The Madras Triple Helix: Origins and Current Status

Manju Bansal

Manju Bansal is a Professor at the Molecular Biophysics Unit, Indian Institute of Science, Bangalore. She is involved in computer modelling of the structure and dynamics of biopolymers: proteins and nucleic acids, as well as the interactions between them. Analysis of possible correlations between sequence and structure, predicting one from the other for these biomolecules, is another area of research being pursued by her group.

Author wishes to thank Anirban Ghosh and Aditi Kanhere for assistance in preparing the figures.

'Madras triple helix' was the name assigned by the scientific community in the West, to the molecular model proposed for the fibrous protein collagen, by G N Ramachandran's group at the University of Madras. As mentioned jocularly in a recent retrospective of this work by Sasisekharan and Yathindra [1], the term was possibly coined due to the difficulty of Western scientists in pronouncing the Indian names of Ramachandran and his associates. The unravelling of the precise nature of collagen structure indeed makes for a fascinating story and as succinctly put by Dickerson [2]: "... to trace the evolution of the structure of collagen is to trace the evolution of fibrous protein crystallography in miniature". This article is a brief review highlighting the pioneering contributions made by G N Ramachandran in elucidating the correct structure of this important molecule and is a sincere tribute by the author to her mentor, doctoral thesis supervisor and major source of inspiration for embarking on a career in biophysics.

What is 'Collagen'?

The term 'collagen' is derived from the Greek word for glue and was initially described as "that constituent of connective tissue, which yields gelatin on boiling". However, it is now known that in some of the tissues, collagen in either heavily cross-linked or covalently bonded to some other stable structure, so that it cannot be extracted by just heating. About one quarter of all the protein in most animals, is collagen. It is the major constituent of all connective tissues in vertebrate as well as invertebrate animals, performing in the connective tissue of animals somewhat the same function as cellulose molecules in plants. Each collagen polypeptide chain contains more than 1000 amino acid residues. Skin, tendon, bone, cartilage, cornea and teeth all

Figure 1. There are many different kinds of collagen assemblies, found in various animal tissues. This picture depicts a tough sheet like surface which is found in basement membranes and supports the skin and other organs. Collagen molecules associate together to form an extended network, seen here in light blue (picture taken from http://www.rcsb. org/pdb/molecules).

contain collagen fibrils. These fibrils may be organized in many different ways: they form molecular cables that strengthen the tendons, large resilient sheets (*Figure* 1) that support the skin and internal organs, as mineralized aggregates in bone and teeth. Today we know that the collagen superfamily of proteins contains at least 19 proteins that are formally defined as collagen and an additional 10 proteins that have collagen-like domains. Collagen, therefore, defies any simple definition and is best characterized on the basis of some constitutive and structural features, which are common to most members of the collagen family.

Two sets of characteristics can differentiate collagen from other proteins, first is the amino acid composition which is distinctive in its very high content of glycine residues (~33% of the total) and the iminoacids residues proline and hydroxyproline (*Figure* 2). Second is the wide angle X-ray diffraction pattern (*Figure* 3), which shows a strong

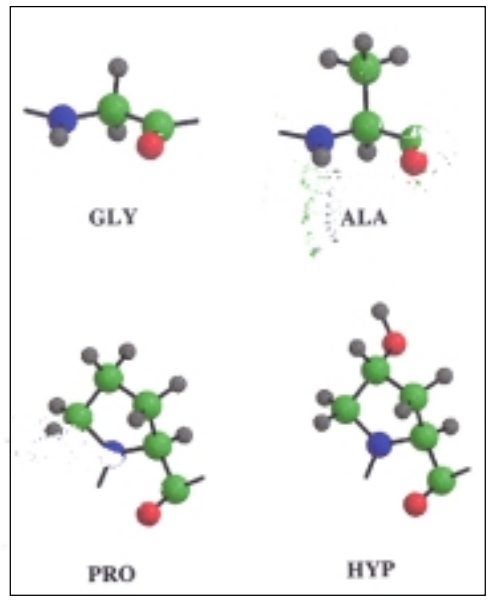


Figure 2. Ball-and-stick models showing the most commonly occurring amino acids in the polypeptide chain constituting a collagen molecule. The residues are represented by their three letter code: glycine (GLY), alanine (ALA), the iminoacids proline (PRO) and 4-hydroxyproline (HYP). The various atoms are shown as spheres (or balls) which are colour coded: Carbon (green); Nitrogen (blue); Oxygen (red); and Hydrogen (grey).

