

Boerdijk-Coxeter helix and biological helices

J.F. Sadoc^{1,a} and N. Rivier²

¹ Laboratoire de Physique des Solides^b, bâtiment 510, Université de Paris-Sud 91405 Orsay Cedex, France

² Laboratoire de Dynamique des Fluides Complexes^c, 3 rue de l'Université, Université Louis Pasteur, 67084 Strasbourg, France

Received 26 November 1998 and Received in final form 12 April 1999

Abstract. Helices and dense packing of spherical objects are two closely related problems. For instance, the Boerdijk-Coxeter helix, which is obtained as a linear packing of regular tetrahedra, is a very efficient solution to some close-packing problems. The shapes of biological helices result from various kinds of interaction forces, including steric repulsion. Thus, the search for a maximum density can lead to structures related to the Boerdijk-Coxeter helix. Examples are presented for the α -helix structure in proteins and for the structure of the protein collagen, but there are other examples of helical packings at different scales in biology. Models based on packing efficiency related to the Boerdijk-Coxeter helix, explain, mainly from topological arguments, why the number of amino acids per turn is close to 3.6 in α -helices and 2.7 in collagen.

PACS. 87.10.+e General theory and mathematical aspects – 36.20.-r Macromolecules and polymer molecules

1 Introduction

Biomolecules have spatial (secondary) structures which result from various interaction forces. For example, covalent interaction: peptide bond for the backbone of proteins, hydrogen bond: whether intrachain (α -helix) or interchain (collagen, β -sheet), and steric repulsion between side groups in proteins. The steric repulsion plays a very important role. Therefore it is not surprising that geometrical considerations could contribute towards the understanding of structures at intermediate scales, such as secondary and tertiary structures for proteins. They could also help in the classification of elementary conformations and in the understanding of chiral relations.

Among elementary conformations, helices occur widely in the biological world. Here we want to relate some biological helices to simple geometrical helices, starting from packing considerations.

This paper discusses structures of biomolecules from a topological viewpoint whereas most of the biological literature takes a metric approach. See, however, Chothia *et al.* [1] and Adzhubei and Sternberg [2]. The covalent bond along the polypeptide chain is the only metric length. The hydrogen bond only comes later and the structure has enough flexibility to accommodate it. The steric interaction is only repulsive and does not impose a length scale a priori. But the paramount importance of

steric interactions involving peptide bonds and side group nearby, is manifest in the Ramachandran plot [3].

It must be emphasized that this apparent topological looseness will be severely controlled by two geometrical constraints:

(i) Folded proteins are close-packed conformations, moulded by the steric repulsion between side groups (Sect. 2.1).

(ii) The rigid peptid unit is, in fact, a stiff, rectangular unit which tiles in a specific way the two-dimensional substrate covered by the secondary structure (α -helix, β -sheet) of the protein (Sect. 3.1 and Fig. 4).

2 The geometry of helices

2.1 The Boerdijk-Coxeter helix

Helices and dense packing of spherical objects are two closely related problems. A very interesting geometrical figure is obtained by stacking regular tetrahedra along one direction. It is called the Boerdijk [4]-Coxeter [5] helix (B-C helix). Select one face of a tetrahedron, on which the next tetrahedron is glued, and proceed on gluing new tetrahedra, with the conditions that no more than three tetrahedra share an edge, and that edges with only one tetrahedron are more or less aligned. A chain of tetrahedra is obtained, on which external edges form three helices (Fig. 1). Surprisingly, this chain is not periodic, owing to an incommensurability between the distances separating centres of neighbouring tetrahedra, and the pitch of the three helices.

^a e-mail: sadoc@lps.u-psud.fr

^b associé au CNRS

^c associé au CNRS

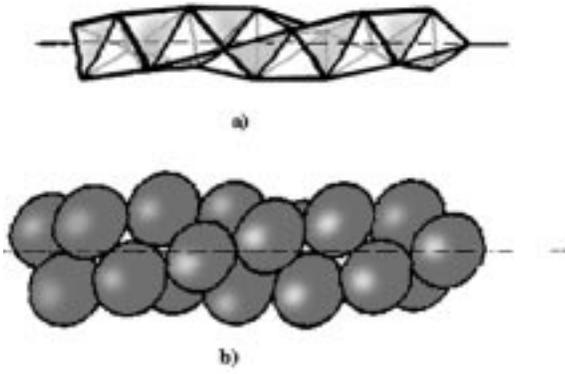


Fig. 1. Boerdijk-Coxeter helix obtained from a necklace of tetrahedra (a). A dense packing of spheres centered on the tetrahedron vertices (b). The B-C helix (a) can be obtained by folding the edges of the triangular lattice of Figure 2, and gluing together the larger sides of the rectangle. A torus is obtained (in curved space) by identification of the smaller sides.

There are different kinds of tetrahedral edges corresponding to the number of tetrahedra sharing a giving edge: Those which appear most parallel to the axis of the Coxeter helix belong to only one tetrahedron. They will be called hereafter type- $\{3\}$. Edges sharing two tetrahedra are called type- $\{2\}$ and edges sharing three tetrahedra, type- $\{1\}$. The number corresponds to the direction of the edge in the phyllotactic representation of the helix (see below). We distinguish several families of helices made of these three types of edges. There are three type- $\{3\}$ helices, but only one type- $\{1\}$ helix and two type- $\{2\}$ helices.

It is useful to describe the Coxeter helix (or any helical structure resulting from close packed units) as a two-dimensional graph on a cylinder. All edges of the graph are geodesic lines on the cylinder. When the cylinder is unfolded on a flat surface, this surface is tiled with triangles. concretely, the Coxeter helix can be built by taking an actual sheet of paper on which a triangular lattice (with equilateral triangles) has been drawn, cutting a strip three triangles-wide, folding the type- $\{2\}$ edges inwards, types- $\{3\}$ and $-\{1\}$ outwards, and gluing.

2.1.1 The Coxeter helix represented on a plane

The B-C helix is related to the problem of packing spheres or tiling by regular tetrahedra, resolved by the $\{3, 3, 5\}$ polytope in curved space. Because it is impossible to tile Euclidean space with regular tetrahedra, space has to acquire a positive curvature. Details can be found in a recent book [6]. In curved space, the helix winds on a torus instead of a cylinder, and it forms a closed curve. The torus can be cut and flattened into a rectangle (or a parallelogram), with identification of opposite sides. Now, folding a rectangle (or a parallelogram) into a torus in curved spherical space S_3 , can be done without any metric distortions. Thus, for the B-C helix, the flattened torus is tiled by tri-

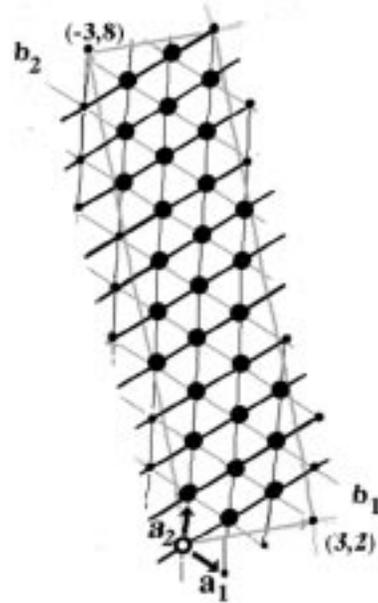


Fig. 2. A flat strip leading to the Boerdijk-Coxeter helix by identification of the two long sides of the rectangle.

