



HAL
open science

The transverse structure of collagen

Nicolas Rivier, J F Sadoc

► **To cite this version:**

Nicolas Rivier, J F Sadoc. The transverse structure of collagen. *Philosophical Magazine*, 2006, 86 (06-08), pp.1075-1083. 10.1080/14786430500355078 . hal-00513623

HAL Id: hal-00513623

<https://hal.science/hal-00513623>

Submitted on 1 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



The transverse structure of collagen

Journal:	<i>Philosophical Magazine & Philosophical Magazine Letters</i>
Manuscript ID:	TPHM-05-May-0266.R1
Journal Selection:	Philosophical Magazine
Date Submitted by the Author:	01-Aug-2005
Complete List of Authors:	Rivier, Nicolas; ACI Biofrustration; Université Louis Pasteur, IPCMS Sadoc, J F; ACI Biofrustration; Université Paris Sud, Laboratoire de Physique des Solides
Keywords:	structure, grain boundaries, quasicrystalline alloys, biological materials, approximants, proteins, modelling, geometry
Keywords (user supplied):	collagen, close-packing, moiré



The transverse structure of collagen

N. RIVIER^{1,2}, J.F. SADOE^{1,3}

¹*ACI Biofrustration*

²*IPCMS, Université Louis Pasteur, 67084 Strasbourg Cedex, France*

³*Laboratoire de Physique des Solides, Université Paris Sud, 91405 Orsay, France*

ABSTRACT

Collagen is the principal constituent of extra-cellular, connective tissue. It is a tightly-packed but flexible bundle of proteins that constitutes a material. The collagen molecule is a triple helix made of three, intertwined polypeptide chains, associated with $\sqrt{3}$ and its rational convergents. Accordingly, the collagen fibril, a long, periodic bundle of finite collagen molecules, has a transverse structure based on two underlying, perpendicular hexagonal lattices. It is the Archimedean, square-triangle tiling $3^2.4.3.4$, in both "overlap" and "gap" regions, with small orthorhombic distortions imposed as rational approximants of the two perpendicular lattices. The three-dimensional structure is a periodic, rotating stack of gap and overlap regions, with a ten-fold rotation symmetry.

Keywords: collagen, twist grain boundary, close-packing, moiré

1. COLLAGEN: CHAIN, MOLECULE, FIBRIL

Collagen is the principal constituent of extra-cellular, connective tissue. It is made of fibrils, that are close-packed bundles of long molecules. Each molecule consists of three intertwined polypeptide chains, forming a right-handed helix (Fig.1b.c). Each chain is a nearly periodic sequence of amino acids (a.a) $\dots-[Gly-X-Y]\dots$, where the a.a X and Y are predominantly *Pro* or *HPro*. The collagen chain is a protein with a periodic (in *Gly*) primary structure and a left-handed helix that is altogether its sole secondary structure and its ternary structure (Fig.1a). Three chains join as a molecule, and the molecules assemble into a material.

The collagen molecules assemble into fibrils. Longitudinally, the molecules are separated by gaps, and the fibril is a periodic alternance of *overlap* and *gap* regions, as indicated in Fig.1d. The transverse structure is topologically, in both regions, the Archimedean square-triangle tiling $3^2.4.3.4$, also known as the main skeleton of the Frank-Kasper sigma phase, and observed by Okuyama et al. [1] in a single crystal of $[Gly-Pro-Pro]_{10}$. The three-dimensional structure of native collagen is a rotating stack of successive overlap-gap-overlap..., separated by boundaries of twist dislocations. The stack is periodic, as is observed.

The essential physical and geometrical features to be included in the structure of the collagen fibril are:

- a) Close-packing of amino acids in a bundle of periodic, polypeptide chains.

b) Flexibility of the fibril, compatible with close-packing. (Long, cylindrical molecules pack tightest as an hexagonal lattice. But such a bundle would be completely rigid. Flexibility requires that i) each molecule has five nearest neighbours instead of six, and that ii) there is a gap between two molecules stacked on top of each other (Fig.1e.)

c) Fibrils of arbitrary length.

d) Intertwining of the chains within the molecule at a constant rate.

Points c) and d) suggest a pattern of molecules along the fibril, and of amino acids within the collagen molecule, that is either periodic or inflationary (quasicrystalline). Diffraction, [1,2,4] indicates that the pattern is periodic, but a chain of amino acids packs naturally as a Boerdijk-Coxeter helix, that contains several approximants of $1+\sqrt{3}$ [5]. These observations are not incompatible. The (longitudinal and transverse) structure of collagen is ultimately periodic, but its unit cell exhibits many successive approximants of $\sqrt{3}$.

2. THE BOERDIJK-COXETER HELIX AND THE COLLAGEN CHAIN

Helices, and dense packing of spherical objects are two closely related geometrical problems. The simplest means of packing tightly a chain of connected spheres (representing amino acids) of arbitrary length is the Boerdijk-Coxeter (B-C) « helix » [5]. It is a stacking of regular tetrahedra, the elementary unit of four close-packed spheres. The stacking can be viewed as an hexagonal lattice (triangular tiling, unit vectors $\mathbf{k} = \mathbf{l} + \mathbf{m}$) on a cylinder. Three families (k, l, m) of helical chains of connected vertices, of different pitch and chirality, are winding around the cylinder. (The natural numbers $k = l + m \geq l \geq m \geq 0$ give the number of helical chains in each family. The B-C helix is (3,2,1) in this – phyllotactic – notation). The cylinder is cut along its axis, the lattice vector \mathbf{b} , and flattened. Figure 2 represents the collagen chain (one of the $l=2$ left-handed helices along \mathbf{l}), but the flattest chain along \mathbf{m} and the three chains along \mathbf{k} joining the *Gly*, *X* and *Y*, are visible. The base of the cylinder is the lattice vector $\mathbf{a} = \mathbf{lm} - \mathbf{ml} = 2\mathbf{m} - \mathbf{l}$.

The regular tetrahedron is not a three-dimensional space-filler. Perfect, close-packing of spheres is only possible in positively curved space [6], where the three helices along \mathbf{k} in Fig. 2 are great circles with 10 vertices each. Moreover, as a one-dimensional structure in Euclidean space, the B-C helix is not periodic, not even quasiperiodic and thus extensible through a sequence of approximants : The number of vertices per turn (rate of winding) of the flattest chain is $2\pi/[\cos^{-1}(-2/3)]$, that is neither a rational number describing a periodic structure, extensible by repetition of the unit, nor a quadratic irrational describing a structure with constant winding rate, extensible by inflation. But the chain with quadratic irrational winding rate $1+\sqrt{3}$ has exactly the same rational approximants through the first 112 vertices and 41 turns ; notably 30/11, represented in Fig. 2.

The cylinder can be unwound on an hexagonal lattice, where the various helical chains are lattice vectors along the directions \mathbf{m} , \mathbf{l} and \mathbf{k} : The flattest helix $A\mathbf{m} = w\mathbf{a} + \mathbf{b}$ has winding rate A/w . The left-handed collagen chain $A'\mathbf{l} = w'\mathbf{a} + \mathbf{b}$ has winding rate A'/w' . The steepest helix winds once around the cylinder : $(w-1)\mathbf{mk} = \mathbf{a} + \mathbf{b}$. Here, the natural integers A , A' , $(w-1)m$ are the numbers of vertices on each helix, and w , w' and 1 are their winding number. All the possibilities are given by the natural integer s , $A = slk$, $A' = smk$, $w = sl + 1$, $w' = sm - 1$, and the cylinder contains $smlk$ vertices ($\mathbf{a} \times \mathbf{b} = smlk \mathbf{m} \times \mathbf{k}$).