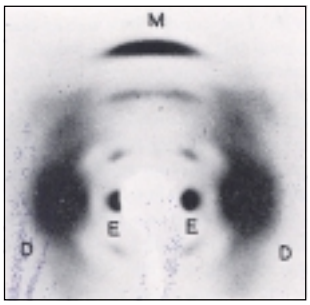


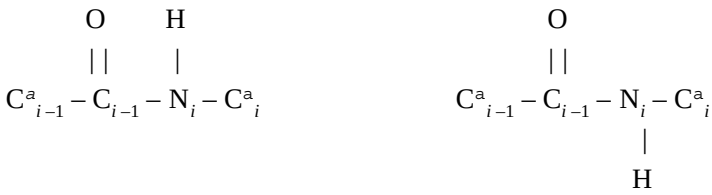
Figure 3. X-ray diffraction pattern of a stretched collagen fibre, taken with the specimen inclined to the X-ray beam (which leads to the pattern being asymmetrical about the equator or horizontal axis). 'M' corresponds to the meridional reflection and its spacing gives information about the rise per residue/unit along the fibre or z-axis (~ 2.9 Å). 'E' refers to the equatorial spots, which provide information about the lateral packing of the molecules in a plane normal to the fibre axis. 'D' refers to diffuse blobs on the equator. The spacings of the equatorial reflections are related to the diameter of the cylindrical molecules as well as the hydration state of the fibres.

meridional arc, corresponding to a repeat axial spacing of about 2.9 Å, and an equatorial spot corresponding to a spacing of 12 Å or more, depending on water content of the fiber or possible reaction with other small molecules. No other protein exhibits these features in combination. Hence their combined presence in an uncharacterized specimen is sufficient evidence that the substance is some form of collagen, but is not enough to give any indication about its molecular structure.

Origin of the 'Madras Triple Helix'

1950s saw the dawn of a 'Golden Period' for structural molecular biology. The correct structures for a-helix and b-sheets in proteins were proposed in 1951, by Linus Pauling's group at Caltech, USA and almost immediately experimentally confirmed to be correct by X-ray diffraction analysis. Several groups were actively working on solving the structure of the fibrous protein collagen, as well as of deoxyribonucleic acid (DNA). More guesses were made about the structure of collagen, than any other fibrous protein, indicating its more complex nature. Structure analysis studies of Pauling and Corey had established that the orientation about the peptide bond connecting two amino acids is essentially planar and that conformational flexibility of protein structures arises due to torsion about the bonds N–C^a and C^a–C. These torsion (or dihedral) angles are given the names j and y. The dihedral angles j_i and y_i for any particular amino acids at position 'i', correspond to the torsion angles (C_{i-1}–N_i–C_i^a–C_i) and (N_i–C_i^a–C_i–N_{i+1}), respectively. The presence of the various side chains attached to the C^a atom, in particular non-glycine amino acids, considerably restricts the possible combinations of j and y, which are sterically allowed and these are described in detail in the accompanying article by C Ramakrishnan. The possible conformations for a collagen chain are further restricted due to the presence of large amounts of iminoacid residues, proline and hydroxyproline, which have their side chains folded back so that the C^d atom in the side chain of proline is covalently bonded to the peptide nitrogen, thus forming a five membered ring *(Figure* 2). Consequently

there is very little freedom of rotation about the N–C^{a} bond and the angle ϕ is restricted to values close to –60°, while ψ is confined to the near-trans (~160°) region, for these residues. However, for the same reasons (viz. presence of the C^{d} atom, in place of the amino hydrogen) the peptide bond preceding an iminoacid at position ' i ' can take up *cis* or *trans* conformation with equal ease, unlike for other amino acids, wherein the *trans* conformation is overwhelmingly favoured:



It was probably for this reason that several early models for collagen incorporated *cis* peptide units in the structure, unlike the all-*trans* models proposed for α and β structures in proteins. Astbury, in 1938, [3] first suggested a helical structure for collagen based on a mixture of *cis* and *trans* type of peptide bonds. This structure did not explain all the features of the collagen X-ray pattern. His suggestion regarding *cis* residues was taken over by Pauling and Corey, who, in 1951 suggested a different structure, based on three co-axial helices, which also was not in good agreement with the observed X-ray pattern. Into this scenario entered G N Ramachandran, a young Indian scientist working at Madras (now Chennai), a crystal physicist by training, but who was fascinated by the beauty and complexity of the novel biomolecular structures. At the suggestion of J D Bernal, during a visit to Madras in 1952, Ramachandran's group started recording X-ray diffraction pictures of collagen fibers from different sources (such as, shark fin, rat tail and kangaroo tail tendon). Less than two years later, in a brief note published in *Nature*, Ramachandran and Kartha [4] proposed the first prototype of the correct structure for collagen. To quote from their report: "… A structure has been obtained which fits the above unit cell (viz from the X-ray diffraction data) and

At the suggestion of J D Bernal, during a visit to Madras in 1952, Ramachandran's group started recording X-ray diffraction pictures of collagen fibers from different sources.

which appears to be in good agreement with infra-red, X-ray and chemical data for collagen. It consists of nine amino-acid residues per unit cell, which corresponds to the observed density. These are linked together to form cylindrical rods, which occur in a hexagonal array. All the residues have the *trans* configuration, and the latest values of Corey and Pauling for the dimension of the amide group were used for the calculation. The residues are arranged in the form of three helical chains, each of pitch 9.5 Å (= *c*) and containing three residues per turn, with the symmetry 3_1. The three helices are also arranged with a 3_1 symmetry about the *c*-axis and they are held together by means of hydrogen bonds to form the cylindrical rods. Of the three *a*-carbon atoms per turn, a hydrogen attached to one of them could be replaced by a general R group to form an amino acid residue such as arginine or lysine; another (P) takes part in forming the proline ring, while the third (G) is in such a position that there is no space for either of its hydrogens to be replaced by any other group, so that it could only form part of a glycine residue". This in a nutshell was the basic triple helical structure (shown schematically in *Figure* 4a) and consists of three parallel helical chains, incorporates only *trans* peptide

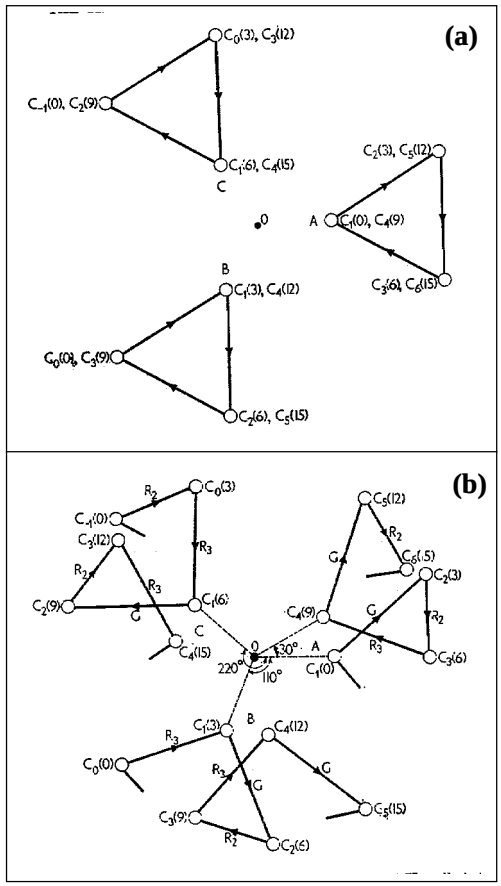


Figure 4. The Madras triple helix, shown projected down the helix axis. (a) The prototype structure, with the three helical chains arranged side by side and observed for some polypeptides. (b) The coiled-coil structure for collagen, in which each of the three helical chains is twisted about the common central axis through an angle of +30° and translated along the axis by ~9 Å, for every three residues. Neighbouring chains are related by a rotation of –110° and a translation of ~3 Å.The diagram is schematic, with the a-carbon atoms being represented as circles and the peptide units by straight lines joining them. The numbers in parentheses indicate the heights of a-carbon atoms, with the residue height being approximated to 3 Å for convenience.

bonds and explains the unusual amino acid composition of collagen.

