Int. J. Peptide Protein Res. **22,** 1983, 83–91

Conformational properties of pairs of amino acids

A.S. KOLASKAR and V. RAMABRAHMAM*

*Center for Cellular and Molecular Biology, RRL Campus, and *School of Life Sciences, University of Hyderabad, Hyderabad, India*

Received 15 July, accepted for publication 4 November 1982

Crystal structure data of 31 different globular proteins were analysed. Potentials for all 400 pairs of 20 amino acid residues in helix, extended structure, chain reversals and coil state were obtained. Analysis of these potentials showed that tripeptides of amino acid residues are not linear combinations of amino acids. The association effect of these tripeptides depends on the types of amino acids associated, the position of amino acids in the pair and the secondary structure in which the pair exists. This built-in information in tripeptides should be used in polypeptide and protein folding studies.

Key words: amino acid pair potentials; data analysis; protein crystal structures; protein folding; secondary structures

Conformational properties of single amino acid residues have been deduced using crystal structure data of globular proteins (1–3). Most of these studies were done to calculate the potential values of single amino acids in three secondary structures, namely, helix, extended structure and chain reversals. Such single residue potential values are used in a large number of algorithms to predict secondary structures (3–10). These studies have pointed out that the results obtained using single residue potentials vary from method to method. This is mainly because the weights given to various amino acids vary considerably with respect to their position in secondary structure. However, as pointed out by Bourgeios *et al.* (11) from their study on *lac* repressor protein, none of these methods can be considered to be good in isolation. Even recent modification of these methods by including hydrophobic interaction term (6,12) does not improve the situation appreciably.

In order to understand the limitations of the methods, we have analysed the crystal structure data of 38 different globular proteins and determined main chain conformations which are significantly affected by side chains (13). This has prompted us to study more carefully the side chain-side chain interactions, the importance of which is never in doubt in polypeptide chain conformations (14–17). As a first step the occurrence of pairs of amino acid residues in four secondary structural states, helix, extended structure, chain reversals and coil has been studied. As observed earlier in case of chain reversals (18) the pairs of amino acids have different conformational properties as compared to constituent amino acid residues. Under Results and Discussion we have shown that pairs of amino acids cannot be treated as linear combinations of single amino acids and the non-linear term is too large to be neglected. It should be mentioned here that there is less data compared

with data on single amino acid residues, when one is working and deriving statistical information at the level of pairs of amino acids, as 400 different pairs have to be considered as against 20 types of residues. Therefore our discussion is restricted to qualitative aspects of the results and less importance is attached to absolute numbers. Further, only those pairs of amino acids which occur a sufficiently large number of times have been used for drawing conclusions.

METHOD

The following 31 globular proteins for which the crystal structures are available were used in our studies: Lamprey cyanamet haemoglobin, bovine ferricytochrome b_5, human haemoglobin (α and β-chains), chicken lysozyme, subtilisin BPN′, bacterial rubredoxin, Jack bean concanavalin A, bacterial thermolysin, bovine chymotrypsinogen A, bacterial high potential protein, Dog-fish apo-lactate dehydrogenase, carp calcium binding protein, carboxypeptidase A complex, bovine ribonuclease S complex, bacterial nuclease complex, sperm whale myoglobin, bacterial ferricytochrome C_2, lobster glyceraldehyde-3-P-dehydrogenase, bacterial ferredoxin, horse alcohol dehydrogenase complex, chicken triose phosphate, Bonito ferrocytochrome C, bacterial cytochrome C_{550}, bovine trypsin-trypsin inhibitor, porcine tosyl elastase, papain, bacterial semiquinone flavodoxin, human Bence-Jones protein REI monomer I, human immunoglobulin G Fab New, human carbonic anhydrase C, human prealbumin monomer II.

As can be seen these proteins belong to all four structural classes: α-proteins, β-proteins, $\alpha+\beta$ proteins and α/β proteins. We have assumed a four-state model for the proteins and thus all the sequences considered are divided into helical, extended structure, chain reversals and coil region. The helical and extended structure regions were taken from crystal structure data as reported in original papers and compiled in AMSOM (19). The chain reversal regions are computed using the algorithm developed by us (18). In present analysis, those chain reversals which are part of either *N*-terminal or *C*-terminal of helix are excluded in order to avoid double counting. Regions of protein sequences other than helix, extended structure or chain reversals are considered as coil region.

In order to arrive at the potential for each of the pairs formed by 20 amino acids in the four different states mentioned above, we have used the following procedure.