The B-C helix and the collagen chain has $(k,l,m) = (3,2,1)$; A/w , an approximant of $1+\sqrt{3}$, has denominator 1, 3, 4, 11, 15, 41, 56, But its winding number $w = 2s + 1$ must be odd ≥ 5 , since $w' = s - 1 \geq 1$, so that, the collagen chain has either $s = 5$, $w = 11$, $A'/w' = 15/4$ (Fig. 2), $A/w = 30/11$, or $s = 7$, $w = 15$, $A'/w' = 21/6 = 7/2$, $A/w = 42/15$ (one vertex must be

1
2
3 added to the next approximant $41/15$ of $1+\sqrt{3}$, to complete the periode of the structure : it is
4 on the collagen chain $21/6$ along **l**, and on the steepest helix $14/1$ along **k**). Moreover, the
5 underlying cylinder is perpendicular to its base, $\mathbf{b}\cdot\mathbf{a} = 0$. The helix $7/2$ ($21/6$ for chemical
6 periodicity), periodic in Euclidean space, is thus the basic helix of collagen (identified by [1]).
7

8 The three chains intertwine to constitute the collagen molecule, a triple helix with a
9 core consisting of the side-groups of the Gly, a B-C helix of H atoms [5].
10

11 12 13 14 3. TRANSVERSE STRUCTURES (GAP, OVERLAP)

15
16 The geometric and numerical interdependence between close-packing, tiling by
17 equilateral triangles, $\sqrt{3}$ and its rational approximants is central to the structure of collagen.
18 The transverse structure of collagen (the packing of the molecules across the fibril of Fig. 1f)
19 is also based on approximants of $1+\sqrt{3}$. Lattices of approximants of $\sqrt{3}$ are readily obtained
20 by superposition of two hexagonal lattices at 90° . One obtains a moiré pattern of coincidence
21 points (Fig.3) which consists of four copies of the gap structure, shown in Fig. 4a. A vertex
22 now represents a collagen molecule. The gap structure is, topologically, the Archimedean
23 square-triangle tiling $3^2.4.3.4$. The axes of the unit cell and the diagonals of the « squares »
24 mark the 6 reticular directions of the two underlying hexagonal lattices. All edges have equal
25 length, but there is a slight distortion (the unit cell is 1% rectangular and the squares are 8%
26 rhombi) imposed by $\sqrt{3}$ (12-fold quasicrystallography). This is the structure observed by
27 Okuyama et al. [1] in a single crystal of $[Gly-Pro-Pro]_{10}$, with an orthorhombic distortion of
28 2%.
29

30
31 Vertex coordination of 5 makes the fibril flexible, provided that there are gaps
32 between molecules stacked on top of each other (Fig.1e), so that the transverse structure is
33 alternatively overlap and gap as one moves along the fibril. It has been established [7] that out
34 of five molecules in the overlap, only four extend into the gap. Bouligand [2] has suggested
35 that gap and overlap have the same square-triangle tiling but at two different scales, and
36 rotated by a topological angle of $\tan^{-1}(1/2)$. He has backed his suggestion with diffraction
37 data, and has given an elegant model (Fig.5) of the gap-overlap interface (a twist grain
38 boundary in the fibril) and of the mechanism whereby the termination of one molecule out of
39 five in the overlap leaves the square-triangle tiling topologically invariant (c to a).
40

41 From the gap structure, it is straightforward to construct the unit cell of the overlap
42 structure on the underlying hexagonal lattice (Fig.4b), given the ratio between the areas of the
43 unit cells ($112/90$, half a triangle in the gap short of $5/4$), and the topological angle of rotation
44 ($\tan^{-1}(1/2)$ seen in Fig.6 and in Fig.5b). The roles of the slightly distorted structures are
45 exchanged : the unit cell is 8% rhombic and the squares are 15% rectangular. It is now the
46 edges (no longer all equal) and the diagonals of the rhombi that are along 6 reticular
47 directions. The overlap structure has two perpendicular mirror axes defining a rectangular unit
48 cell (they bound a fundamental domain). Several points are aligned and regularly spaced : the
49 diagonal of the square is a third of the distance to points (2,-1) or (-1,2) of the unit cell grid.
50 Finally, a similar overlap structure ($\sqrt{3}$ smaller) can be drawn on the same underlying lattice
51 (rotated by 90°) with the «square » of the gap as unit cell.
52

53 The Table below lists the geometric manifestations (square and unit cell of the $3^2.4.3.4$
54 square-triangle tiling) of the various approximants to $\sqrt{3}$ in the overlap and gap structures [8].
55 The vertices are joined by lattice vectors in both underlying hexagonal lattices. If two vertices
56 lie on one of the 6 reticular directions (symmetry axis) of the underlying lattices, an
57 approximant is obtained from their distance measured on both lattices (the sides of a rectangle
58 [...]) on one lattice). If the vertices do not lie on a symmetry axis, they constitute, on one
59
60

lattice, an equilateral rhombus, with nearly equal diagonals on symmetry axes and nearly orthogonal side vectors. This yields two approximants (a) and (b). The last column gives the (orthorhombic or monoclinic) distortion, namely the ratio between $\sqrt{3}$ and its approximant.

j	$\sqrt{3} \approx$	overlap	gap	distortion
1	2/1	[square] unit cell (a)	square (a)	1.155
2	5/3	[mirror unit cell]		1.04
3	7/4	unit cell (b)	square (b) [unit cell]	1.01

For reference:

4	19/11	BC helix
5	26/15	inflated BC helix

The ratio in sizes of the two structures is 1.115, smaller than the topological value of $\sqrt{5/4} = 1.118$, but larger than the experimental ratio of 1.08 [2]. This discrepancy remains unexplained.

Fig.6 shows how the overlap and gap structures match as a twist grain boundary with an important coincidence site lattice, containing points $n(2,-1)$ or $n(-1,2)$ of the overlap unit cell grid and points $n(2,0)$ of the gap unit cell grid (n integer), thereby measuring the angle of rotation in Bouligand's mechanism (Fig.5). The overlap supercell containing 20 vertices and 5 unit cells, with sides $(2,-1)$ and $(-1,-2)$, fits almost exactly with the square of 2×2 gap unit cells, that contains 16 vertices. The rotation angle of $26.996^\circ = \alpha + \gamma$ is the metric equivalent of the topological rotation of $\tan^{-1}(1/2)$. (Direction $(1,0)$ of the overlap grid makes an angle $\alpha = \tan^{-1}(\sqrt{3}/6)$ with the underlying hexagonal lattice ; just as $\beta = \tan^{-1}(\sqrt{3}/9)$ for direction $(2,-1)$, that is also the diagonal of the square, or direction $(1,0)$ in the gap grid.) The rotation between two successive overlap structures across the gap, is double, namely $3\pi/10$. Note that in the unit cell of the rotated structure, the small and large rhombi have exchanged positions. Consequently, the angle of the next rotation is the complementary (to $2\pi/3$) of $3\pi/10$. The collagen fibril, a periodic stack overlap-gap-overlap-gap-..., exhibits both 10 and 12-fold rotations.

The flexibility of a long, tightly packed collagen fibril, regardless of its length or girth, is the effect of the regular alternance of gap and overlap regions with five neighbors to each molecule, that are the main features of our model.

Acknowledgments: The authors acknowledge the financial support of the French Ministère délégué à la Recherche (Action Concertée Incitative 2000), through the ACI Biofrustration. We are indebted to Y. Bouligand for impulsing this work, enlightening discussions, helpful suggestions, and for Figs. 1 and 5.

References

- 1
2
3
4
5 [1] Ke. Okuyama, Ka. Okuyama, S. Arnott, M. Takayanagi, M. Kakudo, 1981, J. Mol. Biol.
6 **152** 427.
7
8
9 [2] Y. Bouligand, 1997, *L'assemblage compact des triples hélices de collagène*, JMC5, poster
10 2108, Soc. Franç. Phys. Orléans 1997.
11 Y. Bouligand, 2001, *Two superposed fibre diagrams in X-ray diffraction patterns of collagen*
12 *tendons: The square-lozenge model*, Unpublished.
13
14
15 [3] Y. Bouligand, J.P. Denèfle, J.P. Lechère, M. Maillard, 1985, Biol. Cell **54** 143.
16
17
18 [4] A. Rich and F.R.C. Crick, 1961, J. Mol. Biol., **3** 483.
19
20
21 [5] J.F. Sadoc and N. Rivier, 1999, Eur. Phys. J. B **12** 309 ; 2000, Mat. Sci. Eng. **294-296**,
22 397.
23
24 [6] J.F. Sadoc and R. Mosseri, 1999, *Geometrical Frustration*, Cambridge Univ. Press.
25
26
27 [7] J.A. Hodge and A.J. Petruska, 1963, in *Aspects of Protein Structure*, G.N. Ramachandran,
28 ed., 289, Acad. Press, London.
29
30
31 [8] N. Rivier and J.F. Sadoc, 2005, in *Topology in Molecular Biology*, M.I. Monastyrsky, ed.,
32 Springer, to appear.
33
34

35
36 Figure captions

37
38 _____ Fig.1 _____

39 Structure of collagen at various scales (from Bouligand [2,3]). a) A single chain *Gly-X-Y*
40 forms a left-handed helix, with a curved axis. Dots indicate the positions of successive amino
41 acids. Their side-groups are pointing outwards; every third one is a glycine (gly) located on
42 the concave side of the axis. b) A collagen molecule is a right-handed triple helix of 3
43 intertwined, left-handed polypeptide chains. c) A simpler representation of the right-handed
44 triple helix. Its width is 1 nm, its length, 300 nm. d) The collagen molecule is often
45 represented by an arrow of length 300 nm. e) The molecules are stacked on top of each other,
46 with a gap of 35 nm, and regularly spaced on a lattice (transverse scale dilated relatively to
47 longitudinal scale). The longitudinal coordinates of molecules *u* and *u'* are the same. One
48 distinguishes gap and overlap levels, *g* and *o* respectively, so that to each five molecules in *o*
49 correspond only four in *g*. There is a regular stagger of molecules by a length $D = g + o = 67$
50 nm. f) A crude representation of the lattice as a cylinder (by identifying *u* and *u'*). This is only
51 schematic as the collagen fibril is a lattice, showing the 67 nm stagger, but with a unit cell that
52 accommodates gap and overlap levels. Notice the right-handed chirality of the *Gly* core and of
53 the triple-helix collagen molecule, opposite to the left-handed chirality of the single collagen
54 chain (Fig. 2).
55
56
57
58
59
60

 Fig.2

The B-C helix represented as a hexagonal lattice on a cylinder, cut along its axis (vertical gray lines) and flattened. The single collagen chain is a left-handed helix winding in the \mathbf{l} direction. The axis $Gly-X$ of helix $7/2$ is perpendicular to the base, so that a stack of three segments $7/2$ constitutes one period of the collagen chain in Euclidean space.

