Acta Cryst. (1977). B33, 1212–1218

A Molecular Crystal Model as a Basis for the Structure of the Collagen Fibril

BY JOHN WOODHEAD-GALLOWAY*

Laboratory of Molecular Biophysics, Zoology Department, Oxford, England and Department of Rheumatology, Medical School, University of Manchester, Manchester M13 9PT, England

(*Received* 2 *August* 1976; *accepted* 21 *September* 1976)

A simple three-dimensional molecular crystal model for the collagen fibril is derived from the schematic suggestion of Hodge & Petruska [*Aspects of Protein Structure* (1963), edited by G. N. Ramachandran, pp. 289–300, New York: Academic Press]. The model is shown to account for certain features of the near-equatorial low-angle X-ray diffraction pattern of wet native rat-tail tendon [Miller & Wray, *Nature, Lond.* (1971), **230**, 437–439]. The relation of this model to other models is explored. Replacing the single molecules by dimers provides a model which gives a fair account of the distribution of the intensity *very close* to the equator of the diffraction pattern. An elaboration of the model is proposed, exploiting the 5/4 ratio of molecules in the 'overlap' and 'gap' regions of the fibril in an attempt to account for the off-equatorial part of the pattern.

1. Introduction

It has recently been shown that a simple disordered molecular model accounts for features of the near-equatorial diffuse scatter in the X-ray diffraction pattern of wet rat-tail tendon (Woodhead-Galloway & Machin, 1976; Woodhead-Galloway & Young, 1977). By a molecular model is intended one in which molecules are not grouped in an intermediate level of order between the molecule itself and the fibril [contrast the microfibril of Smith (1968)]. The success of such a model prompts the exploration of an ordered molecular model to explain the discrete Bragg reflexions also seen in the equatorial region of the diffraction pattern.

2. The model of Hodge & Petruska

Fig. 1(*a*) shows the customary Hodge & Petruska (1963) (H–P) theory of the collagen fibril; successive molecules are staggered with respect to their neighbours by a precisely defined axial displacement of about 67 nm (usually referred to as D and equal to the observed axial period of the fibril). Fig. 1(*b*) is the alternative simple model for the fibril where the axial relationship is a stagger of $2D$ [sometimes called the Bear–Selby (B–S) model – see Bear & Selby (1956) for the origin of the name]. The first question for which an answer is attempted is whether ideas of packing molecules in a plane can immediately be extended to packing in three dimensions.

3. The low-angle X-ray diffraction pattern

The answer to the question is that there is no difficulty in principle in constructing three-dimensional molecular models. It is not intended here, however, to give an ex-

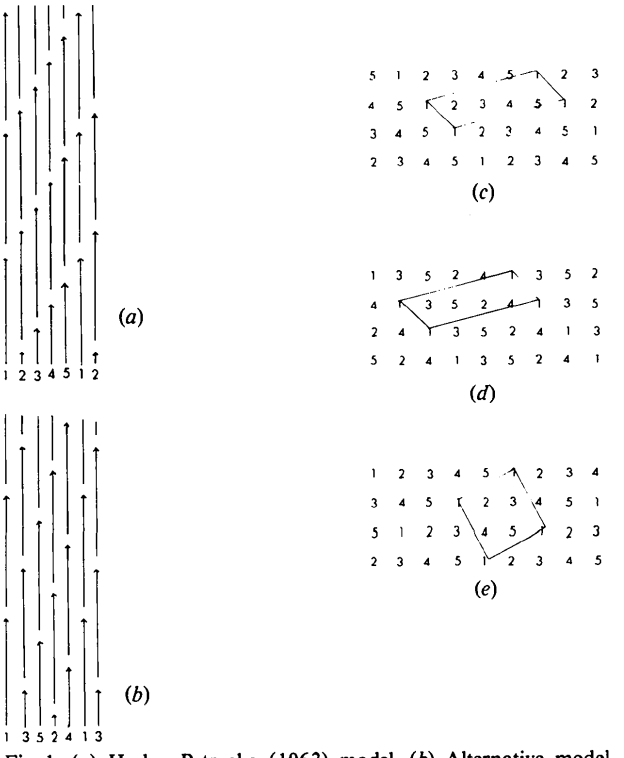

Fig. 1. (*a*) Hodge–Petruska (1963) model. (*b*) Alternative model based on Bear & Selby (1956). (*c*), (*d*), (*e*) Molecules packed on a square lattice to give three-dimensional structure: (*c*) neighbours staggered by D, (*d*) neighbours staggered by $2D$, (*e*) mixture of (*c*) and (*d*) to give a square true unit cell.

