J Math Chem (2009) 45:725–747
DOI 10.1007/s10910-008-9382-7

ORIGINAL PAPER

# Protochirons and protohelices

**Alexandru T. Balaban · Kiran B. Chilakamarri · Douglas J. Klein**

Received: 24 January 2008 / Accepted: 25 February 2008 / Published online: 19 September 2008


**Abstract** A simple "protochiron" is used to systematically develop and describe molecular helices, especially on the cubic and diamond lattices.

**Keywords** Molecular helix · Protohelix · Protochiron

## 1 Introduction

The 1975 Nobel Prize for Chemistry was awarded to Sir John W. Cornforth (for his work on the stereochemistry of enzyme-catalyzed reactions), and to Vladimir Prelog, for his research into the stereochemistry of organic molecules and reactions. A few months before the announcement of the Nobel Prize, on July 19, 1975, V. Prelog had written in his preface to the book on *Chemical Applications of Graph Theory*: "*This book should help to bridge the gap (or even abyss) between chemical and mathematical literature. . . The editor of this first comprehensive monograph on chemical applications of graph theory is himself a pioneer and promoter of reaction graphs. He has been fortunate to find for all chapters authors who have made original contributions to the subject and who are still actively involved in it*" [1]. In his Nobel Lecture entitled "Chirality in Chemistry" presented on December 12, 1975, Vladimir Prelog stated: [2] "*The simplest chiral object of the three-dimensional perceptual space is . . . the chiral three-dimensional simplex, the irregular tetrahedron."* Later, in 1982 in a paper co-authored with Helmchen, Prelog pointed out that four points can be

A. T. Balaban (✉) · D. J. Klein
Texas A&M University at Galveston, 5007 Avenue U, Galveston, TX 77553-1675, USA
e-mail: balalana@tamug.edu

K. B. Chilakamarri
Department of Mathematics, Texas Southern University, Science Building 111G,
3100 Cleburne Street, Houston, TX 77004, USA

connected in a helical arrangement (as in *gauche n*-butane) or a tripodal arrangement (as in a trialkylphosphine with three different alkyl groups); the former structure leads to different *conformations*, and the latter to different *configurations* [3].

## 1.1 Protochirons

In two previous papers, *protochirons* were described as the simplest three-dimensional physical objects able to give rise to enantiomeric structures, which are not superimposable on their mirror image [4,5]. For mathematicians the irregular tetrahedron with six edges all having different lengths qualifies as a protochiron, but for chemists and other scientists interested in real-world nanostructures composed of atoms connected by covalent bonds, there exist two common types of (organic chemical) protochirons, such as discussed by Prelog and Helmchen [3], there called "stereogenic units". In the language of graph theory, atoms are modeled by graph vertices or points, and covalent bonds by graph edges or lines. Different atoms or different types of covalent bonds may be modeled by vertices or bonds with weight or color.

### *1.1.1 Type-1 protochiron*

The first type of the two extant types of protochirons is the familiar asymmetric $sp^3$-hybridized atom with four chirally distinct substituents. That is, this protochiron corresponds to a total of five points with different "weights", represented by a tetrahedral assembly of four differently colored lines emerging from one point—the idea going back to Van't Hoff and Le Bel in 1875. According to the Cahn–Ingold–Prelog rules, the configurations based on precedence rules of the four substituents are designated as *R* and *S* [6]. Extension of such tetrahedral protochirons (ignoring color) to a whole network leads to the diamond network if next-neighbor bonds have the staggered conformation.

### *1.1.2 Type-2 protochiron*

The second type of protochiron (which will be the main tool of the present paper) may also be derived from tetrahedra and most simply involves only three non-planar but sequentially intersecting lines, i.e. a sequence of four points: three non-collinear points determine two lines and a plane; the fourth out-of-plane point adds the third line. Estrada's variable parameter $v$ allows a 3-path graph embedded into Euclidean 3-space to manifest a continuous variation of curvilinear torsion [7]. If one simplifies the picture by asking that all bond angles and lengths be equal, one may illustrate this type of protochiron for any bond angle $\alpha$ such that $0 < \alpha < 180°$. Figure 1 presents the cases for $\alpha = 60°$ and $\alpha = 90°$ (the latter case will be called an *ortho-3-path*). A third example is the carbon skeleton of *n*-butane which can adopt exactly three staggered conformations, two of which correspond to the two enantiomers of *gauche n*-butane, while the third conformer is the achiral *anti* (*transoid*) *n*-butane [8,9]. Bond angles are $\alpha = 109.5°$, as in the ideal diamond network. We call these last chiral conformers *diamond-3-paths*, and denote them as **R** and **S** (boldface and not italicized, in order

to differentiate them from the former designations): in a Newman projection of the 3-path viewed along its central edge, if the smallest rotation of the closer terminal edge till it eclipses the more remote terminal edge is clockwise, then the stereo-descriptor is **R**, and if the rotation is counterclockwise it is **S**.

