Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 18, Issue Number 2, (2000)

Conformational Similarity Indices Between Different Residues in Proteins and α-Helix Propensities

http://www.adeninepress.com

Debnath Pal and Pinak Chakrabarti*
Department of Biochemistry,
P-1/12 CIT Scheme VIIM,
Bose Institute,
Calcutta 700 054 India

Abstract

Various amino acid similarity matrices have been derived using data on physicochemical properties and molecular evolution. Conformational similarity indices, $CS_{XX'}$, between different residues are computed here using the distribution of the main-chain and side-chain torsion angles and the values have been used to cluster amino acids in proteins. A subset of these parameters, $CS_{AX'}$ indicates the extent of similarity in the main-chain and side-chain conformations (ϕ,ψ and χ_1) of different residues (X) with Ala (A) and is found to have strong correlation with α-helix propensities. However, no subset of $CS_{XX'}$ provides any linear relationship with β-sheet propensities, suggesting that the conformational feature favouring the location of a residue in an α-helix is different from the one favouring the β-sheet. Conformationally similar residues (close CS_{AX} values) have similar steric framework of the side-chain (linear/branched, aliphatic/aromatic), irrespective of the polarity or hydrophobicity. Cooperative nucleation of helix may be facile for a contiguous stretch of residues with high overall CS_{AX} values.

Introduction

To understand the physicochemical forces underlying the conversion of a given amino acid sequence into a unique protein fold various amino acid properties have been invoked (1,2). These physical characteristics can be used in sequence alignment methods, especially for the detection of weak sequence homologies (3,4). However, very few of these properties relate to the general conformational features of different residues. Conformational preferences of residues are generally expressed as the propensities to occur in different secondary structural elements (5-6) and are of immense value in structure prediction (7). From an analysis of the amino acid specific main-chain torsion angle distributions Niefiend and Schomburg calculated a set of similarity parameters, which constituted the scoring matrix in protein sequence alignment (8). Likewise, Kolaskar and Kulkarni-Kale have derived a conformational similarity matrix using the ϕ,ψ probability maps of 20 amino acid residues and used it to identify conformationally similar protein fragments (9). We have earlier shown that the ϕ,ψ distribution is dependent on the side chain conformation (10) and it is desirable to have a similarity matrix based on the three-dimensional distribution of ϕ,ψ and χ_1 angles. In this paper we derive such a matrix, whose elements, $CS_{XX'}$, signify the conformational similarity between pairs of amino acids (X and X'), compare it with other known amino acid exchange matrices (11) and show that CS_{AX} indices representing the extent of conformational similarity of different residues to Ala, are correlated to α-helix propensities.

The secondary structure, α-helix, proposed by Pauling *et al.* (12), is characterized by intrahelical CO···NH hydrogen bonding (between residues at position *i* and *i*+4) and main-chain torsion angles, $\phi = -65°$ and $\psi = -40°$ (13). Although these parameters are defined using the main chain atoms, the side-chain has an effect as dif-

*Phone: (91) 33-337-9416;
Fax: (91) 33-334-3886;
E-mail: pinak@boseinst.ernet.in

ferent residues have different helix propensities (5,14). To understand protein folding we need to know what factors determine the tendency of short amino acid sequences to constitute helices in proteins and if such structures can be maintained by the isolated stretch of the polypeptide chain (independent of the rest of the protein molecule) in solution (15-17). The problem is compounded by the fact that helix formation can also be influenced by the context of the sequence and tertiary interactions (18,19). However, unlike the β-sheet propensities, the statistically determined α-helix propensities correlate better with experimentally observed values indicating that the former is tertiary-context dependent (20-22), whereas the latter is more intrinsic to a given residue. Indeed, the most commonly used model to describe the energetics of helix formation by short polypeptide chains, the helix-coil transition, uses only two parameters, a nucleation factor (σ) and an elongation factor (s) for a given amino acid residue and these are assumed to be independent of the sequence environment (23,24). The helix propensity of an amino acid as determined experimentally using short peptides is equated with s (16).