* Present address: MRC, 20 Park Crescent, London W1N 4AL, England.

haustive account of possible models but to be guided by the results of X-ray diffraction to a particular choice.

It is pointed out by Miller & Parry (1973) that the row lines in the diffraction pattern index as the orders of a square unit cell. Their indexing scheme is not without difficulties [see Woodhead-Galloway, Hukins & Wray (1975) and §7 of this paper], but both for simplicity and because of the suggestiveness of the X-ray data, the model will, therefore, be one with a square unit cell.

4. Three dimensional molecular crystal model

Fig. 1(*c*) shows a square packing of molecules where the axial relations between neighbouring molecules are as defined in the Hodge–Petruska model, namely regular displacements by 1 *D*. Fig. 1(*d*) is the analogous model where the displacements are 2 *D*. In neither case is the unit cell square. However, if the two models are 'mixed' so that the relations along one set of planes are H–P and along perpendicular planes B–S, then the configuration of Fig. 1(*e*) emerges where the unit cell is square. In detail, the crystallography of the model is as follows. Intensity is confined to layer lines (l) spaced at 1/5 *D* and row lines which are the orders of square unit cell of side $\sqrt{5}\,a$ where a is the side of the lattice in projection on to the equator.

The interference function, IF, for the unit cell is given by

$$\begin{aligned}\mathrm{IF} = \tfrac{1}{4}[1 &+ \exp i(-akx + 2aky) + \exp i(2akx + 2aky)\\ &+ \exp i(akx + 3aky)] + \exp i(3Dkz + aky)\\ &+ \exp i(Dkz + 2aky)\\ &+ \exp i(4Dkz + akx + aky)\\ &+ \exp i(2Dkz + 2aky + akx).\end{aligned} \tag{4.1}$$

Sampling of this function takes place where

$$\begin{aligned} kz &= l\frac{2\pi}{5D}\\ kx^1 &= k\frac{2\pi}{\sqrt{5}a}\\ ky^1 &= h\frac{2\pi}{\sqrt{5}a}\end{aligned} \tag{4.2}$$

where

$$\begin{aligned} kx &= kx^1\cos\theta + ky^1\sin\theta\\ ky &= -kx^1\sin\theta + ky^1\cos\theta\end{aligned} \tag{4.3}$$

and

$$\theta = \sin^{-1}\frac{2}{\sqrt{5}} = \cos^{-1}\frac{1}{\sqrt{5}}$$

so that

$$\begin{aligned} kx &= (k + 2h)\frac{2\pi}{5a}\\ ky &= (h - 2k)\frac{2\pi}{5a}.\end{aligned} \tag{4.4}$$

Thus, equation (4.1) may be simplified to

$$\begin{aligned}\mathrm{IF} = 1 &+ \exp i(3l + h - 2k)\frac{2\pi}{5}\\ &+ \exp i(l + 2h - 4k)\frac{2\pi}{5} + \exp i(4l + 3h - k)\frac{2\pi}{5}\\ &+ \exp i(2l + 4h - 3k)\frac{2\pi}{5}\end{aligned} \tag{4.5}$$

where, after some rearrangement within the brackets, IF becomes*

$$\mathrm{e}^{-2i\psi} + \mathrm{e}^{-i\psi} + 1 + \mathrm{e}^{i\psi} + \mathrm{e}^{2i\psi} \tag{4.6}$$

and

$$\psi = (-2l + h - 2k)\frac{2\pi}{5}.$$

The systematic absences are, therefore, given by

$$\mathrm{e}^{5i\psi} = 1, \quad \psi \neq 0 \tag{4.7}$$

from which it follows that reflexions are present only when

$$(-2l + h - 2k) = 5m, \quad m \text{ integral}. \tag{4.8}$$

* Purely out of mathematical interest, it is worth noting that if the equation for the systematic absences is rewritten as

$$2\cos 2\psi + 2\cos\psi + 1 = 0,$$

the substitution $2\cos\psi = x$ produces the equation defining the Golden Section (*e.g.* Coxeter, 1969)

$$x^2 + x - 1 = 0.$$

Table 1. *Reflexions present on the first few layer lines (l), spaced at 1/5 D, for selected row lines for the molecular crystal model of this paper consistent with the selection rule $-2l + h - 2k = 5m$ (see text)*