Another familiar network is the simple cubic net. It also gives rise to a pair of enantiomeric chiral 3-paths, with bond angles of 90°, as seen in Fig. 1, which can also have either **R** or **S** stereo-descriptors. Figure 2 presents the protochirons for the three possible diamond-3-paths with their descriptors, and Fig. 3 for the six possible ortho-3-paths with their designations.

The type-2 protochiron involving four points, i.e. involving a 3-path, might be viewed as formally simpler than type-1 protochirons involving five points. In a corresponding chemical context, asymmetrically substituted atoms with tetrahedral geometry are less simple than chiral 3-paths involving covalent bonds, somewhat as the ground state of hydrogen peroxide.

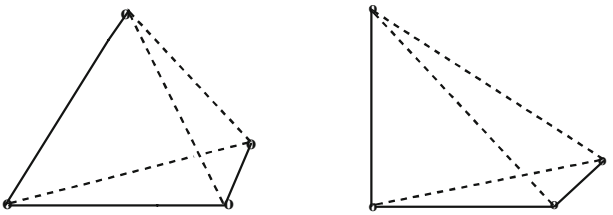

**Fig. 1** Type-2 **S** protochirons (full lines) with bond angles of 60° (a "tight-tetrahedron" protochiron) and 90° (ortho-3-path on the cubic network)

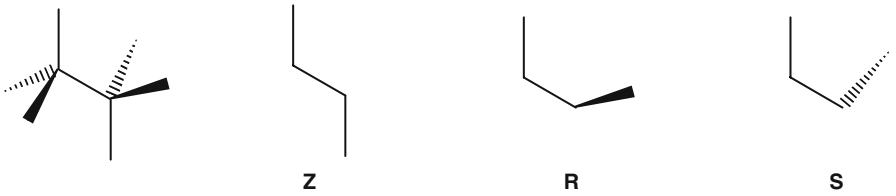


**Fig. 2** The left-most structure corresponds to 2,2,3,3-tetramethylbutane in staggered conformation. The next three are the **Z**, **R**, and **S** protochirons (diamond-3-paths)

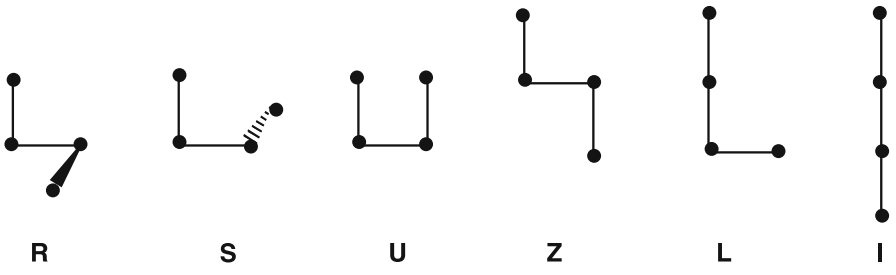


**Fig. 3** The six ortho-3-paths with their designations

A third type of (simple) protochiron entails just three lines radiating from a single point, whence if non-planar leads to chiral structures (much as for the type-1 protochirons), and may be used to describe chiral phosphines or amines. The "invisible" unshared electron pair replaces the fourth tetrahedral ligand.

Again in a chemical context, Prelog and Helmchen noted that somewhat arbitrary carbon-based molecular structures may be factored into type-1 and type-2 protochirons [3]. Asymmetrically substituted atoms with tetrahedral geometry are probably "sturdier" than chiral 3-paths in that the type-1 protochirons, in not so readily rearranging to their mirror image. Still the mathematical description of 3-paths is especially simple, and they do describe different conformations manifested by polymethine (or polyethylene) chains. But other chemical realizations as more rigid structures have been noted [10,11]. Another quite rigid realization is as diamondoid chain molecules (comprised from fused adamantanes)—when we identify these molecules in terms of their dual-net model (with sites in the centers of adamantanes and edges between these central sites corresponding to face-sharing adamantane units). Thence here these 4-vertex 3-paths might be used to describe different helices—indeed Prelog and Helmchen termed this protochiron the "helical stereogenic unit" [3].