If ACDEF is a stretch representing a particular secondary structural region in protein sequence under consideration, pairs AC, CD, DE and EF were formed from them. The number of times a pair formed by i^{th} and j^{th} type of residues occurs in k^{th} secondary structural state (N_{ijk}) is computed for all pairs in four states. Both i and j vary from 1 to 20 and k from 1 to 4. Thus when $i = j$, the pair will be formed by same amino acid. In all other cases it will be formed by two different amino acids. The total number of pairs ($\sum_{k=1}^{4} N_{ijk}$) occurring in 31 different globular proteins are given in Table 1. Then P_{ijk} the potential for a pair of amino acids formed by i^{th} and j^{th} type of amino acids in the k^{th} secondary structure was calculated using following simple relation.

$$S_{ijk} = n_{ijk} \Big/ \sum_{i=1}^{20} \sum_{j=1}^{20} N_{ijk}$$

where $$n_{ijk} = N_{ijk} \Big/ \sum_{k=1}^{4} N_{ijk}$$

$$S_{ij} = \frac{\sum_{k=1}^{4} S_{ijk}}{4}$$

$$P_{ijk} = \frac{S_{ijk}}{S_{ij}} \qquad (1)$$

The values of P_{ijk} obtained for each pair of amino acid residues in each of the four states considered are given in Table 2. We have also determined potentials for single amino acid residues in a way similar to that used for pairs and they are given in Table 3.

RESULTS AND DISCUSSION

As can be seen from Table 1, 6028 pairs of amino acid residues from 31 different proteins distributed among 400 different types have

TABLE 1

Number of occurrences of amino acid pairs in the 31 proteins

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	70	8	35	17	16	44	19	20	41	41	6	18	23	21	11	44	26	48	9	18
C	14	1	5	2	0	18	3	6	13	14	1	8	4	9	7	9	4	7	1	4
D	40	6	18	19	15	36	3	26	21	24	3	15	16	4	7	31	17	27	6	13
E	22	5	14	18	16	24	5	9	25	30	6	13	10	13	10	15	15	23	5	7
F	13	4	11	15	7	18	5	10	17	15	4	7	10	3	8	16	19	11	1	8
G	40	14	35	30	13	38	8	34	41	32	3	19	20	17	19	48	39	49	8	22
H	12	8	2	11	9	15	3	5	11	7	3	4	14	0	1	9	10	12	4	6
I	25	5	20	22	11	28	6	20	23	19	4	20	15	11	8	22	15	23	3	11
K	43	6	29	13	15	31	14	23	33	35	13	20	16	3	6	35	25	42	3	28
L	28	4	22	20	12	31	17	30	39	29	7	16	18	23	17	43	32	35	6	10
M	5	0	6	5	5	6	1	5	12	3	2	5	2	3	1	5	3	11	1	1
N	12	7	10	11	12	20	4	13	14	16	3	14	12	15	8	18	22	19	10	7
P	17	4	22	24	6	25	4	9	14	20	4	12	8	3	5	28	8	24	4	10
Q	24	3	14	8	6	20	8	9	15	12	2	11	9	11	8	15	10	10	4	9
R	7	3	4	6	5	15	3	10	10	20	3	8	6	13	3	20	7	16	2	5
S	46	16	20	22	15	51	9	19	28	37	9	20	14	20	14	50	39	39	14	27
T	32	9	22	19	21	31	10	14	26	28	5	12	23	14	6	30	23	32	8	14
V	54	19	40	22	16	35	18	22	30	42	4	19	14	15	14	42	33	45	6	11
W	8	2	4	4	1	16	2	8	8	4	1	4	2	2	5	7	8	11	1	3
Y	15	5	16	6	3	30	5	8	12	13	2	7	15	9	8	20	20	12	6	10

These pairs can be read by considering the residue first from vertical column and then from horizontal row. Single letter amino acid code has been used.

been used to calculate P_{ijk} values. The number of occurrences for certain types of pairs is considerably small and the values of potentials for such pairs may not be true representatives of the property of these pairs of amino acids as the standard deviation associated with such types of pairs is very large. Therefore, we have obtained the mean value of occurrence of each type of pair ($\bar{N}$) using data given in Table 1. Only those pairs which have $N_{ijk} \geqslant \bar{N}$ ($\bar{N} = 15$) have been used in the following discussion. P_{ijk} values given in Table 2 indicate that there is a large effect on the potential values depending on whether the amino acid residue occurs in first or second position in the pair. This can be seen from the data in Table 2 if one carries out the following simple analysis.