h,k / l	0,0	1,0	2,1	3,2
0	(0,0)		(2,1) (1,$\bar{2}$) ($\bar{1}$,2) ($\bar{2}$,$\bar{1}$)	
1		0,$\bar{1}$	1,2	(3,$\bar{2}$) ($\bar{2}$,3)
2		$\bar{1}$,0	2,$\bar{1}$	(3,2) ($\bar{2}$,$\bar{3}$)
3		1,0	$\bar{2}$,1	(2,3) ($\bar{3}$,$\bar{2}$)
4		0,1	$\bar{1}$,$\bar{2}$	($\bar{3}$,2) (2,$\bar{3}$)
5	(0,0)		(2,1) (1,$\bar{2}$) ($\bar{1}$,2) ($\bar{2}$,$\bar{1}$)	

Table 2. *Calculated intensities for the reflexions of Table* 1

The molecules are treated as solid uniform cylinders of length 4·4 *D*. The layer lines (l) are spaced at 1/5 *D*. The row lines (h,k) are the orders of a square unit cell of side $2R\sqrt{5}$ where $2R$ is the molecular diameter (equal to, for this simple model, the intermolecular spacing, a).

h,k / l	0,0	1,0	1,1	2,0	2,1	2,2	3,0	3,1
0	10 000				330 (4)			93 (4)
1		106	60	15	6			~1
2		89	50	12	5			
3		71	40	10	4			
4		49	28	7	3			
5	47				~1 (4)			(4)
6		13	7	2				
7		4	2					

Table 1 gives some examples of the particular reflexions present in the diffraction pattern and Table 2 the intensities of a number of reflexions calculated for the simplest form of the model where the molecule is treated as a uniform cylinder of length $L = 4{\cdot}4D$ (D = 67 nm, L = 300 nm) and radius R where $R = a/2$. The molecular transform is therefore,

$$2\pi R\left(\frac{J_1 k^1 R}{k^1}\right) 2\sin\frac{kzL}{2}\Big/kz \qquad (4.9)$$

where $k^1 = (kx^2 + ky^2)^{1/2}$.

It is at once apparent that except for the origin, as would be expected, the strongest reflexions in the pattern are the 0,2 1 reflexions corresponding to the first row line of the equatorial projection of the lattice.

5. Non-tetragonal underlying lattices?

The point about this particular model is that in equatorial projection the unit cell appears very simple, merely a square packing of molecules. Such a square packing was assumed in §4 in the derivation of a true unit cell containing five molecular strands which is itself square. It seems not unreasonable to ask whether the squareness of the underlying lattice is necessary for the true unit cell to be square. The answer is that it is necessary for the case in point where the number of molecular strands is five. It is worth pointing out, although the point is rather trivial, that if the number of molecular strands is $n = H^2 + K^2$ where H,K are integers, then a square unit cell can always be built up on an underlying square lattice; thus, n = 2, 4, 5, 8, 9, 10, 13 . . . *etc.*

It is also possible occasionally to construct a square unit cell containing several strands when the underlying lattice is not square but rhomboid, but the number of strands is always large. The smallest value of n is 12 when a lattice with equal sides and an internal angle of ~67° allows the construction.

6. Relation of the model to X-ray scattering data and to other models

The other model for the fibril in which a square unit cell contains five molecular strands is that of the five-stranded microfibril (Smith, 1968; Miller & Wray, 1971; Miller & Parry, 1973) (see Fig. 2*b*). An examination of the equatorial X-ray scattering predicted by the two models appears immediately to rule out the Smith model; up to a point, however, it supports the present model but suggests that some elaboration is needed for a better account of the scattering.

In its simplest form, the microfibril model involves a square (unit-cell side $\simeq 3{\cdot}85$ nm) packing of five-stranded coiled-coils, each of which has the same orientation (Miller & Parry, 1973). Woodhead-Galloway (1976) has shown that in this case, the closest approach of microfibrils is $\xi_s = 2{\cdot}618 \ldots 2R$ where $2R$ is the molecular diameter (the molecule being thought of as a coiled uniform cylinder). ξ_s is then the side of the unit cell and in equatorial projection the predicted X-ray intensities are

$$\propto [J_0(1{\cdot}618k^1R)J_1(k^1R)/k^1R]^2$$

sampled at the orders of the square unit cell of side ξ_s. $k^1 = 2\pi/d$, of course, where d represents measurements in reciprocal space converted to nm say. Table 3(*a*) shows the distribution of equatorial intensity. A slightly more involved model, where alternate fibrils are rotated by π, implies a slight change in the side of the unit cell to $\xi_s = 2{\cdot}701 \ldots 2R$ (Miller & Parry, 1973; Woodhead-Galloway, 1976) but leaves the equatorial indexing unchanged. The intensities are slightly altered (Table 3*b*).

