

## Self-similarity and the assembly of collagen molecules

V. Sasisekharan\* and M. Bansal

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

The Bragg reflections seen on the equator, in the X-ray pattern of collagen fibrils, have been generally interpreted as arising from a quasi-hexagonal packing of the triple-helical collagen molecules. However, these reflections can be 'accurately' indexed only on the basis of a much larger monoclinic cell. In this paper we outline a completely different approach, based on the principle of self-similarity, to describe the assembly of rod-like molecules of collagen. Assuming that the collagen molecules are packed, with local hexagonal site symmetry, but no long range translational order, an aperiodic lattice arrangement is described (with only the inter-molecular separation as a variable parameter). This can account for the observed equatorial X-ray pattern as well as the longer range lateral spacings seen in collagen fibrils under the electron microscope.

THE early 1950s were a major turning point in the history of structural biology. It was during this period that the alpha helical and beta pleated sheet structures for polypeptides and proteins<sup>1,2</sup>, the double-helical structure for DNA<sup>3</sup>, and the triple-helical structure for collagen<sup>4-6</sup> were proposed. The correct clue to the molecular structure of collagen was provided by Prof. G. N. Ramachandran and Dr G. Kartha in their pioneering paper which appeared in *Nature*, 1954, 174, 269. The structure had three parallel polypeptide chains held together by interchain hydrogen bonds<sup>4</sup>. Subsequent refinement and comparison with X-ray data indicated that the three chains are necessarily coiled-coil<sup>5</sup>. Following the then prevailing dogma, that all the available NH groups are to be hydrogen-bonded, a structure with two hydrogen bonds for every three residues (the third residue usually being a proline/hydroxyproline residue) was proposed<sup>6</sup>. The imposition of two hydrogen bonds for every three residues resulted in a few contact distances between non-bonded atoms being less than the sum of their van der Waals radii.

Relieving of these contacts resulted in a structure with one hydrogen bond for every tripeptide of the collagen molecule<sup>7</sup>. However, subsequent theoretical and experimental studies indicated that wherever possible, additional water-mediated hydrogen bonds may be formed between the backbone atoms in neighbouring chains<sup>8</sup>. One of the water molecules can then also hydrogen bond to the hydroxyl group of hydroxyproline at position three in the repeating sequence Gly-R<sub>2</sub>-R<sub>3</sub>, thus providing a rationale for the hydroxylation of proline residues occurring at this position<sup>9</sup>.

It may not be out of place to record in this special issue devoted to biopolymer conformation that the criticism of the two-bonded structure of collagen, namely, that it contained nonbonded contact distances less than the sum of the van der Waals' radii of the atoms involved, led us to investigate, in a detailed manner, equilibrium or minimum contact distances between non-bonded atoms in crystal structures<sup>10</sup>. The outcome of the survey was the conclusion that non-bonded atoms could come as close as 0.89 of the sum of the van der Waals' radii of the corresponding two atoms. The repulsion between the two atoms operates below this distance and generally, conformations that lead to distances between non-bonded atoms below this distance, are disallowed. The above analysis of the crystal structures for allowed contact distances was the basis for the now well-known Ramachandran ( $\phi$ - $\psi$ ) plot for polypeptides and proteins<sup>11</sup>.

This approach to conformational analysis is a negative one. It emphasises the fact that atoms have finite sizes and that no two atoms can approach each other below a certain distance, called the contact distance. Conformations which lead to distances below this distance are 'disallowed' and the remaining conformations are 'allowed'. It does not grade the 'allowed' ones. In other words, it does not tell us which one among the allowed ones is energetically favoured or preferred. This is what, classical, semi-empirical or quantum chemical conformational energy calculations

\*For correspondence.

attempt to do to obtain the minimum energy conformation<sup>12</sup>. But then, the evaluation of global minimum in a problem with multiple variables, is still an unsolved problem. In this context, the application of the contact distance criterion is a powerful one, as it eliminates all the impossible ones, at one step.

