Int. J. Peptide Protein Res. 47, 1996, 110–116
Printed in Belgium – all rights reserved

INTERNATIONAL JOURNAL OF
PEPTIDE & PROTEIN RESEARCH
ISSN 0367-8377

Prediction of conformational states of amino acids using a Ramachandran plot

A.S. KOLASKAR and SANGEETA SAWANT

Bioinformatics Distributed Information Center, University of Pune, Pune, India

Received 24 May, accepted for publication 28 August 1995

(ϕ, ψ) data from crystal structures of 221 proteins having high resolution and sequence similarity cut-off at the 25% level were analysed by dividing the Ramachandran plot in three regions representing three conformational states: (i) conformational state 1: conformations in the (ϕ, ψ) range from (−140°, −100°) to (0°, 0°); (ii) conformational state 2: conformations with (ϕ, ψ) from (−180°, 80°) to (0°, 180°); and (iii) conformational state 3: all the remaining conformations in the (ϕ, ψ) plane which are not included in the above two conformational states.

Normalized probability values of the occurrence of single amino acid residues in conformational regions 1–3 and similar values for dipeptides were calculated. Comparisons of single residue and dipeptide normalized probability values have shown that short-range interactions, although strong, destabilize conformational states of only 44 dipeptides out of the 400×9 possible states. However, dipeptide frequency values provide better resolving power than single-residue potentials when used to predict conformational states of residues in a protein from its primary structure. The simple approach used in the present study to predict conformational states yields an accuracy of >70% for 14 proteins and an accuracy in the range of 50–70% for 247 proteins. Thus these studies point out yet another use of the Ramachandran plot and the role of tertiary interactions in protein folding. © Munksgaard 1996.

Key words: conformational properties; (ϕ, ψ) data analysis; protein structure prediction; Ramachandran plot; tertiary structure

Several attempts have been made in the last few years to predict the three-dimensional structures of proteins which include the use of neural nets, expert systems as well as homology modelling, energy minimization and statistical methods (1–8). Such approaches have given an insight into the problem of protein folding and have aided the implementation of experiments which give information regarding not only the final folded structures but also the intermediates. An impetus for these studies comes mainly from the fact that today one can synthesize a polypeptide of any sequence and length in the laboratory using genetic engineering techniques. However, very few of such synthesized polypeptides become folded and take on a stable three-dimensional structure, thus making the problem of protein folding one of the most challenging problems (9). In recent years, high resolution structural data of globular proteins obtained by using single crystal X-ray diffraction studies have helped protein modellers to gain an insight into the problem of protein folding. The Protein Data Bank (PDB) contains information regarding three-dimensional structures of nearly 1500 proteins (10). Analysis of such large numbers of data helps in formulating rules which can be used in protein folding studies. The (ϕ, ψ) data of proteins have been proved to be useful in developing protein folding algorithms (11–13). In the present study, an attempt has been made to analyze the (ϕ, ψ) data in order to develop a knowledge base which can be used in prediction of protein structure.

The main chain conformations, the (ϕ, ψ) angle values of amino acid residues in a polypeptide, are the result of the properties of amino acid residues, near-neighbour interactions and tertiary interactions. However, if the (ϕ, ψ) data are chosen from a sufficiently large sample of proteins, which are unrelated both in terms of structure and sequence, then the effect due to tertiary interactions and near neighbours will be masked to a very large extent, and the (ϕ, ψ) distribution of each amino acid residue will reflect its conformational property. Similarly, distribution of (ϕ_i, ψ_i) and (ϕ_{i+1}, ψ_{i+1}) of dipeptides will reflect short-range interactions between constituent amino acid residues of the pair in addition to the effect due to the individual amino acid

residues. Therefore, analysis of frequencies of occurrence of single amino acid residues and dipeptides in the various regions of the Ramachandran plot has been carried out on a carefully chosen set of proteins. In the earlier studies (14) the Ramachandran plot has been divided into three parts (Fig. 1). These are

(i) region I: primarily consisting of the closely (or tightly) packed conformations with ϕ ranging from $-140°$ to $0°$ and ψ ranging from $-100°$ to $0°$;
(ii) region II: containing mainly the extended conformations with ϕ ranging from $-180°$ to $0°$ and ψ ranging from $80°$ to $180°$; and
(iii) region III: all the remaining conformations which are not included in regions I and II.

