\beta - \alpha_R - \alpha_R - \alpha_R - \alpha_L$), ($\beta - \alpha_R - \alpha_R - \alpha_R - E'$), and the isolated example, which occurs in the protein enolase[21] (segment 102–106) can be designated as ($\alpha_L - \alpha_R - \alpha_R - \alpha_R - E'$).

The next largest family F2 has only 11 examples and is formed by conformational combination of E, α_L and α_R. This and other groups show that different combinations of (ϕ, ψ) at the three α-carbon atoms do occur in proteins, resulting in the formation of $5 \rightarrow 1$ hydrogen bonds to stabilize the local nonhelical regions. The designations of the families/groups are given in Table 1 *b*. Stereo diagrams of one representative example from each of the families F1, F2, groups g1–g7 and the two isolated examples are shown in Figure 5 *a–k*.

In the sections that follow, the different characteristics and features of α-turns are examined in detail.

End-to-end distances

The end-to-end distance (C_1^α–C_5^α) in the examples of α-turns varies between 4.7 and 6.7 Å. The distribution among the various families/groups in the different ranges is given in Table 2. (There are two examples of α-turns occurring in the protein 2cyp[22] (segment 35–39 belonging to F1) and in the protein 1gox[23] (segment 345–349 belonging to g1), which have end-to-end distances lying outside this range of 4.7–6.7 Å. The end-to-end distance is 4.2 Å in the former and 7.0 Å in the latter.)

The major family F1 has examples throughout the range, indicating flexibility in the chain segment. Groups g2, g3, g4 and g6 have smaller ranges of end-to-end distances (< 5.7 Å) and these groups have served the purpose of bringing the two ends of the segments much closer. On the other hand, the three members of group g5 are characterized by distances in the higher end of the spectrum (5.7–6.7 Å). It is worthwhile to mention that the computation of the corresponding distance in the various α-helical segments occurring in the data set shows that 92% of the examples lie in the range 4.0–7.0 Å (the ideal value for α-helix being 6.2 Å).

Propensity of amino acids to occur in α-helical α-turn

As mentioned earlier, the family F1, which contains α-helical α-turns, comprises a majority of examples. Since this is the shortest helix that can be formed, it