The values

$$\bar{P} = \frac{P_{ijk} + P_{jik}}{2} \qquad (2)$$

when $i \neq j$ are computed. If

$$(\bar{P} - 10\%\,\bar{P}) < (P_{ijk} \text{ and } P_{jik}) < (\bar{P} + 10\%\,\bar{P}) \qquad (3)$$

then only P_{ijk} and P_{jik} were considered to be same. When this analysis was carried out in four states on pairs whose $\sum_{k=1}^{4} N_{ijk} \geqslant \bar{N}$, only one set of pairs Asp-Lys, Lys-Asp was found to satisfy condition (3) in all four states, indicating that the positional effect is minimal for this exceptional case. The sets of pairs for which P_{ijk} and P_{jik} values are different in all four states are given in Table 4. We have also examined the data by increasing the range of inequality in condition (3) to 20% of $\bar{P}$. The results are not qualitatively different.

The significant difference in P_{ijk} and P_{jik} indicates that the side chain of a residue has different effects on polypeptide conformations when it occurs in position one or two of the pair. This has also been shown from conformational energy calculations on tripeptides (20), indicating that this observed asymmetry in potential values is not simply due to the statistics applied. This information is com-

TABLE 2

Potential (P_{ijk}) for each pair of amino acid residues in four secondary structural states are given in the following order: (i) helix (h), (ii) extended structure (e), (iii) chain reversal (b) and (iv) coil region (c)

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
A	1.82	0.46	1.24	2.44	2.16	0.87	1.58	1.80	1.67	2.36	2.70	1.28	0.52	1.75	1.16	1.08	0.88	1.56	1.11	1.31	h
	0.52	1.95	0.55	0.23	1.01	0.84	0.21	1.14	0.59	1.00	0.71	0.39	1.09	0.00	0.82	0.87	1.24	1.30	0.39	1.38	e
	0.65	0.78	1.22	0.75	0.40	1.46	1.00	0.91	0.78	0.32	0.00	1.85	0.58	1.18	0.00	1.11	0.74	0.28	1.86	0.37	b
	1.01	0.81	0.99	0.58	0.42	0.83	1.21	0.16	0.97	0.33	0.59	0.48	1.81	1.07	2.02	0.94	1.15	0.86	0.64	0.95	c
C	0.50	4.00	1.73	1.95	0.00	0.21	0.00	1.29	1.04	0.77	0.00	1.18	0.00	0.41	0.00	1.20	0.96	0.53	4.00	0.00	
	1.58	0.00	0.00	2.05	0.00	1.79	0.00	2.71	0.55	1.91	0.00	0.83	2.61	2.18	1.16	0.42	3.04	1.67	0.00	1.81	
	1.26	0.00	0.00	0.00	0.00	0.71	3.18	0.00	1.74	0.87	0.00	1.99	1.39	0.69	0.92	1.34	0.00	0.88	0.00	1.44	
	0.65	0.00	2.27	0.00	0.00	1.29	0.82	0.00	0.68	0.45	4.00	0.00	0.00	0.72	1.92	1.04	0.00	0.92	0.00	0.75	
D	1.03	1.20	1.37	1.76	0.00	0.39	1.46	1.12	1.22	1.55	0.00	0.22	1.09	0.00	0.44	0.40	0.20	1.38	0.68	0.27	
	0.63	1.27	0.62	0.41	0.88	0.81	0.00	1.33	0.55	0.98	0.00	0.46	0.69	1.81	0.94	0.63	1.06	0.73	2.14	1.43	
	1.59	1.01	1.32	0.98	0.47	1.62	0.00	0.94	1.17	0.78	3.18	2.19	1.46	1.44	2.23	2.35	1.69	0.93	0.00	1.36	
	0.75	0.52	0.69	0.85	2.66	1.18	2.54	0.61	1.06	0.68	0.82	1.14	0.76	0.75	0.39	0.61	1.05	0.96	1.18	0.94	
E	2.34	0.00	1.33	1.19	1.48	0.81	2.16	1.30	2.15	1.52	1.62	0.54	0.40	1.97	3.03	0.70	1.60	1.42	1.61	2.89	
	0.57	1.79	0.84	0.63	1.04	0.68	0.00	0.46	0.00	0.57	0.57	0.00	0.84	0.59	0.00	0.73	0.56	1.50	1.69	0.61	
	0.30	0.00	0.89	1.33	0.41	0.81	1.21	0.73	0.77	1.81	1.81	1.81	0.67	0.95	0.64	1.56	0.45	0.80	0.00	0.00	
	0.79	2.21	0.93	0.86	1.07	1.69	0.63	1.51	1.07	0.00	0.00	1.65	2.09	0.49	0.33	1.01	1.39	0.28	0.70	0.50	
F	2.02	0.00	0.69	1.19	1.70	0.35	0.82	1.87	1.37	1.45	1.95	0.48	0.86	2.17	1.99	0.93	1.24	1.10	0.00	0.93	
	0.00	2.19	1.82	1.00	1.80	0.73	1.74	1.18	0.48	1.53	2.05	1.51	0.90	0.00	1.58	1.47	1.52	1.94	4.00	1.47	
	0.97	0.00	0.58	1.19	0.00	2.32	0.00	0.63	0.76	0.81	0.00	1.60	0.00	1.83	0.00	0.78	0.35	0.00	0.00	0.78	
	1.01	1.81	0.90	0.62	0.50	0.60	1.44	0.33	1.39	0.21	0.00	0.42	2.24	0.00	0.43	0.81	0.90	0.96	0.00	0.81	
G	0.86	0.73	0.54	1.23	1.09	0.38	1.47	0.59	0.96	0.44	0.00	0.84	0.17	0.43	0.55	0.15	0.27	0.61	0.50	0.65	
	0.91	0.77	0.79	0.78	0.58	1.20	0.00	1.49	0.19	0.70	2.23	0.44	1.46	0.45	0.96	1.23	0.77	0.80	2.63	0.86	
	0.80	1.64	1.08	1.03	1.38	1.27	0.82	0.39	1.47	1.49	1.77	0.70	1.46	1.44	1.53	1.47	1.53	1.14	0.00	1.36	
	1.42	0.85	1.59	0.96	0.95	1.15	1.71	1.53	1.38	1.36	0.00	2.01	0.91	1.68	0.96	1.15	1.43	1.45	0.87	1.13	
H	3.05	0.80	0.00	1.08	0.43	1.73	1.13	0.00	0.83	0.49	2.79	1.11	0.45	0.00	0.00	1.98	1.19	2.23	0.00	0.54	
	0.36	0.84	0.00	0.76	1.36	0.78	0.00	2.58	0.59	1.03	0.00	0.00	0.71	0.00	4.00	0.00	0.42	0.67	2.61	1.15	
	0.00	2.01	2.63	0.60	0.72	0.83	1.89	0.00	2.34	1.64	0.00	0.00	2.26	0.00	0.00	1.33	0.67	0.54	1.39	1.83	
	0.59	0.35	1.37	1.56	1.50	0.65	0.98	1.42	0.24	0.85	1.21	2.89	0.59	0.00	0.00	0.69	1.73	0.56	0.00	0.48	

