ELSEVIER

Materials Science and Engineering 294–296 (2000) 397–400

MATERIALS SCIENCE & ENGINEERING A

www.elsevier.com/locate/msea

Boerdijk–Coxeter helix and biological helices as quasicrystals

J.F. Sadoc [a], N. Rivier [b,*]

[a] *Laboratoire de Physique des Solides, Université Paris-Sud, 91405 Orsay cedex, France*
[b] *Laboratoire de Dynamique des Fluides Complexes, Université Louis Pasteur, 67084 Strasbourg cedex, France*

Received 1 September 1999; accepted 17 February 2000

Abstract

Helices and dense packing of spherical objects are two closely related problems. The Boerdijk–Coxeter helix (B–C), a linear stacking of regular tetrahedra, is a very efficient solution to some close-packing problems, including protein folding. The structure of biological helices (α-helix, collagen) is determined chiefly by steric repulsion. Thus, maximum density leads directly to structures related to the B–C helix. We show that the B–C helix is a quasicrystalline structure with $1+\sqrt{3}$ edges per turn, which can be disclinated and inflated. The collagen molecule consists of three intertwined helicoidal chains on the fibres of an inflated Hopf fibration. The central core is a B–C helical packing of the hydrogen atoms side-groups of the amino acid Gly.

Keywords: Helix; Collagen; Protein folding; Inflated Hopf fibration

1. Close-packing and the Boerdijk–Coxeter helix

Our starting point is the close relation between helices and dense packing. The structural unit of dense packing of spherical objects is a (regular) tetrahedron. Tetrahedra cannot fill Euclidean space regularly, but they make up the regular polytope $\{3,3,5\}$ in positively curved space, where each object is in an icosahedral environment. Polytope $\{3,3,5\}$ is a tesselation of S_3 by regular tetrahedra. Now, the (identical) fibres of the Hopf fibration of S_3 are helices [1]. Remains the usual problem of accommodating S_3 into our Euclidean environment, which is done through decurving by topological defects, disclinations, which lengthen, thicken and decompact the helices.

This relationship between dense packing and helices can be seen directly in Euclidean space, through stacking the tetrahedra along one direction. One obtains the Boerdijk–Coxeter (B–C) helix (Fig. 1), a chain of tetrahedra, whose edges constitute helicoidal chains (also called helices, for short) winding in directions defined by the six neighbours of a vertex. There are three helices along the direction $\{3\}$ most nearly parallel to the axis of the B–C helix. One distinguishes also two helices in the $\{2\}$ direction and one single helix in direction $\{1\}$. The B–C helix is labelled $\{3,2,1\}$ in phyllotactic notation (labelling vertices by a natural integer n, in the order of their projection on the axis, vertex n has neighbours $n\pm3$ along helix $\{3\}$, $n\pm2$ along helix $\{2\}$, and $n\pm1$ along the polypeptide chain $\{1\}$) [2].

Surprisingly, the B–C helix is not periodic, because the distances separating the centres of neighbouring tetrahedra, and the pitch of the helical chain, are not commensurate. More surprisingly, the helix made of perfectly regular tetrahedra is not even quasiperiodic (although it fails by an extremely small amount). Obviously, there is no reason why the tetrahedra should be perfectly regular in biological structures as they are made of different amino acids, and inflation of a tight local arrangement is desirable.

It is useful to describe the B–C helix by putting vertices on a cylinder and unfolding the cylinder on a flat surface; this surface is then tiled with equilateral triangles (Fig. 2(a)). Moreover, the B–C helix can be built by taking an actual sheet of paper on which a triangular lattice has been drawn, cutting a strip three triangle-wide, folding the type-$\{2\}$ edges inwards, type-$\{3\}$ and -$\{1\}$ outwards, and gluing. Note that the edges are all straight, and the cylinder is tiled by (flat) equilateral triangles, not quite vertical (the cause of a famous ill-defined limit for calculating the area of a cylinder [3,12]).

Let us map the B–C helix on a cylinder in Euclidean space. The coordinates of the nth vertex A_n of the helix are $(\cos n\theta, \sin n\theta, nc)$. The pitch c is the translational part of the helical motion. The squared distance, or edge length d^2 between vertex n and any neighbour $n+m$ is equal to $d^2(n,n+m)=2-2\cos m\theta+m^2c^2$ for all n, and $m=\pm1,\pm2,\pm3$, here. Eliminating c, we remain with one

* Corresponding author. Tel.: +33-3-88-35-81-41; fax: +33-3-88-35-80-99.
E-mail address: nick@ldfc.u-strasbg.fr (N. Rivier).