angles which are nearly equilateral (some care is needed because only type-3 edges are geodesics of the torus and of S_3 . the other edges are slightly distorted in the flattened torus). the flattened torus is a multiple cell of the triangular lattice with basic vectors $3\mathbf{a}_1$ and $10\mathbf{a}_2$, where \mathbf{a}_1 and \mathbf{a}_2 are the unit vectors of the primitive cell (see Fig. 2). The simplest flat representation of a cylinder is a rectangle, obtained by cutting the cylinder along a generating line. Similarly, a rectangular representation of a torus (in S_3) can be obtained by cutting the torus along its two orthogonal generating circles. Since the torus is tiled by equilateral triangles, we look for a multiple cell $(\mathbf{b}_1, \mathbf{b}_2)$ of the triangular lattice, with the same area as the original cell's $(3\mathbf{a}_1, 10\mathbf{a}_2)$, and hypotenuse $\mathbf{b}_1 + \mathbf{b}_2 = m\mathbf{a}_2$ along the direction \mathbf{a}_2 . We further require that $\mathbf{b}_1 = 3\mathbf{a}_1 + l\mathbf{a}_2$, with integers l, m to be determined (\mathbf{b}_1 is the base of a cylinder around which the helix is wound), and that $\|\mathbf{b}_2\| > \|\mathbf{b}_1\|$. Finally, the multiple cell should be as nearly rectangular as possible, *i.e.* $\mathbf{b}_1 \cdot \mathbf{b}_2 \cong 0$, so that the cut lines \mathbf{b}_1 and \mathbf{b}_2 approximate generating circles of the torus. The best solution, with \mathbf{b}_1 and \mathbf{b}_2 most nearly orthogonal, is given by $\mathbf{b}_1 = 3\mathbf{a}_1 + 2\mathbf{a}_2$, $\mathbf{b}_2 = -3\mathbf{a}_1 + 8\mathbf{a}_2$ and $\mathbf{b}_1 \cdot \mathbf{b}_2 = -2$. Figure 2 shows the multiple cell $(\mathbf{b}_1, \mathbf{b}_2)$ and the primitive basis $(\mathbf{a}_1, \mathbf{a}_2)$.

2.1.2 Images of helices on the flat projection

The three different types of helices can be easily identified on the flat strip $(\mathbf{b}_1, \mathbf{b}_2)$. There are three type- $\{3\}$ winding along \mathbf{a}_2 . With opposite sides of the strip identified, they form closed loops making one turn around the two axes of the torus, with 10 edges and vertices each. In spherical space, they are geodesics, great circles of S_3 , also called fibers of the Hopf fibration of S_3 [7,6]). Two type- $\{2\}$

helices wind along $-\mathbf{a}_1$; they also form close loops, making 4 turns around one axis and one turn around the other, with 15 vertices. Finally, one type- $\{1\}$ helix, the Coxeter helix, winds along $\mathbf{a}_1 + \mathbf{a}_2$. It has 30 edges and vertices, and makes 11 turns around one axis and one around the other. The Coxeter helix has therefore $30/11 = 2.727272\dots$ edges per turn. The helices have opposite chiralities: If type- $\{1\}$ and $\{-3\}$ are right-handed helices, say, type- $\{2\}$ helices are left-handed.

The B-C helix, like the Coxeter helix and the Coxeter chain to be introduced later, is labelled $(3, 2, 1)$ in phyllotactic notation. This is a notation describing triangular lattices on cylinders [8, 9]. It describes economically all the possible structures of composite flowers (phyllotaxis) or of ordered foams inside thin cylindrical tubes. The vertices of the triangular lattice are decorated by close-packed florets or bubbles, which are generated one after the other as the flower grows or as the bubbles rise in the tube. Each vertex is labelled by a natural integer n , in order of increasing altitude on the vertical cylinder, or of increasing age in a flower. The phyllotactic notation (k, l, m) , with $k > l > m$, implies that the vertices labelled $n \pm k$, $n \pm l$, $n \pm m$ are neighbours to vertex n . Consequently, in a triangular lattice, $k = l + m$, since $n + k$ and $(n + l) + m$ label the same neighbour to vertex n . The three types of helix on the cylinder are labelled accordingly: The k helices of type- $\{k\}$ include vertices $\dots n - k, n, n + k, n + 2k, \dots$, and are the steepest ones. The m helices of type- $\{m\}$ are the flattest ones. If there is one single helix going through all the vertices, it is of type- $\{1\}$, and $m = 1$. It is labelled $(k, k - 1, 1)$. The simplest example is the B-C helix $(3, 2, 1)$. Other helices of biological interest are the α -helix $(4, 3, 1)$, the π -helix $(5, 4, 1)$ and Pauling's γ -helix (or 5.1 helix) $(6, 5, 1)$ (see Sect. 3.2.2). We will see (Fig. 6) that collagen is a type- $\{2\}$ helix, including only half the vertices of the triangular lattice.

2.1.3 The Boerdijk-Coxeter helix on an Euclidean cylinder

We can build helices in Euclidean space starting from the flat map of the helix on a torus. We make a long strip by assembling several patching units $(\mathbf{b}_1, \mathbf{b}_2)$ joined by their smaller sides, and fold it into a cylinder.

It is easy in curved space (or on a torus) to count how many turns an helix makes around its axis, as it is a pure topological number. This is not so simple on a cylinder, as we do not know the exact angle between \mathbf{b}_1 and \mathbf{b}_2 after folding.

So, we must use coordinates in Euclidean space (H.S.M., Coxeter, private communication). The coordinates of the n th vertex A_n of an helix are given by:

$$x_n = \cos n\theta \quad y_n = \sin n\theta \quad z_n = nc. \quad (1)$$

The distances between vertex n and vertices $n + 1$, $n + 2$ and $n + 3$ are first neighbours distances. Then

$$\overline{A_n A_{m+n}}^2 = \overline{A_0 A_m}^2 = 2 - 2 \cos m\theta + m^2 c^2. \quad (2)$$

Since the edges $\overline{A_0 A_m}$ all have the same length, for $m = 1, 2, 3$, we find, eliminating c , an equation for $x = \cos \theta$: $3x^3 - 4x^2 - x + 2 = 0$, which factorizes as $(x - 1)^2(3x + 2) = 0$. Discarding the trivial root $x = 1$, we deduce that the angle θ is given by

$$\cos \theta = -2/3, \quad \theta = 131.810^\circ. \quad (3)$$

We can also obtain the translation part of the helical motion, or pitch, $c = \sqrt{10/27}$, or $c/\overline{A_0 A_1} = 1/\sqrt{10} = 0.3162$ in unit of edge length.

The number of edges per turn is given by $\xi = 2\pi/\theta$. It is $\xi = 2.7312$, close to the number $30/11$ on the torus.