Helix propensities have been measured in different monomeric peptide systems (25-29) and small, single-domain proteins (18,30,31), as well as a coiled-coil leucine zipper peptide of *de novo* design (32). Although the different sets of values obtained do not agree numerically, they are significantly correlated between themselves and with the statistical propensity values derived from the structure database (16,33,34). Indeed, one can have a consensus rank order of helix propensities (16). Ala has the highest value, followed by amino acids with long side-chains (Arg, Leu, Lys, Gln, Glu, Met). The other amino acids, except Gly and Pro, have intermediate to low propensities, and Gly and Pro have the lowest. The physical basis for the differences in helix propensities has been provided in terms of electrostatics, specific side-chain - backbone interactions, burial of side-chain apolar surface, loss in conformational entropy etc. (16,17). It is to be noted that there does not exist a single factor or parameter which correlates with the helix propensity value of a residue and which can thus be used to explain the helix forming tendencies of different residues.

As propensities are reflection of the conformational stabilities provided by the residues to helices in both proteins and peptides, we asked if a new descriptor of residue conformation can be defined which correlates with the helix propensity, and thus provides an understanding of the rank order of helix propensities. Ala has the highest α-helix propensity and the degree of conformational similarity of other residues to Ala is shown here to be highly correlated with the helix propensity.

Materials and Methods

A non-redundant and non-homologous (<25% pairwise sequence identity between molecules (35)) set of 294 protein structures determined by X-ray crystallography to a resolution of at least 2.0Å and refined to an R-factor of ≤ 0.20 was selected from the Protein Data Bank (36). φ,ψ and χ_1 angles were calculated; the χ_1 angle of Val was modified to conform to the convention used for Ile and Thr (10). The three rotameric states of the side chains are designated as t (χ_1 in the range 120 to 240°), g^+ (-120 to 0°) and g^- (0 to 120°). The total number and its distribution (%) among the above rotamers are as follows: Ser (4767,23,31,46), Cys (1106,28,56,16), Met (1538,28,64,8), Glu (4419,33,57,10), Gln (2843,32,60,8), Lys (4482,34,58,8), Arg (3378,34,57,9), Leu (6312,32,66,2), Asp (4812,32,50,18), Asn (3722,30,54,16), His (1757,33,54,13), Phe (3199,35,53,12), Tyr (2967,34,53,13), Trp (1246,34,50,16), Val (5472,8,72,20), Ile (4228,10,77,13), Thr (4647,8,44,48), Pro (3709), Ala (6757) and Gly (6418).

$CS_{XX'}$ values were computed by finding out the correlation coefficients between the three-dimensional φ,ψ, χ_1 distributions of the two residues (X and X′):

$$CS_{XX'} = \frac{\sum_i (N_{Xi} - \langle N_X \rangle)(N_{X'i} - \langle N_{X'} \rangle)}{\sqrt{\sum_i (N_{Xi} - \langle N_X \rangle)^2 \sum_i (N_{X'i} - \langle N_{X'} \rangle)^2}}$$

where N_{Xi} is the number of a residue X at grid i (of size $10° \times 10° \times 10°$) and $N_{X'i}$ is the number at the equivalent position for residue X′ and $\langle N_X \rangle$, $\langle N_{X'} \rangle$ are the averages of the numbers of the two residues. The choice of a 10° grid size has been found to be suitable in an earlier study (43). Also, CS_{AX} values (next paragraph) were calculated using various grid sizes; there were no significant differences between values obtained using 10, 15 or 20° grid size.