Reverting to the molecular structure of collagen, it has now become textbook matter that the collagen molecule has a triple-helical coiled-coil structure. This molecular structure satisfies all the available physico-chemical data and is consistent with the wide angle X-ray diffraction data. But collagen fibrils also give rise to meridional as well as equatorial reflections in the small and medium-angle region. These reflections are interpreted as due to higher order periodicity along the molecular axis and lateral assembly of molecules in the horizontal direction. It is the lateral assembly of collagen molecules that we are concerned with, in this paper. The wide-angle pattern of collagen gives an indication that the diameter of the collagen triple helix is about 14 Å and the molecules are probably arranged in a hexagonal lattice with the smallest *d*-spacing corresponding to about 12.6 Å. However the whole series of higher order reflections in the range between 12 and 40 Å can only be explained on the basis of a much larger unit cell of the monoclinic or tetragonal type<sup>13-16</sup>. In an earlier paper it had been shown that these reflections could also arise as subsidiary maxima if the collagen fibril has a cylindrical lattice structure<sup>17, 18</sup>. Several attempts have been made subsequently to accurately index the crystalline equatorial reflections and it is clear that the molecular arrangement is only quasi-hexagonal and open to different interpretations. A detailed analysis has been carried out recently by Miller's group, using data in the medium-angle region, obtained with a focusing mirror monochromator camera<sup>19</sup>. This study, as well as analysis of stained cryosections from collagen fibrils under the electron microscope<sup>20</sup>, have shown that all these data are compatible with the Hulmes-Miller quasi-hexagonal packing model<sup>15</sup>. The observed values reported by these workers, for the equatorial (including near-equatorial) reflections are listed in Table 1. The basic unit cell is monoclinic and has been assigned values of  $a = 39.0 \text{ \AA}$ ,  $b = 26.7 \text{ \AA}$  and  $\gamma = 104.58^\circ$ . Blocks of unit cells, which are assemblies of this unit cell (with dimensions such as  $40m \times 27n$  where  $m$  and  $n$  have small integer values) have been invoked to account for the occurrence of periodicities in the range 56.0–116.5 Å, as seen in the electron microscope<sup>20</sup>.

However, Miller and Tocchetti<sup>19</sup> have also noted that only some of the fibrils have 3-D crystallinity indicative of specificity in the lateral aggregation of collagen molecules. The fact that the crystalline reflections are accompanied by diffuse scatter and that the sharp reflections are not always observed, leads one

**Table 1.** The observed reflections in the equatorial, (including near-equatorial) region of the X-ray pattern of rat tail tendon collagen<sup>19</sup> are listed, along with the spacings seen in the electron micrographs of collagen fibrils<sup>20</sup>.

$R_{\text{obs}}$ ( $\text{\AA}^{-1}$ ) (X-ray)	$d = 1/R_{\text{obs}}$ ( $\text{\AA}$ )
0.104	9.64
0.097	10.31
0.088	11.41
0.082	12.24
0.079	12.64
0.075	13.28
0.073	13.69
0.057	17.48
0.053	18.90
0.041	24.61
0.038	26.51
0.026	38.00
—	73.8(6.1)*
—	99.2(6.2)*

\*These *d*-spacings are obtained from electron micrographs<sup>20</sup>. The standard deviations for these spacings are given in parentheses.