0921-5093/00/$ – see front matter
PII: S0921-5093(00)01136-9

Fig. 1. B–C helix as a stacking of tetrahedra.

equation for $x=\cos\theta$, which factorises as $(x-1)^2(3x+2)=0$. The trivial roots $x=1$ can be discarded, so that $\cos\theta=-2/3$, $\theta=131.810^\circ$, and the pitch, in units of edge length is $c/d=1/\sqrt{10}=0.3612$.

The number of edges per turn is given by $\xi=2\pi/\theta=2.7312=[2,1,2,1,2,1,1,2,1,7,6,1,1,\ldots]$, where the sequence of integers $[\ldots, c_i, \ldots]$ is shorthand for the continuous fraction representation $c_0+1/\{c_1+1/[c_2+1/(c_3+1/\ldots)]\}$. We note several rational approximants, notably $30/11=[2,1,2,1,2]=2.727272\ldots$, which is the B–C helix on a torus in curved space (Clifford torus in the Hopf fibration of $\{3,3,5\}$ [1,4]).

The B–C helix is not a quasiperiodic structure, because ξ is not a quadratic irrational. It is not context-free inflatable (i.e. going from one finite structure (a rational approximant of ξ) to the next, in an automatic fashion, independent of the length of the helix). To make it inflatable, the sequence $[\ldots, c_i, \ldots]$ should be (ultimately) periodic (the period is underlined), and ξ, a quadratic irrational [5].

The quasiperiodic, inflatable helix $(3,2,1)$ has $\omega=[\underline{2,1}]=1+\sqrt{3}=2.73205$ edges per turn. It is identical to the B–C helix through 112 edges. Both helices have the same convergent 30/11, and thus, originate from the same B–C helix in curved space; but generation of the quasicrystalline helix has been by inflation, a symmetry, whereas the construction of the B–C helix in Euclidean space has been a geometrical accident. We note that both helices have three, principal or intermediate convergents in common beyond 30/11, namely 41/15, 71/26 and 112/41, so that inflation is obviously a generating factor, even of the B–C helix. The B–C helix 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 densest packing of 120 spheres in curved space. Polytope $\{3,3,5\}$ has a discrete Hopf fibration (the

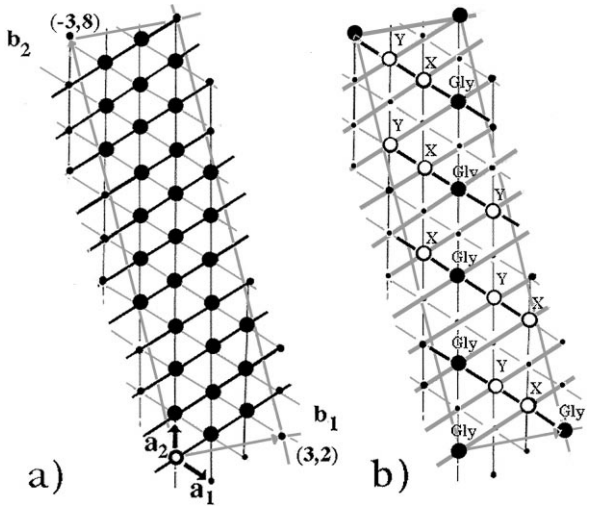

Fig. 2. (a) The B–C helix on a flat strip tiled with equilateral triangles. The long sides should be identified. (b) A B–C helix with the collagen sequence –Gly–X–Y–. The polypeptide chain PPII is a left-handed helix of type-{2}, with twice the pitch of the B–C helix. Direction-{3} along $\boldsymbol{a}_2$, direction-{2} along $-\boldsymbol{a}_1$, direction-{1} along $\boldsymbol{a}_1+\boldsymbol{a}_2$.

fibres are 12 great circles with 10 vertices each) (Fig. 3). Points on the same fibre are projected on the base space as a single point, which is representative of a whole fibre. The base space of the Hopf fibration of S_3 is a sphere S_2. Here, the 12 fibres are represented by the 12 vertices of an icosahedron on S_2. Note that the base space is a set of projected points, not actual points of the polytope. In particular, the points of Fig. 3 are not at the same altitude on the fibre. Nevertheless, the configuration on the base reflects somewhat the local configuration in total space $\{3,3,5\}$. For instance, a fibre is surrounded by five parallel fibres; but the parallelism is in the Clifford sense (being at constant distance of each other) and the five fibres wind around the 'central' one.