For comparing residues (Gly, Ala and Pro) with no (or restricted) χ_1, only the two-dimensional ϕ,ψ distribution was used. When comparing the three-dimensional ϕ,ψ,χ_1 distribution of a residue with the two-dimensional ϕ,ψ distribution of Gly/Ala/Pro, the former was divided into three ϕ,ψ distributions corresponding to the three rotameric states of χ_1, and each of them was independently compared to the latter. The weighted average (on the basis of the relative population of X in the three χ_1 states) provided the similarity index. CS_{AX} values relating Ala to all other residues were thus calculated. This method of calculation, using three χ_1-dependent ϕ,ψ maps, takes into account the effect of the side-chain on the ϕ,ψ distribution of X. For comparisons, we also calculated CS_{AX} values by considering only those points that lie in two distinct regions of the Ramachandran plot, broadly encompassing the two secondary structural elements – α-helical (ϕ = -180 to 0°, ψ = -120 to 60°) and β-sheet (ϕ = -180 to 0°, ψ = 60 to 240°) regions, and also by considering only those residues that are not located on any regular secondary structure.

Chou-Fasman (5) type propensity values, $P_{\alpha X}$, were derived using the same dataset, after identifying the residues located in the helices (of all types) using the program, DSSP (37).

Results and Discussion

(a) Conformational Similarity and Clustering of Amino Acid Residues

$CS_{XX'}$ values are presented in Figure 1. A larger value signifies a greater correlation between the maps of the two residues. A complete-linkage cluster analysis was performed (with distances between residues being 0.35 or less (*i.e.*, $CS_{XX'}$ of 0.65 or more) and the results (Figure 2) provide a pictorial representation of the residue clusters based on conformation. We had earlier classified amino acid residues (with atoms up to the γ position and beyond, but excluding Pro) into 5 classes based on a rather simple method of analysis of two-dimensional ψ,χ_1 and ϕ,χ_1 plots (only the negative range of ϕ was considered) of individual residues (10). The classes and their members are: (I) Ser, Cys, Met, Glu, Gln, Lys and Arg; (II) Leu; (III) Asp and Asn; (IV) Phe, Tyr, His and Trp; and (V) Val, Ile and Thr. The present analysis is based on three-dimensional ϕ,ψ,χ_1 maps (full range of ϕ and all residues included). While essentially confirming the earlier groupings it also shows how individual residues differ within a class. Ser which was found to be a constituent of class (I) is now shown to have a distribution of torsion angles fairly distinct from the other members, which must have been caused by its ability to form short range hydrogen bonded contact due to the presence of a hydroxyl group rather close to the main-chain atoms. Though Leu was quite alike other class (I) members, it was not put in the same group as it did not have significant presence in the g^- state of the side chain (10). Based on the present result Leu can be put in class (I), which is also justified as its α-helix propensity is quite similar to other long-chain members of the class (discussed later). Additionally, residues (with no or resticted χ_1, Ala, Gly and Pro), which were earlier left out,

are also placed relative to other residues. It is interesting to note that Ala ca indeed be placed along with other class (I) members; the relevance of which discussed later in the context of α-helix propensity.

SER	CYS	MET	GLU	GLN	LYS	ARG	LEU	ASP	ASN	HIS	PHE	TYR	TRP	VAL	ILE	THR	PRO	ALA	
0.58																			CYS
0.60	0.76																		MET
0.63	0.75	0.91																	GLU
0.59	0.76	0.90	0.94																GLN
0.60	0.74	0.84	0.91	0.91															LYS
0.61	0.74	0.84	0.90	0.91	0.93														ARG
0.57	0.78	0.90	0.93	0.93	0.92	0.90													LEU
0.57	0.73	0.85	0.82	0.81	0.75	0.73	0.80												ASP
0.50	0.70	0.77	0.74	0.76	0.73	0.70	0.75	0.86											ASN
0.54	0.63	0.64	0.71	0.70	0.78	0.77	0.72	0.60	0.63										HIS
0.49	0.61	0.58	0.68	0.68	0 80	0.79	0.73	0.48	0.52	0.78									PHE
0.51	0.60	0.58	0.67	0.67	0.78	0.78	0.71	0.47	0.51	0.77	0.90								TYR
0.53	0.59	0.58	0.68	0.68	0.78	0.77	0.72	0.50	0.53	0.69	0.85	0.82							TRP
0.48	0.53	0.65	0.60	0.60	0.51	0.52	0.58	0.61	0.50	0.38	0.35	0.34	0.32						VAL
0.48	0.55	0.68	0.62	0.63	0.53	0.54	0.60	0.64	0.53	0.38	0.34	0.33	0.31	0.97					ILE
0.63	0.40	0.48	0.48	0.45	0.39	0.41	0.41	0.46	0.34	0.31	0.26	0.28	0.25	0.74	0.73				THR
0.51	0.38	0.36	0.39	0.38	0.41	0.41	0.36	0.37	0.35	0.34	0.29	0.30	0.39	0.20	0.22	0.28			PRO
0.74	0.72	0.90	0.91	0.91	0.90	0.90	0.91	0.74	0.71	0.74	0.71	0.68	0.74	0.65	0.70	0.57	0.46		ALA
0.48	0.46	0.55	0.56	0.56	0.56	0.56	0.56	0.47	0.49	0.47	0.44	0.43	0.46	0.38	0.42	0.37	0.32	0.61	GLY