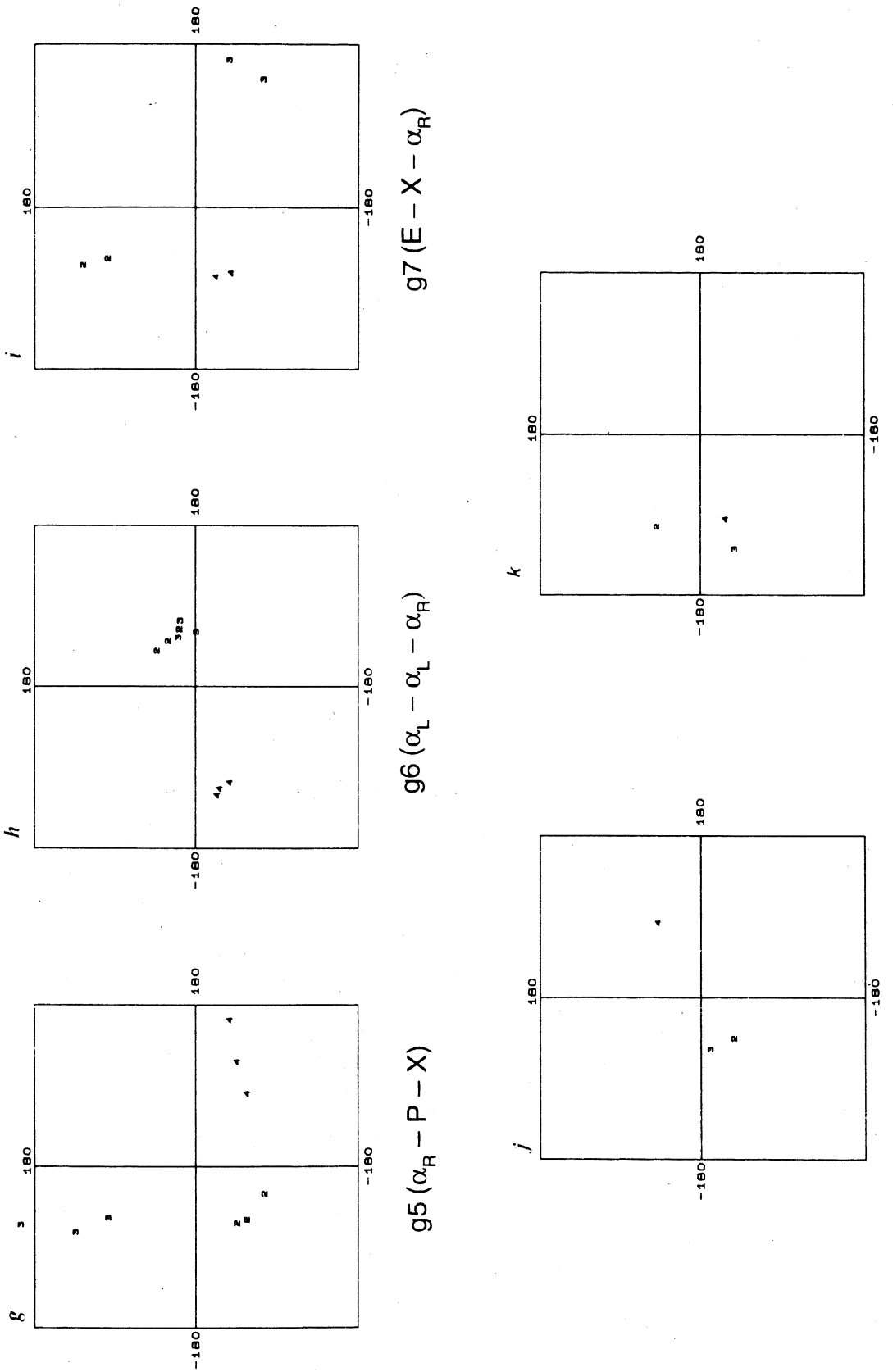

Figure 3. *a–k*, The plot of (ϕ, ψ) at the three α-carbon atoms in the α-turns in proteins grouped into different categories, based on conformational similarity. The conformational points corresponding to C_2^α, C_3^α and