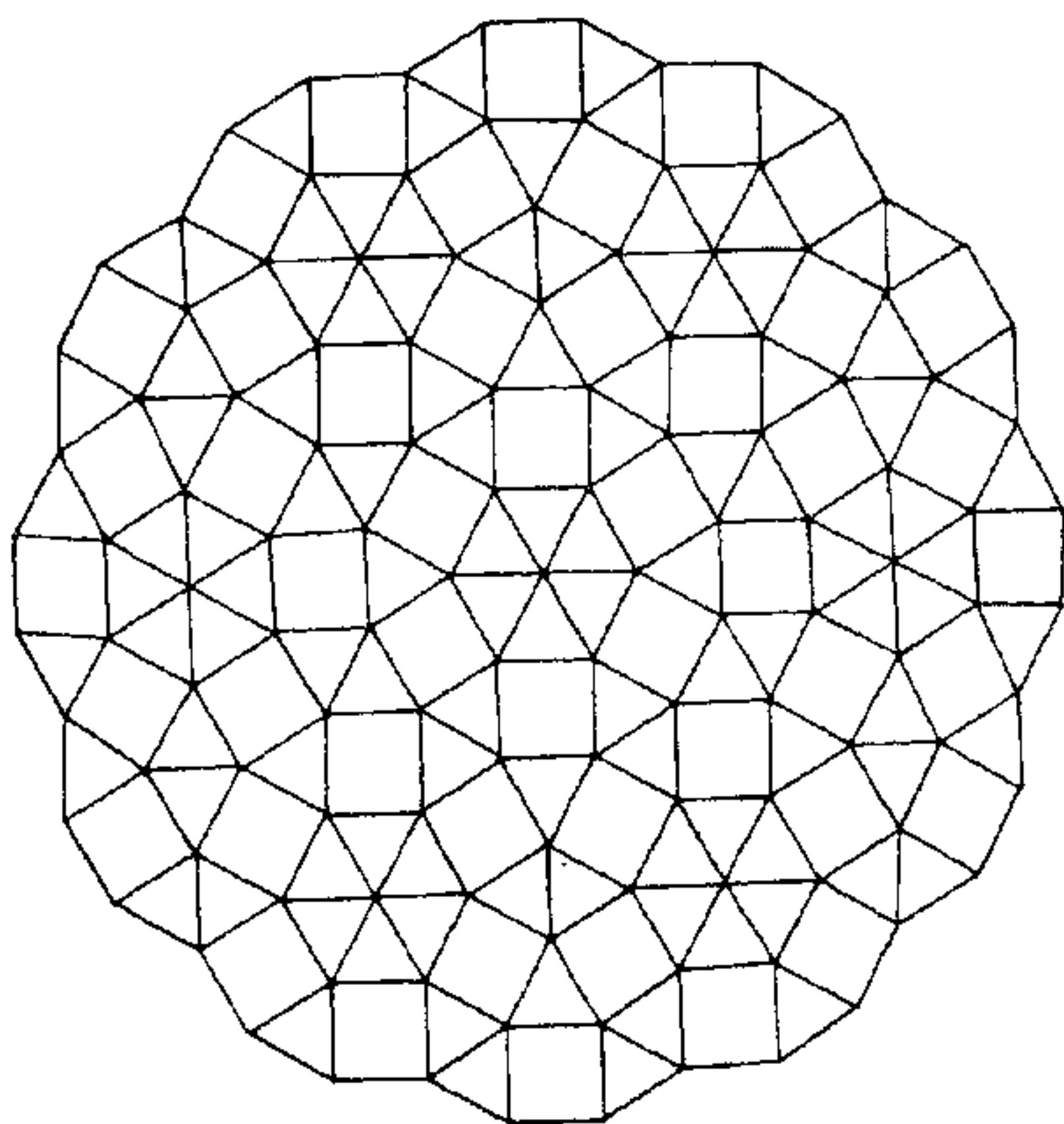
to conclude that the lateral geometric order is not always regular and also breaks down easily.

