Int. J. Peptide Protein Res. **24**, 1984, 392–401

Are secondary structures secondary?

A.S. KOLASKAR[1] and V. RAMBRAHMAM[2]

[1] *Centre for Cellular & Molecular Biology, Regional Research Laboratory Campus, and* [2] *School of Life Sciences, University of Hyderabad, Hyderabad, India*

Received 4 January, accepted for publication 26 March 1984

Crystallographically observed secondary structures in globular proteins are divided into two categories based on calculated potential values for these structures. Type I secondary structures are those which have maximum potential in observed structural state. These type I structures consist of those amino acids which intrinsically prefer that structural state. On the other hand type II secondary structures have a calculated potential value which is not maximum for the observed state. Type II structures are composed of those amino acid residues which do not intrinsically prefer the observed state and thus seem to have been formed due to tertiary interactions. All the observed secondary structures have been identified as type I or type II in 31 different globular proteins considered. It is suggested that type II structures are formed at a latter stage of protein folding mainly due to favourable tertiary interactions.

Key words: amino acid pair potentials, application; globular protein structure data analysis; protein folding; secondary structure analysis

We have undertaken an analysis of crystal structure data of a large number of globular proteins in order to gain some insight into the folding process of proteins and to prepare an algorithm that would predict the three-dimensional structure of a protein from its primary structure. During this process the main chain conformational similarity among amino acid residues as well as main chain conformations affected significantly by corresponding side chains have been studied (1,2). It was also shown that a pair of amino acid residues behaves differently as compared to its constituents at the secondary structure level, mainly because the constituent amino acid residues interact non-linearly to form the pair (3). Therefore, the potentials obtained at the level of pairs of amino acid residues for four secondary states: helix (h), extended structure (e), chain reversal (t) and coil (c), have been used to analyze the crystallographically observed secondary structures in globular proteins. This analysis has been aimed at reaching an understanding of the formation of secondary structures as well as finding the reasons for the failures of secondary structure predictions made using statistical methods (4,5). It may be mentioned here that most of the secondary structure prediction algorithms do not give a prediction accuracy of more than 50–60%. Attempts to modify these algorithms for better results are not encouraging (6,7) and suggest that an upper limit for such a prediction accuracy might be about 60%.

In our analysis it has been pointed out that observed secondary structures fall under two

categories. The first type of secondary structures (called type I) have amino acid residues which have intrinsic preference for that secondary state. The other secondary structures (called type II) seem to have been formed due to favorable tertiary structural interactions even though the constituent amino acid residues may not intrinsically prefer that secondary structural state. This seems to us one of the main reasons for the failure of the commonly used statistical methods that are used to predict secondary structures. They assume that all regular structures are formed due to the interactions among amino acid residues which are closer sequentially and thus secondary structures can be predicted using the probability of occurrence of each of the amino acid residues in various secondary structural states. Though secondary structure prediction schemes (4–5, 8–11) have without a doubt given very useful results and the development of these methods is a major step in predicting protein structure, their accuracy seems to be limited, for the reasons we discuss here. The succeeding sections deal with the method used for the present analysis, the results obtained and their implications.

METHOD

A set of *31* globular proteins which has been used to derive the potential of pairs of amino acids (3) is used in this analysis. The helical and extended structure regions of these proteins are taken from the original papers as compiled in AMSOM (12) and chain reversals are computed using a simple algorithm developed by us, based on the coordinates of C^{α} atoms (13). The remaining amino acid residues were assumed to be in the unordered or the coil state. For analysis, each of these observed secondary structural stretches was considered. If ACDEF is a single letter coded amino acid sequence of a particular stretch, then its potential A_i to exist in any of the secondary state i was calculated making use of our amino acid pair potential data by the simple relation:

$$A_{\alpha} = 1/2\ [P_{\alpha}(AC) + P_{\alpha}(CD) + P_{\alpha}(DE) + P_{\alpha}(EF)]$$

$$A_{\beta} = 1/2\ [P_{\beta}(AC) + P_{\beta}(CD) + P_{\beta}(DE) + P_{\beta}(EF)]$$