C_4^α are marked as 2, 3 and 4, respectively. The designation of each category is also marked.

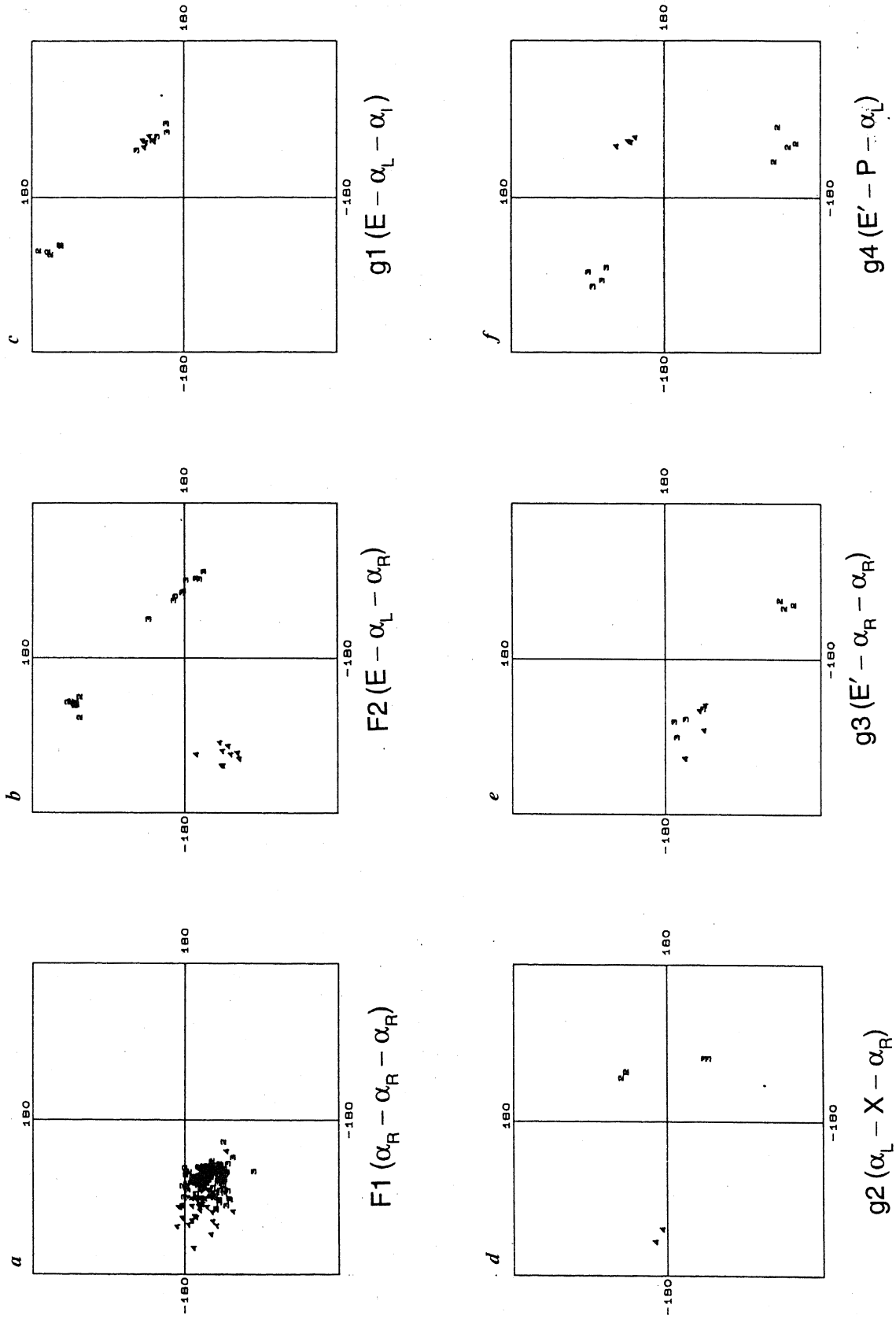
a
F1 (αR − αR − αR)
b
F2 (E − αL − αR)
c
g1 (E − αL − αL)
d
g2 (αL − X − αR)
e
g3 (E′ − αR − αR)
f
g4 (E′ − P − αL)
180
−180