An altogether different approach, based on the principle of self-similarity could also be thought of to explain the observed reflections in the medium- and low-angle equatorial region of the collagen X-ray pattern. The problem of horizontal assembly of rod-like molecules, such as collagen, could be treated as similar to 2-D tiling. In general, the application of the principle of self-similarity to space-sharing problems leads to both periodic and aperiodic lattice arrangements. The possibility of tiling a plane with any given symmetry was first shown by Sasisekharan<sup>21, 22</sup>. In 2-D space the application of crystallographic symmetries, such as 2, 3, 4 and 6, lead to regular lattice arrangements, with the vertices being identical everywhere. Noncrystallographic symmetries such as 5, 7, 8 and higher integers always lead to aperiodic tiling arrangements that are self-similar. These symmetries do not have translational lattice periodicity and hence there is no long-range order. On the other hand they can have bond orientational order and possess local site symmetry. It was also demonstrated that self-similar aperiodic tilings with no long-range translational order could also be obtained for the crystallographic symmetries cited above<sup>22</sup>.

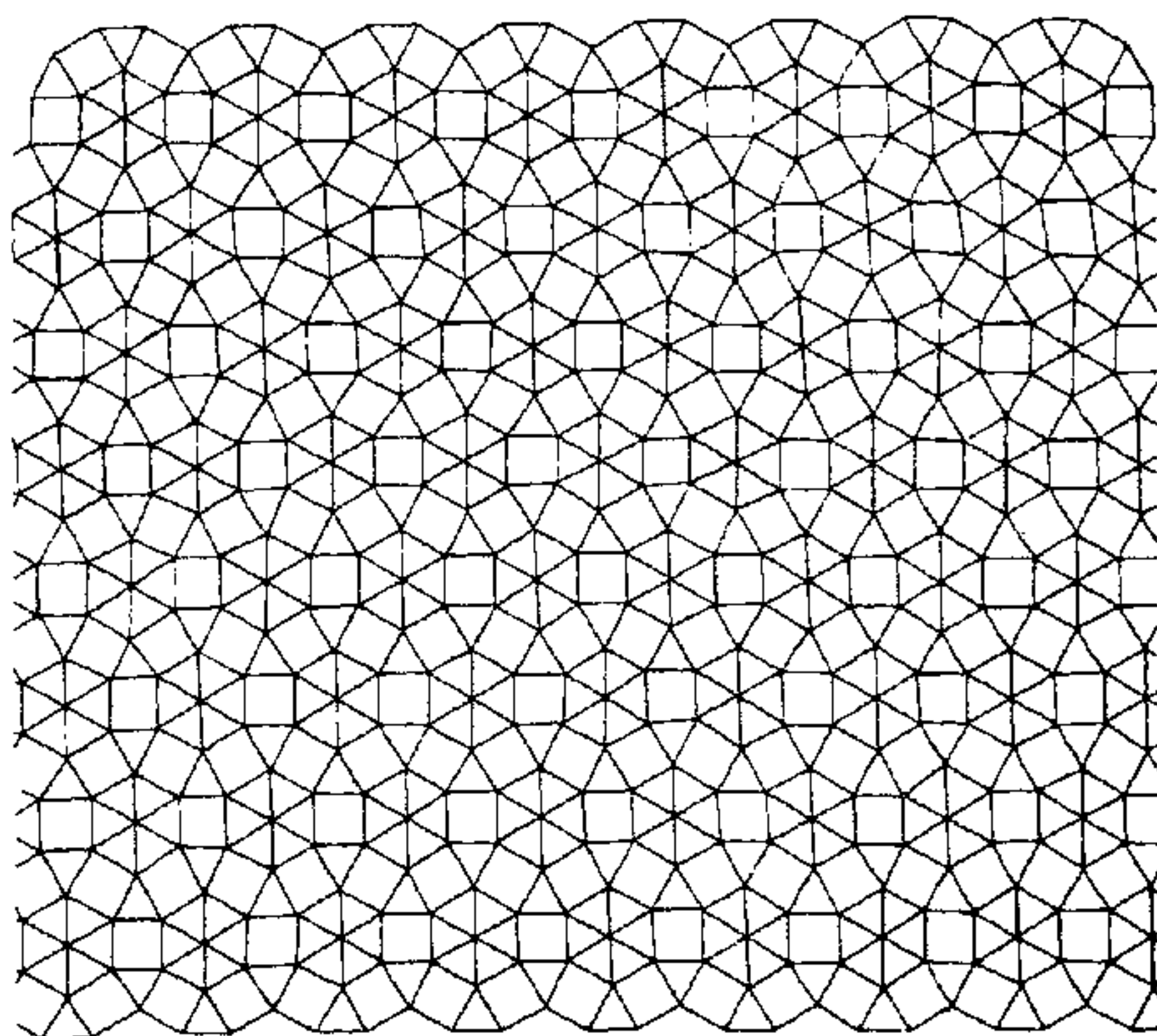
Assuming that the collagen molecules are arranged in a pseudo-hexagonal pattern, we have generated, periodic hexagonal lattices as well as aperiodic lattices with hexagonal site-symmetry. In order to arrive at an aperiodic arrangement with local hexagonal symmetry, based on the self-similarity principle, the basic motifs are a square and an equilateral triangle. The vertices of