2.1.4 The Coxeter chain as a quasicrystal

The Coxeter chain is the B-C helix in Euclidean space, decorated by spheres on its vertices. It is not a periodic structure, because ξ is not a rational number. The question arises: is it a quasicrystal, with the usual property of quasicrystals of having an inflation-deflation symmetry, and thus well approximated by crystals with larger and larger unit cells? The sequence of integers in the continuous fraction expansion of ξ is not periodic, so that ξ is not a quadratic irrational (a theorem of Lagrange), necessary condition for context-free inflation-deflation symmetry [10, 11], and the infinite Coxeter chain constructed above has no inflation-deflation symmetry. It is interesting to search for a chain close to the Coxeter helix which has approximant structures.

This modified quasicrystalline Coxeter helix is described in the appendix. It has a number of edges per turn $\varpi = 1 + \sqrt{3} = 2.73205$, a quadratic irrational, very close to $\xi = 2.7312$, for the Coxeter chain, and to $30/11 = 2.7272\dots$ for the B-C helix in curved space.

2.2 The α -helix: a disclinated Boerdijk-Coxeter helix

The α -helix is one of the important secondary structures found in proteins. The number of elementary steps of the backbone is given to be close to 3.6 units per turns. This is about one larger than $\xi = 2.7312$ for the B-C helix, so that we must increase its diameter.

Disclinations are the natural defects associated with rotational or helicoidal symmetry. In the case of helices, disclinations are characterized by an axis, which is the axis of the cylinder on which is drawn the helix, an angle $\delta\theta$ of rotation, and a vector of translation $\delta\mathbf{s}$ parallel to the axis. Such a wedge disclination combined with a translation is sometimes called a dispiration [12]. The effect of a disclination on a cylinder is explained in Figure 3: the perimeter of the cylinder of radius r is changed from $2\pi r$ to $(2\pi + \delta\theta)r$, and one of the lips of the cut cylinder is translated by $\delta\mathbf{s}$ before regluing. If there is a discrete geometrical structure supported by the cylinder surface, as a discrete helix, the displacement which is the combination of the rotation $\delta\theta$ and the translation $\delta\mathbf{s}$ must be an element of the symmetry group of the structure. If the helix

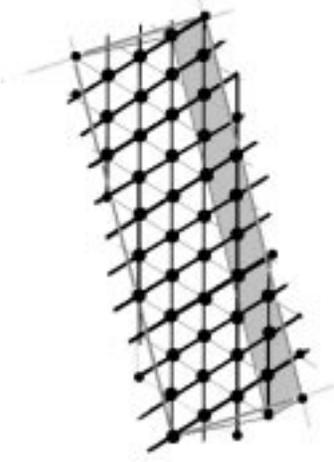


Fig. 3. The strip for an α -helix, by identification of the longer sides. The resulting cylinder is obtained by disclinating the cylinder supporting the B-C helix. It has one additional row of triangles (shaded).

is represented on a strip, the strip is sheared and its width is increased (for positive $\delta\theta$). This is the case whether the helix is drawn on a cylinder or, in curved space S_3 , on a torus, changing the parallelogram patch of the triangular lattice: its width is increased by one triangular unit and it is sheared in order to ensure identification of the longer sides. The new parallelogram is defined by vectors \mathbf{b}_1^α and \mathbf{b}_2^α : $\mathbf{b}_1^\alpha = 4 \mathbf{a}_1 + 3 \mathbf{a}_2$ and $\mathbf{b}_2^\alpha = -3 \mathbf{a}_1 + 8 \mathbf{a}_2$ (unchanged). The type-1 helix running along edges parallel to $\mathbf{a}_1 + \mathbf{a}_2$ consists of 41 edges; it turns 11 times around one axis of the torus and once around the other, leading to a number of edges per turns $\xi = 41/11$. This helix is (4, 3, 1) in phyllotactic notation.

In order to obtain the number of edges per turn ξ in Euclidean space, we use the coordinates defined in equation (1) for the B-C helix and set equal the distances between neighbours $\overline{A_0A_1} = \overline{A_0A_3} = \overline{A_0A_4}$. Note that the helix is no longer a chain of face-on-face tetrahedra, and that all other distances between vertices, notably $\overline{A_0A_2}$, are larger than $\overline{A_0A_1}$. Eliminating c , we obtain a quadratic equation for $x = \cos\theta$, $(x-1)^2(16x^2 + 17x + 2) = 0$. The trivial roots $x = 1$ can be discarded. The root $x = (-17 - \sqrt{161})/32$ gives a distance between non-neighbouring vertices $\overline{A_0A_2}$ smaller than $\overline{A_0A_1}$ and is incompatible with steric repulsions. The only geometrically relevant root is thus,

$$\cos\theta = (-17 + \sqrt{161})/32, \quad \theta = 97.74^\circ. \quad (4)$$

The number of edges per turn, given by $\xi = 2\pi/\theta$, is $\xi = 3.6831$, close to $41/11 = 3.727272\dots = [3, 1, 2, 1, 2]$, obtained on the torus. The corresponding quasicrystalline helix, infinitely extensible by inflation, has $\varpi = 2 + \sqrt{3} = 3.73205 = [3, 1, 2, 1, 2, 1, 2, \dots]$; it is indeed a quadratic irrational. These numbers are close to the value 3.6 observed in the α -helix of real proteins. In biological α -helices, there is no reason a priori for the distances between two succes-

sive central carbons along the chain (determined by the peptide bond) to be the same as those between neighbours in directions 3 and 4 (determined by hydrogen bond and steric repulsion, respectively). *A posteriori*, however, if steric interactions play an important role, it is natural that closed-packed conformations are adopted, with all distances adjusted on the only length scale, that of the covalent peptide bond. Indeed, the number of amino acids per turn in biological α -helices is 3.6, which is a two dimensional structure stabilized by hydrogen bonds, a celebrated result of L. Pauling in 1951, who “let the models fold naturally into any screw they were comfortable with” [13].

The translation parallel to the helix axis, per step, is $c = 0.3637$ or $c/\overline{A_0A_1} = 0.2347$ in units of edge length.

3 Proteins on the triangular lattice

A protein is a stiff chain of amino acids. The sequence of amino acids is called its primary structure. The protein folds tightly into a complex and specific arrangement (tertiary structure) of regular structural elements. These regular units, which constitute the secondary structure of the protein, are the α -helix, and the parallel or antiparallel β -strand. They can all be drawn on a triangular lattice in order to emphasize their regularity. Here, we focus on those which are related to the α -helix.

3.1 The peptide unit

The covalent peptide bond between a C' atom and a N atom lead to a stiff, planar unit of six atoms, which can be schematically represented by a rectangle with the two C_α atoms on opposite corners, the C'-N pair inside the rectangle parallel to its larger side. Next to the remaining corners are an H atom bonded to the N atom, and an O atom bonded to the C' atom (see [3]). The polypeptide chain is constituted by a necklace of rectangles connected by their C_α atoms. It is a chain of rectangles linked by covalent bonds across one of their diagonals. The hydrogen bond takes place between an H atom of one unit and the nearest O atom belonging to a different rectangular unit. It couples rectangular units across their second diagonal (Fig. 4). In this way, covalent and hydrogen bonds impose a two dimensional structure.