The strongest equatorial intensity in the experimental pattern is at $d \simeq 1{\cdot}25$ nm (*e.g.* Miller & Wray, 1971; Wray, 1972). This clearly then corresponds to, roughly speaking, the 3,0 or 3,1 reflexions for the unit cell described above. Yet, in this region, the intensity

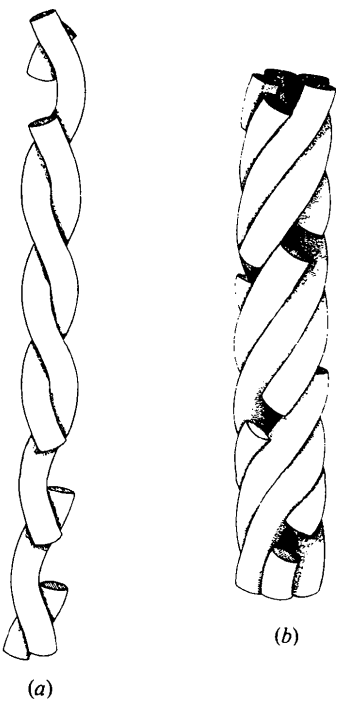

Fig. 2. (*a*) Two-standard coiled-coil (Woodhead-Galloway, Hukins & Wray, 1975). (*b*) Five-stranded coiled-coil (Smith, 1968; Miller & Wray, 1971).

predicted by the microfibril model is essentially zero. It seems most unlikely that the microfibril model can be correct.

The model derived in this paper is, in the simple form presented so far, inadequate to deal with the diffraction pattern in detail. However, if we limit attention for the moment to the strictly equatorial intensity, one correspondence with experiment can be seen. A square packing of molecules predicts that the strongest equatorial intensity is at $d \simeq 2R$. Thus, an immediate interpretation of the strong equatorial at $\sim 1{\cdot}25$ nm is that it is simply the intermolecular spacing $\simeq 2R$. Since here, both theory and the interpretation of experiment are limited to a consideration of structure in equatorial projection, a model with any distribution of axial displacements will predict this feature alone identically, for example the random-stagger model treated in a preliminary way by Woodhead-Galloway & Young (1977).

Table 3. *Calculated equatorial intensities for square packing of five-stranded microfibrils*

(*A*) All microfibrils have the same azimuthal relation. (*B*) Alternate microfibrils are rotated by π. For the purposes of accounting for the equator only, the microfibrils are treated as regular continuous five-stranded coiled-coils. The distance of approach of near neighbours is taken to be minimal, $2{\cdot}618 \ldots 2R$ in *A*; $2{\cdot}7013 \ldots 2R$ in *B* (see Woodhead-Galloway, 1976). Numbers in parentheses represent the number of reflexions on each row line.

h,k	0,0	0,1	1,1	2,0	2,1	2,2	3,0	3,1
A	10 000	460 (4)	125 (4)	305 (4)	139 (8)			
B	10 000	605 (4)	81 (4)	341 (4)	185 (8)	~1		

If the 1·25 nm strong equatorial is arising from the intermolecular spacing, then the position of the first row line predicted by the model is at $1{\cdot}25\sqrt{5}$ nm $\simeq$ 2·8 nm. Although there is a weak row line at about this spacing, the first row line observed in the pattern is at $d \simeq 3{\cdot}8$ nm ($\simeq 2{\cdot}8\sqrt{2}$). In addition to the equatorial reflexion alluded to above, however, there is also a rather weaker one at $d \simeq 1{\cdot}75$ nm $\simeq 1{\cdot}25\sqrt{2}$ which indexes as the 1,2 reflexion of a unit cell of side ~3·8 nm. Thus, both experiment and theory have the feature that the ratio of the spacing of the first

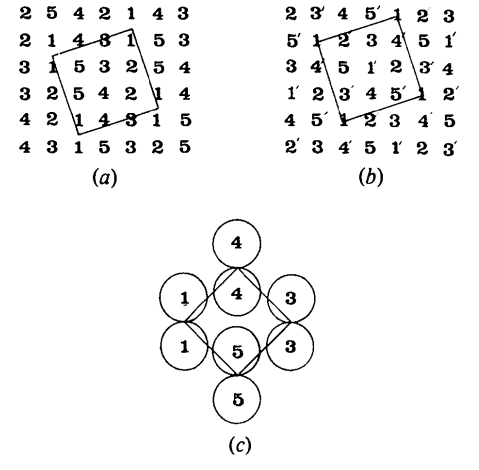

Fig. 3. Two methods of doubling the unit-cell size from five molecules to ten molecules. (*a*) Replacing each molecule by a dimer. (*b*) By 'alternation'. For the model of Woodhead-Galloway, Hukins & Wray, alternation is of supercoil between left and right hand. (*c*) Actual arrangement used in the calculation of Table 4 with $2R/a = 0{\cdot}9$.