$$A_t = 1/2\ [P_t(AC) + P_t(CD) + P_t(DE) + P_t(EF)]$$

$$A_c = 1/2\ [P_c(AC) + P_c(CD) + P_c(DE) + P_c(EF)]$$

where, P_{α}, P_{β}, P_t and P_c are potential values of a pair of amino acids in four secondary structural states. Thus, the potential values were calculated for all four states for each secondary structural segment. For example, crystal structure data of Hen egg white lysozyme show that regions 5–15, 24–34, 80–85 and 88–96 exist in helical state. Calculated A_i values for the 5–15 region are $A_{\alpha} = 47.2$, $A_{\beta} = 17.1$ $A_t = 19.2$ and $A_c = 16.1$ while for 24–34 region these values are $A_{\alpha} = 25.3$, $A_{\beta} = 28.7$, $A_t = 19.0$ and $A_c = 26.8$, indicating that 5–15 segment has intrinsic preference for helical conformation but not for the segment 24–34.

In order to check that the results obtained above are not artifacts due to the usage of amino acid pair potentials, calculations were repeated using the single amino acid residue potentials derived by Chou & Fasman (4), Garnier *et al.* (5) and us (3).

RESULTS AND DISCUSSION

Analysis of the calculated potential values for secondary structural stretches have brought out an important observation, namely that the calculated potential is not always maximum for secondary structures which are present in active conformation of the protein molecule. These results are not altered by the use of amino acid pair potentials or single residue potentials. We thus termed the secondary structural stretches for which the average potential is a maximum for the observed secondary structural state, type I, and those stretches for which the average potential is a maximum for a secondary structural state other than observed state, type II. The observed secondary structures in 31 different globular proteins divided into type I or type II based on our analysis and obtained using amino acid pair potentials are given in Table 1.

The 31 proteins are grouped in Table 2 into the various structural classes to which they belong. This data suggests there are more type I helices for all structural types of

TABLE 1

Secondary structural regions, h (helix), e (extended structure) and t (chain reversals), observed in globular proteins along with their categorization, namely type I and type II (*) secondary structure. Structural type of the protein is given in brackets. The "other" category also includes proteins which have one domain belonging to one type of structural protein and the other to the other type

Protein									
Human hemoglobin α-chain (α)	h	3–18	20–35	52–71	80–88	94–112	36–43*	118–141*	
	e	–							
	t	73–76	75–78	72–75*	89–92*	113–116*			
Human hemoglobin β-chain (α)	h	57–76	85–93	99–117	123–146	35–41*	50–56*		
	e	–							
	t	42–45	43–46	77–80	78–81	118–121	80–83*	94–97*	
Lamprey hemoglobin (α)	h	12–29	45–52	62–66	67–88	92–106	111–127	132–148	30–44*
	e	–							
	t	55–58	53–56*						
Spermwhale Met-myoglobin (α)	h	1–19	20–35	36–42	51–57	58–77	86–94	100–118	125–148
	e	–							
	t	43–46	44–47	46–49	78–81	79–82	95–98	121–124	82–85*
		119–122*							
Bacterial ferricytochrome C_2 (α)	h	2–10	64–71	75–80	96–110				
	e	–							
	t	12–15	15–18	21–24	32–35	43–46	49–52	84–87	14–17*
		26–29*	35–38*	39–42*	53–56*	54–57*	56–59*		
Bonito ferrocytochrome C (α)	h	1–11	61–69	90–103					
	e	–							
	t	12–15	14–17	15–18	35–38	43–46	50–53		
		51–54	53–56	70–73	71–74	75–78	21–24*		
		26–29*	27–30*	32–35*	73–76*				
Bacterial cytochrome C_{550} (α)	h	55–56	72–80	106–119	5–16*	81–91*			
	e	–							
	t	33–36	38–41	92–95	22–25*	23–26*	27–30*		
		40–43*	41–44*	45–48*	121–124*	129–132*	130–133*		
Bacterial rubredoxin (β)	h	–							
	e	1–16	35–46	17–27*	28–34*	47–53*			
	t	–							