We can decorate a triangular lattice by rectangular units: The longer sides of the rectangles are heights of the triangles, the smaller sides, one half of their edges. This constitutes a kind of rectangular chess-board tiling, represented in Figure 4. The grey rectangles are peptides units. In helices, the C_α - C_α diagonals are all parallel; they are the edges of the triangles in direction $\{1\}$. Incidentally, the side groups on the C_α all point outwards on the helix. The C_α atoms are on the vertices of the triangular lattice. Another family of triangle edges, in direction $\{k-1\}$ support the smaller sides of the rectangles, connected by the hydrogen bond H-O. The third family of triangle edges,

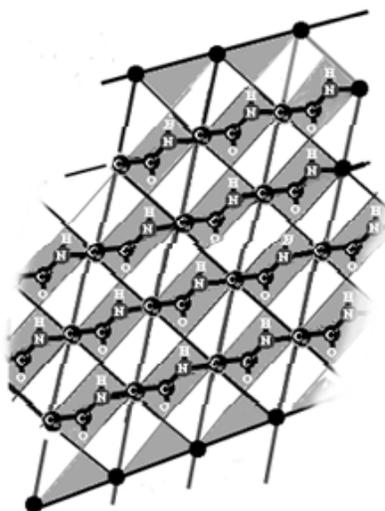


Fig. 4. Peptide units on the triangular lattice. There are three types of edges corresponding to covalent bonds (thick, black diagonals of grey rectangles in direction $\{1\}$), hydrogen bonds (thin, black lines in direction $\{k-1\}$) and steric interactions (thick, grey lines in direction $\{k\}$). This patch folds into a cylinder as in Figure 3.

in direction $\{k\}$, are the diagonals of the white rectangles which are not peptide units. Thus, each of the three edge directions corresponds to a different type of interaction: covalent interactions along $\{1\}$, hydrogen bonds along $\{k-1\}$ and steric interactions between side groups along direction $\{k\}$, for the helix $(k, k-1, 1)$.

3.2 Helices in proteins

3.2.1 The α -helix

Consider the helix $(4, 3, 1)$ described previously as a rolled triangular lattice. If we decorate the triangular lattice by peptide units as a chess-board tiling, we obtain a very interesting model of protein folded into the α -helix secondary structure. Notice that the three types of interactions along the three directions of edges in the triangular lattice are related to the three indices (w, v, u) in phyllotactic notation: $u = 1$ corresponds to the covalent bonding of the polypeptide chain, $v = 3$ to the hydrogen bonds and $w = 4$, to the steric repulsions.

To avoid any confusion, let us emphasize that the hydrogen bond HO does not link the C_α atoms directly, but the N atom on the backbone $N-C_\alpha-C$ of the i th amino acid, with the C atom of amino acid $(i+3)$. The hydrogen bond completes a full turn of the helix. There are thus, 3 complete amino acids, plus one distance $N-C_\alpha-C$ (approximately $2/3$ of an amino acid), namely 3.6 amino acids per turn. In biological papers, it is often stated that the hydrogen bond takes place between amino acids i and

$(i+4)$, thereby labelling the peptide rectangular units instead of the central carbon atoms C_α .

Can such a structure remain flat, instead of folding into a cylinder? The answer is no, for two reasons: With this arrangement of the peptide units and the standard L-chirality of the central C_α atoms, the side groups are all on the same side of the triangulated surface and have strong steric repulsion. The surface buckles into a cylinder. Moreover, a flat structure would distort too much the sp^3 tetrahedral symmetry of the C_α atoms. This buckling is manifest in the Ramachandran plot [3].

3.2.2 Other helices in proteins

Other helices that the classical α -helix are sometimes observed in proteins. Collagen will be discussed in Section 4. All helices which cover a rolled triangular lattice are labelled $(k, k-1, 1)$ in phyllotactic notation.

- The B-C helix, $(3, 2, 1)$ in phyllotactic notation, has hydrogen bonds represented by edges between sites i and $i+2$ sites. Topologically, it is identical to the so-called 3_{10} -helix [14,16]. This helix is not commonly observed in proteins as a secondary structural element. But α -helices sometimes begin or end with one single turn of a 3_{10} -helix (one hydrogen bond). There are also indications that long $(3, 2, 1)$ helices are observed in biopolymers produced in mushrooms. Hydrogen bonds in a 3_{10} -helix link the N atom of the backbone of amino acid i to the C atom of amino acid $(i+2)$.

- The next possibility is the α -helix $(4, 3, 1)$. There are hydrogen bonds represented by triangle edges between sites i and $i+3$, thus connecting peptide units i and $i+4$.

- The $(5, 4, 1)$ helix, obtained by folding a strip with one additional row of triangles compared to the α -helix is called the π -helix.

- The $(6, 5, 1)$ helix corresponds to the Pauling 5.1-helix (or γ -helix).

Increasing k further would yield helices on flatter cylinders; the steric repulsion between side groups becomes too important and there are no proteins with $k > 6$.

Several polypeptide synthetic helices have phyllotactic structures, as was noticed by Frey-Wyssling [15]

3.3 Structural transition from the $(4, 3, 1)$ helix to the $(3, 2, 1)$ helix

We have seen in Section 3.2.2 that an α -helix $(4, 3, 1)$ can terminate with one single turn of a 3_{10} -helix $(3, 2, 1)$. The structural transition is achieved very simply on the strip of triangular lattice by one single dislocation [8], as shown in Figure 5. The width of the strip is smoothly changed by one triangle.

With the same construction, one can connect an α -helix to a π -helix, or with two dislocations, to a γ -helix.

Such connection are present in real proteins, as suggested originally by Harris *et al.* [18]. Motion of the dislocation (glide or climb) can play a part during refolding of a protein, this involves conformational changes from

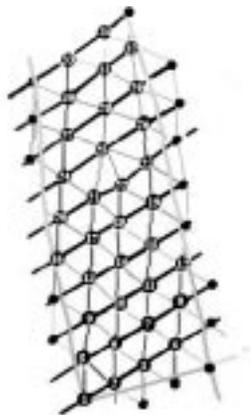


Fig. 5. Strip of triangular lattice containing a dislocation. At the dislocation, the type- $\{1\}$ helix, in bold lines, changes from an α -helix (4, 3, 1) to a B-C helix (3, 2, 1); one type- $\{4\}$ helix and one type- $\{3\}$ helix are interrupted.



Fig. 6. Ribbon representation of a triple-helix collagen molecule.

one helix to another, with a smaller or a larger diameter. The energy barrier between initial and final helical conformations is very high if all H-bonds have to be cut. The climb of a dislocation, involving only one H-bond at any one time, has a much lower energy barrier. This is indeed the standard mechanism for plastic deformation of solids, and for the growth of epithelial tissues [17]. It has been observed in foams in a cylindrical tube [8].