Table 4. *Intensities of reflexions present on the first few layer lines (l) for the dimer model of Fig. 3(a) for* $2R/a = 0{\cdot}9$

Intensities less than 10 (scaled against the origin = 10 000) are neglected. The values *d* (experimental) are taken from Miller & Parry (1973). Although they record a 3,0 row line, it is extremely difficult to observe in the diffraction pattern. The model predicts scarcely any intensity on the 2,2 and 3,0 row lines. As a description of the intensity distribution *very* close to the equator it seems good. *E* denotes a truly equatorial reflexion in the experimental pattern.

h,k / *l*	0,0	1,0	1,1	2,0	2,1	2,2	3,0	3,1
0	10 000				(1,2) 81 ($\bar{1}$,$\bar{2}$) 81 (2,$\bar{1}$) 81 ($\bar{2}$,1) 81			(3,1) 720 ($\bar{3}$,$\bar{1}$) 720 (1,$\bar{3}$) 800 ($\bar{1}$,3) 800
1		(0,1) 133	($\bar{1}$,$\bar{1}$) 21	(2,0)	(1,$\bar{2}$) 54	($\bar{2}$,2)	($\bar{3}$,0) 12	(1,3)
2		($\bar{1}$,0) 55	(1,$\bar{1}$) 72	(0,2) 46	(2,1)	($\bar{2}$,$\bar{2}$)	(0,$\bar{3}$)	($\bar{3}$,1)
3		(1,0) 43	($\bar{1}$,1) 56	(0,$\bar{2}$) 36	($\bar{2}$,$\bar{1}$)	(2,2)	(0,3)	(3,$\bar{1}$)
4		(0,$\bar{1}$) 61	(1,1)	($\bar{2}$,0)	($\bar{1}$,2) 25	(2,$\bar{2}$)	(3,0)	($\bar{1}$,$\bar{3}$)
5	47				(1,2) ($\bar{1}$,$\bar{2}$) (2,$\bar{1}$) ($\bar{2}$,1)			(3,1) ($\bar{3}$,$\bar{1}$) (1,$\bar{3}$) ($\bar{1}$,3)
d (experimental) (nm)		3·8	2·65	1·89	1·75 *E*	1·37	?	1·26 *E*

equatorial to the first row line is $1:\sqrt{5}$ and that the first equatorial is the 2,1 reflexion of the square unit cell. This observation suggests a simple modification of the proposed model involving a doubling of the size of the unit cell so that the 1·25 nm strong equatorial spot represents the 3,1 reflexions.

7. Square unit cell with ten molecular strands

Fig. 3 shows two ways in which a simple doubling of the unit-cell size can be achieved. In Fig. 3(*a*), doubling is achieved by replacing the molecules by pairs of molecules; in Fig. 3(*b*), by alternation, allowing the 'molecules' to exist in two slightly different forms. Here this is a formal device only, but a combination of the two sorts of doubling where the dimer is a coiled-coil, allowed to exist with each of two hands, has been suggested as a model by Woodhead-Galloway *et al.* (1975). The combination of methods of doubling allowed a further doubling to a unit cell with 20 molecules. Attention here will be concentrated on the first method. Analysis exactly similar to that performed in the earlier part of the paper shows that the crystallographic features of such a model are (i) if the intermolecular spacing is a, the side of the true cell is $\sqrt{10}\,a$, and (ii) the systematic absences for the model are

$$\begin{aligned}&\text{(i) } h+k \text{ odd}, \quad 4l+k+3h = 1,3,7,9,11, \textit{etc.}\\ &\text{and (ii) } h+k \text{ even}, \quad 4l+k+3h = 2,4,6,8,12, \textit{etc.}\end{aligned} \tag{7.1}$$

A model of this sort meets two of the deficiencies of the earlier simple model. First, the size of the unit cell can be made approximately correct (but see §8), and second, in practice, a weak equatorial reflexion might be predicted as the 2,1 reflexions of the large unit cell. A simple calculation illustrates the point (see Fig. 3*c*). Again, the molecules are treated as uniform cylinders of radius R. If $a = 2R$, there is no equatorial at the 2,1 position (where $d = \sqrt{2}a = \sqrt{2}2R$). However, if $2R/a < 1$, approximating to a situation in which electron density tends to be concentrated more in the middle of the molecule, then a weak equatorial is predicted. Table 4 gives the distribution of intensities for reflexions on and near the equator for such a model where $2R/a = 0{\cdot}9$. As before, there are four reflexions present at (2,1) and (3,1).