A useful way to picture the Hopf fibration is to take a torus, obtained by folding a rectangle, on which a diagonal and two lines parallel to it have been drawn; these three lines are Hopf fibres. They fold into three intertwined great circles, each with winding number 1×1 on the torus

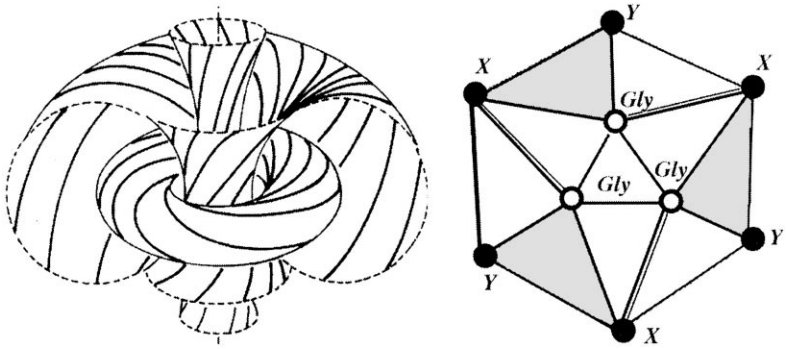

Fig. 3. Left hand side: Hopf fibration of S_3 by great circles (stereographic projection). The base of the Hopf fibration of $\{3,3,5\}$ is an icosahedron. Right hand side: on the icosahedron is represented one collagen molecule. Double lines are the projection of hydrogen bonds.

(Fig. 2). This is actually a representation of the B–C helix in curved space. The three great circles, of 10 vertices each, are the three type-$\{3\}$ helices made of edges most parallel to the generator of the torus. The polypeptide chain, for α-helices, is the single type-$\{1\}$ helix (thick line) winding 30×3 times around the torus, with $30/11=[2,1,2,1,2]$ edges per turn, which is indeed a convergent of both ξ (for the B–C helix in Euclidean space), and ω, its quasiperiodic, inflatable realisation. Clearly, gentle decurving of $\{3,3,5\}$ yields the latter. Space is effectively flattened (Euclidean) after only a few rounds of decurving operations (see [1]), so that it is impossible to distinguish the two Euclidean structures. To construct the quasicrystalline B–C helix on the inflatable torus, it is necessary to define a flat strip (Fig. 2(a)), a parallelogram tiled by equilateral triangles of the required area, which forms a torus by identification of opposite sides. The sides of the strip are the lattice vectors $\boldsymbol{b}_1$ and $\boldsymbol{q}$, which should be as orthogonal as possible. The original Clifford torus has unit cell $(3\boldsymbol{a}_1, 10\boldsymbol{a}_2)$, and strip $\boldsymbol{b}_1=3\boldsymbol{a}_1+2\boldsymbol{a}_2$, $\boldsymbol{q}=\boldsymbol{b}_2=-3\boldsymbol{a}_1+8\boldsymbol{a}_2$ [6]. The idea is to construct a sequence of lattice vectors $\boldsymbol{q}$, of increasing length and more and more orthogonal to the given $\boldsymbol{b}_1$. One can use a crystallographic procedure similar to the cut-and-projection method, or, more simply, arithmetic: Write $\boldsymbol{q}=c\boldsymbol{a}_1+d\boldsymbol{a}_2$, with c, d integers, and the rational $x=-d/c$. The chain of length L can be drawn in the extended triangular lattice as a lattice vector $L(\boldsymbol{a}_1+\boldsymbol{a}_2)$. $L=(3d-2c)$ is also the area of the strip, in units of $\sqrt{3}/2$. The winding number ω of the chain around the cylinder is the component $L(\boldsymbol{a}_1+\boldsymbol{a}_2)-\boldsymbol{q}=\omega\boldsymbol{b}_1$ of the chain vector along $\boldsymbol{b}_1$. Thus, $\omega=d-c$. The rationals $L/\omega=y=(3x+2)/(x+1)$ are the number of edges per turn for each approximant of the sequence of inflated helices. This is the standard inflation relation for y, with first convergents $y=2/1=[2]$ (for $x=0/1$) and $y=3/1=[2,1]$ (for $x=1/0$). Also, for $x=8/3$, $y=30/11=[2,1,2,1,2]$ is the helix on the Clifford torus in $\{3,3,5\}$. The limit of the sequence of helices on tori increasing in length, is the quasicrystalline B–C helix, with $y=\omega=\overline{[2,1]}=1+\sqrt{3}$ edges per turn, solution of the equation $\omega=(3\omega+2)/(\omega+1)$. The structures of the helices can be read immediately as the sequence of inflated words, A→B→BBA→BBAB→BBABBBABBBA→BBABBBA-BBBABBAB→BBAB BBAB BBA BBAB BBAB BBAB BBA BBAB BBAB BBAB BBA…, obtained according to the concatenation rule $w_i=[w_{i-1}\ldots w_{i-1}]\ w_{i-2}$. (The bracket $[\ldots]$ contains c_i (w_{i-1})). Here, $c_i=2$ or 1, alternatively. The order of the words w_i is obtained by noticing that its successive principal convergents are alternatively larger (thus, ending as BA) and smaller (ending as AB) than ω. (cf. representation (F. Klein [5]) of ω as a line of slope $1/\omega$ in a square lattice, and of the sequences as staircases of horizontal (B) and vertical (A) steps). The words can also be obtained from the substitution rule A→BBA, B→BBAB at each (double: period [2,1]) step. Words are translated into pattern (Fig. 2(a)): starting from the origin, B is an ordinary step along the chain, AB a step which crosses the border and completes a winding turn around the cylinder.