4 The collagen

4.1 An important protein

Collagen is another helicoidal structure very common in biology. It is helicoidal at different scales, submolecular, molecular, through to the entire organism in the submarine worm studied by Gaill and Bouligand [19]. A collagen

molecule consists of three polypeptide chains intertwined in a triple helix. Each individual polypeptide chain is itself a helix, but it takes the helical conformation when associated with the other two chains (Fig. 6). Hydrogen bonds connect different helices and stabilize the triple helix collagen molecule. For this reason, the individual helical chain is completely different from the α -helix, even though it is sometimes called α -chain [20,21].

Diffraction patterns of collagen show ambiguities in the periodicities of the structure, both along the molecule [22, 23] and in the perpendicular plane for collagen fibrils [24]. This strongly suggests a quasiperiodic structure in both directions, as discussed, along the molecule, in the appendix.

The helical conformations are due to the facts that every third amino acid of each chain is a glycine (Gly), and that the sequence is rich in proline (Pro). Gly is the amino acid with the smallest side group (H). A larger side group than Gly would prevent close contact between the three chains. About half of the Pro side groups are hydroxylated; the resulting hydroxyproline is referred to as “Hyp”. Each chain is a periodic sequence of repeating units Gly-X-Y, where either X or Y is, almost always, Pro or Hyp. The molecule is stabilized by hydrogen bonds between the backbone amide (N atom) of a Gly amino acid and the backbone carbonyl (C atom) of amino acid X, as indicated in Figure 9b.

4.2 Hopf fibration and the collagen

4.2.1 The PPII helix

The structure of a single chain of collagen, procollagen, is called polyproline II (PPII). PPII is an artificial biopolymer, but the PPII helix occurs in a number of proteins [2]. The PPII helix has a preference to proline, but almost any natural amino acid unit can sometimes be found in this conformation. PPII helices are left-handed helices with a rise per peptide unit, almost twice as large as in α -helices. In comparison with α -helices, they are therefore longer, for the same number of peptide units, less tightly packed and their main chain is more accessible.

In a collagen molecule three left-handed PPII helices are intertwined and wrap around each other in a right-handed helix. We must take these two opposite chiralities into account in modelling the structure of collagen.

4.2.2 The Boerdijk-Coxeter helix and the PPII helix

The B-C helix (3, 2, 1) can be described on a triangular lattice. The edges of the triangles support three different types of helices, type- $\{1\}$, $\{-2\}$ and $\{-3\}$. If we choose a right-handed Coxeter helix (the type- $\{1\}$ helix of the B-C helix), then the type- $\{3\}$ helices are also right-handed, whereas the type- $\{2\}$ helices are left-handed. Intrinsically, the tetrahedron chain has the two chiralities.

Let us suppose that the left-handed PPII helix is of the type- $\{2\}$, winding on the strip which folds into the B-C

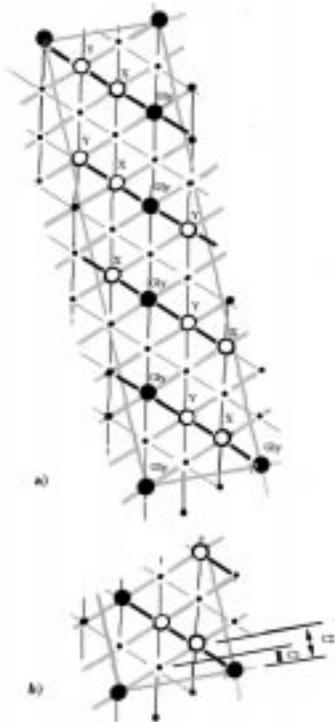


Fig. 7. A Boerdijk-Coxeter helix with the collagen sequence *Gly-X-Y* along one of the type- $\{2\}$ helices (a). This represents a single proto-collagen helix, with the PPII structure. The pitch (per amino acid) for PPII helix is twice that of B-C helix (b).

helix. There are two type- $\{2\}$ helices. Thus, only one half of the sites (vertices) must be visited by the PPII helix (Fig. 7a). This choice is supported by the fact that the rise per peptide unit in a PPII helix is twice that of an α -helix or of the Coxeter helix (Fig. 7b).

The collagen amino acid sequence is $(\text{Gly-X-Y})_n$, so all the Gly amino acids are gathered on a right-handed type- $\{3\}$ helix occupying half of its sites.

4.2.3 The collagen triple-helix molecule and the Hopf fibration of polytope $\{3, 3, 5\}$

The Coxeter chain of tetrahedra can be extracted from polytope $\{3, 3, 5\}$, which is a regular scaffolding of S_3 (the hypersphere) made of 600 tetrahedra, or the closest packing of 120 spheres in curved space. The 120 vertices of this polytope can be distributed, 10 each, on 12 non-intersecting great circles of S_3 . These great circles constitute the fibers of the Hopf fibration of S_3 [6]. They have surprising properties associated with the curved space in which they are living. They are parallel in the Clifford sense (being at constant distance of each other) but they are also entangled, winding once around each other. A useful way to picture the Hopf fibration is to take a torus, obtained by folding a rectangle, on which a diagonal and one line parallel to the diagonal have been drawn: these two lines fold into two intertwined circles like those in S_3

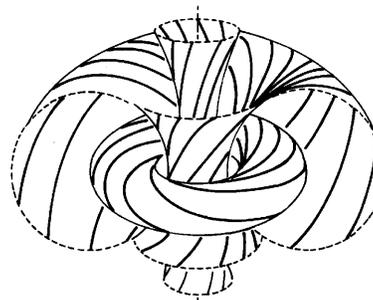


Fig. 8. Stereographic projection of a Hopf fibration of a hypersphere by great circles.

(Fig. 8 shows a stereographic projection of the Hopf fibration).

In our case, we have a discrete Hopf fibration (by 12 great circles) of the discrete scaffolding of S_3 , polytope $\{3, 3, 5\}$. A fibration is specified by a fiber (a great circle with 10 vertices) and a base space, or base. A point on the base is representative of a whole fiber. The base space of the Hopf fibration of S_3 is a sphere S_2 . Here, the 12 fibers are represented by the 12 vertices of an icosahedron on S_2 (Fig. 9a). Without going into details, the configuration on the base reflects the local configuration in total space S_3 or $\{3, 3, 5\}$. For instance, a fiber is surrounded by 5 parallel fibers; but the parallelism is in the Clifford sense and the 5 fibers wind around the “central” one. The full B-C chain is represented by an equilateral triangle on the base. If we let the B-C chain close on itself, with its helices winding on a torus instead of a cylinder, the triangle on the base is representative of the three type- $\{3\}$ helices. In curved space these three helices are three intertwined great circles of the Hopf fibration of S_3 .

Consider now three triangles of the base icosahedron, representative of three B-C helices, in a configuration described (in grey) in Figure 9b. The B-C helices are intertwined in a right-hand screw, like the strands of a rope. Three vertices on the base (open circles in Fig. 9b), one from each triangle, define the core of the 3-helices screw.