%(should be put on the right of Fig.1)%

 Fig.3

Superposition of two hexagonal lattices at 90° yields a moiré pattern of coincidence points. By construction, they are the rational approximants of $\sqrt{3}$. The brightest points (corners of a square unit cell, with eight inner points) constitute four, translated copies of the gap structure, shown in Fig. 4a. The corners of the square unit cell belong to all four copies, the corners of the central square, to one copy each, and the corners of the intermediate square tilted at 45° , to two copies. Alternatively, one can view the square cell as decorated with (1/4 of) a dodecagonal wheel around each corner. Each copy of the gap structure contributes five points to the wheel.

%%[[[The frame and the horizontal and vertical scales can be removed from Fig.3]]]]%

 Fig.4

The unit cells of the square-triangle $3^2.4.3.4$ transverse structures of the collagen, drawn on an hexagonal lattice. A vertex represents the collagen molecule. a) Gap. b) Overlap. c) A smaller overlap structure, with the square of the gap structure as unit cell, drawn on the hexagonal lattice rotated by 90° . It is similar to b), but smaller by $\sqrt{3}$. Also drawn (grey) is the associated Voronoi tiling,

%(should be put on the right of Fig.3)%

 Fig.5

The gap-overlap interface in a crystalline collagen fibril (schematic) (from Bouligand [2]). Gap (a) and overlap (c) structures are topologically identical, square-triangle patterns. (b) indicates how the gap structure deforms to intercalate new molecules (o) in the overlap, or, reciprocally, how one molecule (o) out of every 5 in the overlap terminates leaving the pattern invariant.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

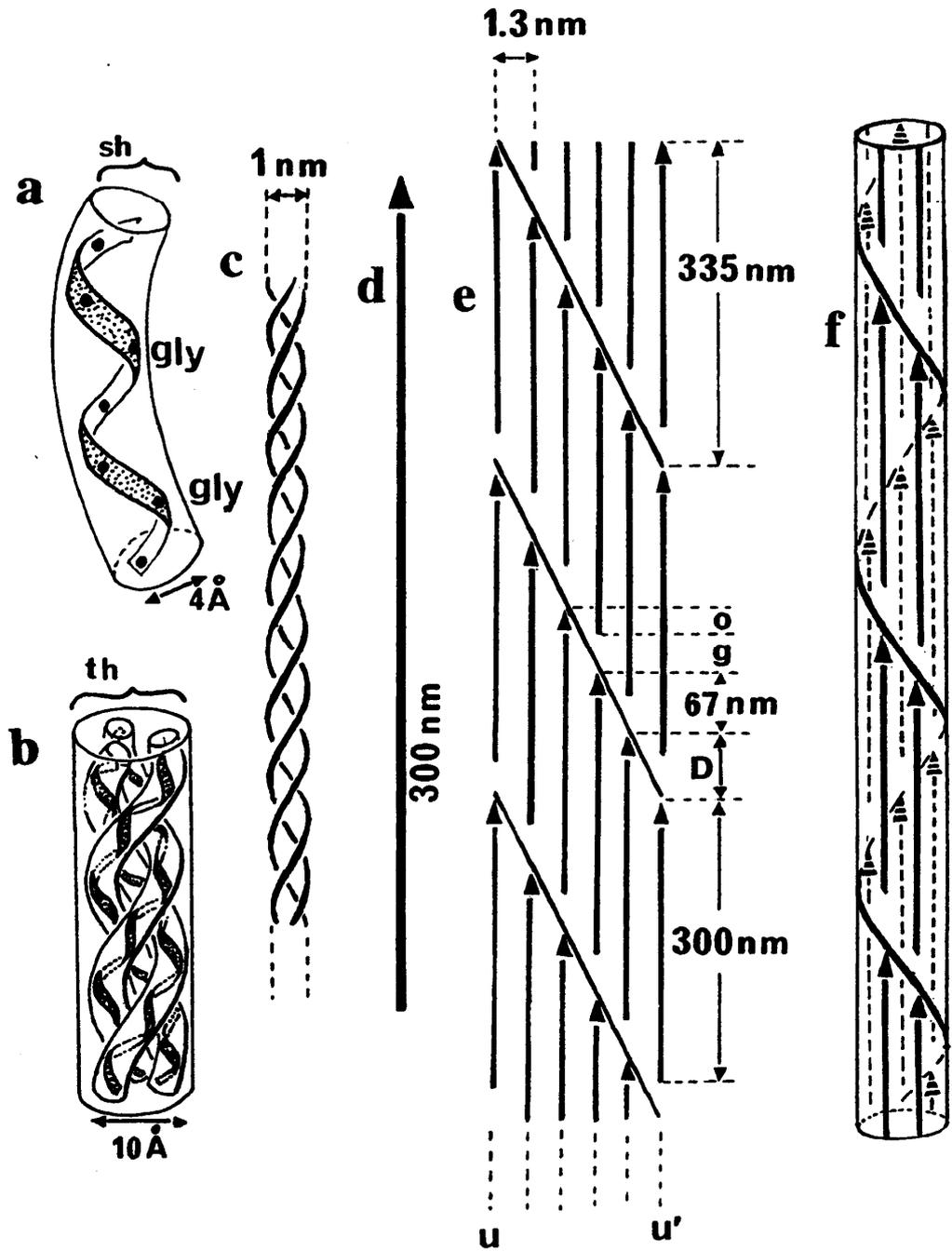
Fig.6

Two successive twist grain boundaries overlap-gap-overlap. Note the 5/4 ratio between the areas of the overlap and gap unit cells, and the coincidence site lattice of equidistant, aligned vertices in the two overlap structures, rotated by $3\pi/10$.