2. Biological helices

Among biological molecules arranged in helices, proteins, i.e. polypeptide chains [7], wind around the B–C helix (e.g. [6], the α-helix, which is the most common secondary structure of native proteins). Here, we describe collagen, which is a highly inflatable molecule, it can be found at almost any scale, from the molecule to the whole animal (e.g. in marine worms) [8,13]. Our purpose is to show how and why these helices can be inflated. They are, indeed, biological quasicrystals.

A protein is a stiff polymer chain of amino acids. The amino acid consist of a backbone –[N–C_α–C]–. Attached to the central carbon atom C_α, are, besides the carboxy and amino groups of the backbone, one hydrogen atom, and one side-group which specifies the amino acid. Glycine (Gly), with H as side group, is the smallest amino acid, and the only one which is not chiral. We will see that the size and symmetry of Gly govern the structure of the collagen molecule. In their native state, proteins are very tightly and uniquely folded. The backbone chain is fairly stiff (the peptide bond, and thus, the link between two central carbons can be considered as rigid), so that steric repulsion (between different side-groups, and between side-group and the backbone) plays an essential part in the native structure.

3. Collagen

Collagen is a protein –[Gly–X–Y]–, where either X or it Y are almost always the proline amino acid (Pro). It forms helicoidal structures very common in biology. The structure of artificial collagen –[Gly–Pro–Pro]– has been extensively studied, at the levels of molecule, fibril, and crystal of fibrils [9,10]. A collagen molecule consists of three polypeptide chains intertwined in a triple helix. Each individual polypeptide chain takes the helical conformation when associated with the other two chains. Hydrogen bonds connect different helices and stabilize the triple helix collagen molecule. The helical conformations are due to the periodicity of the chain, every third amino acid is a Gly, the amino acid with the smallest side group (H). A larger side group would inhibit close contact between the three chains. The molecule is stabilized by (extrinsic, horizontal) hydrogen bonds between the backbone amide (N) of a Gly, of one chain and the backbone carbonyl (C) of residue X of another (Fig. 3). The structure of a single chain of collagen, procollagen, is called PPII. It is a left-handed helical chain with a pitch twice as large as that of the B–C chain. These facts, together with the chemical (Gly) periodicity three, makes the structure of PPII evident (Fig. 2(b)). Period 3 imposes the same strip and triangular tiling (3, 2, 1) as for B–C helices. We will see why the PPII chain covers only half the vertices of the tiling. It is a consequence of close-packing. The molecule of collagen is maximally packed, even if its constituting helices are looser.