Protein							
Jack bean concanavalin A (β)	h	–					
	e	24–30	46–48	49–56	60–67	73–79	89–97
		104–108	124–131	169–179	189–199	4–10*	35–38*
		109–115*	139–145*	147–149*	153–157*	208–215*	
	t	14–17	34–37	80–83	117–120	160–163	183–186
		201–204	203–206	226–229	15–18*	31–34*	81–84*
		82–85*	134–137*	216–219*	222–225*	227–230*	229–232*
		230–233*					
Bovine trypsin (β)	h	235–245*					
	e	29–37	41–46	50–55	63–69	80–93	101–108
		132–141	179–184	196–202	212–217	155–166*	224–229*
	t	23–26	48–51	56–59	70–73	95–98	96–99
		115–118	167–170	175–178	186–189	16–19*	25–28*
		72–75*	116–119*	119–122*	192–195*		
Porcine tosyl elastase (β)	h	244–254*					
	e	29–36	44–52	53–58	68–73	141–147	186–193
		214–221	222–226	86–89*	108–115*	204–210*	233–240*
	t	22–25	59–62	99–102	136–139	157–160	160–163
		195–198	196–199	227–230	1–4*	8–11*	10–13*
		12–15*	61–64*	116–119*	122–125*	134–137*	154–157*
		161–164*	164–167*	166–169*	169–172*	178–181*	229–232*
Human Bence Jones protein REI (monomer I) (β)	h	–					
	e	1–7	16–26	31–40	42–50	69–77	84–92
		61–67*	95–105*				
	t	55–58	8–11*	79–82*	80–83*		
Human immunoglobulin I (β)	h	–					
	e	1–5	7–12	16–25	31–38	83–91	132–140
		147–151	202–208	41–46*	53–66*	68–75*	94–108*
		116–120*	160–169*	173–182*	193–199*		
	t	27–30	126–129	127–130	128–131	154–157	155–158
		186–189	188–191	77–80*			
Human immunoglobulin II (β)	h	–					
	e	1–13	15–25	33–41	43–53	67–75	76–85
		92–101	103–114	115–118	155–163	178–188	198–205
		208–216	54–58*	120–129*	141–150*	166–176*	

Table 1 (continued)

Human immunoglobulin II (β)	t	60–63	61–64	63–66	28–31*	29–32*	87–90*	
		132–135*	135–138*	136–139*	189–192*	191–194*	193–196*	
		194–197*						
Human prealbumin monomer II (β)	h	75–83						
	e	54–56	115–123	11–19*	29–35*	42–49*	67–75*	
		89–97*	104–112*					
	t	1–4						
Chicken lysozyme (α + β)	h	5–15	80–85	24–34*	88–96*			
	e	42–46*	50–54*	59–60*				
	t	17–20	19–22	36–39	66–69	74–77	79–82	
		99–102	103–106	106–109	113–116	115–118	124–127	
		69–72	20–23*	82–85*	108–111*	119–122*	122–125*	
Bovine chymotrypsinogen A (α + β)	h	164–173	230–245					
	e	28–35	42–49	50–56	101–110	133–141	212–219	63–69*
		81–92*	155–162*	177–186*	193–201*	223–229*		
	t	108–111	115–118	4–7*	11–14*	13–16*	16–19*	23–26*
		95–98*	96–99*	116–119*	125–128*	202–205*	203–206*	
Bacterial high potential protein (α + β)	h	12–16	28–31					
	e	69–73	48–51*	57–58*	59–64*			
	t	3–6	8–11	20–23	37–40	38–41	40–43	42–45
		22–25*	23–26*	43–46*	53–56*	77–80*		
Carp calcium binding protein (α + β)	h	8–18	26–33	40–51	61–70	99–107	79–90*	
	e	56–60*	95–98*					
	t	2–5	20–23	1–4*	34–37*	35–38*	71–74*	
Bovine ferricytochrome b_5 (α + β)	h	33–38	42–50	64–74	8–16*	54–62*	80–87*	
	e	3–7	19–25	28–31	51–53*	75–79*		
	t							
Bovine ribonuclease S complex (α + β)	h	3–13	24–34	50–59				
	e	41–48	71–75	105–111	118–124	80–86*	96–104*	
	t	16–19	36–39	65–68	87–90	91–94	92–95	17–20*
		66–69*	112–115*					