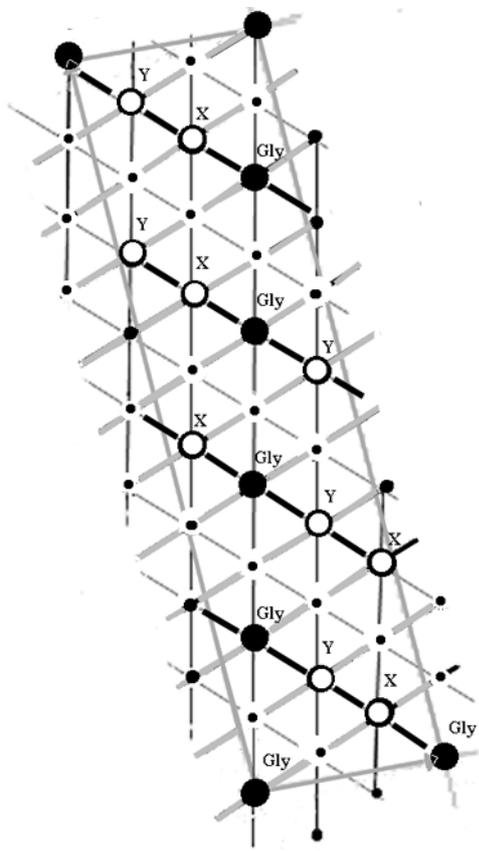
%Fig.5 and Fig.6 (can be put together)%

The transverse structure of collagen

N. RIVIER^{1,2}, J.F. SADOE^{1,3}

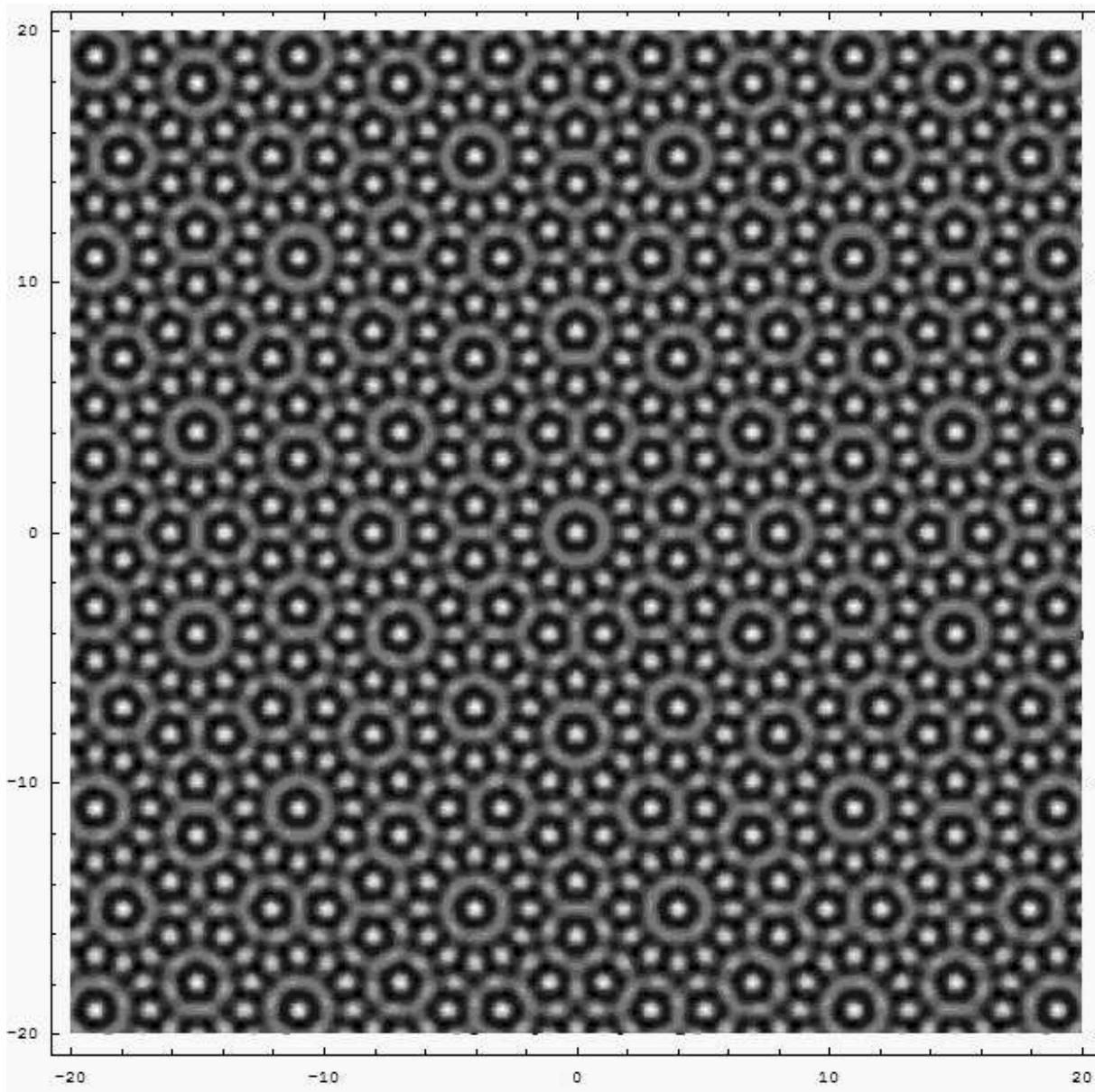


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



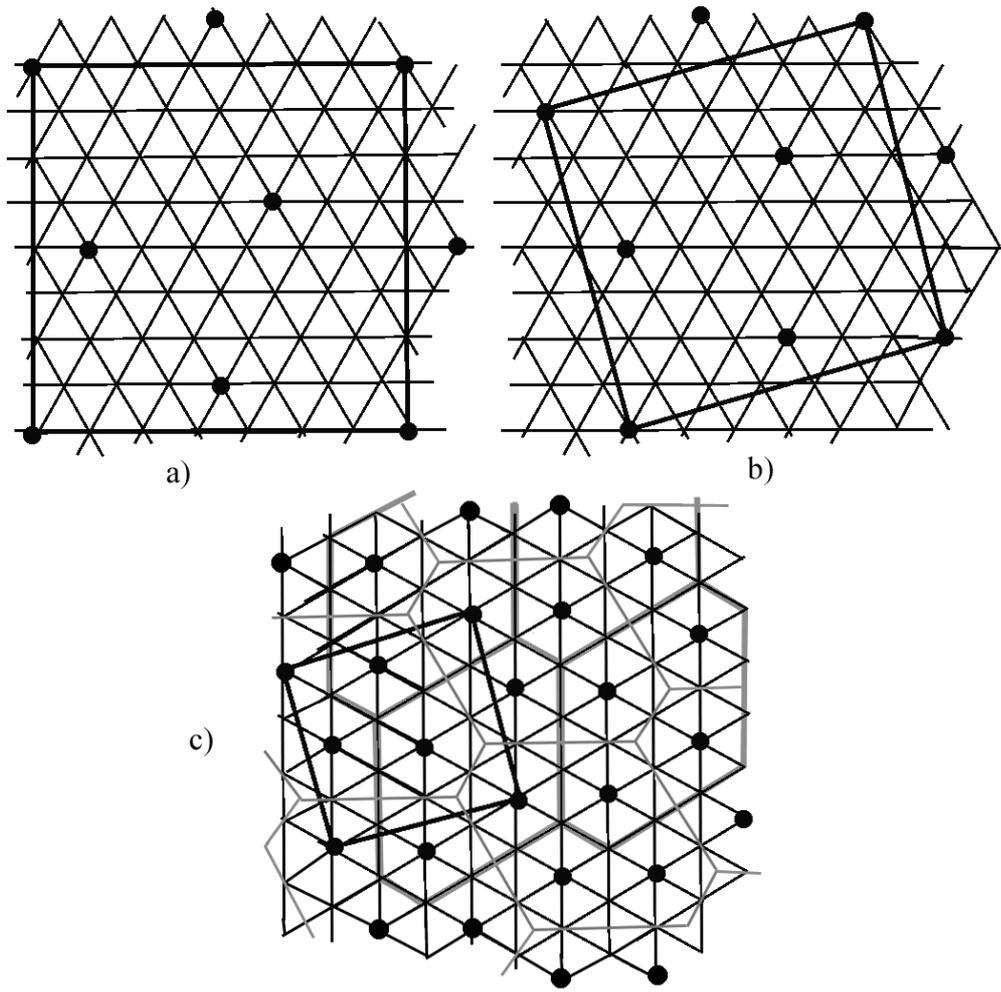
FC
ew Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

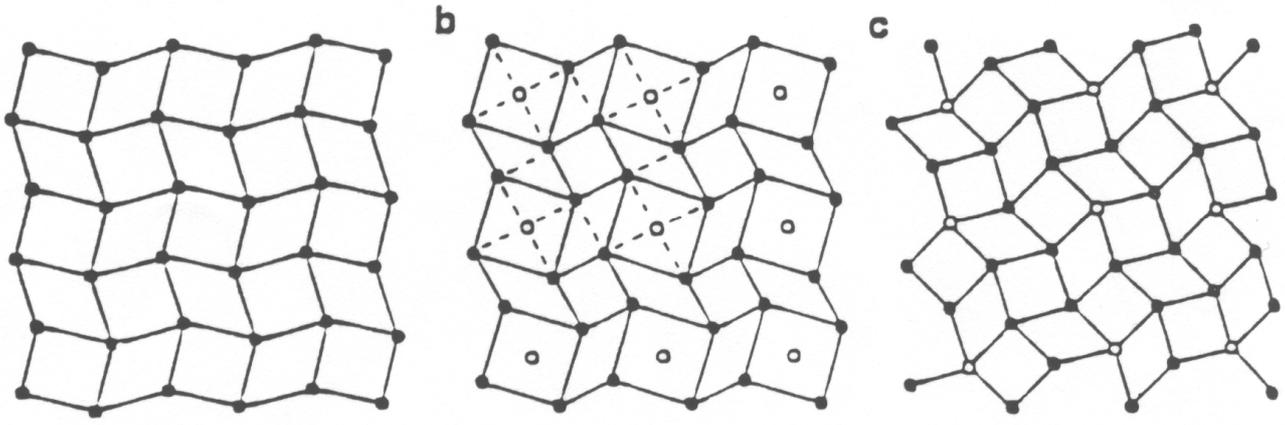


Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

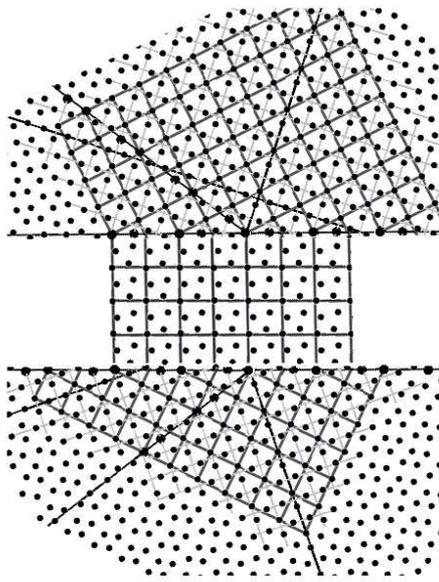


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

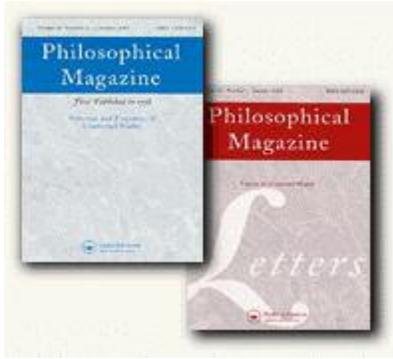


er Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For Peer Review Only



see file (i am trying to send you a ciorrected file)