These calculated intensities are intended to be illustrative only. A very much more refined approach would be needed for a real comparison with experiment. Nevertheless, the features of that part of the experimental pattern very close to the equator are recovered quite well; for example, the strongest reflexion after the 3,1 equatorials is present on the first (3·8 nm) row line. Without a detailed analysis of the diffraction data and rather more serious calculation, there is little point in pursuing this aspect of the problem further.

8. The off-equatorial intensity

The presence of strong off-equatorial intensity in the experimental pattern suggests that the molecules in the fibril are not wholly straight as has been presumed so far. For part of their length they must coil or tilt or do both. Supercoiling is invoked in the five-stranded microfibril model by Miller & Wray (1971) and straight tilting by Nemetschek & Hosemann (1973) and Hosemann, Dreissig & Nemetschek (1974) in their model. Supercoiling is an important feature of the two-stranded coiled-coil model of Woodhead-Galloway *et al.* (1975) but introduced to explain the equatorial intensity distribution, not the non-equatorial.

Since the model of this paper is one of the packing of essentially single, straight molecules, it seems consistent, initially, to suppose that molecules are straight for part of their length and tilted for part, and since the fibril (in the axial direction) consists of an alternation of 'gap' and 'overlap' regions, the simplest assumption is that the molecules are straight in the overlap region and tilted in the gap (or *vice versa* of course). Since the ratio of the density of molecules in the overlap region to those in the gap (Fig. 1*a*) is 5:4, it again seems arguable that tilting takes place in the low-density region, *i.e.* in the gap; this conjecture is supported by the following train of thought.

In §6, attention was concentrated particularly on the strong equatorial reflexion at $d \simeq 1{\cdot}25$ nm. The strongest off-equatorial reflexions whose intensity is similar to this equatorial one are observed at $d \simeq 1{\cdot}37$ nm, equi-spaced above and below the equator, corresponding closely to the position of 2,2 reflexions arising from a unit cell of side ~3·8 nm. The simplest explanation for this second reflexion is that it also corresponds to an intermolecular spacing. The numerology here is interesting; in the overlap region there are ten molecular strands in the unit cell; in the gap region there are only eight. If the supposition were made that in the gap region a square lattice is main-

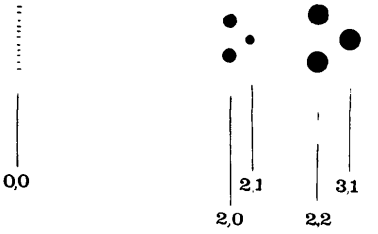

Fig. 4. Theoretical intensity distribution for the suggestion of §7, based on equatorial projections with a dimer model where $2R/a = 0{\cdot}9$. The 2,2 and 3,1 reflexions of a 3·8 nm square cell are not particularly sensitive to the value of $2R/a$ chosen. The 2,0 and 2,1 reflexions are sensitive. In particular, for the simple cylinder model used, the 2,1 reflexions disappear if $2R/a = 1$. 'Tilt' in the gap region is ~5°. Area of spots is proportional to the calculated intensity.

tained but that the molecular strands move apart so as to fill the available space uniformly, then the intermolecular spacing in the gap is not a, but rather a^1 where $a^1 = \sqrt{(5/4)}\, a$, and a strong reflexion would be expected at a^1. Further, the reflexion corresponding to this spacing would fall at the position of the 2,2 reflexions of the 3·8 nm square cell which corresponds very well to the indexing of this reflexion (Miller & Parry, 1973; Woodhead-Galloway *et al.*, 1975). If the molecules in the gap are also tilted, then the reflexion(s) would be tilted off the true equator. Fig. 4 shows the intensity distribution calculated on the basis of 'equatorial' projections alone, for the gap and overlap regions for a model where $2R/a = 0{\cdot}9$. By increasing this ratio, the intensity of the 2,0 and 2,1 reflexions may be decreased, although the intensity of the 2,2 and 3,1 reflexions is not very sensitive to the value of $2R/a$ chosen.