Figure 1: Matrix of conformational similarity indices relating different residues.

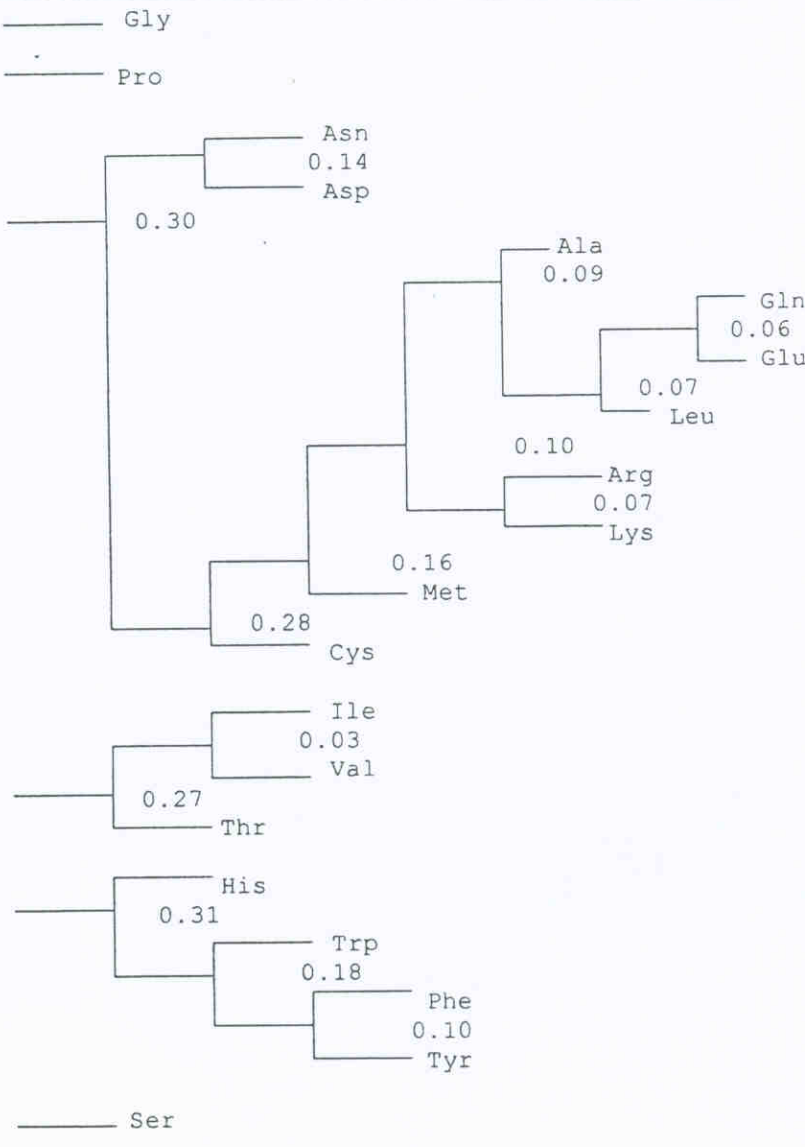

Figure 2: Minimum spanning tree obtained for ($1-CS_{XX'}$) values using complete-linkage cluster analysis with a threshold distance of 0.35. The distance between two residues or the maximum of all the distances between two clusters is indicated when they are below the threshold value.