Journal:	<i>Philosophical Magazine & Philosophical Magazine Letters</i>
Manuscript ID:	draft
Journal Selection:	Philosophical Magazine
Date Submitted by the Author:	n/a
Complete List of Authors:	Rivier, Nicolas; Universite Louis Pasteur, IPCMS
Keywords:	grain boundaries, structure
Keywords (user supplied):	

powered by ScholarOne
Manuscript Central™

The transverse structure of collagen

N. RIVIER^{1,2}, J.F. SADOE^{1,3}

¹ACI Biofrustration

²IPCMS, Université Louis Pasteur, 67084 Strasbourg Cedex, France

³Laboratoire de Physique des Solides, Université Paris Sud, 91405 Orsay, France

ABSTRACT

Collagen is the principal constituent of extra-cellular, connective tissue. **It is a tightly-packed but flexible bundle of proteins that constitutes a material.** The collagen molecule is a triple helix made of three, intertwined polypeptide chains, associated with $\sqrt{3}$ and its rational convergents. Accordingly, the collagen fibril, a long, periodic bundle of finite collagen molecules, has a transverse structure based on two underlying, perpendicular hexagonal lattices. It is the Archimedean, square-triangle tiling $3^2.4.3.4$, in both "overlap" and "gap" regions, with small orthorhombic distortions imposed as rational approximants of the two perpendicular lattices. The three-dimensional structure is a periodic, rotating stack of gap and overlap regions, with a ten-fold rotation symmetry.

Keywords: collagen, twist grain boundary, close-packing, moiré

1. COLLAGEN: CHAIN, MOLECULE, FIBRIL

Collagen is the principal constituent of extra-cellular, connective tissue. It is made of fibrils, that are close-packed bundles of long molecules. Each molecule consists of three intertwined polypeptide chains, forming a right-handed helix (Fig.1b.c). Each chain is a nearly periodic sequence of amino acids (a.a)-[Gly-X-Y]-..., where the a.a *X* and *Y* are predominantly *Pro* or *HPro*. **The collagen chain is a protein with a periodic (in *Gly*) primary structure and a left-handed helix that is altogether its sole secondary structure and its ternary structure (Fig.1a). Three chains join as a molecule, and the molecules assemble into a material.**

The collagen molecules assemble into fibrils. Longitudinally, the molecules are separated by gaps, and the fibril is a periodic alternance of *overlap* and *gap* regions, as indicated in Fig.1d. The transverse structure is topologically, in both regions, the Archimedean square-triangle tiling $3^2.4.3.4$, also known as the main skeleton of the Frank-Kasper sigma phase, and observed by Okuyama et al. [1] in a single crystal of [Gly-Pro-Pro]₁₀. The three-dimensional structure of native collagen is a rotating stack of successive overlap-gap-overlap..., separated by boundaries of twist dislocations. The stack is periodic, as is observed.

The essential physical and geometrical features to be included in the structure of the collagen fibril are:

- a) Close-packing of amino acids in a bundle of periodic, polypeptide chains.

1
2
3 b) Flexibility of the fibril, compatible with close-packing. (Long, cylindrical
4 molecules pack tightest as an hexagonal lattice. But such a bundle would be completely rigid.
5 Flexibility requires that i) each molecule has five nearest neighbours instead of six, and that
6 ii) there is a gap between two molecules stacked on top of each other (Fig.1e.)

7
8 c) Fibrils of arbitrary length.

9 d) Intertwining of the chains within the molecule at a constant rate.

10 Points c) and d) suggest a pattern of molecules along the fibril, and of amino acids within the
11 collagen molecule, that is either periodic or inflationary (quasicrystalline). Diffraction, [1,2,4]
12 indicates that the pattern is periodic, but a chain of amino acids packs naturally as a Boerdijk-
13 Coxeter helix, that contains several approximants of $1+\sqrt{3}$ [5]. These observations are not
14 incompatible. The (longitudinal and transverse) structure of collagen is ultimately periodic,
15 but its unit cell exhibits many successive approximants of $\sqrt{3}$.
16
17

20 21 2. THE BOERDIJK-COXETER HELIX AND THE COLLAGEN CHAIN

22
23 Helices, and dense packing of spherical objects are two closely related geometrical
24 problems. The simplest means of packing tightly a chain of connected spheres (representing
25 amino acids) of arbitrary length is the Boerdijk-Coxeter (B-C) « helix » [5]. It is a stacking of
26 regular tetrahedra, the elementary unit of four close-packed spheres. The stacking can be
27 viewed as an hexagonal lattice (triangular tiling, unit vectors $\mathbf{k} = \mathbf{l} + \mathbf{m}$) on a cylinder. Three
28 families (k, l, m) of helical chains of connected vertices, of different pitch and chirality, are
29 winding around the cylinder. (The natural numbers $k = l + m \geq l \geq m \geq 0$ give the number of
30 helical chains in each family. The B-C helix is (3,2,1) in this – phyllotactic – notation). The
31 cylinder is cut along its axis, the lattice vector \mathbf{b} , and flattened. Figure 2 represents the
32 collagen chain (one of the $l=2$ left-handed helices along \mathbf{l}), but the flattest chain along \mathbf{m} and
33 the three chains along \mathbf{k} joining the *Gly*, *X* and *Y*, are visible. The base of the cylinder is the
34 lattice vector $\mathbf{a} = \mathbf{lm} - \mathbf{ml} = 2\mathbf{m} - \mathbf{l}$.
35
36

37
38 The regular tetrahedron is not a three-dimensional space-filler. Perfect, close-packing
39 of spheres is only possible in positively curved space [6], where the three helices along \mathbf{k} in
40 Fig. 2 are great circles with 10 vertices each. Moreover, as a one-dimensional structure in
41 Euclidean space, the B-C helix is not periodic, not even quasiperiodic and thus extensible
42 through a sequence of approximants : The number of vertices per turn (rate of winding) of the
43 flattest chain is $2\pi/[\cos^{-1}(-2/3)]$, that is neither a rational number describing a periodic
44 structure, extensible by repetition of the unit, nor a quadratic irrational describing a structure
45 with constant winding rate, extensible by inflation. But the chain with quadratic irrational
46 winding rate $1+\sqrt{3}$ has exactly the same rational approximants through the first 112 vertices
47 and 41 turns ; notably 30/11, represented in Fig. 2.
48

49
50 The cylinder can be unwound on an hexagonal lattice, where the various helical chains
51 are lattice vectors along the directions \mathbf{m} , \mathbf{l} and \mathbf{k} : The flattest helix $A\mathbf{m} = w\mathbf{a} + \mathbf{b}$ has
52 winding rate A/w . The left-handed collagen chain $A'\mathbf{l} = w'\mathbf{a} + \mathbf{b}$ has winding rate A'/w' . The
53 steepest helix winds once around the cylinder : $(w-1)\mathbf{mk} = \mathbf{a} + \mathbf{b}$. Here, the natural integers A ,
54 A' , $(w-1)m$ are the numbers of vertices on each helix, and w , w' and 1 are their winding
55 number. All the possibilities are given by the natural integer s , $A = slk$, $A' = smk$, $w = sl + 1$,
56 $w' = sm - 1$, and the cylinder contains smk vertices ($\mathbf{a} \times \mathbf{b} = smk \mathbf{m} \times \mathbf{k}$).
57

58
59 The B-C helix and the collagen chain has $(k,l,m) = (3,2,1)$; A/w , an approximant of
60 $1+\sqrt{3}$, has denominator 1, 3, 4, 11, 15, 41, 56, But its winding number $w = 2s + 1$ must be
odd ≥ 5 , since $w' = s - 1 \geq 1$, so that, the collagen chain has either $s = 5$, $w = 11$, $A'/w' = 15/4$
(Fig. 2), $A/w = 30/11$, or $s = 7$, $w = 15$, $A'/w' = 21/6 = 7/2$, $A/w = 42/15$ (one vertex must be

1
2
3 added to the next approximant $41/15$ of $1+\sqrt{3}$, to complete the periode of the structure : it is
4 on the collagen chain $21/6$ along **l**, and on the steepest helix $14/1$ along **k**. Moreover, the
5 underlying cylinder is perpendicular to its base, $\mathbf{b}\cdot\mathbf{a} = 0$. The helix $7/2$ ($21/6$ for chemical
6 periodicity), periodic in Euclidean space, is thus the basic helix of collagen (identified by [1]).
7