	1.69	0.82	0.93	1.72	0.76	0.77	1.20	2.06	0.85	1.60	2.96	0.95	0.66	0.68	1.38	1.00	0.74	0.33	0.00	0.69
	1.13	1.74	1.96	0.61	1.60	1.49	1.27	0.99	1.25	1.90	1.04	1.41	0.70	2.15	1.45	1.58	1.57	2.10	1.17	1.82
I	0.51	0.00	0.62	0.00	0.00	1.08	1.01	0.63	0.57	0.34	0.00	0.64	1.86	0.57	0.77	0.84	0.83	0.56	1.86	0.58
	0.67	1.44	0.49	1.67	1.65	0.67	0.52	0.33	1.33	0.17	0.00	1.00	0.77	0.59	0.40	0.58	0.86	1.01	0.97	0.90
	1.37	0.56	1.28	1.23	1.75	0.81	1.48	1.00	1.28	1.07	1.23	0.55	0.68	2.79	1.20	1.01	1.55	1.49	1.37	0.78
	0.81	0.59	0.54	0.65	0.53	0.43	0.00	1.41	0.45	1.36	0.65	0.78	1.19	0.00	1.27	0.47	0.33	0.98	1.44	1.10
K	1.00	1.88	1.07	0.52	0.84	2.22	1.66	0.56	1.43	0.72	0.52	1.23	1.14	0.00	1.01	0.75	0.78	0.47	0.00	1.09
	0.82	0.97	1.11	1.61	0.87	0.53	0.86	1.02	0.84	0.84	1.61	1.44	0.99	1.21	0.52	1.76	1.35	1.06	1.19	1.02
	2.16	0.90	1.23	1.88	0.88	0.97	1.29	1.07	1.69	1.75	1.57	0.48	0.00	0.99	1.76	1.29	0.88	1.66	1.13	0.67
	0.57	0.00	0.16	0.99	0.62	0.64	1.81	2.12	0.94	0.66	1.10	1.02	0.94	1.04	0.93	1.09	0.93	1.87	0.00	1.06
L	0.68	1.52	1.81	0.63	1.48	1.02	0.72	0.00	0.50	1.05	0.88	0.81	0.74	0.83	0.74	0.87	0.42	0.19	1.89	1.69
	0.59	1.58	0.80	0.49	1.02	1.37	0.19	0.82	0.87	0.54	0.46	1.69	2.32	1.15	0.58	0.75	1.76	0.29	0.98	0.58
	0.86	0.00	1.46	2.44	1.55	1.87	0.00	1.67	1.82	1.29	4.00	1.28	0.00	1.07	4.00	1.47	0.51	1.12	0.00	0.00
	0.90	0.00	0.00	0.86	2.45	0.00	0.00	0.88	0.64	2.71	0.00	0.00	0.00	1.13	0.00	0.00	2.15	1.58	0.00	4.00
M	0.00	0.00	0.00	0.00	0.00	1.05	0.00	0.00	1.02	0.00	0.00	2.16	4.00	1.80	0.00	1.24	0.00	0.00	0.00	0.00
	2.24	0.00	2.54	0.71	0.00	1.09	4.00	1.45	0.53	0.00	0.00	0.56	0.00	0.00	0.00	1.29	1.34	1.30	4.00	0.00
	0.58	2.10	0.00	1.09	1.35	0.53	1.00	0.74	0.75	0.72	0.00	0.55	1.16	1.60	0.00	0.18	0.60	1.52	0.84	0.82
	1.22	0.56	2.38	0.38	0.00	0.00	2.12	0.78	1.32	0.00	2.83	0.58	0.61	0.56	0.38	0.58	1.11	1.20	1.33	0.43