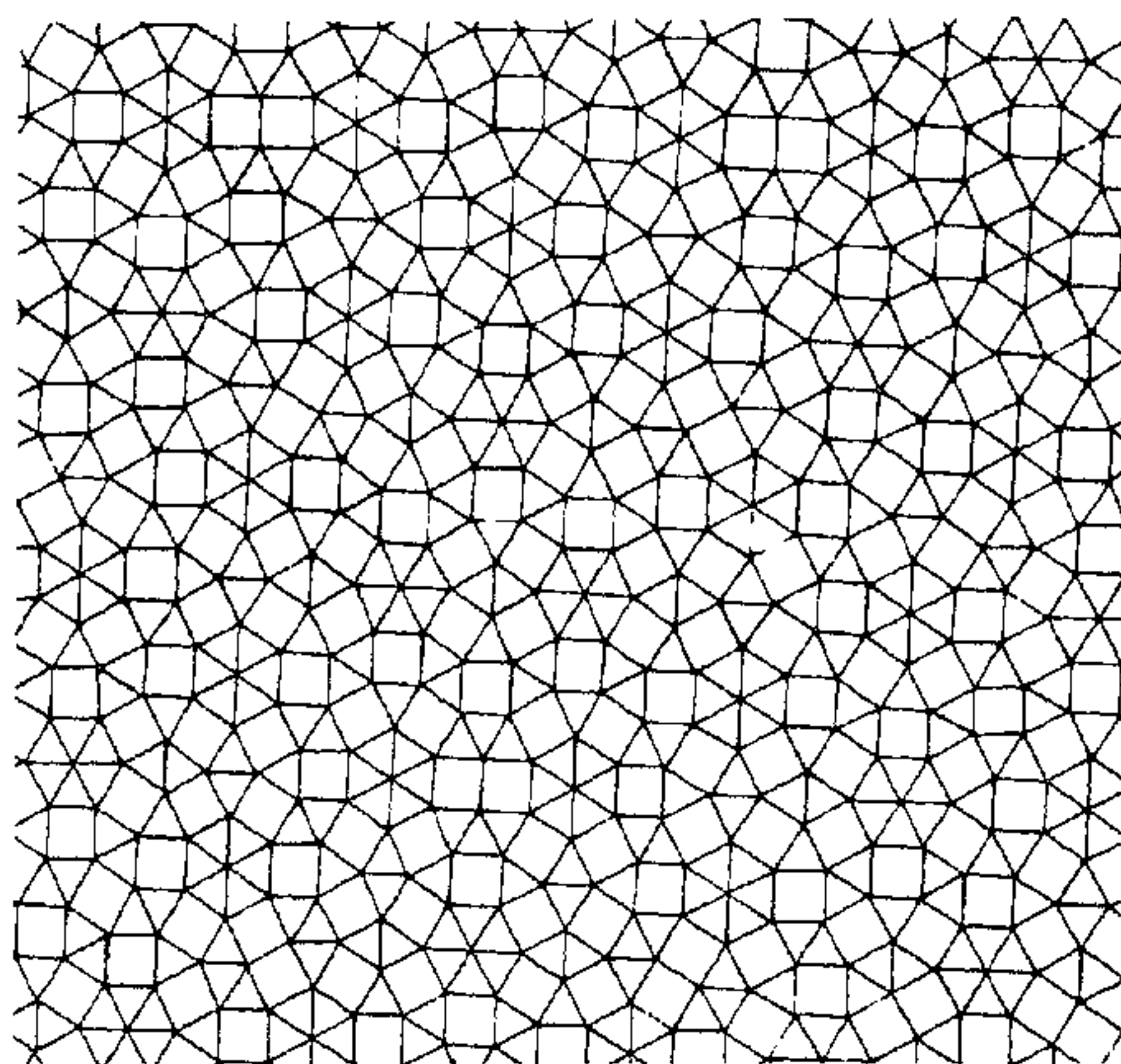
these two basic motifs can be taken to represent the triple-helical collagen molecule. Inflation/deflation procedures then lead to both periodic and aperiodic arrangements. Figure 1 shows a small portion of the aperiodic arrangement while Figure 2 shows an extended region with periodic arrangement. It may be noticed that the hexagonal symmetry in Figure 1, in contrast to Figure 2, does not extend laterally, though the arrangement has local six-fold symmetry. This can