8 The three chains intertwine to constitute the collagen molecule, a triple helix with a
9 core consisting of the side-groups of the Gly, a B-C helix of H atoms [5].
10

11 12 13 14 3. TRANSVERSE STRUCTURES (GAP, OVERLAP)

15
16 The geometric and numerical interdependence between close-packing, tiling by
17 equilateral triangles, $\sqrt{3}$ and its rational approximants is central to the structure of collagen.
18 The transverse structure of collagen (the packing of the molecules across the fibril of Fig. 1f)
19 is also based on approximants of $1+\sqrt{3}$. Lattices of approximants of $\sqrt{3}$ are readily obtained
20 by superposition of two hexagonal lattices at 90° . One obtains a moiré pattern of coincidence
21 points (Fig.3) which consists of **four** copies of the gap structure, shown in **Fig. 4a**. A vertex
22 now represents a collagen molecule. The gap structure is, topologically, the Archimedean
23 square-triangle tiling $3^2.4.3.4$. The axes of the unit cell and the diagonals of the « squares »
24 mark the 6 reticular directions of the two underlying hexagonal lattices. All edges have equal
25 length, but there is a slight distortion (the unit cell is 1% rectangular and the squares are 8%
26 rhombi) imposed by $\sqrt{3}$ (12-fold quasicrystallography). This is the structure observed by
27 Okuyama et al. [1] in a single crystal of $[Gly-Pro-Pro]_{10}$, with an orthorhombic distortion of
28 2%.
29
30
31

32 Vertex coordination of 5 makes the fibril flexible, provided that there are gaps
33 between molecules stacked on top of each other (Fig.1e), so that the transverse structure is
34 alternatively overlap and gap as one moves along the fibril. It has been established [7] that out
35 of five molecules in the overlap, only four extend into the gap. Bouligand [2] has suggested
36 that gap and overlap have the same square-triangle tiling but at two different scales, and
37 rotated by a topological angle of $\tan^{-1}(1/2)$. He has backed his suggestion with diffraction
38 data, and has given an elegant model (Fig.5) of the gap-overlap interface (a twist grain
39 boundary in the fibril) and of the mechanism whereby the termination of one molecule out of
40 five in the overlap leaves the square-triangle tiling topologically invariant (c to a).
41
42

43 From the gap structure, it is straightforward to construct the unit cell of the overlap
44 structure on the underlying hexagonal lattice (Fig.4b), given the ratio between the areas of the
45 unit cells ($112/90$, half a triangle in the gap short of $5/4$), and the topological angle of rotation
46 ($\tan^{-1}(1/2)$ seen in Fig.6 and in Fig.5b). The roles of the slightly distorted structures are
47 exchanged : the unit cell is 8% rhombic and the squares are 15% rectangular. It is now the
48 edges (no longer all equal) and the diagonals of the rhombi that are along 6 reticular
49 directions. **The overlap structure has two perpendicular mirror axes defining a
50 rectangular unit cell (they bound a fundamental domain). Several points are aligned and
51 regularly spaced : the diagonal of the square is a third of the distance to points (2,-1) or
52 (-1,2) of the unit cell grid.** Finally, a similar overlap structure ($\sqrt{3}$ smaller) can be drawn on
53 the same underlying lattice (rotated by 90°) with the «square » of the gap as unit cell.
54
55

56 The Table below lists the geometric manifestations (**square and unit cell of the**
57 **$3^2.4.3.4$ square-triangle tiling**) of the various approximants to $\sqrt{3}$ in the overlap and gap
58 structures [8]. The vertices are joined by lattice vectors in both underlying hexagonal lattices.
59 If two vertices lie on one of the 6 reticular directions (symmetry axis) of the underlying
60 lattices, an approximant is obtained from their distance measured on both lattices (**the sides of
a rectangle [...] on one lattice**). If the vertices do not lie on a symmetry axis, they constitute,

on one lattice, an equilateral **rhombus**, with nearly equal diagonals on symmetry axes and nearly orthogonal side vectors. This yields two approximants (a) and (b). The last column gives the (orthorhombic or monoclinic) distortion, namely the ratio between $\sqrt{3}$ and its approximant.

j	$\sqrt{3} \approx$	overlap	gap	distortion
1	2/1	[square] unit cell (a)	square (a)	1.155
2	5/3	[mirror unit cell]		1.04
3	7/4	unit cell (b)	square (b) [unit cell]	1.01

For reference:

4	19/11	BC helix
5	26/15	inflated BC helix

The ratio in sizes of the two structures is 1.115, smaller than the topological value of $\sqrt{(5/4)} = 1.118$, but larger than the experimental ratio of 1.08 [2]. This discrepancy remains unexplained.

Fig.6 shows how the overlap and gap structures match as a twist grain boundary **with an important coincidence site lattice, containing points $n(2,-1)$ or $n(-1,2)$ of the overlap unit cell grid and points $n(2,0)$ of the gap unit cell grid (n integer), thereby measuring the angle of rotation in Bouligand's mechanism (Fig.5).** The overlap supercell containing 20 vertices and 5 unit cells, with sides $(2,-1)$ and $(-1,-2)$, fits almost exactly with the square of 2×2 gap unit cells, that contains 16 vertices. **The rotation angle of $26.996^\circ = \alpha + \gamma$ is the metric equivalent of the topological rotation of $\tan^{-1}(1/2)$. (Direction $(1,0)$ of the overlap grid makes an angle $\alpha = \tan^{-1}(\sqrt{3}/6)$ with the underlying hexagonal lattice ; just as $\beta = \tan^{-1}(\sqrt{3}/9)$ for direction $(2,-1)$, that is also the diagonal of the square, or direction $(1,0)$ in the gap grid.)** The rotation between two successive overlap structures across the gap, is double, namely $3\pi/10$. Note that in the unit cell of the rotated structure, the small and large rhombi have exchanged positions. Consequently, the angle of the next rotation is the complementary (to $2\pi/3$) of $3\pi/10$. The collagen fibril, a periodic stack overlap-gap-overlap-gap-....., exhibits both 10 and 12-fold rotations.

The flexibility of a long, tightly packed collagen fibril, regardless of its length or girth, is the effect of the regular alternance of gap and overlap regions with five neighbors to each molecule, that are the main features of our model.

Acknowledgments: The authors acknowledge the financial support of the French Ministère délégué à la Recherche (Action Concertée Incitative 2000), through the ACI Biofrustration. We are indebted to Y. Bouligand for impulsing this work, enlightening discussions, helpful suggestions, and for Figs. 1 and 5.

References

- [1] Ke. Okuyama, Ka. Okuyama, S. Arnott, M. Takayanagi, M. Kakudo, 1981, *J. Mol. Biol.* **152** 427.
- [2] Y. Bouligand, 1997, *L'assemblage compact des triples hélices de collagène*, JMC5, poster 2108, Soc. Franç. Phys. Orléans 1997.
Y. Bouligand, 2001, *Two superposed fibre diagrams in X-ray diffraction patterns of collagen tendons: The square-lozenge model*, Unpublished.
- [3] Y. Bouligand, J.P. Denèfle, J.P. Lechère, M. Maillard, 1985, *Biol. Cell* **54** 143.
- [4] A. Rich and F.R.C. Crick, 1961, *J. Mol. Biol.*, **3** 483.
- [5] J.F. Sadoc and N. Rivier, 1999, *Eur. Phys. J. B* **12** 309 ; 2000, *Mat. Sci. Eng.* **294-296**, 397.
- [6] J.F. Sadoc and R. Mosseri, 1999, *Geometrical Frustration*, Cambridge Univ. Press.
- [7] J.A. Hodge and A.J. Petruska, 1963, in *Aspects of Protein Structure*, G.N. Ramachandran, ed., 289, Acad. Press, London.
- [8] N. Rivier and J.F. Sadoc, 2005, in *Topology in Molecular Biology*, M.I. Monastyrsky, ed., Springer, to appear.

Figure captions

Fig.1

Structure of collagen at various scales (from Bouligand [2,3]). a) A single chain *Gly-X-Y* forms a left-handed helix, with a curved axis. Dots indicate the positions of successive amino acids. Their side-groups are pointing outwards; every third one is a glycine (gly) located on the concave side of the axis. b) A collagen molecule is a right-handed triple helix of 3 intertwined, left-handed polypeptide chains. c) A simpler representation of the right-handed triple helix. Its width is 1 nm, its length, 300 nm. d) The collagen molecule is often represented by an arrow of length 300 nm. e) The molecules are stacked on top of each other, with a gap of 35 nm, and regularly spaced on a lattice (transverse scale dilated relatively to longitudinal scale). The longitudinal coordinates of molecules *u* and *u'* are the same. One distinguishes gap and overlap levels, *g* and *o* respectively, so that to each five molecules in *o* correspond only four in *g*. There is a regular stagger of molecules by a length $D = g + o = 67$ nm. f) A crude representation of the lattice as a cylinder (by identifying *u* and *u'*). This is only schematic as the collagen fibril is a lattice, showing the 67 nm stagger, but with a unit cell that accommodates gap and overlap levels. Notice the right-handed chirality of the *Gly* core and of the triple-helix collagen molecule, opposite to the left-handed chirality of the single collagen chain (Fig. 2).