Such a broad division of the (ϕ, ψ) plane, although reducing the resolution, allows one to calculate statistically relevant potential or normalized frequency values for each amino acid residue as well as dipeptides.

Such potentials have been used in the present study to develop a simple algorithm which enables assignment of conformational state to each amino acid residue in a given primary structure of a protein.

MATERIAL AND METHODS

Choice of data

A set of 221 proteins was selected from the PDB, using the algorithm of Hobohm *et al.* (15), with 25% sequence similarity cutoff. In other words, the best resolved structures with little sequence similarity among proteins were chosen. For those proteins which had breaks in the chain (due to missing amino acid residues), fragments (with length greater than 40) were treated as separate entries. Thus the chosen data set has 244 entries containing 51 998 amino acid residues.

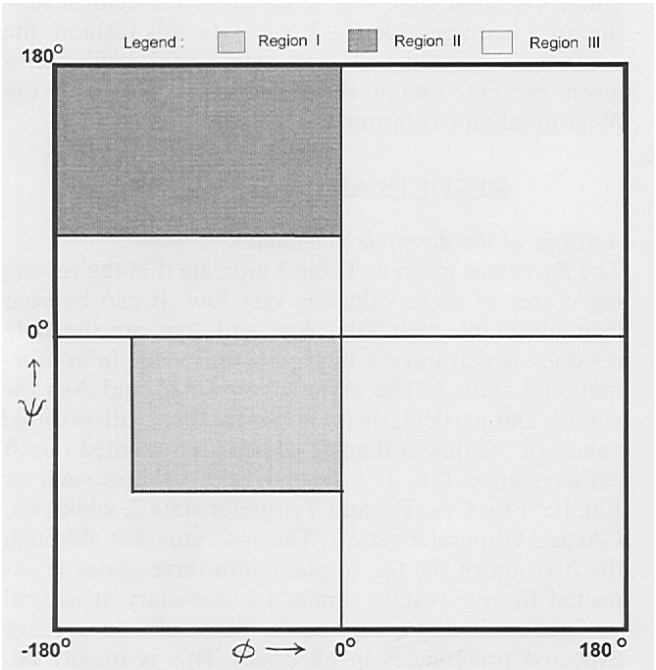

FIGURE 1

Ramachandran plot showing three conformational regions I, II and III.

Calculations of potentials

(ϕ, ψ) values were calculated for all the amino-acid residues in each protein in the set. Conformational state 1, 2 or 3 corresponding to regions I, II or III of the Ramachandran plot, respectively, was assigned to each amino-acid residue on the basis of its (ϕ, ψ) value. In the same fashion, conformational states 1–1, 1–2, 1–3, 2–1, 2–2, 2–3, 3–1, 3–2 and 3–3 were assigned to every overlapping dipeptide. Frequencies of single residues in three states were calculated and normalized using the following simple formula:

$$P_{ik} = \frac{n_{ik} N}{\sum_{k=1}^{3} n_{ik} \sum_{i=1}^{20} n_{ik}} \qquad (1)$$

where n_{ik} is the number of times the amino acid residue of type i occurs in state $k = 1$–3; N is the total number of residues, and P_{ik} is the potential value of amino acid of type i in state k.

Potential values P_{ijk} for dipeptides made up of amino acid residues of the types i and j, in state k (where k varies through nine states 1–1 to 3–3) were calculated in a similar fashion by using the relation:

$$P_{ijk} = \frac{n_{ijk} N}{\sum_{k=1}^{9} n_{ijk} \sum_{i=1}^{20} \sum_{j=1}^{20} n_{ijk}} \qquad (2)$$

where n_{ijk} is the number of occurrences of the dipeptide in the conformational state $k = 1$–1 to 3–3, and N is the total number of dipeptides in the data set.

P_{ik} and P_{ijk} values were divided into five class intervals in order to minimize the errors in the analysis, and weight values W_{ik} or W_{ijk} were assigned for each of the classes which were used in the prediction algorithm (Table 1). W_{ijk} values are not given but are available on request.