N	1.45	0.88	0.63	0.61	1.89	1.78	0.00	2.06	1.26	1.20	0.00	0.93	1.46	0.45	3.00	2.13	1.77	0.95	0.00	2.75
	0.75	0.46	0.98	1.91	0.98	1.69	0.88	0.43	0.66	2.08	1.17	1.93	0.76	1.39	0.62	1.11	0.52	0.33	1.83	0.00
	1.24	0.00	0.34	0.65	1.37	0.00	2.14	0.94	0.75	0.33	1.00	0.57	0.40	1.46	0.00	0.38	0.89	0.86	1.69	0.40
	0.65	0.80	0.18	0.27	1.44	0.91	0.00	1.00	0.26	1.04	2.12	0.30	0.85	0.00	0.79	0.93	0.47	1.64	0.89	0.84
P	1.39	2.54	1.42	2.40	0.00	1.45	0.00	0.00	1.68	1.92	0.00	1.90	2.04	0.00	1.26	1.48	1.49	0.00	1.42	0.67
	0.72	0.66	2.06	0.68	1.19	1.63	1.86	2.06	1.31	0.71	0.88	1.23	0.71	2.54	1.96	1.21	1.16	1.50	0.00	2.09
	1.37	0.00	0.93	2.81	0.00	0.85	1.34	0.36	1.73	1.60	1.95	1.55	0.00	0.91	0.97	0.79	0.60	1.64	0.74	1.31
	0.80	1.17	0.49	0.00	3.43	0.71	0.00	0.91	0.78	1.02	2.05	0.41	0.94	1.29	0.51	1.39	0.63	1.29	0.78	2.31
Q	1.02	1.86	1.96	0.79	0.00	1.70	1.50	0.72	0.83	0.54	0.00	0.00	0.74	1.53	0.82	0.44	2.51	0.00	2.48	0.00
	0.80	0.97	0.61	0.41	0.57	0.74	1.17	1.50	0.65	0.84	0.00	2.03	2.32	0.27	1.70	1.38	0.26	1.07	0.00	0.38
	1.80	1.13	2.04	1.82	1.38	0.24	2.79	0.72	1.56	0.91	4.00	0.45	0.00	1.14	1.07	1.05	0.54	0.70	2.14	0.73
	0.63	0.00	1.07	0.64	1.46	1.03	0.00	2.67	0.41	1.54	0.00	0.47	0.78	0.90	1.13	0.56	1.14	1.71	0.00	0.77
R	0.00	1.89	0.00	1.02	1.16	1.23	0.00	0.61	0.66	0.92	0.00	1.51	0.00	0.96	1.80	1.47	0.91	0.78	0.00	1.23
	1.57	0.98	0.89	0.53	0.00	1.49	1.21	0.00	1.37	0.64	0.00	1.57	3.22	1.00	0.00	0.92	1.41	0.81	1.86	1.27
	1.54	0.88	1.40	1.43	0.24	0.36	0.81	0.78	0.60	0.50	1.09	0.50	1.04	1.02	0.75	0.29	0.26	0.68	1.11	0.26
	0.94	0.69	0.37	0.00	1.51	1.08	0.43	1.23	0.76	1.60	0.77	0.53	0.28	0.36	0.26	0.83	1.11	1.83	1.46	1.25
S	0.82	1.47	1.47	1.60	1.20	1.10	1.36	0.98	1.81	0.85	1.83	1.96	1.31	1.72	1.68	1.44	1.62	0.65	0.46	1.33
	0.71	0.96	0.76	0.97	1.04	1.46	1.41	1.02	0.84	1.05	0.32	1.02	1.37	0.89	1.31	1.44	1.00	0.84	0.97	1.15