Chiral objects can be obtained from 3-path protochirons by three operations: *concatenation* (association of two 3-paths such that two vertices merge into one), *assembly* (association of two 3-paths such that two edges merge into one), and *overlap* (association such that the last two edges of one 3-path merge with the first two edges of a second 3-path) [2]. Helices result by suitable repeated overlap operations. The helix is described by the stereo-descriptors corresponding to each of the central edges added sequentially by the overlap. Thus, among the six possible staggered conformers of *n*-pentane, two are helical, namely **RR** and **SS**, and they are formed by overlap of two diamond-3-paths having the corresponding stereo-descriptors. The present note discusses in detail helices formed from the three types of 3-paths illustrated in Fig. 2, for the diamond network ($\alpha = 109.5°$). But first, some briefer attention is paid to helices formed from the three orthopaths **R**, **S**, and **Z** of the cubic network ($\alpha = 90°$) illustrated in Fig. 3, and also helices based on the tetrahedral protochiron with $\alpha = 60°$. In principle, with angles smaller than, or equal to 60° one can obtain an infinity of helices as shown by two examples in Fig. 4.

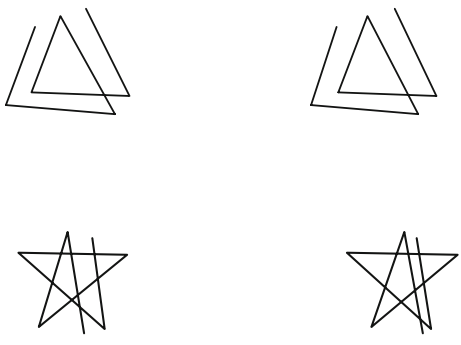

**Fig. 4** Stereo-views of helices with small bond angles leading to triangles or stellated pentagons when viewed along the helix axis

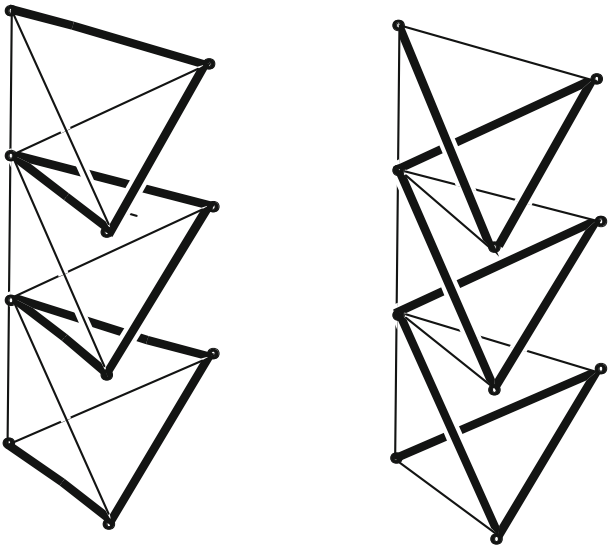

**Fig. 5** Enantiomeric tight-tetrahedron helices

### 1.2 Helices based on tight-tetrahedron-3-paths

If several tetrahedra are assembled such that two tetrahedra share a vertex, two tetrahedron edges involving this vertex are aligned along a fixed line, and the other edges are parallel, then a simple helix may be obtained as indicated in Fig. 5. Viewed along the helix axis, the overlapping lines give rise to an equilateral triangle.

An interesting helical face-sharing tetrahedron chain, called tetrahelix by Buckminster Fuller, has been analyzed in detail by Zheng et al. [12]. Its twist is irrational and never repeats itself.

### 1.3 Helices based on ortho-3-paths

We present in Fig. 6 the simplest enantiomeric helices formed from repeated overlap of the **R** or **S** rectangular-3-paths. It can be seen in Fig. 7 that the repeating unit of such helixes has a period (pitch) of three, and that in a side-view along the axis of the helix, one can see a triangular arrangement of the overlapping bonds in the simple cubic network.

It is easy to assign integer coordinates to points on a cubic network by assuming an interatomic distance of one arbitrary unit along the $x$, $y$, and $z$ axes. Table 1 presents coordinates for two enantiomeric helices on the cubic network shown in Fig. 7.