Algorithm for assigning conformational states to amino-acid residues in a given polypeptide

A given protein sequence was divided into successive overlapping heptapeptides. The heptapeptides were treated as independent units to fix the conformational state of the central amino acid residue. This was done by computing the interactions among the amino acid residues in a heptapeptide consisting of residues numbered ($l-3$) to ($l+3$). The following pairwise interactions were considered:

(a) interactions of the central amino acid residue

TABLE 1

Ranges of potential values and respective assigned weights

Range of potential values, P_{ik} or P_{ijk}	Weight assigned, W_{ik} or W_{ijk}
0.00–0.50	0
0.51–0.75	1
0.76–1.25	2
1.26–2.00	3
2.01 and higher	4

with all other residues of the heptapeptide; i.e. $(l, l-3)$, $(l, l-2)$, $(l, l-1)$, $(l, l+1)$, $(l, l+2)$ and $(l, l+3)$;

(b) interactions between the (near-neighbour) residues of dipeptides, $(l-3, l-2)$, $(l-2, l-1)$, $(l+1, l+2)$ and $(l+2, l+3)$;

(c) interactions between residues other than near neighbours, i.e. interactions of the type $(l-3, l\text{-}1)$, $(l-1, l+1)$, and $(l+1, l+3)$.

For each heptapeptide, all the possible conformations were enlisted. The total weight value W_c was computed for each conformation 'c' of a heptapeptide by calculating various interactions between pairs of amino acid residues as given in eqn. (3).

$$W_c = \sum_{m=l-3}^{l+2} A_m (W_{ijk})_{m,m+1} + \sum_{m=l-3, m\neq l-1, l+1}^{l+3} B_m (W_{ik})_m + (W_{ik})/2 + \sum_{m=l-3, m\neq l-2, l, l+2, l+3}^{l+3} C_m \{[(W_{ik})_m + (W_{ik})_{m+2}]/2\} \quad (3)$$

where A_m are the scaling factors for the dipeptide interactions of the type $(m,m+1)$ where m is in the range $(l-3, l+2)$; B_m are the scaling factors for interactions of the central amino acid residue with the residue m, where m assumes the values $l-3$, $l-2$, $l+2$ and $l+3$; C_m are the scaling factors for dipeptide interactions of the type $(m,m+2)$, where m assumes the values $l-3$, $l-1$ and $l+1$; $(w_{ijk})_{m,m+1}$ is the weight of the dipeptide made of residues i and j in state k, at positions m and $m+1$ in the heptapeptide where m varies from $l-3$ to $l+2$; $(w_{ik})_m$ is the weight of amino acid i in state k occurring at position m in the heptapeptide $(l-3, l+3)$; and $(w_{ik})_l$ is the weight of amino acid i in state k where l is the central residue of heptapeptide $(l-3, l+3)$.

The interaction terms would be directly proportional to weight values, and thus the proportionality constants (A_m, B_m and C_m) for each pairwise interaction term were calculated to give a maximum W_c for experimentally observed conformational states of the heptapeptide. The scaling factors A_m, B_m and C_m in the equation were assigned by trial and error. The values of these constants depend on the position of the amino-acid residue in the heptapeptide, but are independent of the type of amino acid residue. On the other hand, W_{ijk} values are dependent on the type of residue and its conformational state. The values of the scaling factors are listed in Table 2.

TABLE 2

Values of the constants A_m, B_m and C_m *used in eqn. (3)*

Symbol of constant	Value of constant for conformation of type	
	I	II
A_{l-3}, A_{l+1}	0.50	0.25
A_{l-2}	0.50	0.50
A_{l-1}, A_l	1.00	1.00
A_{l+2}	0.50	0.125
B_{l-3}	0.25	0.25
B_{l-2}	0.50	0.50
B_{l+2}	0.25	0.25
B_{l+3}	0.25	0.125
C_{l-3}	0.25	0.25
C_{l-1}	0.50	0.50
C_{l+1}	0.25	0.125

Type I conformations of heptapeptides are those in which all seven residues are in the same conformational state (either 1 or 2 or 3) and *type II* conformations are those in which the states of all seven residues are not identical.

Using the steps depicted in the flowchart in Fig. 2, conformational state was assigned to the central residue of a heptapeptide $(l-3, l+3)$. In this fashion, the conformational state of each amino acid residue in a given protein, except three residues each from the *N*-terminal and *C*-terminal ends was fixed.