TABLE 2 (*Contd.*)

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
T	1.42	0.00	0.49	1.23	0.52	0.57	1.45	1.00	0.85	1.21	0.00	1.63	0.90	1.14	0.00	0.60	0.51	0.58	0.50	0.54
	0.75	3.24	0.69	0.65	1.84	0.97	0.38	1.31	0.60	1.42	1.79	0.69	0.79	0.90	1.51	0.63	1.60	1.72	2.63	1.14
	0.80	0.00	1.38	0.69	0.88	1.35	1.22	1.25	1.19	0.68	0.00	0.55	1.52	0.48	0.00	1.41	0.57	0.78	0.00	0.91
	1.03	0.76	1.43	1.43	0.76	1.10	0.95	0.43	1.36	0.70	2.21	1.14	0.79	1.49	2.49	1.36	1.32	0.91	0.87	1.41
V	2.02	0.57	0.96	0.70	0.49	0.64	0.45	0.88	1.65	1.92	1.05	0.99	0.52	1.44	0.77	0.54	0.94	0.78	0.70	0.70
	0.99	1.81	1.31	2.23	1.81	1.46	2.37	2.22	1.16	0.91	1.11	1.47	0.55	0.51	1.91	2.29	1.24	2.10	1.47	1.48
	0.36	0.96	0.64	0.30	0.41	0.89	0.00	0.29	0.00	0.16	0.00	0.67	1.32	1.21	0.87	0.46	0.59	0.29	0.00	0.59
	0.63	0.66	1.09	0.77	1.28	1.02	1.18	0.61	1.20	1.01	1.84	0.87	1.60	0.84	0.45	0.71	1.23	0.83	1.83	1.23
W	2.56	0.00	0.90	0.00	0.00	0.73	2.14	1.81	0.43	1.95	4.00	0.00	0.00	1.95	0.00	0.98	0.55	0.67	0.00	0.00
	0.54	0.00	0.00	1.15	0.00	0.77	0.00	1.43	1.37	2.05	0.00	3.14	0.00	2.05	0.93	0.52	0.58	2.48	0.00	4.00
	0.00	0.00	1.52	0.00	0.00	0.81	0.00	0.76	1.45	0.00	0.00	0.00	0.00	0.00	0.00	1.65	0.00	0.56	0.00	0.00
	0.89	4.00	1.58	2.85	4.00	1.69	1.86	0.00	0.75	0.00	0.00	0.86	4.00	0.00	3.07	0.86	2.87	0.29	4.00	0.00
Y	0.24	0.79	0.72	0.60	2.62	0.54	0.89	0.44	0.64	2.09	4.00	0.64	0.48	0.46	0.87	0.77	0.17	0.98	0.61	0.72
	1.75	2.51	2.03	1.89	1.38	0.80	0.00	2.81	0.34	0.63	0.00	0.00	0.25	1.94	0.92	1.23	1.63	1.04	1.29	2.67
	1.19	0.00	0.40	1.00	0.00	2.00	0.00	0.75	1.07	0.50	0.00	0.00	1.60	0.00	1.46	0.65	1.44	0.55	1.03	0.61
	0.83	0.69	0.84	0.52	0.00	0.66	3.11	0.00	1.95	0.78	0.00	3.36	1.67	1.60	0.76	1.35	0.75	1.43	1.07	0.00

TABLE 3

Single residue potentials obtained using the crystal structure data of proteins mentioned in the text