Figure 3 *a–f.*

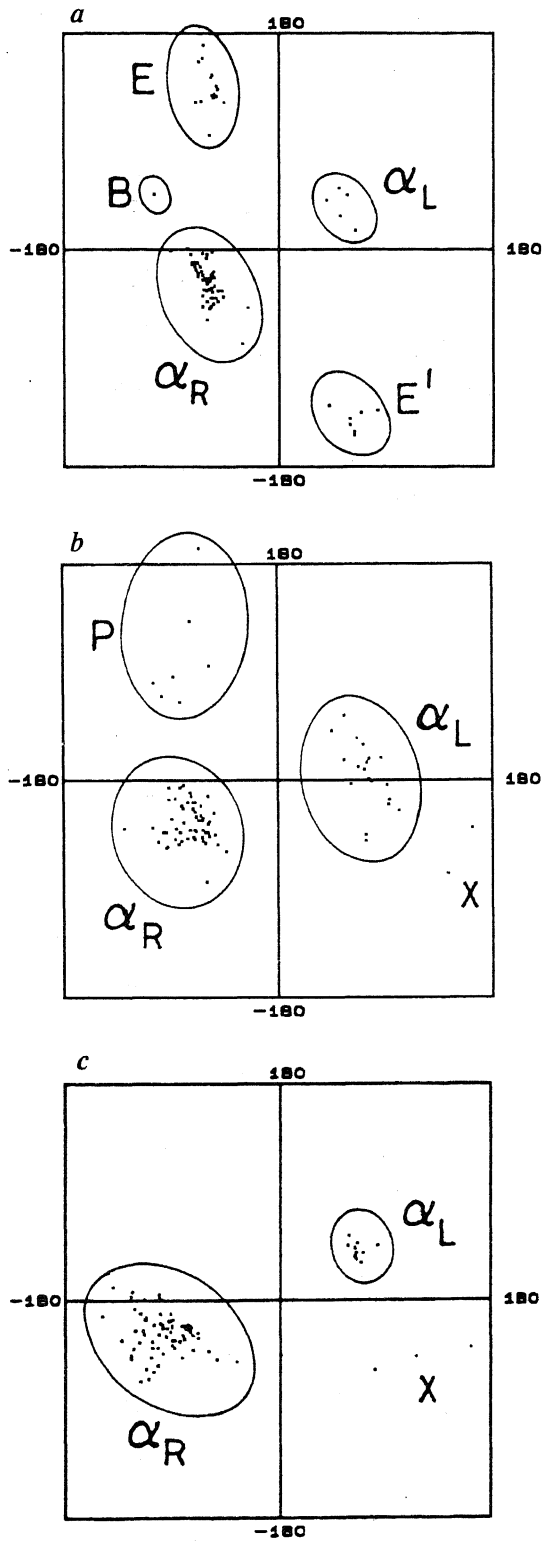

Figure 2. The (ϕ, ψ) plot of the conformational points at the three α-carbon atoms in the examples of isolated α-turns: *a*, at C_2^α; *b*, at C_3^α; *c*, at C_4^α. The clusters in each of the plots are also marked, with their designations indicated.

(1ova)). In addition to this, there is a solitary example at (–106°, 46°) (referred to as B, as this occurs in the bridge region of ($\phi - \psi$) map[20]). Thus, the distribution at C_2^α points to the possibility of at least four major families.

Figure 2 *b*, which corresponds to the ($\phi - \psi$) distribution at C_3^α, shows three prominent clusters. The one which lies in the α_R region comprises a majority of the points. The other two clusters are slightly more scattered as compared to the α_R region. One of these occurs in the right half of the map spanning across the ϕ-axis. This is denoted as α_L. The third one occurs in the left top quadrant of the map, in which the extended conformations normally occur. But five of the seven points are clustered closer to the poly proline region of the map (–60°, 120°) than the $\psi = 180°$ axis. Hence, we have preferred to use the symbol P rather than E to denote this cluster (see Figure 2 *a*). In addition to these, there are two isolated points occurring near the right edge of the map which are designated as X.

Figure 2 *c*, which shows the distribution of (ϕ, ψ) at C_4^α, has two clusters, a major one in α_R and a minor one in α_L region. In addition to these there are three isolated points (X) occurring in the right bottom quadrant of the map. It may be worthwhile to note that, there are no points in the 'extended' region of the map as contrasted with Figure 2 *a* and *b*.

Classification into families

As the formation of $5 \rightarrow 1$ hydrogen depends on the *combined set* of (ϕ, ψ) values at the three α-carbon atoms *in a collective way*, the examples have been carefully looked into and grouped into different families such that members of the same family have nearly the same combination of (ϕ_2, ψ_2), (ϕ_3, ψ_3), (ϕ_4, ψ_4). This exercise has revealed that a large fraction of examples (60%) form one major family (designated as F1). The other families contain relatively small number of examples. The family which has the next highest number (9 examples) is designated as F2. Apart from these, seven more groups can be identified, having 2–5 examples. In view of the small number of examples, these will be referred to as groups g1–g7. In addition, there are two isolated examples. The details pertaining to the members in various families and groups are given in Table 1 *a*. The (ϕ, ψ) plots are shown in Figure 3 *a–k* and the average (ϕ, ψ) values at C_2^α, C_3^α and C_4^α for each category are given in Table 1 *b*.

The family F1, which has a bulk of examples, is characterized by α_R conformations at all the three carbon atoms C_2^α, C_3^α and C_4^α. Thus, this family, which can be designated as 'α-helical α-turn' represents the shortest helix that can possibly occur. The (ϕ, ψ) values at the two α-carbon atoms immediately preceding and succeeding the turn, i.e. at C_1^α and C_5^α, are shown in Figures 4 *a* and *b* (absence of points in the α_R region in this figure confirms the isolated nature of the α-turn in these

Table 1a (*contd.*)

PDB code	Segment	Residues				
1gcr	136–140	M	P	S	Y	R
2cyp	35–39	E	Y	D	N	Y
2fbj	L48–52	Y	E	I	S	K
2prk	212–216	W	I	G	G	S
Group g2 (2)						
2fbj	H100–104	H	Y	Y	G	Y
2rhe	50–54	Y	Y	N	D	L
Group g3 (4)						
1fkf	82–86	Y	G	A	T	G
1ova	A69–73	K	D	S	T	R
3grs	55–59	G	G	T	C	V
4bp2	25–29	Y	G	C	Y	C
Group g4 (4)						
2act	188–192	W	G	E	E	G
2lzm	27–31	I	G	I	G	H
3apr	E11–15	Y	G	N	D	I
4fxn	56–60	M	G	D	E	V
Group g5 (3)						
1fxl	71–75	F	I	P	L	F
1ypi	A25–29	E	R	L	N	T
4fxn	40–44	I	D	E	L	L
Group g6 (3)						
1gdl	O300–304	I	D	G	K	M
2ltn	A100–104	Y	L	G	V	F
4pep	9–13	Y	L	D	T	E
Group g7 (2)						
1ak3	A129–133	H	P	G	S	G
1fxl	72–76	I	P	L	F	D
Isolated examples						
1rbp	63–67	L	L	N	N	W
2fbj	L166–170	D	S	K	D	S