**Figure 1.** The aperiodic arrangement of vertices (each vertex corresponds to a collagen molecule), obtained on the basis of self-similarity principle, is shown in detail. Note the presence of local hexagonal site-symmetry.



**Figure 2.** An extended region showing a periodic hexagonal arrangement, using a square and an equilateral triangle as the basic motifs. The unit cell corresponds to  $(1 + \sqrt{3})$  or  $\approx 2.732$  times the inter-vertex distance  $d$ . Hence if  $d \approx 14 \text{ \AA}$  the hexagonal unit cell edge  $a \approx 38 \text{ \AA}$ .



**Figure 3.** An extended region showing an aperiodic arrangement, using a square and an equilateral triangle as the basic motifs. Here also the inter-vertex distance is constant, and the environment of all vertices are self-similar and non-random. The local site-symmetry highlighted in Figure 1 is preserved throughout this extended region but there is no long-range translation order.

be seen more clearly in Figure 3, which shows a larger region with aperiodic arrangement. However, the near-neighbour contacts are self-similar and uniform throughout, so that the form of the assembly is preserved in the extended pattern. Since each vertex in the lattice is the site of a collagen molecule, the inter-vertex distance should correspond to the inter-molecular separation which generally lies between 14 and 15 Å. The side of the unit cell for the periodic hexagonal arrangement shown in Figure 2 is then about 38–40 Å and therefore will not account for all the observed equatorial reflections in the collagen X-ray pattern. On the other hand, there is no obvious unit cell, with a repeat of this size, in the aperiodic arrangement (Figure 3), which is characterized only by the vertices being self-similar and equidistant from each other. A Fourier transform of the two lattices will therefore be different, with the periodic lattice giving rise to a conventional reciprocal lattice pattern. The Fourier transform computed for an aperiodic arrangement of the type shown in Figures 1 and 3 also gives rise to sharp maxima, but with no lateral periodicity. A cylindrically averaged Fourier transform of such a lattice, with an inter vertex spacing of  $a = 14.8 \text{ \AA}$  is shown in Figure 4. A comparison of the  $d$ -spacings of the maxima in Figure 4 with the spacings of the equatorial reflections listed in Table 1 shows reasonably good agreement. The maxima corresponding to 12.6, 13.5 and 38.0 Å are particularly strong. The calculated transform also shows remarkably sharp maxima in the low-angle region ( $R < 0.02 \text{ \AA}^{-1}$ ) and could be attributed