Refinement of the Triple Helical Structure

However, Ramachandran was not one to rest on his laurels and continued studying the X-ray diffraction patterns of stretched collagen fibres, which indicated that the true repeat for collagen was not 3 residues per turn but had a rather unusual non-integral value of $3^1/_3$ or 10 residues in 3 turns. Ramachandran and Kartha in 1955 [5] promptly revised their structure on the basis of this new data and came up with the entirely novel concept of a rope-like 'coiled-coil structure' for collagen. In this modified structure (also published in *Nature*) the three chains, instead of being arranged with their axes parallel to the fibre axis, are all wound around the common central axis, thus following a helical path. Every third *a*-carbon atom (corresponding to the glycine position in the original structure) occurs on the surface of a cylinder of radius 1 Å, with successive ones being displaced by 8.58 Å along the axis and rotated by an angle of +36° (subsequently refined to a value of +30°). The two other chains occur in such a manner that the three chains (which individually take up a left-handed helical structure for *L*-amino acids) are symmetrically disposed with respect to each other and related by a rotation of ±108° (±110° in the refined model) about the axis of the cylinder and a translation of ±2.86 Å parallel to the axis (*Figure* 4b). The major helix is therefore wound in a direction opposite to that of the individual minor helices, with every third residue being similarly situated with respect to the major helix axis. This fundamental idea of a coiled-coil structure has been confirmed by recent crystal structure analysis of oligotripeptides of well-defined sequences of –Gly–R_2–R_3– type (*Figure* 5).

Ramachandran and Kartha in 1955 came up with the entirely novel concept of a rope-like 'coiled-coil structure' for collagen.

G N Ramachandran during one of his lectures mentioned that, he got the idea of a coiled-coil model for collagen, from astronomy. The moon, while it rotates, also revolves around the earth and always presents the same side to the earth because of

G N Ramachandran during one of his lectures mentioned that, he got the idea of a coiled-coil model for collagen, from astronomy.

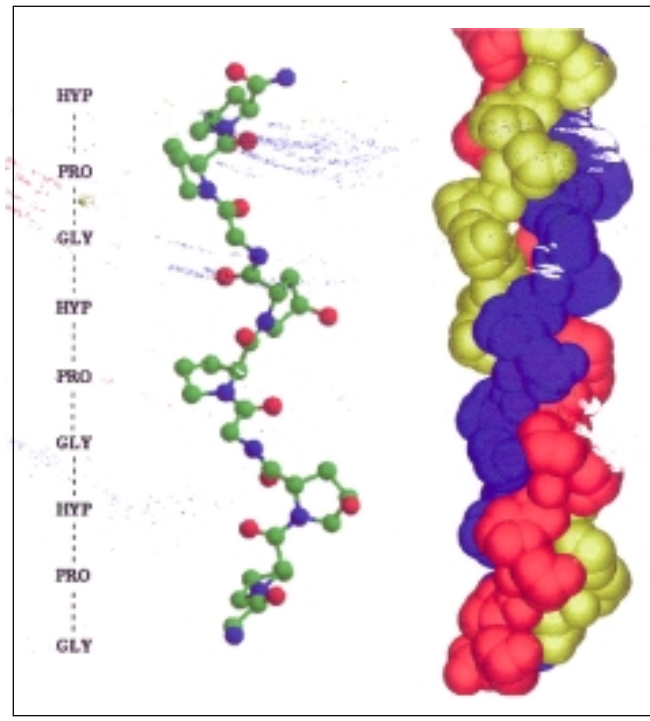


Figure 5. The molecular structure of an oligotripeptide, determined recently using X-ray crystallography (Bella and others, 1995), is shown here, for a fragment with the amino acid sequence –Gly–Pro–Hyp– repeated three times. Ball-and-stick model of the extended single chain, which takes up a left handed helical strcuture, is shown in the middle, while a space-filling model of the triple helical assembly is shown on extreme right, with each chain being represented by a different colour.

their coordinated movements. This idea was incorporated into the collagen structure, in which the glycyl residues always face the center of the triple helix. Thus, G N Ramachandran was the first to make the insightful link that the unusually high content of glycine and iminoacids in collagen, must have a structural basis. In particular, a unique structure specific role was assigned to the glycine residue at every third position in the repeating tripeptide sequence $(Gly–R_2–R_3)_n$. The presence of large amounts of the bulky iminoacids also has important implications for the collagen structure, leading to the rather extended helices, while the absence of amino groups restricts the possibility of hydrogen bonds. Both the prototype parallel, as well as the coiled-coil, triple helical structures were postulated to be stabilized by two N-H…O hydrogen bonds per tripeptide, between neighbouring chains. These involved the NH group of glycine and the NH group of the residue in R2 position when this was not an iminoacid, being hydrogen bonded to the oxygen atoms of CO groups in a neighbouring chain. This structure was criticized by

G N Ramachandran was the first to make the insightful link that the unusually high content of glycine and iminoacids in collagen, must have a structural basis.