In those proteins which contain two or more examples, the segments are well separated from one another. The two examples occurring in protein flavodoxin (1fx1) are somewhat unique in that the two α-turn segments 71–75 and 72–76 overlap, resulting in a longer segment 71–76. The individual segments are, however, conformationally different and the full segment 71–76 is not α-helical.

For those segments that have been picked up as α-turns, the (ϕ, ψ) angles at the three intervening α-carbon atoms C_2^α, C_3^α, and C_4^α have been computed and analysed.

Conformational analysis

The (ϕ, ψ) angles at each of the three C^α atoms have been plotted on the (ϕ − ψ) plane and these are shown in Figure 2 *a–c* corresponding to C_2^α, C_3^α, and C_4^α, respectively. These plots are first analysed independently and then used collectively to arrive at families of α-turns.

In Figure 2 *a*, which corresponds to the conformation at C_2^α, four groups of points are plainly seen, one of them being a dense cluster having a very large number of examples near the right-handed α-helical region designated as α_R. Of the remaining three clusters, one which occurs in the left-handed α-helical region is designated as α_L. The third, having few examples, occurs near the extended conformational region and has been designated as E. The fourth cluster which has eight examples occurs in a region that can be considered as inverse to the cluster E and has been denoted as E′. (Seven out of these eight examples are Gly and the odd member out is the residue Asp, A70 in the protein ovalbumin[19]

Table 1b. Average (ϕ, ψ) values (rounded off to the nearest 5°) at the three intervening α-carbon atoms of the α-turn segment along with their designations

Family/ group	Average (ϕ, ψ) at C_2^α	C_3^α	C_4^α	Designation*
F1	(−65, −25)	(−75, −35)	(−100, −20)	$\alpha_R-\alpha_R-\alpha_R$
F2	(−60, 130)	(75, 0)	(−115, −50)	E–$\alpha_L-\alpha_R$
g1	(−65, 155)	(70, 30)	(60, 40)	E–$\alpha_L-\alpha_L$
g2	(0, 95)	(70, −15)	(−40, 25)	α_L – X–α_R
g3	(60, −145)	(−75, −25)	(−80, −40)	E′–α_R–α_R
g4	(60, −140)	(−95, 75)	(60, 40)	E′–P–α_L
g5	(−50, −60)	(−70, 105)	(120, −50)	α_R–P–X
g6	(50, 30)	(60, 10)	(−120, −30)	α_L–α_L–α_R
g7	(−65, 110)	(150, −60)	(−80, −30)	E–X–α_R
#	(−50, −40)	(−60, −15)	(80, 45)	α_R–α_R–α_L
#	(−105, 45)	(−130, −40)	(−100, −30)	B–α_R–α_R

*For the location of regions α_R, α_L, E, E′, B, P, and X see Figure 2.
– Isolated examples.

α-turn, the proteins are examined to find out all tetrapeptide segments which have this hydrogen bond. The hydrogen bond criteria used for this purpose is as follows.

The hydrogen bond length l (N . . O distance) must lie in the range 2.6–3.6 Å and the hydrogen bond angle θ (H–$\hat{N}$ . . O) should be less than or equal to 40°. (The criteria used here are slightly relaxed compared to those that have been normally used in other conformational studies. For example, in small molecules, where the crystal structures are accurately determined, the hydrogen bond length is in the range 2.6–3.2 Å and hydrogen bond angle is usually less than 30° (ref. 17)). The reason for relaxing the criteria is the lesser accuracy in protein structure determination as well as the lack of information on the positional coordinates of hydrogen atoms.)

As the hydrogen positions are not available in the coordinate data of Protein Data Bank (PDB) (with few exceptions, however), the H-bond angle θ could not be readily evaluated. For this purpose, the positional coordinates of the amide hydrogens have been computed by geometrically fixing the hydrogen atoms at the corresponding nitrogen atoms in the plane of the peptide unit using the ideal bond angles around the nitrogen atom[18].