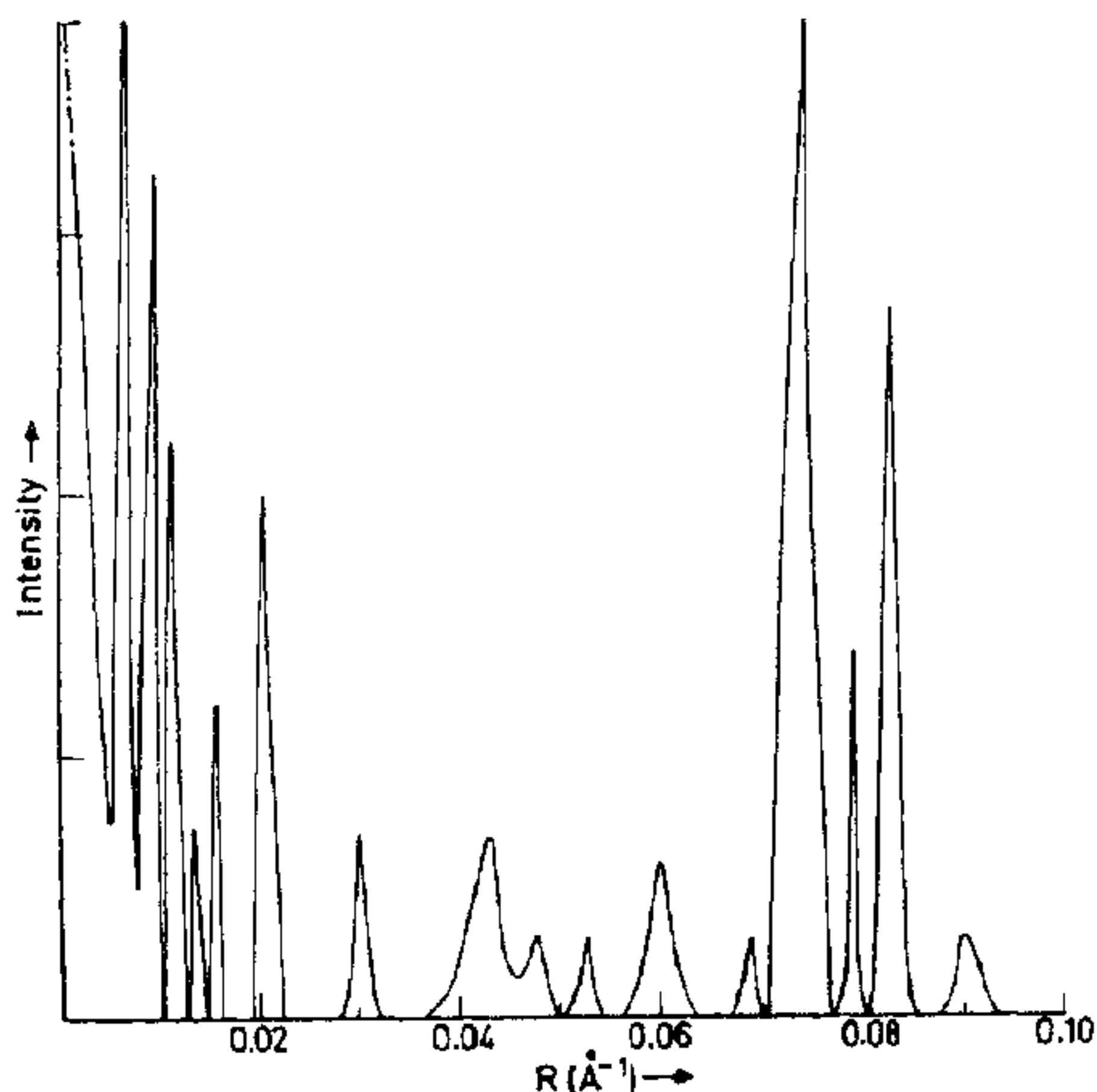


Figure 4. The cylindrically averaged Fourier transform of an aperiodic arrangement, shows sharp maxima in the medium as well as low angle region. The transform has been calculated for an intervertex separation of 14.8 Å and gives maxima at R-values close to those listed in Table 1.