Bacterial nuclease S complex ($\alpha + \beta$)	h	99–107	122–134	54–67*				
	e	12–19	21–27	30–36	38–41	70–78	89–95	108–113*
	t	3–6	46–49	47–50	49–52	83–86	137–140	139–142
		1–4*	115–118*	117–120*				
Human carbonic anhydrase C ($\alpha + \beta$)	h	34–38	155–162	219–229	16–20*			
	e	30–33	39–41	45–53	55–62	65–71	87–96	116–124
		206–212	215–218	229–231	239–241	76–82*	108–110*	125–126*
		140–151*	171–176*	191–196*	256–259*			
	t	7–10	8–11	11–14	12–15	99–102	128–131	129–132
		163–166	167–170	180–183	198–201	234–237	249–252	21–24*
		24–27*	104–107*	130–133*	131–134*	135–138*	178–181*	232–235*
Subtilisin BPN′ (α/β)	h	14–20	64–73	103–117	132–147	223–238	268–275	5–10*
		242–252*						
	e	28–32	120–124	148–152	205–209	45–50*	89–94*	213–217*
	t	23–26	36–39	39–42	51–54	56–59	60–63	97–100
		98–101	157–161	159–162	193–196	219–222	263–266	83–86*
		85–88*	166–169*	168–171*	171–174*	181–184*	187–190*	259–262*
		261–264*						
Bovine carboxypeptidase A complex (α/β)	h	14–28	82–88	112–122	215–231	285–306	72–80*	94–103*
		173–187*	254–262*					
	e	32–46	265–271	49–53*	60–66*	104–109*	190–196*	200–204*
		239–241*						
	t	8–11	29–32	56–59	89–92	123–126	142–145	148–151
		150–153	153–156	159–162	169–172	206–209	232–235	273–276
		275–278	277–280	3–6*	4–7*	67–70*	162–165*	242–245*
		244–247*	245–248*	250–253*	278–281*			
Lobster GPD (α/β)	h	9–22	36–46	101–112	147–162	250–264	316–332	201–216*
		279–286*						
	e	1–7	27–32	90–97	115–120	126–128	142–146	168–179
		224–233	236–246	270–274	289–302	303–312	56–61*	62–67*
		185–191*	192–194*					
	t	47–50	78–81	79–82	83–86	84–87	85–88	138–141
		163–166	218–221	219–222	221–224	265–268	48–51*	76–79*
		122–125*	129–132*	130–133*	133–136*			
Chicken triose phosphate isomerase monomer (α/β)	h	79–87	95–102	105–120	130–137	138–154	177–196	197–204
		17–31	44–55*	213–223				

Table 1 (continued)

Chicken triose phosphate	e	5–11	36–42	89–93	122–129	205–209	60–63*	159–167*
isomerase monomer		227–231*						
(α/β)	t	237–240	12–15*	56–59*	57–60*	74–75*	169–172*	234–237*
		243–246*						
Bacterial semiquinone	h	10–27	66–74	93–107	124–138			
flavodoxin (α/β)	e	1–6	30–35	108–110	115–119	48–55*	80–89*	111–114*
	t	62–65	39–42*	42–45*	43–46*	56–59*	57–60*	77–80*
Bacterial thermolysin (others)	h	137–150	235–246	260–274	281–296	301–312	67–87*	160–179*
	e	27–30	37–46	52–54	60–63	97–106	112–116	3–13*
		15–25*	31–36*	55–58*	119–123*			
	t	107–110	126–129	130–133	132–135	151–154	181–184	209–212
		210–213	216–219	217–220	249–252	250–253	276–279	297–300
		92–95*	128–131*	133–136*	187–190*	189–192*	190–193*	192–195*
		194–197*	197–200*	204–207*	205–208*	207–210*	300–303*	311–314*
Dogfish apolactate dehydro-	h	2–6	55–70	120–130	165–181	249–263	308–329	33–44*
genase complex (others)		107–109*	227–236*	237–245*				
	e	23–28	77–81	92–97	159–161	188–192	281–295	48–53*
		134–139*	200–207*	267–279*	298–303*			
	t	88–91	140–143	184–187	208–211	213–216	19–22*	84–87*
		86–89*	102–105*	103–106*	142–145*	152–155*	196–199*	209–212*
		216–219*	221–224*					
Bacterial ferrodoxin (others)	h	–						
	e	1–7	8–19	35–46	20–26*	27–34*	47–54*	
	t	–						
Horse alcohol dehydrogenase	h	47–54	229–236	250–259	275–283	353–365	170–187*	202–212*
complex (others)		305–310*	324–336*					
	e	34–40	62–65	72–78	128–132	145–146	148–152	193–199
		218–224	287–293	312–318	369–374	9–14*	22–29*	41–45*
		68–71*	86–92*	135–138*	156–160*	238–243*	262–269	347–352*
	t	1–4	102–105	105–108	116–119	123–126	244–247	271–274
		295–298	296–299	319–322	342–345	3–6*	55–58*	80–83*
		100–103*	115–118*	139–142*	140–143*	161–164*	165–168*	338–341*
Papain (others)	h	24–43	49–57	67–78	117–128	137–143*		
	e	5–7	130–131	169–175	179–183	185–191	206–208	111–112*
		162–167*						
	t	19–22	58–61	61–64	62–65	82–85	96–99	195–198
		201–204	8–11*	84–87*	98–101*	198–201*	199–202*	