Amino acid residue	Helix	Extended structure	Chain reversals	Coil region
Ala	1.46	0.80	0.81	0.93
Cys	0.75	1.31	1.10	0.83
Asp	0.98	0.91	1.28	0.82
Glu	1.47	0.74	0.96	0.83
Phe	1.20	1.21	0.75	0.84
Gly	0.62	0.88	1.28	1.22
His	1.26	0.77	1.11	0.86
Ile	1.08	1.43	0.66	0.82
Lys	1.15	0.71	0.98	1.16
Leu	1.24	1.06	0.84	0.86
Met	1.50	0.97	0.74	0.79
Asn	0.78	0.77	1.30	1.14
Pro	0.58	0.80	1.25	1.36
Gln	1.16	0.92	1.08	0.84
Arg	0.99	1.04	1.04	0.93
Ser	0.74	0.97	1.30	0.98
Thr	0.75	1.08	0.97	1.21
Val	1.03	1.47	0.53	0.98
Trp	0.81	1.24	0.72	1.23
Tyr	0.74	1.28	1.03	0.95

Three letter amino acid code has been used.

TABLE 4

Pairs of amino acid residues whose P_{ijk} and P_{jik} values differ significantly in all four states (see Text)

Ala-Pro, Pro-Ala;	Ala-Val, Val-Ala;	Asp-Ser, Ser-Asp
Asp-Thr, Thr-Asp;	Glu-Lys, Lys-Glu;	Glu-Leu, Leu-Glu
Glu-Val, Val-Glu;	Phe-Leu, Leu-Phe;	Lys-Ser, Ser-Lys
Leu-Asn, Asn-Leu;	Leu-Thr, Thr-Leu;	Asn-Val, Val-Asn

Three letter amino acid code has been used.

pletely lost when one is working at the single amino acid residue level.

Further, looking at the values of potentials in both Table 2 and Table 3 it would appear that the pairs formed by amino acid residues which individually have a higher potential to exist in a particular secondary structure, also have great potential to exist in the same secondary structure as their constituents. This is not always true and the following pairs are such examples in the four secondary structural states considered.

Helix: Ile-Gln; Ile-Val; Lys-Ile; Leu-Gln; Gln-Gln; Val-Glu; Val-Phe; Val-Ile; Ile-Val.

Extended structure: Ile-Ile; Leu-Leu; Leu-Arg; Leu-Thr; Val-Leu.

Chain reversals: Cys-Gly; His-Gly; Asn-Gln; Gln-Ser; Tyr-Asp; Tyr-Ser.

Coil: Lys-Gly; Lys-Lys; Lys-Pro; Asn-Thr; Thr-Pro.

Tables 2 and 3 also illustrate that even though the constituent amino acid residues have less potential for a particular secondary structural state, the pair formed from the residues may have a great potential in that secondary structural state, as shown by the examples given below.

Helix: Asp-Asp; Asp-Pro, Arg-Ser; Ser-Asp.

Extended structure: Ala-Pro; Gly-Gly; Gly-Pro; Lys-Pro; Gln-Ser; Ser-Gly.

Chain reversals: Ala-Ile; Glu-Glu; Glu-Leu; Lys-Ala; Lys-Lys; Leu-Leu; Thr-Lys.

Coil: Ala-His; Ala-Gln; Asp-Phe; Glu-Phe; Ile-Glu; Leu-Gln; Gln-Ser; Ser-Phe; Ser-Leu; Ser-Ser; Ser-Tyr; Val-Asp; Val-Phe; Val-His; Tyr-Ser.

The above discussion suggests that the pairs of amino acids have different conformational properties from their constituent single amino acid residues. However, it does not clearly indicate whether one can reconstitute the pair potentials from single residue potentials. The single amino acid residue potential given in Table 3 for all four states have been used to determine the potentials for pairs of amino acids both by addition and multiplication of these individual potential values. These calculated potential values were compared with the observed potentials for pairs of amino acids given in Table 2. The comparison shows that the association effect between residues in the pair, which is included in the values of Table 2, is considerable and varies not only with the type of amino acids and the position they take in the pair but also with the secondary structure. In other words, comparing the ratio of potentials calculated for the pair of amino acids with the observed value in all four secondary structural states, the two vary considerably, indicating that short-range interactions are not incorporated in single residue potentials used in protein folding studies. This can be further seen by assuming that the observed potential for a given pair depends on (i) the types of amino acids, (ii) the secondary structure in which it is occurring, and (iii) the position of the residue in the pair, thus obtaining

$$P_{ijk} = (SS_{ik} \cdot X_{i1})(SS_{jk} \cdot X_{j2}) \qquad (4)$$

X_{i1} and X_{j2} are single amino acid potentials when i^{th} type of residue is in position one and j^{th} type of residue in position two. Note that X_{i1} and X_{j2} not only represent the type of amino acid but also the position of amino acid in the pair. Thus for a residue i, X_{i1} and X_{i2} are assumed to be different. SS_{ik} and SS_{jk} are secondary structural contributions to the association effect.