Rich and Crick [6] as being untenable, due to some close van der Waal's atomic contacts and its inability to accommodate an iminoacid at the R2 position, even though –Gly–Pro–Hyp– sequences are known to occur in collagen. Rich and Crick [6] as well as Cowan and others [7] both proposed a slightly modified alternative model, with only one set of interchain hydrogen bonds for every tripeptide repeat. Ramachandran accepted this criticism and subsequently suggested that while only one set of direct N–H…O hydrogen bonds are possible, additional hydrogen bonds could be formed via water molecules [8]. This structure was a happy blend of good stereochemistry with maximum possible number of hydrogen bonds. Inter-chain C^{a}–H…O hydrogen bonds, involving glycine residues (similar to those reported for poly-glycine II structure) were also first proposed for the collagen structure by Ramachandran and Sasisekharan (1965) [9].

4-hydroxyproline is only found preceding glycine and known to lend additional stability to the collagen structure, but how it accomplishes this was not clear for quite some time.

Another interesting aspects of the primary structure is the fact that while the iminoacid proline is incorporated in locations R2 and R3 with almost equal frequency, the residues at location R3 alone are hydroxylated to give 4-hydroxyproline. Thus 4-hydroxyproline is only found preceding glycine and known to lend additional stability to the collagen structure, but how it accomplishes this was not clear for quite some time. In 1972, investigation of this puzzling problem was assigned to me (then a fresh graduate student) by Ramachandran. To our great delight, we found that hydroxyproline residues in the R3 position can form additional hydrogen bonds via the water molecules located in the inter-chain space (as shown in *Figure* 6) as well as between neighbouring triple helical molecules in the collagen fibrils. This very nicely explained the role of hydroxyproline residues in providing additional stability to the collagen structure, when they occur at the R3 position in the collagen tripeptide sequence [10]. Some of the other amino acids also show a preference for one or the other location and the unequal distributions of charged residues gives rise to favourable intra and inter molecular interactions (through ion-pairs), while non-polar residues

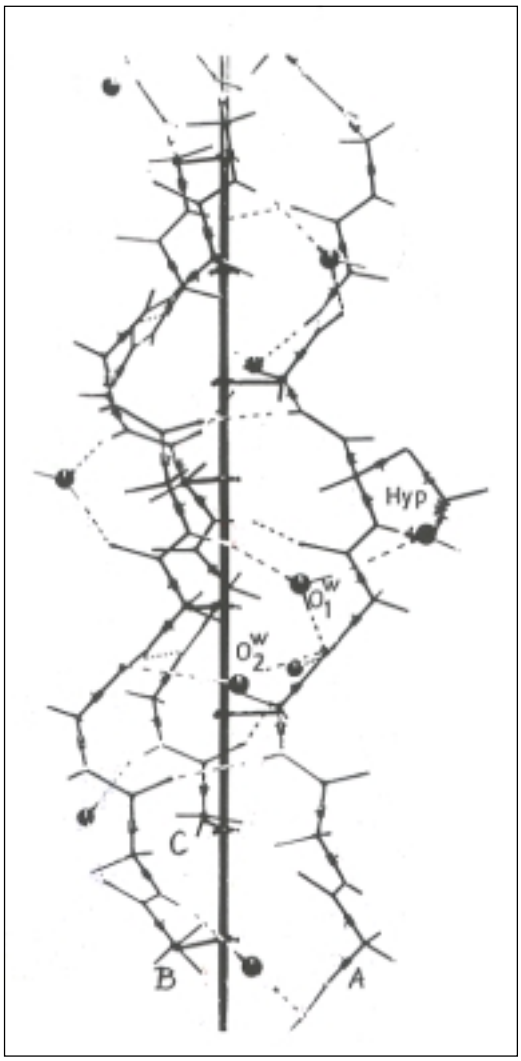


Figure 6. A model of the collagen triple-helical structure built in Ramachandran's laboratory (in 1975), using 'Kendrew type' skeletal models. The sequence shown here is a repeat of the tripeptide –Gly–Ala–Ala– in all three chains A, B and C. The dark colour rods represent covalent bonds connecting various atoms, while hydrogen bonds are shown by the light and dark banded connectors. Spheres represent the oxygen atoms of the water molecules linking the chains. The iminoacid hydroxyproline (Hyp) is shown occurring at one place in the molecule, with its hydroxyl group forming a hydrogen bond via a water molecule (O^w_1), thus providing additional stability to the triple helical structure.

may favour either of the two positions due to stereochemical considerations.