TABLE 2

Total number of each type of observed secondary structure in different structural types of globular proteins along with the division into type I and type II. The "others" category also includes proteins which have one domain belonging to one type of structural protein and the other to the other type

Structural type of protein	Helix			Extended structure			Chain reversal		
	Observed	Type I	Type II	Observed	Type I	Type II	Observed	Type I	Type II
α	42	36	6	–	–	–	64	35	29
β	3	–	3	96	59	37	84	40	44
$\alpha + \beta$	33	22	11	57	30	27	90	51	39
α/β	38	27	11	44	26	18	80	43	37
Others	30	20	10	57	32	25	75	35	40
Total	146	105	41	254	147	107	393	204	189

proteins. The number of type II secondary structures is non-negligible and particularly large in the case of chain reversals and extended structures. In fact, inaccuracies in predicting secondary structures using statistical methods are quite large for extended and chain reversal structures as opposed to α-helical structures.

It is generally suggested that residues in the sequential neighbourhood of secondary structures are very important and decide the structure of the segment. In order to study the influence of these residues on the secondary structure-forming capacity, potential values were calculated by incorporating the four residues present before the *N*-terminus and after the *C*-terminus of the secondary structural segment. The average potential values calculated after the addition of these eight residues indicate that the assignment of one secondary structure over the other for a given stretch becomes less easy, in very few cases is the calculated potential the maximum one for the observed secondary structure. This study indicates that the addition of few residues at either the *N*-terminus or the *C*-terminus or both does not improve the results.

To understand the formation of type I and type II secondary structures, three different aspects were studied: composition of the amino acids of these structures; occurrence of type I and type II structures in different portions of the primary structure of a protein; and length of the secondary structural stretch.

Composition of amino acids

The study of the composition of observed secondary structures was undertaken mainly because it is inherently assumed in all statistical prediction schemes that the secondary structural stretches found by crystallographic analysis always have residues which either prefer or are indifferent to that structure. However, as mentioned above, we have found a considerable number of observed secondary structures which have submaximal potentials for the state in which they exist. The calculated average potential for the type II secondary structures can be less if (i) most of the constituent amino acids have low potentials for the observed secondary structures, or (ii) most of the residues present are indifferent to that structure, or (iii) residues having a high and low potential for the observed secondary structural state are almost equal in number. Our analysis of type II secondary structures has revealed that these structures consist of amino acids which do not have an intrinsic preference for the observed secondary structural state. This analysis also points out that type II structures are not formed due to clustering of amino acids having maximum tendency to exist in that secondary structure and followed by those residues which have least preference for that state. Also type II structures are not made from amino acid residues which are indifferent to that structural state. This suggests that type II structures are formed due to the existence of tertiary interactions which force these segments to assume the secondary structure not necessarily favored by constituent amino acid residues. In other words, the segment or its polymer which assumes the type II structure in protein

will not fold into a secondary structure as observed in protein molecules under similar conditions.

However, our analysis shows that type I structures comprise mostly residues which have a high potential for the observed secondary structural state. This could be expected since the calculated average potential for the observed secondary structural state is maximum and thus the segment or its polymer which takes the type I structure will fold under similar conditions into a structure as observed in protein, indicating that tertiary interactions are not very important.

Occurrence of secondary structures, in N*-terminal, central and* C*-terminal parts of protein*

Further, it was of interest to see whether type I or type II structures are localized in any particular region of the protein such as the *N*-terminal, central and *C*-terminal parts. The frequency of occurrence of type I and type II structures, shown in Table 3, suggests that type I and type II structures are distributed throughout the primary structure of the protein and are not localized in any particular part of the protein.