If the above assumptions are valid, and if the observed potentials of pairs can be partitioned into potentials of single amino acids, as is done in eqn. 4, then, using Table 2, for a given secondary structure

$$\frac{P_{AA}}{P_{CA}} = \frac{P_{AC}}{P_{CC}} = \frac{P_{AD}}{P_{CD}} \cdots = \frac{P_{AY}}{P_{CY}}$$

The ratio will also be equal to

$$\frac{P_{AA}}{P_{AC}} = \frac{P_{CA}}{P_{CC}} = \frac{P_{DA}}{P_{DC}} \cdots = \frac{P_{YA}}{P_{YC}}$$

where P_{AA} etc. are P_{ijk} values for a particular secondary structure. However, these ratios differ considerably even considering statistical fluctuations.

It is clear from the results and discussion above that single amino acid potentials, though statistically more reliable, will not account for the interactions between two neighbouring side chains. These interactions vary depending upon the postion of amino acid residue. They are also a function of the secondary structure in which these residues are present. Probably the single most important result of our studies, which may have appeared obvious at the outset, is that the pairs of amino acids cannot be regarded as a linear combination of constituent amino acids. Even at the secondary structural level the conformational properties of pairs of amino acids differ from the conformational properties of constituent amino acids. Secondary structure

prediction algorithms must, therefore, incorporate this aspect if better results are to be expected.

REFERENCES

1. Levitt, M. (1978) *Biochemistry* **17**, 4277–4285
2. Argos, P. & Palau, J. (1982) *Int. J. Peptide Protein Res.* **19**, 380–393
3. Chou, P.Y. & Fasman, G.D. (1978) in *Adv. Enzymol.* **47**, 45–148
4. Garnier, J., Osguthorpe, D.J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120
5. Busetta, B. & Hospital, M. (1982) *Biochim. Biophys. Acta* **701**, 111–118
6. Palau, J., Argos, P. & Puigdomenech, P. (1982) *Int. J. Peptide Protein Res.* **19**, 394–401
7. Lim, V.I. (1974) *J. Mol. Biol.* **88**, 873–894
8. Ptitsyn, O.B. & Finkelstein, A.V. (1970) *Biophysika* **15**, 785–796
9. Bunting, J.R., Athey, T.W. & Cathou, R.E. (1972) *Biochim. Biophys. Acta* **285**, 60–70
10. Wu, T.T. & Kabat, E.A. (1971) *Proc. Natl. Acad. Sci. US* **68**, 1501–1506
11. Bourgeios, S., Jernigan, R.L., Szu, S.C., Kabat, E.A. & Wu, T.T. (1979) *Biopolymers* **18**, 2625–2643
12. Rose, G.D. (1978) *Nature* **272**, 586–590
13. Kolaskar, A.S. & Ramabrahmam, V. (1982) *Int. J. Peptide Protein Res.* **19**, 1–9
14. Krigbaum, W.R. & Rubin, H.B. (1971) *Biochim. Biophys. Acta* **229**, 368–383
15. Krigbaum, W.R. & Komoriya, A. (1979) *Biochim. Biophys. Acta* **576**, 204
16. Meirovitch, H., Rackovsky, S. & Scheraga, H.A. (1980) *Macromolecules* **13**, 1398–1405
17. Wu, T.T., Szu, S.C., Jernigan, R.L., Bilofsky, H. & Kabat, E.A. (1978) *Biopolymers* **17**, 555–572
18. Kolaskar, A.S., Ramabrahmam, V. & Soman, K.V. (1980) *Int. J. Peptide Protein Res.* **16**, 1–11
19. Feldman, R. (1976) *Atlas of Macromolecular Structure on Microfiche*, 1st edn. Tracer Jitco Inc., Rockville, MD
20. Brown III, F.R., Hopfinger, A.J. & Blout, E.R. (1972) *J. Mol. Biol.* **63**, 101–115

Address:

Dr. *A.S. Kolaskar*
Center for Cellular and Molecular Biology
RRL Campus, Taranaka
Hyderabad 500-009
India