9. Summary of features of a model for the collagen fibril

(1) The analysis of the present paper leads to the conclusion that a unit cell which is (approximately?) square contains ten molecular strands in the overlap region and eight in the gap region. By contrast, the five-stranded microfibril model contains only five strands in the same size unit cell.

(2) The ten strands are arranged as five dimers in the overlap region and the eight strands as possibly four dimers in the gap region. The intensity calculations of the paper suggest that the dimers are straight in the overlap and tilted in the gap. Without detailed intensity calculations, the possibility that the dimers are supercoiled cannot be entirely ruled out. If they are supercoiled, then the arrangement must be with alternate two-stranded coiled coils having opposing hands (Woodhead-Galloway *et al.*, 1975; Woodhead-Galloway, 1976). Further it is worth pointing out that the word 'dimer' implies that the two constituent molecules are in register. Woodhead-Galloway *et al.* (1975) suggested two-stranded coiled-coils where molecules are related by a stagger of $1D$ (and $4D$) (see Fig. 2*a*). A further, and the only other, possibility is with the molecules staggered by $2D$ (and consequently $3D$). If the true unit cell has a side of 5·5 nm ($=\sqrt{2} \times 3{\cdot}8$ nm) (Miller & Parry, 1973; Woodhead-Galloway *et al.* 1975), then this is suggestive of supercoiling and alternation of hand.

(3) The detailed arrangement of dimers within the unit cell must be essentially as in Fig. 3(*a*) to account for the indexing of the equatorials and the distribution of intensity shown in Table 4. However, it is by no means obvious that such a precisely defined lateral arrangement is necessary in the model suggested in §7. The point requires clarification.

(4) The details of a model with 'straight' molecules in the overlap region and tilted molecules in the gap region have not yet been worked out. It seems likely, however, that such a model predicts that the row lines are split, one vertical arm arising from the overlap region and one tilted arm arising from the gap region [see also Nemetschek & Hosemann (1973) and Hosemann, Dreissig & Nemetschek (1974)], corresponding roughly to experimental observations.

(5) It is possible that the 'unit cells' in the gap and overlap regions differ slightly in size, and this difference may depend on the state of hydration of the tendon. If this is true, then it is very possible that drying reduces the size of the unit cell in the gap region with a consequent shift in the position of the 2,2 off-equatorial reflexion (see Fig. 4) towards smaller values of d.

(6) The presence of the α_2 chain with its preponderance of large hydrophobic residues may well favour in-register assembly of two molecules [although, see Hukins & Woodhead-Galloway (1976) for a different view]. It would be valuable now to seek the origins of square packing in the amino acid sequence. It may well be that square packing is facilitated by the prior assembly into dimers. For example, square packing is closest packing in the suggestion of Woodhead-Galloway *et al.* (1975). Even if the formation of some sort of dimer is the origin of the square packing, the arrangement of dimers in Fig. 3(*a*) is quite complicated, and its origins are likely to pose a rather subtle problem involving the pseudo D periodicity of the amino acid sequence (Hulmes, Miller, Parry, Piez & Woodhead-Galloway, 1973).

(7) The indexing of the experimental X-ray diffraction pattern row lines is not particularly good. Some errors in excess of 2% are found between the best estimated positions and those of experiment. This may be in part due to different unit cell sizes in the gap and overlap regions (see above) but might either or in addition be due to some distortion of the lattice from truly tetragonal. This point also requires further investigation.

10. Miscellaneous comments

(i) A numerological point regarding the relation of the disordered molecular model of Woodhead-Galloway & Machin (1976) to the considerations of this paper is worth noting. In this model, in the overlap region, there are ten molecules in a unit cell of side ~3·8 nm. Thus the number density of strands is $10/(3{\cdot}8)^2 \simeq 0{\cdot}69$ nm^{-2}. In the gap region, the corresponding number density is 0·55 nm^{-2} since there are only eight strands. This is the number density estimated by Woodhead-Galloway & Machin (1976) from the equatorial profile of the diffuse scattering, suggesting that disorder might be largely confined to the gap region. This observation is

suggestive, but the number density is difficult to estimate, and the point needs further work.

(ii) A strong triplet of reflexions is observed in the X-ray diffraction pattern of the silk of *Apis Mellifera* (Atkins, 1967), although it is 'reversed' – the off-equatorial reflexions are at smaller values of d than the truly equatorial one. Atkins's interpretation of his diffraction pattern is a packing of four stranded ropes. The presence of the triplet strongly suggests that the silk must contain structural features involving straight and tilted segments (and probably gap and overlap regions), as is proposed in this paper. It seems very likely that a model involving single molecules, or at most, dimers, will be found for that structure as well.