Current Status

It is interesting to note that, the original triple helical structure, proposed for collagen in 1954, consisting of an assembly of three helical chains without further coiling, while not found in collagen, turned out to be the prototype structure for a variety of polypeptides and is also observed in some regions of globular proteins. Thus, poly-glycine, poly-*L*-proline and poly-*L*-hydroxyproline are stereochemically similar to the simple triple helix. Also crystal structures of several proteins have revealed that proline rich fragments in various proteins, including a crucial element in proteins that are produced by oncogenes, take up the left-handed, three residues per turn helical structure, very similar to the individual helices in the 'prototype' triple helical structure. A variety of collagen related polypeptides and oligopeptides, with a tripeptide repeat of the –Gly–R_2–R_3– type, have however been shown to take up the typical coiled-coil triple helical structure, but they show considerable variation in their helical parameters. Thus, the term

'collagen structure' really defines a family of structures, related by certain common features.

The coiled-coil triple helical structure of collagen, the interchain (rather than intrachain) direct hydrogen bonds involving the glycine NH group, additional water mediated N–H…O=C and hydroxyproline-water hydrogen bonded networks, as well as C^{a}-H…O hydrogen bonds have all been visualized in recent high resolution crystal structures of triple helical peptides of defined sequences [11]. The essential requirement of glycine, and only glycine, at every third position in the sequence and its critical structural role have been proven unambiguously. It is pertinent to mention that over 400 mutations in different collagens have been identified to cause a variety of human diseases. An overwhelming majority of these involve the substitution of glycine residues by non-glycyl residues and manifest themselves as defects in the stability, extensibility or assembly of collagen molecules/fibrils in various tissues. This is entirely expected on the basis of the Ramachandran coiled-coil triple helical structure of collagen, wherein substitution of a glycine, at the centrally located position in the core of the triple chain assembly, would considerably perturb the molecular structure (for example by introducing a kink), which in turn affects its stability or its susceptibility to various enzymes. In addition, the structure places the sidechains of residues at positions R2 and R3 on the surface of the molecule. This arrangement explains the ability of many collagens to polymerize, since the clusters of hydrophobic and charged sidechains direct self-assembly into precisely ordered structures.

Thus, most of the salient features of the molecular structure proposed for collagen by G N Ramachandran, have been fully validated by recent structural studies, more than 40 years later. It is a matter of great satisfaction that the first major contribution from this most distinguished Indian biophysicist (who followed up the collagen structure elucidation with his more well-known work on the Ramachandran map), has at last received its due recognition.

Suggested Reading

[1] V Sasisekharan and N Yathindra, In *Trends in Collagen* (G Chandrakasan, N Yathindra, and T Ramasami, Eds.), Indian Academy of Sciences, Bangalore, 1999.

[2] R E Dickerson, *The Proteins*, Vol. 2, 604-778, 1964.

[3] W T Astbury, *Trans. Faraday Soc.* Vol. 34, 378, 1938.

[4] G N Ramachandran and G Kartha, *Nature*, Vol. 174, 269-270, 1954.

[5] G N Ramachandran and G Kartha, *Nature*, Vol. 176, 593-595, 1955.

[6] A Rich and F Crick, *Nature*, Vol.176, 915-916, 1955.

[7] P M Cowan, S McGavin and A C T North, *Nature*, Vol. 176, 1062-1064, 1955.

[8] G N Ramachandran and R Chandrasekharan, *Biopolymers*, Vol. 6, 1649-1658, 1968.

[9] G N Ramachandran and V Sasisekharan, *Biochim. Biophys. Acta.*, Vol. 109, 314-316, 1965.

[10] G N Ramachandran, M Bansal and R S Bhatnagar, *Biochim. Biophys. Acta*, Vol. 322, 166-171, 1973.

[11] J Bella, B Brodsky and H M Berman, *Structure*, Vol. 3, 893-906, 1995.

Address for Correspondence
Manju Bansal
Molecular Biophysics Unit
Indian Institute of Science
Bangalore 560 012, India.
Email: mb@mbu.iisc.ernet.in