to the presence of long-range lateral assemblies, such as are seen under the electron microscope<sup>20</sup>. Thus, an aperiodic arrangement, based on self-similarity, and described by a single parameter corresponding to the inter-molecular separation, is sufficient to explain the dominant features of the observed equatorial reflections in the X-ray pattern as well as the lateral assembly seen in the electron micrographs of collagen fibrils. A detailed analysis of the different aperiodic arrangements has to be carried out to compare the exact spacings of the equatorial reflections observed in the collagen X-ray pattern with the predicted maxima in the calculated Fourier transform.

1. Pauling, L. and Corey, R. B., *Proc. Natl. Acad. Sci. USA*, 1951, **31**, 235.
2. Pauling, L. and Corey, R. B., *Proc. Natl. Acad. Sci. USA*, 1953, **39**, 253.
3. Watson, J. D. and Crick, F. H. C., *Nature*, 1953, **171**, 737.
4. Ramachandran, G. N. and Kartha, G., *Nature*, 1954, **174**, 269.
5. Ramachandran, G. N. and Ambady, G. K., *Curr. Sci.*, 1954, **23**, 349.
6. Ramachandran, G. N. and Kartha, G., *Nature*, 1955, **176**, 593.
7. Rich, A. and Crick, F. H. C., *Nature*, 1955, **176**, 915.
8. Ramachandran, G. N. and Chandrasekaran, R., *Biopolymers*, 1968, **6**, 1649.
9. Ramachandran, G. N., Bansal, M. and Bhatnagar, R. S., *Biochem. Biophys. Acta*, 1973, **322**, 166.
10. Sasisekharan, V., in *Collagen* (ed. Ramanathan, N.), Wiley, New York, 1962, p. 39.
11. Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V., *J. Mol. Biol.*, 1963, **7**, 95.
12. Ramachandran, G. N. and Sasisekharan, V., *Adv. Protein Chem.*, 1968, **23**, 283.
13. North, A. C. T., Cowan, P. M. and Randall, J. T., *Nature*, 1954, **174**, 1142.
14. Miller, A., in *Biochemistry of Collagen* (eds. Ramachandran, G. N. and Reddi, A. H.), Plenum Press, New York, 1976, p. 85.
15. Hulmes, D. J. S. and Miller, A., *Nature*, 1979, **282**, 878.
16. Woodhead-Galloway, J., Hukins, D. W. L. and Wray, J. S., *Biochem. Biophys. Res. Commun.*, 1975, **64**, 1237.
17. Ramachandran, G. N. and Sasisekharan, V., *Arch. Biochem. Biophys.*, 1956, **63**, 255.
18. Sasisekharan, V. and Ramachandran, G. N., *Proc. Indian Acad. Sci.*, 1957, **A45**, 363.
19. Miller, A. and Tocchetti, D., *Int. J. Biol. Macromol.*, 1981, **3**, 9.
20. Chew, M. W. K. and Squire, J. M., *Int. J. Biol. Macromol.*, 1986, **8**, 27.
21. Sasisekharan, V., *Pramana—J. Phys.*, 1986, **26**, 283.
22. Sasisekharan, V., Baranidaran, S., Balaguruswamy, V. S. K., Srinivasan, A. and Gopal, E. S. R., *Pramana—J. Phys.*, 1989, **33**, 405.

ACKNOWLEDGEMENT. V. S. thanks Fogarty International Center, National Institutes of Health, USA. Part of this work was carried out when he was a Fogarty scholar.