I am indebted to Dr A. Miller and Dr J. S. Wray, who suggested that I might work on collagen, and to Dr B. Brodsky, Dr D. J. S. Hulmes and Dr D. W. L. Hukins for discussion at various times. I was supported in Oxford by the Guinness Research Fellowship at New College and in Manchester by the Sir Henry Royce Research Fellowship. I am most grateful for both.

References

ATKINS, E. D. T. (1967). *J. Mol. Biol.* **24**, 139–141.
BEAR, R. S. & SELBY, C. C. (1956). *J. Biophys. Biochem. Cytol.* **2**, 55–69.
COXETER, H. S. M. (1969). *Introduction to Geometry*, 2nd ed. New York: John Wiley.
HODGE, J. A. & PETRUSKA, A. J. (1963). *Aspects of Protein Structure*, edited by G. N. RAMACHANDRAN, pp. 289–300. New York: Academic Press.
HOSEMANN, R., DREISSIG, W. & NEMETSCHEK, T. (1974). *J. Mol. Biol.* **83**, 275–280.
HUKINS, D. W. L. & WOODHEAD-GALLOWAY, J. (1976). *Biochem. Biophys. Res. Commun.* **70**, 413–417.
HULMES, D. J. S., MILLER, A., PARRY, D. A. D., PIEZ, K. A. & WOODHEAD-GALLOWAY, J. (1973). *J. Mol. Biol.* **79**, 137–148.
MILLER, A. & PARRY, D. A. D. (1973). *J. Mol. Biol.* **75**, 441–447.
MILLER, A. & WRAY, J. S. (1971). *Nature, Lond.* **230**, 437–439.
NEMETSCHEK, T. & HOSEMANN, R. (1973). *Kolloid Z.* **251**, 1044.
SMITH, J. W. (1968). *Nature, Lond.* **219**, 157–158.
WOODHEAD-GALLOWAY, J. (1976). *Acta Cryst.* B**32**, 1880–1884
WOODHEAD-GALLOWAY, J., HUKINS, D. W. L. & WRAY, J. S. (1975). *Biophys. Biochem. Res. Commun.* **64**, 1237–1244.
WOODHEAD-GALLOWAY, J. & MACHIN P. A. (1976). *Acta Cryst.* A**32**, 368–372.
WOODHEAD-GALLOWAY, J. & YOUNG, W. H. (1977). *Acta Cryst.* To be published.
WRAY, J. S. (1972). D.Phil. Thesis, Univ. of Oxford.

Acta Cryst. (1977). B**33**, 1218–1223

The Crystal and Molecular Structure of α-5-Acetyl-2′-deoxyuridine

BY T. A. HAMOR, M. K. O'LEARY AND R. T. WALKER

*Chemistry Department, The University, Birmingham B*15 2*TT, England*

(*Received* 6 *September* 1976; *accepted* 8 *October* 1976)

The title compound crystallizes in the orthorhombic space group $P2_12_12_1$ with $a = 16{\cdot}40\,(1)$, $b = 10{\cdot}34\,(1)$, $c = 7{\cdot}01\,(1)$ Å, $Z = 4$. The structure was refined to $R = 3{\cdot}95\%$ for 1535 observed counter amplitudes. The pyrimidine ring is essentially planar with the acetyl group inclined at 13° to it. There is evidence of conjugation between the pyrimidine and acetyl moieties. The sugar takes up the C(2′)-*endo* conformation and the arrangement about C(4′)–C(5′) is *trans–gauche*. The glycosidic torsion angle O(1′)–C(1′)–N(1)–C(6) is −12° (*anti* conformation), with the angle between the pyrimidine and sugar planes 76°.

Introduction

The group of analogues of the nucleic acid base thymine, the 5-substituted uracils, and the corresponding nucleosides have recently aroused considerable interest with regard to antiviral activity (Muraoka, Seto & Ueda, 1970; Muraoka, Takada & Ueda, 1970; Muraoka & Ueda, 1973) and also the possibility of replacement of thymine in bacteriophage DNA (Pietrzykowska & Shugar, 1967) and bacterial DNA (Piechowska & Shugar, 1965).

During the preparation of one such deoxynucleoside, 5-acetyl-2′-deoxyuridine, by a modified Hilbert–Johnson procedure, two anomers were obtained. It is