(b) Similarity to Other Matrices

Tomii and Kanehisa (11) have collected 42 published similarity (or mutational) matrices derived using different physicochemical and biochemical properties of amino acids and which have been used for protein sequence alignments and similarity searches. On comparison we find that our matrix is quite different from others, the closest resemblance being with the one based on ϕ,ψ probability maps (9). But even with this the correlation coefficient is quite low (0.42), suggesting that the third dimension, χ_1, and consequently the side chain has made a significant contribution to $CS_{XX'}$ values. Even though the matrix from this study is quite unlike other matrices, residues in many of the conformationally similar clusters are found to be highly exchangeable within evolutionarily related proteins. For example, by analyzing the replacement pattern between amino acids in structurally similar proteins Risler *et al.* (38) delineated four strong clusters: (i) Ile and Val, (ii) Leu and Met, (iii) Lys, Arg and Gln, and (iv) Tyr and Phe. These residues are also shown here to be conformationally similar (Figure 2), thus suggesting that during evolution the substitution of residues is strongly dictated by conformational consideration. To be used in sequence alignment programs, the distance or probability matrix is usually converted into a somewhat arbitrary weight matrix (8,9,38). We are now investigating the utility of using the $CS_{XX'}$ matrix directly in comparing protein sequences to identify conformationally similar fragments.

(c) Correlation with α-Helix Propensity

The row for Ala in Figure 1 contains the CS_{AX} values indicating the degree of conformational similarity (CS_{AX}) between Ala (A) and all other residues (X). This measure lies between 0 (no similarity) and 1 (identical conformation). CS_{AX} values show a very high degree of correlation with α-helix propensities, $P_{\alpha X}$ (Figure 3), which means that the latter values match with how similar the conformation of the residue is with Ala.

We also examined if the magnitude of CS_{AX} values is just a reflection of the similarity of the distribution of ϕ,ψ (and χ_1) angles in the α-helical region. To rule out the possibility, we carried out two types of calculations. First, a modified CS_{AX} parameter was calculated using distributions where residues with any regular secondary structure were excluded. Such distributions would not suffer from any bias imposed by the secondary structure (helix in particular). These modified CS_{Ax} values also have a high correlation coefficient (0.8, excluding Pro and Ser) with $P_{\alpha X}$. Second and from a converse point of view, CS_{Ax} values were computed using only the α-helix and (for comparison) β-sheet regions of the ϕ,ψ map. These sets of values have lower correlation coefficients with $P_{\alpha X}$ values (Table I), and not unexpectedly, all residues show an increase in CS_{AX} values when the calculation is restricted to the helical region only. Thus the conformational similarity of X to Ala over the whole ϕ,ψ (and χ_1) space is what determines its helix propensity, and a stretch of residues having a high overall CS_{AX} value can, in a concerted way, form a helix.

Another point we considered was if residues with similar propensities for other secondary structures also have similar distribution of points in the ϕ,ψ,χ_1 space. For that we used the parameter, CS_{IX} (Figure 1), the conformational similarity defined relative to Ile, and found that it has a poor correlation (0.43) with the β-sheet propensity. This means that the propensity of a residue to be in β-sheet does not depend on how similar its conformational map is with that of Ile, the residue with one of the highest β-sheet propensity. These results indicate that the structural requirements for the formation of different secondary structures are different. For helix formation a contiguous stretch in the polypeptide chain should contain residues with high overall CS_{AX} values, which is not true for β-sheet formation where residues involved are from non-contiguous regions of the chain. Indeed, we have recently shown that there are other residue characteristics which correlate with β-sheet propensities (39).

Other helix propensity scales based on both experimental data and theoretical consideration were also compared with. Pace and Scholtz (34) have derived a scale using the available experimental data on 11 systems, including both proteins and peptides. A scale based only on data from peptides was developed by Muñoz and Serrano (28). Other scales considered were the structure-based one by Luque *et al.* (40) and another one by Koehl and Levitt (41) generated using computer-designed sequences. Results given in Table II show that CS_{AX} values are in excellent agreement with all but one (marked 'Design') of the scales.