RESULTS AND DISCUSSION

Analysis of the dipeptide potentials

The P_{ik} values given in Table 3 indicate that the resolving power of these values is very low. It can be seen from this table that Gly, Asp and Asn are the only residues which have a high potential value in conformational state 3. The preference by Asp and Asn for state 3, and particularly for conformations with positive values of the dihedral angle ϕ, has been pointed out in earlier studies (16, 17). Amino acid residues such as Val, Ile, Pro, Cys, Thr and Tyr prefer state 2, while Ala, Glu and Gln prefer state 1. This indicates that although the division of the (ϕ, ψ) plane into three states is expected to give results similar to secondary structural preferences by the amino-acids, our results do not give expected preferences in all cases. This is mainly because the long-range and medium-range interactions which are responsible for formation of secondary structures are masked in our studies.

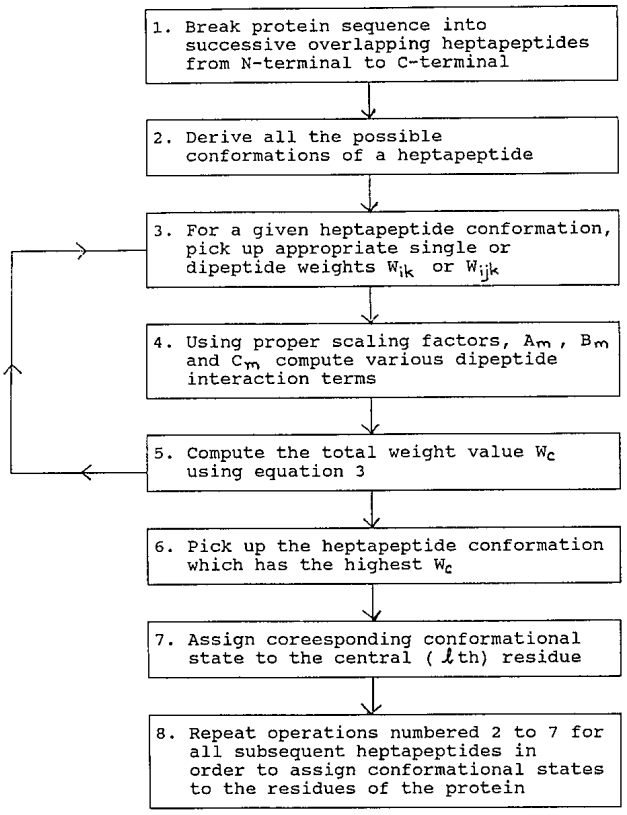

FIGURE 2

Flowchart for the assignment of conformational states based on short-range interactions.

TABLE 3

Values of potentials of single amino acid residues in three conformational states

Serial number i	Amino acid	Number of amino acids in the data set	Values of potentials P_{ik} in states (k)		
			1	2	3
1	Gly	4190	0.46	0.37	4.08
2	Ala	4371	1.32	0.81	0.62
3	Val	3556	0.86	1.43	0.29
4	Leu	4293	1.17	1.03	0.45
5	Ile	2785	0.98	1.31	0.27
6	Pro	2470	0.86	1.36	0.45
7	Met	1077	1.20	0.94	0.61
8	Cys	862	0.79	1.34	0.70
9	Ser	3229	0.97	1.09	0.85
10	Thr	3161	0.87	1.25	0.71
11	Asp	3109	0.99	0.85	1.42
12	Glu	3180	1.35	0.77	0.61
13	Arg	2313	1.19	0.93	0.66
14	Lys	3084	1.17	0.92	0.75
15	Asn	2409	0.79	0.86	1.93
16	Gln	1876	1.21	0.88	0.73
17	Phe	2136	0.92	1.18	0.73
18	Tyr	1940	0.87	1.25	0.72
19	His	1181	0.92	1.01	1.19
20	Trp	776	1.09	1.07	0.58

Note high values for $P_{1,3}$, $P_{11,3}$, $P_{15,3}$ and low vlues for $P_{1,1}$, $P_{3,3}$, $P_{4,3}$, $P_{5,3}$ and $P_{6,3}$.

The P_{ijk} values (not given) have a better spread over a range of 0.00–10.53 and provide an insight into the interactions between near neighbour residues and conformational properties of dipeptides; although some of the P_{ijk} values may have large error owing to a smaller number of observations in the state. Analysis of P_{ijk} values and corresponding single-residue potentials P_{ik} indicates that interactions do exist between amino-acid residues of type i and j in most cases; however, they do not destabilize the conformations preferred by the individual residues. Exceptions to this were observed which are given in Table 4. As can be seen from this table, even though individual residues i and j of the dipeptides have high P_{ik} and P_{jk} values, the P_{ijk} value is low.