In a collagen molecule, three left-handed PPII helical chains 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. The Gly of each chain occupy half the sites of a type-$\{3\}$ helix, a geodesic, and fibre of the Hopf fibration of $\{3.3.5\}$ (Fig. 3). The other two type-$\{3\}$ helices of the B–C helix or of the torus in curved space have half their sites occupied one by the X, the other by the Y amino acids of the polymer. Each PPII chain is made of three fibres, and is represented as one triangle on the base of the fibration (shaded in Fig. 3). The collagen molecule is made of three, intertwined PPII helices, nine fibres, represented on the base of the Hopf fibration by the three shaded triangles of Fig. 3. The core of the collagen molecule is itself a three helices screw, represented on the base by the central triangle made of one vertex from each shaded triangle. It is a Clifford torus made of three interwound geodesics, right-handed type-$\{3\}$ helices of five Gly and five empty vertices. The side-groups of Gly are inside the core. Only Gly, the smallest amino acid, is small enough to be thus accommodated. Moreover, since each Gly has two H atoms as side-groups (one is its distinctive side-group, the other is the H atom bound to the central C_α of all amino acids), the core of the collagen molecule is a perfect B–C helix of 30 close-packed H atoms. Finally, since the collagen molecule is based on the B–C helix, it can be inflated, by the multiplier $1+\sqrt{3}$, as we have seen earlier on.

This explains why collagen is not an α-helix, why it is a periodic sequence, why Gly, why period 3, why it can be found at scales so different, in an entirely natural way. It will come to no surprise that the transverse array of collagen molecules should also be based on the same inflation multiplier $1+\sqrt{3}$, since the whole structure proceeds from decurving the isotropic polytope $\{3,3,5\}$, through its Hopf fibration. The molecules are vertices of a 2D tiling by squares and equilateral triangles, with 12-fold rotation symmetry. This tiling must be an approximant. It cannot be a fully inflatable 12-fold quasicrystal, because the collagen molecule is itself inflatable in the longitudinal direction, and a space-filling structure based on 12-fold rotation symmetry cannot be inflatable in all the 3 space directions. Discrete subgroups of SO(3) involving rotations by $2\pi/12$ are C_{12} and D'_{12}, have a single axis of 12-fold rotations [11]. This is why 12-fold quasicrystals are periodic in the third directions, like 10- and eight-fold quasicrystals. Diffraction data suggest [9,10] a tiling dual of the main skeleton of the sigma phase (β-uranium), with vertex figure $(3^2.4.3.4)$ and symmetry group p4gm. It is based on the approximants 8/3 (square diagonals) and 11/4 (period of the square unit cell) of $1+\sqrt{3}$. The sigma phase is indeed one of the simplest tetrahedrally close-packed phase, and the only one to be realized unambiguously in a mono-atomic substance, which justifies our contention that the secondary structure of biological helices can be understood as a close packing of identical amino acids. This is, of course, a first approximation, but it is surprisingly accurate.

References

[1] J.F. Sadoc, R. Mosseri, Frustration Géométrique, Eyrolles, Paris 1997, Geometrical Frustration, Cambridge University Press, Cambridge, 1999.
[2] N. Pittet, P. Boltenhagen, N. Rivier, D. Weaire, Europhys. Lett. 35 (1996) 547.
[3] H. A. Schwarz, Gesammelte Mathematische Abhandlungen, Vol. 2, p. 309.
[4] M. Kléman, J. Phys. Lett. 46 (1985) L–723.
[5] H. Davenport, The Higher Arithmetic, Hutchinson, London, 1952.
[6] J.F. Sadoc, N. Rivier, Eur. Phys. J. B 12 (1999) 309.
[7] C. Branden, J. Tooze, Introduction to Protein Structure, Garland, NY/London, 1991.
[8] F. Gaill, Y. Bouligand, Tissue and Cell 19 (1987) 625.
[9] K. Okuyama, N. Tanaka, T. Ashida, M. Kakudo, S. Sakakibara, Y. Kishida, J. Mol. Biol. 72 (1972) 571.
[10] Y. Bouligand, poster 2108 SFP-JMC5 Orléans, 1996.
[11] B. Doubrovnine, S. Novikov, A. Fomenko, Géométrie Contemporaine, 1ère. partie, Mir, Moscou, 1982, §20.
[12] H. Lebesgue, Sur la Mesure des Grandeurs, L'enseignement Mathématique, Genève, 1935, p. 95.
[13] Y. Bouligand, Revue du Palais de la Découverte 23 (221) (1994) 39.