On each B-C helix, represented by a grey triangle in Figure 9b, we construct one, left-handed PPII helix. On this helix with the collagen sequence $(\text{Gly-X-Y})_n$, all Gly amino acids are on the same type- $\{3\}$ helix (Fig. 7a), which is a fiber of the Hopf fibration. On each fiber of $\{3, 3, 5\}$ with 10 sites, there are 5 Gly. The hydrogen bonds occur between (atom N of) the Gly amino acid of one helix and (atom C of) the X amino acid of another, at the same “altitude”. The individual collagen helices are left-handed, but the Gly core is right-handed, as is the whole molecule. In fact, the 15 Gly of the core are on the 30 sites of a right-handed B-C helix, which constitutes the core of the triple-helix collagen molecule. More precisely, because Gly is the only non-chiral amino acid (its side group is a H atom), the core of the collagen molecule is constituted by 30 H atoms on a B-C helix. There are 2 H atoms per Gly: one is its distinctive side group, the other is the H atom which binds to the central C_α of all amino acids.

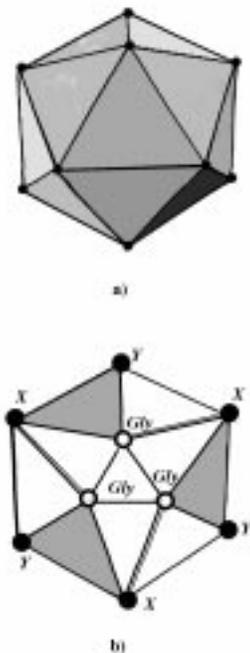


Fig. 9. (a) The base of the Hopf fibration of the $\{3,3,5\}$ -polytope is an icosahedron. Each vertex of the icosahedron is only a representation (a projection) of one fiber (type-3 helix in our notation). Three fibers, represented by a triangular face of the icosahedron, constitute a Coxeter helix. (b) A collagen molecule is represented by three grey triangles each decorated by one PPII helix (see Fig. 7). With some care, (b) can be seen as a local representation of a flat cross section of the triple helix molecule. Hydrogen bonds (double lines) are bridges between the Gly of one helix and the X of another, approximately horizontal. By contrast, in an α -helix, the hydrogen bonds are intra-helical and almost vertical. Notice the right handed chirality of the Gly core and of the triple collagen molecule opposite to the left handed chirality of the PPII helices.

Because we have chosen a right-handed Hopf fibration, all the type- $\{3\}$ and type- $\{1\}$ helices are right-handed like the triple-helix, but the PPII helices are left-handed.

Is such a model molecule realistic? Because it is based on the geometrical properties of spherical curved space we may have some doubt. Let us reverse the construction. We begin by packing tightly an even number of H atoms on a B-C helix. This is a highly efficient means of packing spherical atoms [6]. Then, the symmetry of this helix extends, through the C_α of the Gly, to the three procollagen chains. In curved space, these three chains would have the same length as that of the core, made of Gly only. But, in Euclidean space, the three external chains are greatly extended. This can be seen easily in a stereographic projection of the base. Moreover, the whole molecule can be extended ad infinitum, through inflation, as was done for a single B-C helix in Euclidean space, in Section 2.1.4.

Thus, if the curved space model is an abstract, but ideal template, its projection in real space is a realistic model which takes into account the difference in size between the side group of Gly and that of other amino acids

(mainly Pro), has the proper length of H bonds, and explains the diverse chiralities in the triple-helix collagen molecule.

It thus seems that the PPII helix of one individual collagen chain is caused by the abundance of proline in the sequence of amino acids. But the intertwined, triple helix structure of the collagen molecule is due to the fact that every third amino acid is a glycine. The H atoms, half of which are the side groups of Gly, pack tightly (as a B-C helix) in the core of the triple helix, and interhelical H bonds stabilize the molecule. Whether the collagen molecule is a tertiary or a secondary structure is a matter of semantics. But it exhibits helical structures, of opposite chirality, at two different scales.

5 Conclusions

Biopolymers are, at a first level of organization, one-dimensional sequences like any polymer chains; but, at a second level, various interactions impose organized structures in space. The different types of structures in proteins: primary, secondary or tertiary are related to close-packed structures in one, two or three dimensions. The α -helix, considered here as a 2D close-packing, a triangular lattice rolled on a cylinder, is an essential step in protein folding. This is, incidentally, one of the reasons for the success of the hydrophobic cluster analysis (H.C.A.), which predicts the folding pattern of several proteins [25]. Viewing the α -helix as a 2D structure makes clear the geometrical and topological constraints required. For instance, the number of residues observed per turn is imposed by the geometry of the triangular packing on a cylinder. Chiralities are also well described from this point of view.

In 3 dimensions, tetrahedral packing is the tiling corresponding to triangular packing on a surface. It is only undistorted and defect-free in curved space: This is the $\{3,3,5\}$ polytope, and its fiber, the B-C helix. The B-C helix can be put into Euclidean space, and extended (inflated), if necessary. With small distortions of the tetrahedra, it can also coexist, tightly packed, with other B-C helices. This finely tuned local geometry with minimal distortion is extendable to longer helices without increasing the distortion.

While the exact geometry of biological helices may appear complicated, their topology is determined simply and directly by steric considerations. We have shown mathematical helices which have exactly the topology of the α -helix, or of the collagen helix and its triple helix molecule.

In fact, nearly fifty years after Pauling's paper on the α -helix [13], we still do not know why the α -helix is near-universal as *the* helical structure in proteins (topological universality, i.e. about 3.6 aa per turn, and not about 2.6 or 4.6), why collagen is different, why collagen should be a periodic sequence ...-Gly-X-Y-..., why Gly, why period three. These are general questions, of fundamental interest, at least to a physicist. This very fundamental difference between α -helices and collagen (in their occurrence, and in the scale of the structure) may well tell us that the biological function of collagen lies in its (particular)

structure, whereas, in other proteins, the α -helix, as an universal secondary structure, is only a supramolecular element, with the (specific) function of the protein in the ternary structure. These are the kind of questions, which we have attempted to answer in this paper, or at least to justify.

We have mainly discussed the α -helix and the collagen, but there are other biopolymers related to the B-C chain: DNA triple helices can be viewed as a packing of base pairs along a B-C helix, the three strands being the three type- $\{3\}$ helices. Then, the classic double helix is simply the triple helix which has lost a strand, leaving the major groove in the gap. Actually, the biological world is very rich in helices moulded chiefly by close-packing. These close-packed helices occur at all sizes, not only at the molecular level, but also at intermediate, through to macroscopic dimensions. Microtubules in cells and phylotactic helices in plants are two good examples.

Appendix: The quasicrystalline Coxeter helix

The Coxeter chain is the B-C helix in Euclidean space, decorated by spheres on its vertices. It is not a periodic structure, because ξ is not a rational number. We construct a modified quasicrystalline Coxeter helix, very close to the Coxeter chain. This chain would have a must have a number of edges per turn ϖ close to ξ which is a quadratic irrational. This quasicrystalline Coxeter helix will have a sequence of approximant structures, related by inflation.

We follow a procedure similar to the “cut and projection” method used to construct one-dimensional quasicrystals. Consider the triangular lattice and the vector \mathbf{b}_1 , base of the cylinder. Recall the construction of the vector \mathbf{b}_2 , approximatively perpendicular to \mathbf{b}_1 , in Figure 2. We would like to construct a sequence of vectors \mathbf{q}_i , of increasing length, becoming more and more orthogonal to \mathbf{b}_1 as i increases.