Out of the total 44 dipeptides listed in the table, 36 dipeptides have the conformational state of either 1–2 or 2–1; and only 8 dipeptides are in conformational states 1–3, 2–3, 3–1 or 3–2. There are no dipeptides which are in states 1–1 or 2–2 or 3–3. These observations can be explained by the rationale that the conformational states 1 and 2 represent, respectively, closely packed structures such as α-helices, and extended structures like β-sheets; while state 3 represents coils and chain reversals. Hence, conformational states 1–1, 2–2 and 3–3 are well stabilized; whereas conformational states like 1–2 or 2–1 may become unfavourable and also may cause a sudden break in a contiguous stretch of a regular conformation.

Further analysis of the P_{ijk} values revealed that only two dipeptides, Met-Trp and Trp-Met, have high P_{ijk} even though individual residues i and j have low P_{ik} values. P_{ijk} of Met-Trp and Trp-Met in state 3–3 are 2.01 and 4.75, respectively, although the potential values of P_{ik} value of the residue Met and Trp in state 3 are 0.61 and 0.58, respectively. It may be noted that the occurrences of these individual residues and the dipeptides in state 3 and 3–3 are very low. They are Met – 104, Trp – 71, Met-Trp – 1 and Trp-Met – 3. Therefore, the percentage error in observed P_{ik} and P_{ijk} values will be large.

Assignment of conformational states

The values of the scaling factors A_m, B_m and C_m have been obtained to get the prediction accuracy as high as possible. However, predicted results should not be too sensitive to the values of these factors. In other words, the robustness of the scaling factors has been considered as one of the criteria. It may be mentioned that scaling factor values were determined using a small subset of the original data set, consisting of only 15 proteins which belong to different structural classes. The refinement of the scaling factors was then carried

TABLE 4

List of dipeptides and their conformational states in which they have low potential values even though constituent single residues have high potential values in the states concerned

Dipeptide (*i*–*j*)	Potential of (*i*–*j*)	State of (*i*–*j*)	Single residue potentials	
			Res. *i*	Res. *j*
A–I	0.47	1–2	1.32	1.31
A–P	0.00	1–2	1.32	1.36
V–M	0.39	2–1	1.43	1.20
V–R	0.46	2–1	1.43	1.19
L–P	0.00	1–2	1.17	1.36
L–W	0.47	1–2	1.17	1.07
I–M	0.00	2–1	1.31	1.20
I–W	0.00	2–1	1.31	1.09
P–W	0.33	2–1	1.36	1.09
M–V	0.24	1–2	1.20	1.43
M–I	0.40	1–2	1.20	1.31
M–P	0.00	1–2	1.20	1.36
C–D	0.36	2–3	1.34	1.42
S–P	0.31	1–2	0.97	1.36
S–W	0.00	1–2	0.97	1.07
T–H	0.46	2–3	1.25	1.19
D–W	0.48	3–2	1.42	1.07
E–P	0.00	1–2	1.35	1.36
R–P	0.00	1–2	1.19	1.36
R–C	0.00	1–2	1.19	1.34
R–W	0.00	1–2	1.19	1.07
K–P	0.00	1–2	1.17	1.36
Q–P	0.25	1–2	1.21	1.36
F–A	0.34	2–1	1.18	1.32
F–L	0.34	2–1	1.18	1.17
F–M	0.00	2–1	1.18	1.20
F–E	0.34	2–1	1.18	1.35
F–R	0.46	2–1	1.18	1.19
F–Q	0.19	2–1	1.18	1.21
F–W	0.00	2–1	1.18	1.09
Y–M	0.49	2–1	1.25	1.20
Y–K	0.45	2–1	1.25	1.17
Y–Q	0.44	2–1	1.25	1.21
Y–H	0.41	2–3	1.25	1.19
H–S	0.47	3–1	1.19	0.97
H–K	0.44	3–1	1.19	1.17
H–F	0.00	3–2	1.19	1.18
W–V	0.40	1–2	1.09	1.43
W–P	0.00	1–2	1.09	1.36
W–M	0.00	2–1	1.07	1.20
W–S	0.00	1–2	1.09	1.09
W–R	0.45	2–1	1.07	1.19
W–N	0.47	1–3	1.09	1.93
W–Q	0.45	2–1	1.07	1.21