Results and discussion

Using the H-bond criteria mentioned above, it has been possible to identify 2525 tetrapeptide segments in the proteins which have $5 \rightarrow 1$ hydrogen bonds. These examples obviously include those segments which are part of α-helices, since an α-helix is a succession of $5 \rightarrow 1$ hydrogen-bonded segments. The removal of the segments that are part of an α-helix has been done by a combination of (φ, ψ) criteria and by inspection. The conformations at (φ, ψ) at the five α-carbon atoms are designated as R or X, whether or not it lies in the range -130 to 0° for φ and -130 to 30° for ψ. If there are four **successive** R's (i.e. the designations at the four α-carbon atoms i, $i+1$, $i+2$ and $i+3$ or at $i+1$, $i+2$, $i+3$ and $i+4$ are R's) then it is treated as forming a part of α-helix. The examples which still remain are further examined by inspection to remove any possible borderline cases. On removing those examples which form part of an α-helix, there were 84 examples which remained, and these can be considered as isolated α-turns. (The corresponding numbers of examples for β- and γ-turns are 383 and 210, respectively). The PDB codes of the proteins and the segments in which these α-turns occur are listed in Table 1 *a*. These examples occur in 51 of 107 proteins of the total set. The number of examples of these α-turns in a protein varies from one to four except in the protein phosphate dehydrogenase (1gdl), where there are five examples.

Table 1a. α-Turn segments and the residues in them occurring in the data set used, grouped into different families and groups.The number of examples in each family/group is given within parentheses

PDB code	Segment	Residues				
Family F1 (50)						
1aap	A24–28	D	V	T	E	G
1acx	82–86	D	C	A	T	D
1ak3	A149–153	D	D	L	T	G
1cox	391–395	N	A	G	S	G
1drf	152–156	D	L	E	K	Y
1eca	38–42	F	T	Q	F	A
1gd1	O47–51	D	S	V	H	G
1gd1	O138–141	D	P	K	A	H
1gox	345–349	S	L	K	E	I
1mba	43–47	F	A	D	F	K
1omd	2–6	I	T	D	I	L
1ova	A277–281	S	S	N	V	M
1ova	A318–322	L	S	G	I	S
1rdg	6–10	C	T	V	C	G
1rdg	14–18	D	P	A	K	G
1rdg	39–43	C	P	V	C	G
1sn3	7–11	K	K	S	D	G
1sn3	31–35	A	K	N	Q	G
1thb	A113–117	L	P	A	E	F
2act	85–89	T	E	E	N	Y
2alp	140–156	G	R	T	T	G
2aza	A40–44	A	K	S	A	M
2ca2	34–38	D	T	H	T	A
2cpp	77–81	D	Y	R	H	F
2cpp	328–332	N	E	R	E	N
2csc	59–63	D	P	D	E	G
2cyp	58–62	D	K	H	D	N
2er7	E240–244	S	S	S	V	G
2fcr	94–98	D	A	E	G	Y
2fcr	148–152	D	M	V	N	D
2ltn	A54–58	D	R	E	T	G
2ltn	A125–129	N	A	A	W	D
2ltn	A167–171	N	A	A	T	N
2rhe	93–97	N	D	S	L	D
2rsp	A46–50	S	E	E	D	W
2sga	140–156	G	S	T	T	G
2sga	219–223	N	C	R	T	G
2trx	A59–63	N	I	D	Q	N
3blm	50–54	D	T	K	S	G
3cla	97–101	H	Q	E	T	E
3cla	194–198	H	H	A	V	C
3fgf	67–71	G	V	C	A	N
3grs	164–168	H	E	S	Q	I
4cha	A95–99	N	S	L	T	I
4enl	102–106	K	S	K	L	G
4fd1	35–39	H	P	D	E	C
5cpa	3–7	S	T	N	T	F
5cpa	29–33	H	P	E	L	V
5p21	145–149	S	A	K	T	R
6ldh	181–185	H	S	C	S	C
Family F2 (9)						
1ak3	A137–141	N	I	E	F	N
1cox	453–457	L	L	N	S	A
1fkf	87–91	H	P	G	I	I
1gdl	O129–133	V	M	G	V	N
1gdl	O267–271	L	K	G	I	L
2tec	E261–265	G	T	G	T	Y
2trx	A49–53	Y	Q	G	K	L
4fdl	8–12	C	I	K	C	K
5cpa	89–93	N	Y	G	Q	N
Group g1 (5)						
1gcr	47–51	R	P	N	Y	Q