One can observe from Table 1 that the sequences of each coordinate repeat with a period of three, after which the pattern with each coordinate increased by 1 is repeated, to continue this trend. In the above example, the $z$ coordinates are the same for the two enantiomers, but the $x$ and $y$ coordinates are permuted. In Figs. 4 and 5 the corresponding triangles of overlapping edges viewed alongside the helix axis share

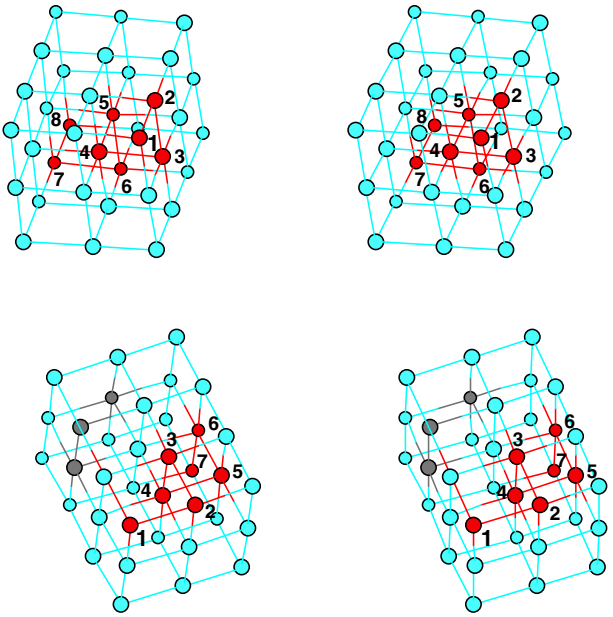


**Fig. 6** Stereo-views of an all-**R** helix (top) and all-**S** helix (bottom) on the cubic network

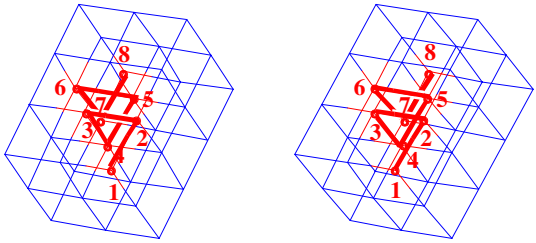


**Fig. 7** Stereo-view of a cubic **RRRRRRR** helix involving points 1 through 8

**Table 1** Coordinates for 8 helical vertices of enantiomeric helices (all-**R** and all-**S**) in the simple cubic net

| Vertex | **RRRRRRR** | | | **SSSSSSS** | | |
|---|---|---|---|---|---|---|
| | *x* | *y* | *z* | *x* | *y* | *z* |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 | 1 |
| 3 | 0 | 1 | 1 | 1 | 0 | 1 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 1 | 2 | 1 | 1 | 2 |
| 6 | 1 | 2 | 2 | 2 | 1 | 2 |
| 7 | 2 | 2 | 2 | 2 | 2 | 2 |
| 8 | 2 | 2 | 3 | 2 | 2 | 3 |

an edge for two enantiomeric helices. This results in the two enantiomeric helices differing in point coordinates for every third point, but sharing the two other points.

### 1.4 Helices based on diamond-3-paths

It is also easy to assign integer coordinate to points on the diamond net, as shown by Balaban et al. [13]. This entails a smallest interatomic distance of $\sqrt{3}$. A triamantane carbon skeleton with such coordinates is shown in Fig. 8. The origin of the coordinate system is at the center of the left-most adamantane subunit. It can be easily observed that this center is equi-distant from the two pairs of vertices with coordinates (±2, 0, 0), (0, 0, ±2), and (0, ±2, 0).

Table 2 displays integer coordinates for a pair of enantiomeric helices with nine points on the diamond network, one of which is presented in Fig. 9. Looking at Table 2 one sees that the $x$-coordinates are the same for both enantiomers. On looking along the helix axis, one can observe that the overlapping bonds form a square, in agreement with the period of four for all three coordinates. The enantiomeric helices are seen to share an edge of the square formed by the apparently overlapping bonds. The helix