Table I

Conformational similarity indices, CS_{AX}, for various residues and their correlation coefficients with Chou-Fasman type α-helix propensity values, $P_{\alpha X}$.

Ser	Cys	Met	Glu	Gln	Lys	Arg	Leu	Asp	Asn	His	Phe	Tyr	Trp	Val	Ile	Thr	Pro	Ala	Gly	Correlation
(a) Using the whole φ,ψ-space																				
0.74	0.72	0.90	0.91	0.91	0.90	0.90	0.91	0.74	0.71	0.74	0.71	0.68	0.74	0.65	0.70	0.57	0.46	1.00	0.61	0.92
(b) Using the whole φ,ψ-space, but excluding residues with any regular secondary structure																				
0.68	0.51	0.59	0.69	0.61	0.67	0.69	0.74	0.45	0.38	0.53	0.60	0.61	0.60	0.38	0.46	0.41	0.88	1.0	0.39	0.46[a]
(c) Using α-helical region only																				
0.77	0.84	0.96	0.92	0.94	0.93	0.94	0.95	0.84	0.83	0.81	0.85	0.81	0.85	0.81	0.80	0.66	0.56	1.00	0.98	0.64
(d Using β-sheet region only																				
0.75	0.56	0.66	0.71	0.69	0.70	0.73	0.63	0.51	0.51	0.62	0.60	0.59	0.64	0.42	0.40	0.50	0.71	1.00	0.64	0.45

[a]0.80, excluding Pro and Ser.

Table II

Comparison (using correlation coefficients) between CS_{AX} and some α-helix propensity scales.

	Pace (34)	Agadir (28)	Luque (40)	Design (41)
CS_{AX}	-0.86	-0.81	-0.83	-0.60

Pro was excluded in all the published scales.

The ranking of residues in terms of CS_{AX} (Figure 3) is chemically intuitive. Aliphatic side-chains (not branched before the γ position) have high CS_{AX} values; only Ser and Cys, with oxygen/sulfur atom at the γ position which can participate in short range interactions, have lower values. Other classes (10) of residues (like aromatic, β-branched etc.) have values in distinct ranges of CS_{AX}. This suggests that the topology of the side chain (linear, β-branched, γ-branched-aliphatic, γ-branched-aromatic, etc.) has the strongest bearing on the CS_{AX} values and in turn on the helix propensities, as they are highly correlated. Asp and Glu, which are similarly charged and should have similar electrostatic effects when placed in a helix, have very different helix propensities. Interestingly, these two residues have quite