 Fig.2

The B-C helix represented as a hexagonal lattice on a cylinder, cut along its axis (vertical gray lines) and flattened. The single collagen chain is a left-handed helix winding in the \mathbf{l} direction. The axis $Gly-X$ of helix $7/2$ is perpendicular to the base, so that a stack of three segments $7/2$ constitutes one period of the collagen chain in Euclidean space.

%(should be put on the right of Fig.1)%

 Fig.3

Superposition of two hexagonal lattices at 90° yields a moiré pattern of coincidence points. By construction, they are the rational approximants of $\sqrt{3}$. The brightest points (**corners of a square unit cell, with eight inner points**) constitute **four, translated** copies of the gap structure, shown in Fig. 4a. **The corners of the square unit cell belong to all four copies, the corners of the central square, to one copy each, and the corners of the intermediate square tilted at 45° , to two copies. Alternatively, one can view the square cell as decorated with (1/4 of) a dodecagonal wheel around each corner. Each copy of the gap structure contributes five points to the wheel.**

%[[[The frame and the horizontal and vertical scales can be removed from Fig.3]]]]%

 Fig.4

The unit cells of the square-triangle $3^2.4.3.4$ transverse structures of the collagen, drawn on an hexagonal lattice. A vertex represents the collagen molecule. a) Gap. b) Overlap. c) A smaller overlap structure, with the square of the gap structure as unit cell, drawn on the hexagonal lattice rotated by 90° . It is similar to b), but smaller by $\sqrt{3}$. Also drawn (grey) is the associated Voronoi tiling,

%(should be put on the right of Fig.3)%

 Fig.5

The gap-overlap interface in a crystalline collagen fibril (schematic) (from Bouligand [2]). Gap (a) and overlap (c) structures are topologically identical, square-triangle patterns. (b) indicates how the gap structure deforms to intercalate new molecules (o) in the overlap, or, reciprocally, how one molecule (o) out of every 5 in the overlap terminates leaving the pattern invariant.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

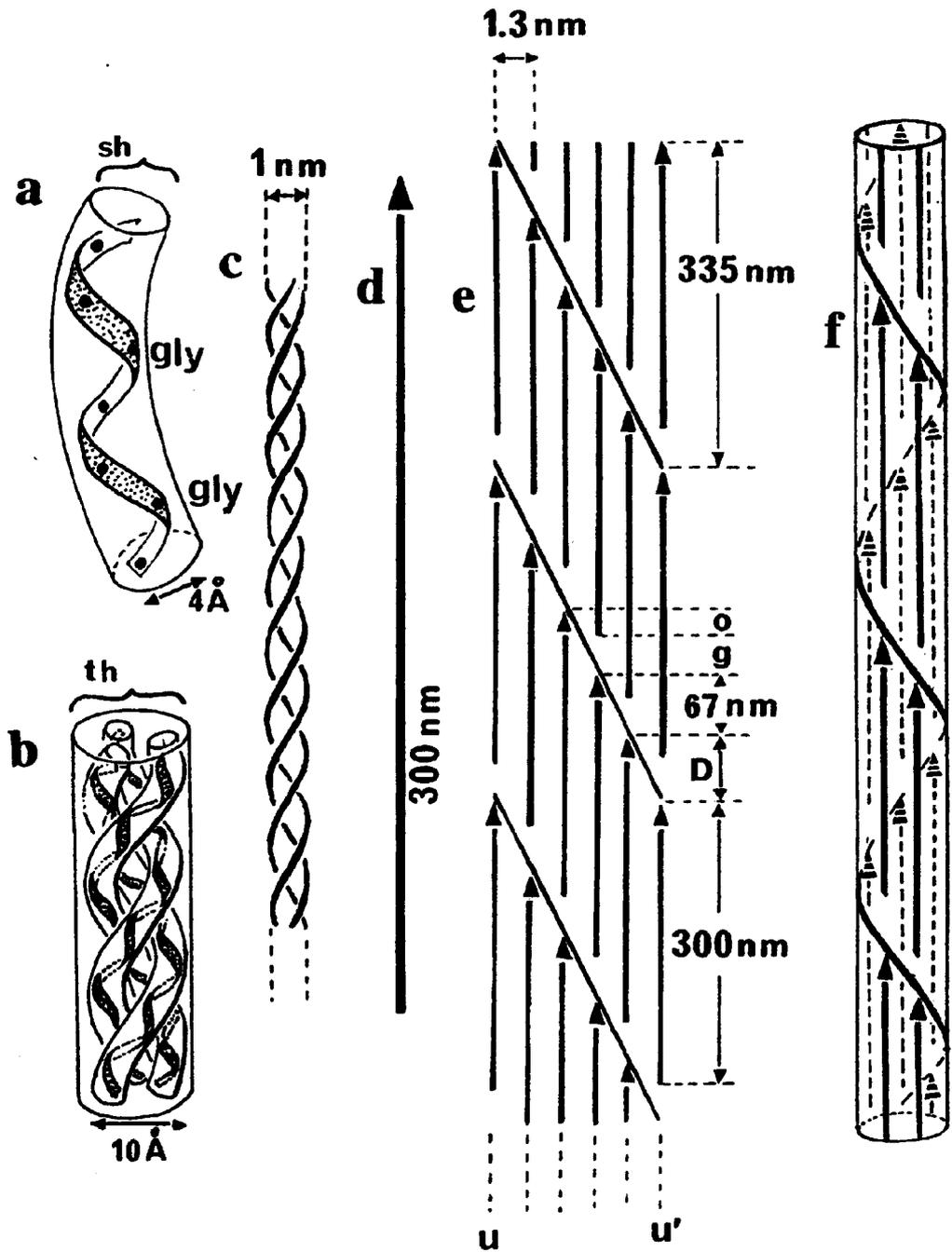
Fig.6

Two successive twist grain boundaries overlap-gap-overlap. Note the $5/4$ ratio between the areas of the overlap and gap unit cells, and the coincidence site lattice of equidistant, aligned vertices in the two overlap structures, rotated by $3\pi/10$.