out on the whole set of 221 proteins (used for calculation of P_{ijk} values). Refined constants were used in the program to predict the conformational states of each amino acid in 221 protein entries as well as additional 100 proteins. These 100 proteins were selected using the same criteria as discussed in the method. The results show that one could predict conformational states of amino acids of 14 proteins with an accuracy of 70% or more (Table 5). Such a high accuracy of predicted results by a method which uses only short-range and medium-range interactions but no long-range or tertiary interactions indicates that the observed three-dimensional structures of these 14 proteins may be having an insignificant effect on tertiary interactions.

On the other hand, for 247 proteins, which form the bulk of the data set, the prediction accuracy has been observed in the range 51–70% (Table 6). The least prediction accuracy was observed for 5 proteins falling in the range 30–40%. One may therefore conclude that in these proteins tertiary interactions are not only important but critical in deciding the three-dimensional structure. Thus these results re-emphasize that tertiary interactions must be incorporated in prediction algorithms in order to achieve improved accuracy of prediction.

Recent studies have shown considerable improvement in α-helix and extended structure predictions (18, 19). The high accuracy of such methods, although commendable, does not allow one to predict (ϕ, ψ) values for every residue in the protein. Although prediction accuracy of the method discussed here seems lower than the secondary structure prediction algorithms,

TABLE 5

List of proteins for which the accuracy of prediction is > 70%

Name of the protein with biological source	PDB code	Accuracy obtained (%)
Apolipoprotein E4 (LDL binding domain) (human)	1LE4	75
Bilin Binding Protein (chain A) (cabbage butterfly)	1BBPA	73
Cytochrome C (chain A) (*Rhodospirillum molischianum*)	2CCYA	72
Des-(ile 318-arg 417)-tyrosyl-transfer RNA synthetase (chain A) (*Bacillus stearothermophilus*)	4TS1A	71
Engrailed homeodomain complex with DNA (chain C) (fruit fly)	1HDDC	90
Ferritin (chain H) (human)	1FHAH	74
Guanylate kinase (*Saccharomyces cerevisiae*)	1GKY	71
High-potential iron–sulfur protein (phototrophic bacteria)	1ISU	71
Immunoglobulin FAB fragment (mouse)	1MFBH	71
MHC class I H-2K (chain B) (mouse)	1VAAB	72
Myoglobin (sperm whale)	1FCS	74
Subtilisin Carlsberg inhibitor	1CSEI	71
Triosephosphate isomerase (chain A) (*E. coli*)	1TREA	71
Tryptophan synthase (chain A fragment) (*salmonella typhimurium*)	1WSYA	73

TABLE 6

Prediction results in the various ranges of accuracy

Range of accurcy (%)	Number of proteins in the range
41–50	71
51–60	197
61–70	50
71–80	12
81–90	2

such a comparison has little meaning. In secondary-structure prediction algorithms one is predicting the conformations of a small percentage of residues of a protein. Even in strong α-proteins such as myoglobin, only 67% of residues are in an α-helical conformation, and thus even if the prediction accuracy is 100% for such proteins, conformations of only 67% residues are actually predicted. Secondly, prediction of conformations by making use of long-range interactions can force the molecule to fold in one of the local minimum conformations, far away from the native conformation. Thus, comparison of the predicted conformational states in the Ramachandran plot with secondary structure predictions may not be valid but one may qualitatively compare the results given here with secondary-structure data. This is mainly due to the fact that conformational state 1 includes α-helices while conformational state 2 includes β-sheets. It must be mentioned that states 1 and 2 also include bends as well as many other conformations belonging to the coiled state. The contiguous stretch of residues in conformational states 1 and 2 may more often represent α-helices and β-sheets, respectively. Under this assumption and using the criteria by Rost *et al.* (20), the comparison carried out of amino-acid residues in α-helices and β-sheets from X-ray diffraction data (PDB data) and present results showed an agreement in the range 45–55%.