Length

Type I and type II helices and extended structures are further divided into long, medium and short stretches. To make such a classification, first the average length ($\bar{L}$) (i.e. the average number of residues in the stretch) and the standard deviation (σ) are calculated. If the number of residues in the stretch is greater than ($\bar{L} + \sigma$), the stretch is termed "long"; if the number of residues is less than ($\bar{L} - \sigma$), the stretch is "short"; and if the number of residues is in between ($\bar{L} + \sigma$) and ($\bar{L} - \sigma$), the stretch is of "medium" length. In Table 3, $\bar{L}$ and σ are given for helix and extended structure. Data given in Table 3 shows that most of the long helices are of type I (about 75% long helices are type I) while nearly 50% of all long extended structures belong to type II. On the other hand, nearly 50% of medium length helices and extended structures are of type II. Thus, at least on the basis of length one can say that type I and type II structures are not distinguishable, particularly extended structures.

Our analysis of the length of secondary structures and the distribution of secondary structural segments along the sequence has not helped us to distinguish type I structures from type II structures. We have, therefore, studied their spatial locations in various proteins. Our preliminary results indicate that most type II secondary structures (particularly β-strands and α helical structures) are located on the surface of the proteins. Type I structures seem to be

TABLE 3

Distribution of type I and type II secondary structures

	Helix		Extended structure		Chain reversals	
	Type I	Type II	Type I	Type II	Type I	Type II
Along protein primary structure						
N-terminus	30	14	62	27	58	64
Central part	37	11	40	42	85	74
C-terminus	38	16	45	34	61	51
Lengthwise						
Long	71	24	20	20	–	–
Medium	34	16	115	84	–	–
Short	–	1	12	3	–	–
Average length ($\bar{L}$)	11.73		7.03			
Standard (σ) deviation	4.4		2.79			

buried inside the protein molecule. More details about these studies as well as their significance in constructing the folding pathway will be discussed later.

CONCLUSIONS

Type I secondary structures are those stretches of polypeptide chain which have an in-built tendency to assume the observed secondary structure. Type I structures would thus be good candidates to act as nucleation sites during the protein folding process.

Type II secondary structures are formed as a result of interactions among amino acid residues that are close not only at the primary structure level (i.e. covalently) but also at the tertiary structure level. This suggests that all secondary structures need not be formed at the same stage of the protein folding process. Further, these data also point out that a few secondary structures formed at an early stage of protein folding might break due to tertiary interactions.

Secondary structure prediction schemes thus should take into consideration tertiary interactions also, since a sufficiently large number of the secondary structures are formed and some are broken due to tertiary interactions. But the presently available crystal structure data seem to be inadequate to derive statistically significant weights for tertiary interactions, as these interactions vary with the structural class of the protein. Hence, a direct prediction of the tertiary structure, without being concerned about the prediction of secondary structures, seems to be a better alternative at present.

REFERENCES

1. Kolaskar, A.S. & Ramabrahmam, V. (1981) *Int. J. Biol. Macromol.* **3**, 171–178
2. Kolaskar, A.S. & Ramabrahmam, V. (1981) *Int. J. Peptide Protein Res.* **19**, 1–9
3. Kolaskar, A.S. & Ramabrahmam, V. (1983) *Int. J. Peptide Protein Res.* **22**, 83–91
4. Chou, P.Y. & Fasman, G.D. (1978) *Advan. Enzymol.* **47**, 45–148
5. Garnier, J., Osguthorpe, D.J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120
6. Busetta, B. & Hospital, M. (1982) *Biochim. Biophys. Acta* **701**, 111–118
7. Palau, J., Argos, P. & Puigdomenech, P. (1982) *Int. J. Peptide Protein Res.* **19**, 394–401
8. Tanaka, S. & Scheraga, H.A. (1976) *Macromolecules* **9**, 168–182
9. Lim, V.I. (1974) *J. Mol. Biol.* **88**, 873–894
10. Kabat, E.A. & Wu, T.T. (1973) *Biopolymers* **12**, 751–774
11. Bounting, J.R., Athey, T.W. & Cathou, R.E. (1972) *Biochim. Biophys. Acta* **285**, 60–70
12. Feldman, R. (1976) *Atlas of Macromolecular Structure on Microfiche*, 1st edn. Tracer Jitco Inc. Rockville, MD
13. Kolaskar, A.S., Soman, K.V. & Ramabrahmam, V. (1980) *Int. J. Peptide Protein Res.* **16**, 1–11

Address:
Dr. *A.S. Kolaskar*
Centre for Cellular and Molecular Biology
Regional Research Laboratory Campus
Hyderabad – 500 007
India