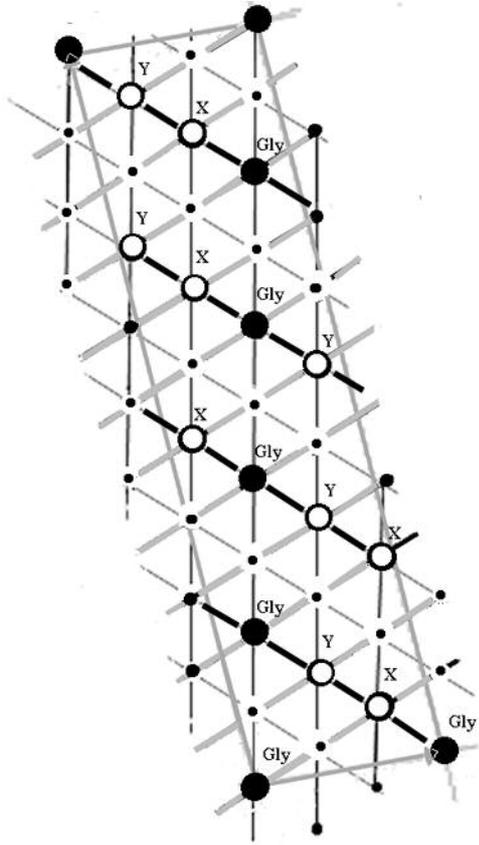
%Fig.5 and Fig.6 (can be put together)%

For Peer Review Only

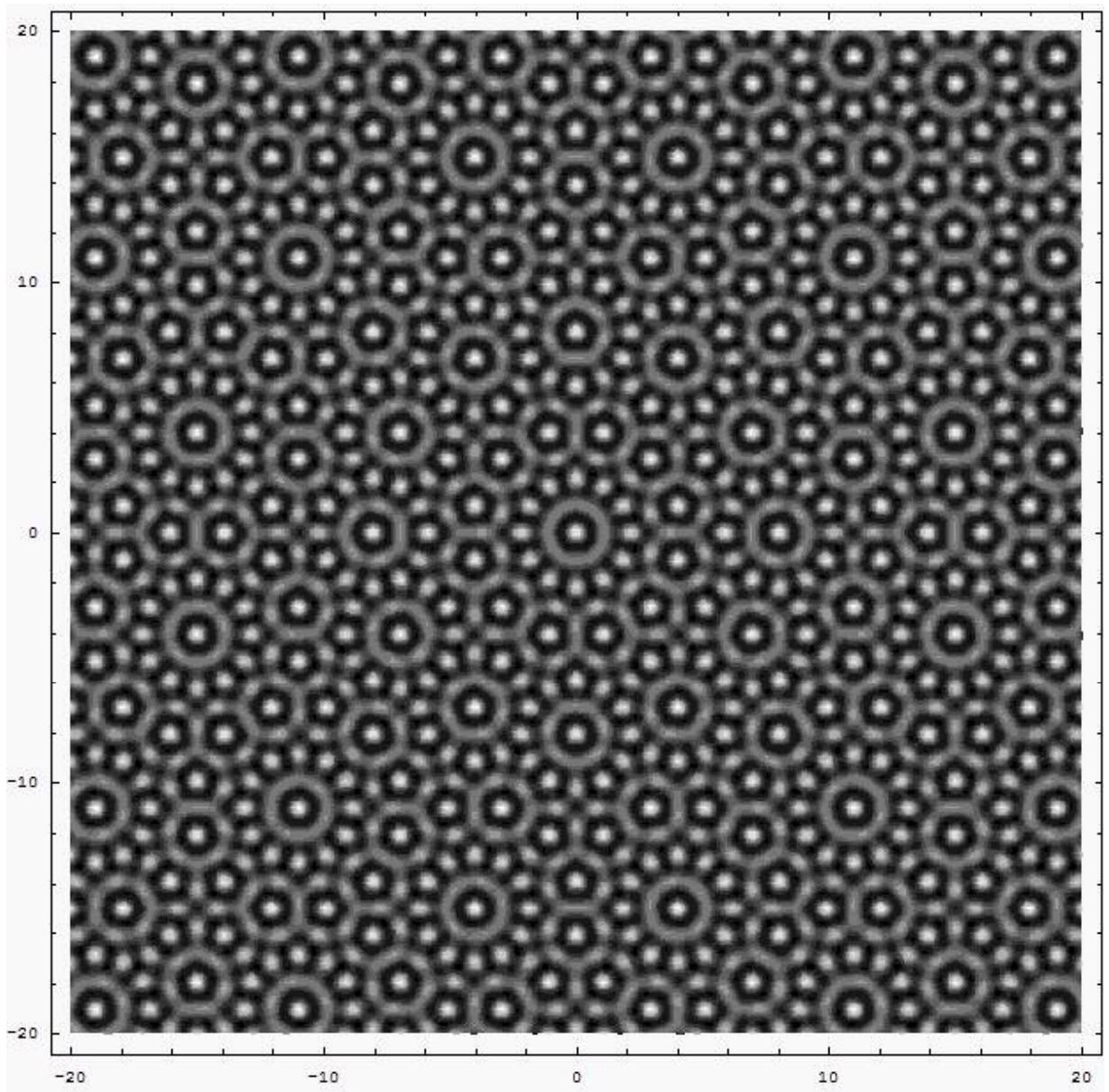
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

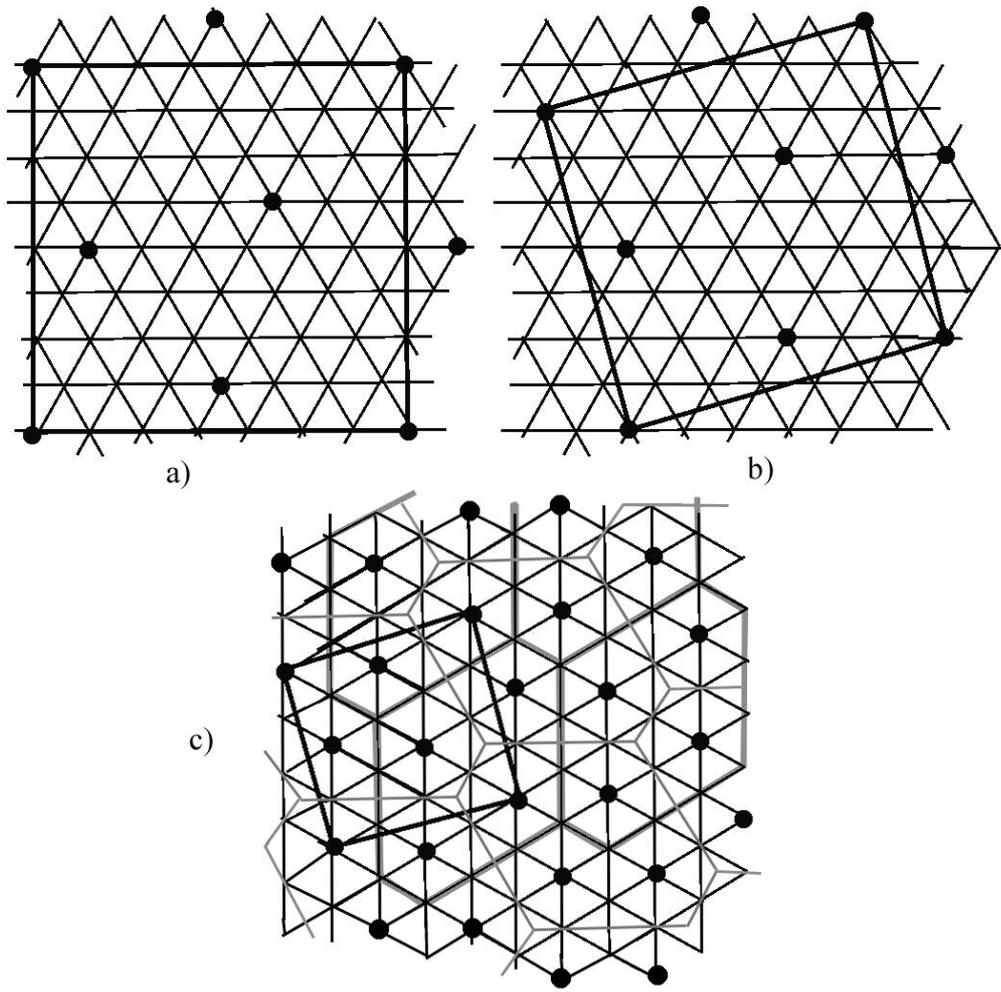


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



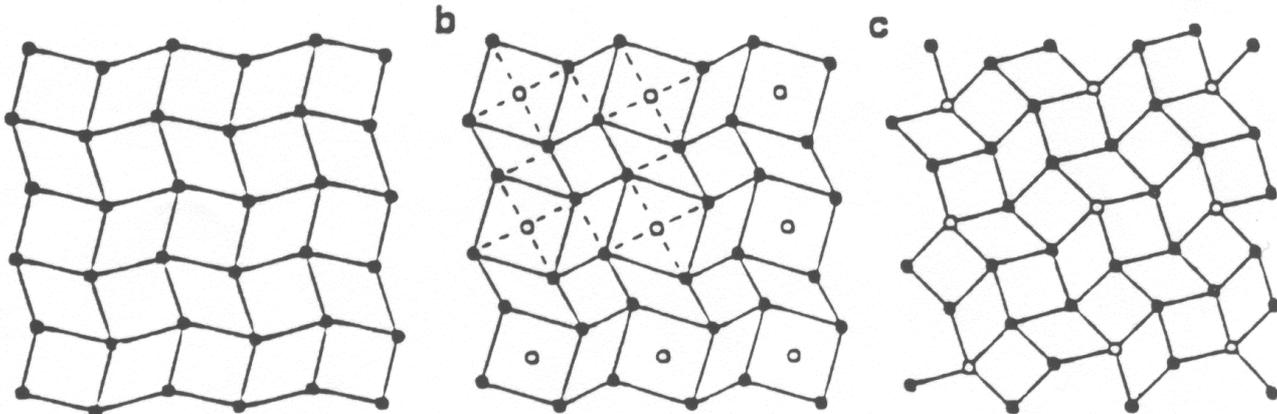
arXiv

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



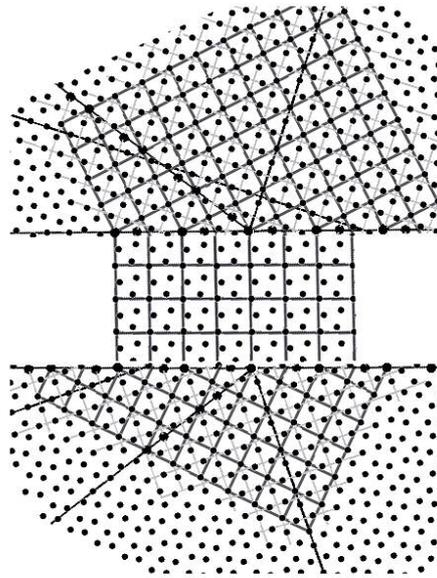
Pre-proof

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Peer Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For Peer Review Only