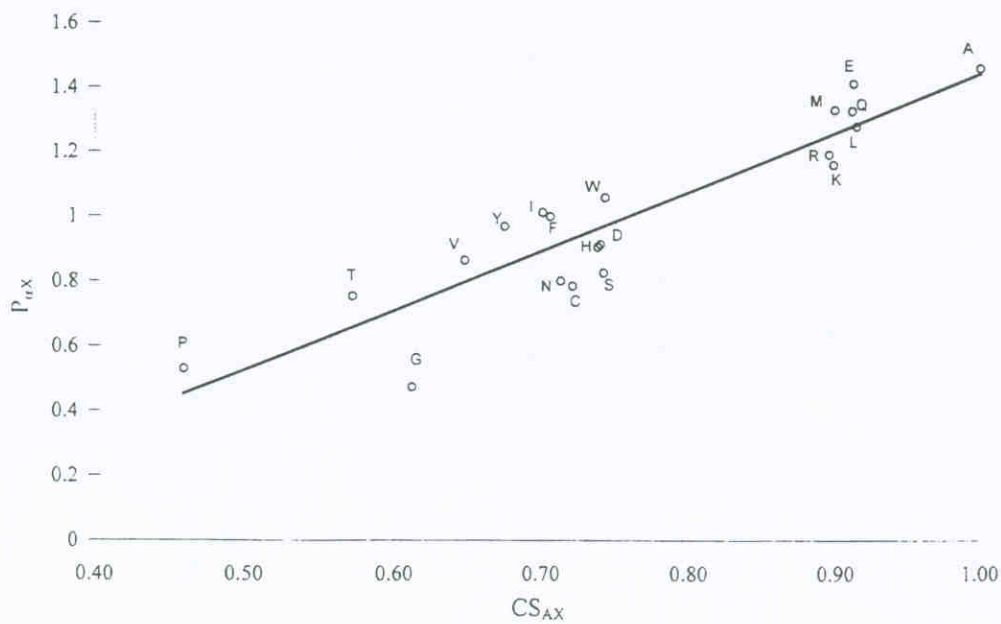

Figure 3: Plot of $P_{\alpha X}$ against CS_{AX} (details given in Table I). The fitted line has an equation $P_{\alpha X} = 1.827\ CS_{AX} - 0.388$.

different conformational features and belong to different clusters (Figure 2). Residues with close CS_{AX} values have similar ϕ,ψ,χ_1 distribution and should have similar potential energy surface. For the formation of an α-helix, four residues have to be in the α-helix conformation before an $i\cdots i+4$ hydrogen bond can stabilize the structure. Conformationally similar residues occurring in sequence can act cooperatively to form the nucleation centre (23,24) for the α-helix, which, once formed, can further be stabilized by additional interactions involving the side-chains and the main-chain (15,16,42). Gly and Pro with low CS_{AX} would oppose this cooperative process and thereby impede helix formation. Residues with high CS_{AX} values (>0.8) (Ala, Glu, Gln, Lys, Arg, Leu and Met) constitute a repertoire of both hydrophilic and hydrophobic groups, and can be used to form helices with polar, nonpolar and amphipathic characteristics to meet the diverse packing requirements of proteins.

Conclusions

Parameters based on conformational angles ϕ,ψ and χ_1 are useful in explaining thermodynamic properties of residues (39,43). Here the distribution of these angles has been used to derive the conformational similarity indices ($CS_{XX'}$) between residue pairs (X and X′) and the subsequent clustering of amino acids. The α-helix propensity is correlated with CS_{AX} which is a measure of the extent of similarity of the main-chain conformation of a residue, X, (as modulated by the side-chain χ_1) with that of Ala. CS_{AX} is a distinctive characteristic of the steric features (rather than the charge or hydrophobicity) of the side-chain (as the chain is extended beyond the C_β position of Ala) and residues with high CS_{AX}, in tandem, can form a helix.

Acknowledgements

We thank Prof. M. Kanehisa for sending us the program to compare similarity matrices in the AAindex database. This work was supported by grants from CSIR (to PC) and DBT (to Bose Institute) and a fellowship from CSIR to (to DP).