22. Finzel, B. C., Poulos, T. L. and Kraut, J., *J. Biol. Chem.*, 1984, **259**, 13027–13036.
23. Lindqvist, Y., *J. Mol. Biol.*, 1989, **209**, 151–166.
24. Chou, P. Y. and Fasman, G. D., *Biochemistry*, 1974, **13**, 211–222.
25. Leszczynski, J. F. and Rose, G. D., *Science*, 1986, **234**, 849–855.
26. Ramakrishnan, C. and Srinivasan, N., *Curr. Sci.*, 1990, **59**, 837–847.
27. Eisenberg, D., Weiss, R. M., Terwilliger, T. C. and Wilcox, W., *Faraday Symp. Chem. Soc.*, 1982, **17**, 109–120.
28. Soman, K. V. and Ramakrishnan, C., *Int. J. Biol. Macromol.*, 1986, **8**, 89–96.
29. Milner-White, E. J., and Poet, R., *Biochem. J.*, 1986, **240**, 289–292.
30. Pavone, V., *Int. J. Biol. Macromol.*, 1988, **10**, 238–240.
31. Sibanda, B. L., Blundell, T. C. and Thornton, J. M., *J. Mol. Biol.*, 1989, **206**, 759–777.
32. Vrielink, A., Lloyd, L. F. and Blow, D. M., *J. Mol. Biol.*, 1991, **219**, 533–554.
33. Summers, L., Wistow, G., Narebor, M., Moss, D., Lindley, P., Slingsby, C., Blundell, T., Bartunik, H. and Bartels, K., *Pept. Protein Rev.*, 1984, **3**, 147–168.
34. Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D. and Davies, D. R., *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 7009–7013.
35. Bhat, T. N., Padlan, E. A. and Davies, D. R., (private communication through PDB).

ACKNOWLEDGEMENTS. We thank Prof. P. Balaram and Dr N. V. Joshi for their suggestions and comments.

Received 24 July 1995; accepted 31 July 1995

MEETINGS/SYMPOSIA/SEMINARS

Winter School on 'NMR in Structural Biology' (an event for the Golden Jubilee Celebrations of TIFR)

Date: 5–18 November 1995
Place: TIFR, Bombay

Topics include: Multidimensional high resolution NMR spectroscopy in liquids; Introduction to pulse/field gradients; Applications to nucleic acids and proteins; Biomembranes/cellular NMR; NMR imaging; Molecular dynamics/molecular simulations.

Contact: Dr (Mrs) Sudha Srivastava
National Facility for High Field NMR
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 400 005, India
Phone: 22-215 2971 Ext. 2271
Telex: 011-83009 TIFR IN
Fax: 091-22-2152110
e-mail: sudha@tifrvax.tifr.res.in

Sixth All India Conference on Cytology and Genetics

Date: 23–25 February 1996
Place: Rohtak

Contact: Dr M. S. Chennaveeraiah
Hon. Secretary
Society of Cytologists and Geneticists
No. 9, Byrasandra Main Road
First Block East, Jayanagar
Bangalore 560 011, India

International Conference on Natural and Technological Coastal Hazards

Date: 2–6 December 1995
Place: Tirupati, India

Contact: Prof. K. L. Narasimha Rao
Department of Geology
Sri Venkateswara University
Tirupati 517 502, India
Fax: +91-8574-24111/25211

Second Congress on Traditional Sciences and Technologies of India

Date: 27–31 December 1995
Place: Anna University, Madras

Subjects include: Agriculture and livestock management; Forestry, water management, fisheries, bio-diversity; Health-care, food and nutrition; Architecture and building materials; Metallurgy and metal-working; Textiles and natural dyes; Pottery, leather, wood and bamboo, other industries; Ship-building and navigational technologies; Indian mathematics.

Themes include: Role of communities; Education and training; Local markets; Role of women; Economic policies for traditional S&T.

Contact: The Madras Secretariat
Traditional S&T Congress
Students' Centre, Anna University
Madras 600 025, India
Phone: (044) 2351126-Extn: 3113
Fax: (044) 2368403/2350397