The “cut and projection” method requires the construction of a band, or acceptance domain. The band is a region inside the cylindrical strip defined in Figure 2, bounded by two lines parallel to the axis of the cylinder. All vertices of the triangular lattice lying within the band, are end points of lattice vectors which are more or less perpendicular to \mathbf{b}_1 . The band is constructed by translating the unit cell $(-\mathbf{a}_1, \mathbf{a}_2)$, keeping its origin on a straight line. The direction δ of the straight line is determined by successive rational approximants, according to the procedure detailed below, that is by inflation. δ is not exactly orthogonal to \mathbf{b}_1 . The basis vectors $(-\mathbf{a}_1, \mathbf{a}_2)$ are chosen so that the band straddles the straight line of direction δ .

The basis vectors $-\mathbf{a}_1 = (1, 0) = \mathbf{q}_{-1}$ and $\mathbf{a}_2 = (0, 1) = \mathbf{q}_0$ are the smallest vectors in the band. The next couple of vectors approximating direction δ , nearly orthogonal to \mathbf{b}_1 are $\mathbf{q}_1 = -\mathbf{a}_1 + 2\mathbf{a}_2 = (1, 2)$ and $\mathbf{q}_2 = -\mathbf{a}_1 + 3\mathbf{a}_2 = (1, 3)$, so that $\mathbf{b}_2 = \mathbf{q}_1 + 2\mathbf{q}_2$. Since $\mathbf{q}_1 \wedge \mathbf{q}_2 = \mathbf{q}_{-1} \wedge \mathbf{q}_0$, these two vectors constitute also a primitive unit cell of the triangular lattice. The transfor-

mation is defined by the matrix

$$S = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

Repeated applications of the transformation yield successive primitive cells $(\mathbf{q}_{-1+2n}, \mathbf{q}_{2n})$ by

$$\begin{pmatrix} \mathbf{q}_{-1+2n} \\ \mathbf{q}_{2n} \end{pmatrix} = S \begin{pmatrix} \mathbf{q}_{-1+2(n-1)} \\ \mathbf{q}_{2(n-1)} \end{pmatrix} = S^n \begin{pmatrix} \mathbf{q}_{-1} \\ \mathbf{q}_0 \end{pmatrix}. \quad (\text{A.1})$$

The original unit cell $(-\mathbf{a}_1, \mathbf{a}_2) = (\mathbf{q}_{-1}, \mathbf{q}_0)$ becomes more and more elongated, with its basis vectors stretched along the direction δ . Notably, $\mathbf{q}_3 = \mathbf{b}_2 = \mathbf{q}_1 + 2\mathbf{q}_2 = -3\mathbf{a}_1 + 8\mathbf{a}_2$ (fig. 2), and $\mathbf{q}_4 = \mathbf{q}_1 + 3\mathbf{q}_2 = -4\mathbf{a}_1 + 11\mathbf{a}_2$.

Accordingly, successive strips are constructed as parallelograms of edges \mathbf{b}_1 and \mathbf{q}_i . The strip is folded into a cylinder by identification of the two sides parallel to \mathbf{q}_i . The B-C helix is a helix of type- $\{1\}$, winding around the cylinder, as illustrated in Figure 2 for $i = 3$.

Alternatively, the strip $(\mathbf{b}_1, \mathbf{q}_{2n+1})$ on the triangular lattice can be constructed as a succession of three strips, two $(\mathbf{b}_1, \mathbf{q}_{2n})$, followed by one $(\mathbf{b}_1, \mathbf{q}_{2n-1})$, because $\mathbf{q}_{2n+1} = 2\mathbf{q}_{2n} + \mathbf{q}_{2n-1}$. For example, the strip in Figure 2, $(\mathbf{b}_1, \mathbf{q}_3)$, is composed of two $(\mathbf{b}_1, \mathbf{q}_2)$ followed by one $(\mathbf{b}_1, \mathbf{q}_1)$. This stacking of three parallelograms can be folded into a cylinder by identifying opposite sides. On the triangular lattice, this stacking is the same as the cylinder $(\mathbf{b}_1, \mathbf{q}_3)$. Similarly, the strip $(\mathbf{b}_1, \mathbf{q}_{2n})$ is a succession of two strips, one $(\mathbf{b}_1, \mathbf{q}_{2n-1})$, followed by one $(\mathbf{b}_1, \mathbf{q}_{2n-2})$, because $\mathbf{q}_{2n} = \mathbf{q}_{2n-1} + \mathbf{q}_{2n-2}$. This construction is the inflation transformation in quasicrystals. Here, inflation is applied on a triangular lattice, on the cylinder around which the B-C helix is wound.

Here is the recipe for obtaining the number of edges per turn for the helix of type- $\{1\}$. Simply count the number of edges and turns for each successive parallelogram, and add up. Successive strips $(\mathbf{b}_1, \mathbf{q}_i)$ yield, as i increases, rational approximants of the winding number (number of edges per turn) of the infinite, quasicrystalline Coxeter helix. The first approximant is the parallelogram $(\mathbf{b}_1, \mathbf{q}_{-1})$, with 2 edges for one turn. Following approximants are $(\mathbf{b}_1, \mathbf{q}_0)$, with 3 edges for one turn, and $(\mathbf{b}_1, \mathbf{q}_1) = (\mathbf{b}_1, 2\mathbf{q}_0 + \mathbf{q}_{-1})$ has $3 + 3 + 2$ edges for $1 + 1 + 1$ turns. Because $\mathbf{q}_2 = 2\mathbf{q}_0 + \mathbf{q}_{-1} + \mathbf{q}_0$, $(\mathbf{b}_1, \mathbf{q}_2)$ has $3 + 3 + 2 + 3 = 11$ edges for 4 turns. This procedure can be iterated by inflation, replacing at each (double) step, 2 by $3 + 3 + 2$, and 3 by $3 + 3 + 2 + 3$. The numbers of 2's and 3's, at level n of the iteration, are denoted by n_2^n and n_3^n , respectively. Thus, $n_2^{-1} = 1$, $n_3^{-1} = 0$; $n_2^0 = 0$, $n_3^0 = 1$; $n_2^1 = 1$, $n_3^1 = 2$; $n_2^2 = 1$, $n_3^2 = 3$. Then, since

$$\begin{pmatrix} n_2^{n+1} \\ n_3^{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} n_2^n \\ n_3^n \end{pmatrix} \quad (\text{A.2})$$

there are $3n_3^n + 2n_2^n$ edges for $n_3^n + n_2^n$ turns at level $(n+1)$. Thus, the number of edges per turn $\varpi = \frac{3n_3^n + 2n_2^n}{n_3^n + n_2^n} = \frac{n_3^{n+1}}{n_2^{n+1}}$, which is $\varpi = 1 + \sqrt{3}$, in the limit of large n . This number is close to $30/11$, but it is now a quadratic irrational,

which confirms the quasicrystalline, inflatable nature of this modified Coxeter helix. The largest eigenvalue of this matrix (the Perron root) $\lambda = 2 + \sqrt{3}$, gives the inflation multiplier $n_3^{n+1}/n_3^n = n_2^{n+1}/n_2^n$ in the large n limit.