References and Footnotes

1. D.D Jones, *J. Theor. Biol. 50*, 167-183 (1975).
2. I. Ladunga and R.F. Smith, *Prot. Engng. 10*, 187-196 (1997).
3. Y. Kubota, K. Nishikawa, S. Takahashi and T. Ooi, *Biochim. Biophys. Acta 70*, 242-252 (1982).
4. P. Argos, *J. Mol. Biol. 193*, 385-396 (1987).
5. P.Y. Chou and G.D. Fasman, *Advan. Enzymol. 47*, 45-148 (1978).
6. M. Levitt, *Biochemistry 17*, 4277-4285 (1978).
7. F. Eisenhaber, B. Persson and P. Argos, *Crit. Rev. Biochem. Mol. Biol. 30*, 1-94 (1995).
8. K. Niefind and D. Schomburg, *J. Mol. Biol. 219*, 481-497 (1991).
9. A.S. Kolaskar and U. Kulkarni-Kale, *J. Mol. Biol. 223*, 1053-1061 (1992).
10. P. Chakrabarti and D. Pal, *Protein Engng. 11*, 631-647 (1998).
11. K. Tomii and M. Kanehisa, *Protein Engng. 9*, 27-36 (1996).
12. L. Pauling, R.B. Corey and H.R. Branson, *Proc. Natl. Acad. Sci. USA. 37*, 205-210 (1951).
13. G.N. Ramachandran and V. Sasisekharan, *Adv. Prot. Chem. 23*, 283-437 (1968).
14. A.V. Finkelstein, A.Y. Badretdinov and O.B. Ptitsyn, *Proteins 10*, 287-299 (1991).
15. V. Muñoz and L. Serrano, *Nat. Struct. Biol. 1*, 399-409 (1994).
16. A. Chakrabartty and R.L. Baldwin, *Adv. Prot. Chem. 46*, 141-176 (1995).
17. R. Aurora, T.P. Creamer, R. Srinivasan and G.D. Rose, *J. Biol. Chem. 272*, 1413-1416 (1997).
18. A. Horovitz, J.M. Matthews and A.R. Fersht, *J. Mol. Biol. 227*, 560-568 (1992).
19. R. Aurora and G.D. Rose, *Protein Sci. 7*, 21-38 (1998).
20. C. A. Kim and J.M. Berg, *Nature 362*, 267-270 (1993).
21. D.L Minor, Jr. and P.S. Kim, *Nature 371*, 264-267 (1994).
22. C.K. Smith and L. Regan, *Science 270*, 980-982 (1995).
23. B.H. Zimm and J. K. Bragg, *J. Chem. Phys. 31*, 526-535 (1959).
24. S. Lifson and A. Roig, *J. Chem. Phys. 34*, 1963-1974 (1961).
25. P.C. Lyu, M.I. Liff, L.A. Marky and N.R. Kallenbach, *Science 250*, 669-671 (1990).
26. S. Padmanabhan, S. Marqusee, T. Ridgeway, T.M. Laue and R.L. Baldwin, *Nature (London) 344*, 268-270 (1990).
27. S.H. Park, W. Shalongo and E. Stellwagen, *Biochemistry 32*, 7048-7053 (1993).
28. V. Muñoz and L. Serrano, *J. Mol. Biol. 245*, 275-296 (1995).

29. J. Yang, E.J. Spek, Y. Gong, H. Zhou and N.R. Kallenbach, *Protein Sci. 6*, 1264-1272 (1997).
30. M. Blaber, X.-J. Zhang and B.W. Matthews, *Science 260*, 1637-1640 (1993).
31. J.K. Myers, C.N. Pace and J.M. Scholtz, *Biochemistry 36*, 10923-10929 (1997).
32. K.T. O'Neil and W.F. DeGrado, *Science 250*, 646-651 (1990).
33. M. Blaber, X.J. Zhang, J.D. Lindstrom, S.D. Pepiot, W.A. Baase and B.W. Matthews, *J. Mol. Biol. 235*, 600-624 (1994).
34. C.N. Pace and J.M. Scholtz, *Biophys. J. 75*, 422-427 (1998).
35. U. Hobohm and C. Sander, *Protein Sci. 3*, 522-524 (1994).
36. F. Bernstein, T.K. Koetzle, G.J.B. Williams, E.F. Meyer, Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol. 112*, 535-542 (1977).
37. W. Kabsch and C. Sander, *Biopolymers 22*, 2577-2637 (1983).
38. J.L. Risler, M.O. Delorme, H. Delacroix and A. Henaut, *J. Mol. Biol. 204*, 1019-1029 (1988
39. D. Pal and P. Chakrabarti, *Acta Crystallogr. D56*, 589-594 (2000).
40. I. Luque, O.L. Mayorga and E. Friere, *Biochemistry 35*, 13681-13688 (1996).
41. P. Koehl and M. Levitt, *Proc. Natl. Acad. Sci. USA. 96*, 12524-12529 (1999).
42. K. Sun and A. J. Doig, *Protein Sci. 7*, 2374-2383 (1998).
43. D. Pal and P. Chakrabarti, *Proteins 36*, 332-339 (1999).

Date Received: July 13, 2000

Communicated by the Editor Ramaswamy H. Sarma