CONCLUSIONS

The approach of predicting conformational states is an alternative to the prediction of secondary structures and seems more beneficial in building tertiary structures, as it completely excludes the consideration of long-range interactions and does not assume that during protein folding, regular structures are formed first, which then interact with each other to form the final three-dimensional structure. The *de novo* synthesis experiments on the computer by Rawn and Feldman (personal communication) support our approach that during folding, few regular structures may get unfolded, and many of the irregular structures may form α-helices or β-sheets due to tertiary interactions. The approach described here can be refined to obtain better results by incorporating available knowledge based on protein three-dimensional structure. This has been avoided in order to point out the effect of only short-range interactions in protein folding studies.

Secondly, in the secondary-structure prediction approach, one cannot build the three-dimensional structure of the molecule, as the relative orientation of the secondary structural elements as well as conformations of the coiled part cannot be assigned. On the other hand, in the present approach, one can assign (ϕ, ψ) values ($-60°$, $-60°$), ($-140°$, $140°$) and ($60°$, $80°$) for residues in conformational states 1, 2 and 3, respectively. Protein molecules built using these (ϕ, ψ) values and predicted conformational states were found to be more open and loosely structured than the native molecules. The conformation of the protein thus assigned may act as a good starting point in an energy-minimization approach to achieve native conformation having a minimum free energy. Studies in this direction are in progress.

ACKNOWLEDGEMENTS

We acknowledge financial support provided by the Department of Biotechnology, Government of India for carrying out this work.

REFERENCES

1. Xin, Y., Carmeli, T.T., Liebman, M.N. & Wilcox, G.L. (1992) *Proceedings of The Second International Conference On Bioinformatics, Supercomputing and Complex Genome Analysis* (Lim, H.A. *et al.*, eds.), pp. 359–375
2. Chen, S. (1992) *Proceedings of The Second International conference On Bioinformatics, Supercomputing and Complex Genome Analysis* (Lim, H.A. *et al.*, eds.), pp. 391–399
3. Holm, L. & Sander, C. (1992) *Proteins* **14**, 213–223
4. Levitt, M. (1992) *J. Mol. Biol.* **226**, 507–533
5. Bryant, S.H. & Lawrence, C.E. (1993) *Proteins* **16**, 92–112
6. Sippl, M.J. (1990) *J. Mol. Biol.* **213**, 859–883
7. Sippl, M.J. (1993) *J. Comput. Aided Mol. Design* **7**, 473–501
8. Blundell, T.L., Sibanda B.L., Sternberg, M.J.E. & Thornton, J.M. (1987) *Nature (London)* **326**, 347–352
9. Hoffman, M. (1991) *Science* **253**, 1357–1358
10. Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanoushi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542
11. Rooman, M., Kocher, J. & Wodak, S (1991) *J. Mol. Biol.* **221**, 961–979.
12. Kang, H.L., Kurochkina, N.A. & Lee, B. (1993) *J. Mol. Biol.* **229**, 448–460
13. Vajda, S. (1993) *J. Mol. Biol.* **229**, 125–145
14. Kolaskar, A.S. & Rambrahmam, V. (1981) *Int. J. Biol. Macromol.* **3**, 171–178
15. Hobohm ,U., Scharf R., Schneider R. & Sander C. (1992) *Protein Sci.* **1**, 409–417
16. Kolaskar, A.S. & Kulkarni-Kale, U. (1992) *J. Mol. Biol.* **223**, 1053–1061
17. Srinivasan, N., Anuradha, V.S., Ramakrishnan, C., Sowdhamini, R. & Balaram, P. (1994) *Int. J. Peptide Protein Res.* **44**, 112–122
18. Gromiha, M.M. & Ponnuswamy, P.K. (1995) *Int. J. Peptide Protein Res.* **45**, 225–240

19. Geourjon, C. & Deleage, G. (1994) *Protein Eng.* **7**, 157–164
20. Rost, B., Sander, C. & Schneider, R. (1994) *J. Mol. Biol.* **235**, 13–26
21. Kolaskar, A.S. & Rambrahmam, V. (1984) *Int. J. Peptide Protein Res.* **24**, 392–401
22. Kabsch, W. & Sander, C. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1075–1078

Address:

Dr. *A.S. Kolaskar*
Bioinformatics Distributed Information Center
University of Pune
Pune 411007
India

Tel.: 350195/355039
Fax: (+91) 212350087
Email: kolaskar@bioinfo.ernet.in