In fact, the two numbers ξ and ϖ are extremely close. The number ϖ , being a quadratic irrational, has a periodic sequence of integers c_i (alternatively 1 and 2) in its continued fraction expansion:

$$\begin{aligned}\varpi &= c_0 + 1/\{c_1 + 1/(c_2 + 1/[c_3 + \dots])\} \quad (\text{A.3}) \\ \varpi &= [c_0, c_1, c_2, c_3, \dots] = [2, 1, 2, 1, 2, \dots].\end{aligned}$$

This compares to $\xi = [2, 1, 2, 1, 2, 1, 1, 2, 1, 7, 6, 1, 1, 5, 4, \dots]$, the corresponding infinite expansion of ξ in continuous fraction, which is nonperiodic, so that ξ is an irrational but not a quadratic irrational (it is not even algebraic). The two $[c_i]$ sequences are the same through to c_5 , and differ seriously at c_9 . Their first rational convergents (approximants) are identical, and the helices indistinguishable through 112 steps. For reference, $30/11 = [2, 1, 2, 1, 2]$ is the rational number of edges per turn for the B-C helix.

The difference between the Coxeter chain and its quasiperiodic relative can be understood by considering the rolling of the strip tiled by triangles into a cylinder. For the Coxeter chain, it is necessary to fold the strip along the edges of the tiling, in order to have straight edges of flat triangles on the cylinder. The normal to the surface of the cylinder is not a continuous function of the position and the mean curvature is not defined on the edges of the tiling. In the quasicrystalline chain, the strip is gently rolled into a cylinder with a constant mean curvature everywhere, so that the edges of the triangles are curved lines, which affect slightly the metric properties (distances, areas) of the tiling.

The quasicrystal approach presented above is based chiefly on crystallography. It is also possible to use arithmetic. Let the lattice vector $\mathbf{q} = c\mathbf{a}_1 + d\mathbf{a}_2$ with c, d integers, and the rational $x = -d/c$. The chain can be drawn in the extended triangular lattice as the lattice vector $L(\mathbf{a}_1 + \mathbf{a}_2)$ in the direction $(\mathbf{a}_1 + \mathbf{a}_2)$. Its length $L = (3d - 2c)$ is also the area of the strip (in units of $\sqrt{3}/2$). The winding number w (of the helix around the cylinder) is the component along \mathbf{b}_1 of the chain vector, namely $L(\mathbf{a}_1 + \mathbf{a}_2) - \mathbf{q} = w\mathbf{b}_1$. Thus $w = d - c$. The rational $L/w = y = (3x + 2)/(x + 1)$ is the number to be inflated. We have the B-C solution $y = 30/11$ (or $y = [2, 1, 2, 1, 2]$, in continued fraction expansion), for $x = 8/3$. But also, for $x = 0/1$ ($\mathbf{q}_{-1} = -\mathbf{a}_1$), $y = 2/1$ (or $y = [2]$), and for $x = 1/0$ ($\mathbf{q}_0 = \mathbf{a}_2$), $y = 3/1 = [2, 1]$. The approximants are therefore the rational convergents of the quadratic irrational $\varpi = 1 + \sqrt{3} = [2, 1, 2, 1, 2, 1, \dots]$, solution of the equation $\varpi = 2 + 1/[1 + (1/\varpi)] = [3\varpi + 2]/[\varpi + 1]$. This produces a sequence of inflated words w_i , $A \rightarrow B \rightarrow BBA \rightarrow BBAB \rightarrow BBABBBABBBA \rightarrow BBABBBABBABBAB \rightarrow BBAB BBAB BBA BBAB BBAB BBAB BBA BBAB BBAB BBAB BBA \dots$, accord-

ing to the rule $w_i = (w_{i-1})^{c_i} \cdot w_{i-2}$, where the c_i are the alternating integers 2 and 1 in the continued fraction expansion (7) of ϖ . The words can also be obtained by using the substitution rule $A \rightarrow BBA, B \rightarrow BBAB$, at each successive (double) step. The double inflation step betrays the period 2, 1 in the continued fraction expansion. Thus AA , isolated B and $BBBB$ are forbidden. The words can be translated directly into the pattern of Figure 2: Starting from the origin, B is a step along the helix, and AB a step which crosses the border of the strip and completes one winding turn around the cylinder.

References

1. C. Chothia, M. Levitt, D. Richardson, Proc. Natl. Acad. Sci. USA **74**, 4130 (1977).
2. A. Adzhubei, M. Sternberg, Protein Sci., **3** 2395 (1994).
3. C. Branden, J. Tooze, *Introduction to Protein Structure* (Garland Publ. New York, London, 1991).
4. A.H. Boerdijk, Philips Res. Rep. **7**, 303 (1952).
5. H.S.M. Coxeter, Can. Math. Bull. **28**, 385 (1985).
6. J.F. Sadoc, R. Mosseri, *Frustration géométrique* (Eyrolles, Paris 1997); *Geometrical Frustration* (Cambridge Univ. Press, 1999).
7. H.S.M. Coxeter, *Regular Complex Polytopes* (Cambridge Univ. Press, 1973).
8. N. Pittet, P. Boltenhagen, N. Rivier, D. Weaire, Europhys. Lett. **35**, 547 (1996).
9. R.O. Erickson, Science (New York) **181**, 705 (1973).
10. P.A.B. Pleasants, in *Elementary and Analytic Theory of Numbers* (Banach Center Publications, PWN, Warsaw, 1984), pp. 439-461.
11. W.F. Lunnon, P.A.B. Pleasants, J. Math. Pures Appl. **66**, 217 (1987).
12. W.F. Harris, Sc. Am. **237**, 130 (1977).
13. F. Crick, *What Mad Pursuit*, Alfred P. Sloan Foundation Series, Basic Books 1988, p. 58.
14. S.M. Miück, G.V. Martinez, W.R. Fiori, A.P. Todd, G.L. Milhauser, Nature **359**, 653 (1992).
15. A. Frey-Wyssling, Nature **173**, 596 (1954).
16. L. Zhang, J. Hermans, J. Am. Chem. Soc. **116**, 11915 (1994).
17. M.B. Pyshnov, J. Theor. Biol. **87**, 189 (1980).
18. W.F. Harris, H.D. Chandler, H.R. Hepburn, South Afr. J. Sci. **72**, 25-26 (1976).
19. F. Gaill, Y. Bouligand, Tissue & Cell **19**, 625-642 (1987).
20. A.G. Walton, J. Blackwell, Biopolymers (Academic Press, New York 1973).
21. Z. Szabo, <http://iona.cryst.bbk.ac.uk/assignments/projects/szabo>.
22. K. Okuyama, M. Takayanagi, T. Ashida, M. Kakudo, Polymer J. **9**, 341 (1977).
23. Y. Bouligand, SFP-JMC5 Orléans, poster (1996).
24. V. Sasisekharan, M. Bansal, Curr. Sci. **59**, 863 (1990).
25. I. Callebaut, G. Labesse, P. Durand, A. Poupon, L. Canard, J. Chomilier, B. Henrissat, J.P. Mornon, CMLS, Cell. Mol. Life Sci. **53**, 621 (1997).