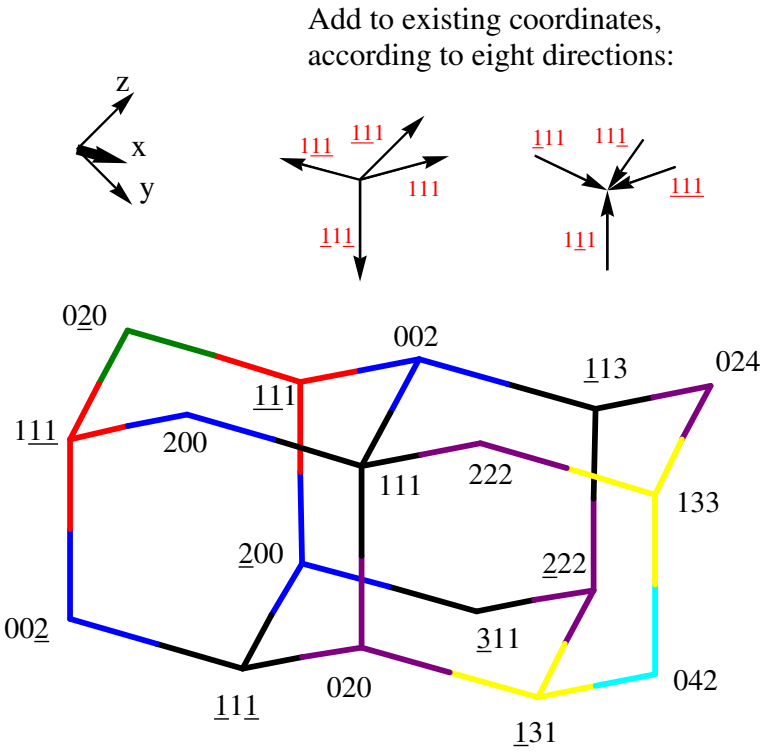


**Fig. 8** Triamantane skeleton with integer coordinates for its carbon atoms. Colors characterize the $y$ coordinate. Following the direction of an arrow, the algebraic increments of coordinates are indicated in red. Underlined digits have negative sign

**Table 2** Coordinates for helical points in enantiomeric helices on the diamond net

| Vertex | **RRRRRRR** | | | **SSSSSSS** | | |
|---|---|---|---|---|---|---|
| | *x* | *y* | *z* | *x* | *y* | *z* |
| 1 | −4 | 0 | 2 | −4 | 0 | 2 |
| 2 | −3 | 1 | 3 | −3 | −1 | 1 |
| 3 | −2 | 0 | 4 | −2 | −2 | 2 |
| 4 | −1 | −1 | 3 | −1 | −1 | 3 |
| 5 | 0 | 0 | 2 | 0 | 0 | 2 |
| 6 | 1 | 1 | 3 | 1 | −1 | 1 |
| 7 | 2 | 0 | 4 | 2 | −2 | 2 |
| 8 | 3 | −1 | 3 | 3 | −1 | 3 |
| 9 | 4 | 0 | 2 | 4 | 0 | 2 |
| 10 | 5 | 1 | 3 | 5 | −1 | 1 |
| 11 | 6 | 0 | 4 | 6 | −2 | 2 |
| 12 | 7 | −1 | 3 | 7 | −1 | 3 |
| 13 | 8 | 0 | 2 | 8 | 0 | 2 |
| 14 | 9 | 1 | 3 | 9 | −1 | 1 |
| 15 | 10 | 0 | 4 | 10 | −2 | 2 |
| 16 | 11 | −1 | 3 | 11 | −1 | 3 |

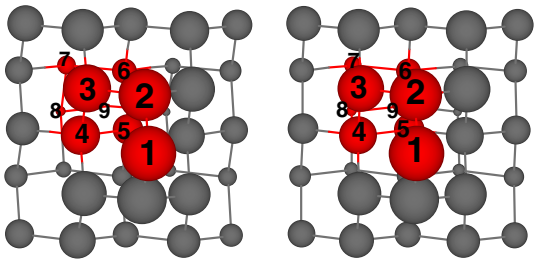


**Fig. 9** Stereo-view of a helix on the diamond network (red spheres)

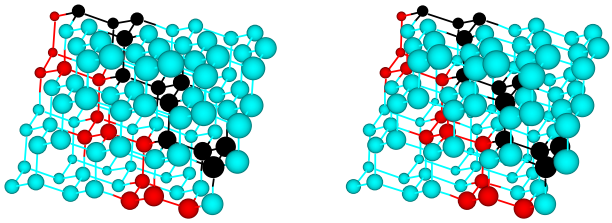

**Fig. 10** Two enantiomeric helices (all-S in red, and all-R in black) on the diamond lattice

axis is oriented along the *x* axis, as seen by the continually increasing *x* coordinate (by 1) for each successive step of the helix. In agreement with the observation that the helix in Fig. 9 forms a square when viewed along its axis, one can see that the *y* and *z*-coordinates in Table 2 repeat themselves with a period of 4.

As will be seen in Table 2 and Fig. 10, integer coordinates for all-**R** and all-**S** helices on the diamond lattice in a particular direction have the pitch (period) of 4, and for the *y* and *z* coordinates they provide a continuous repetition, whereas the x coordinate increases steadily along natural numbers.

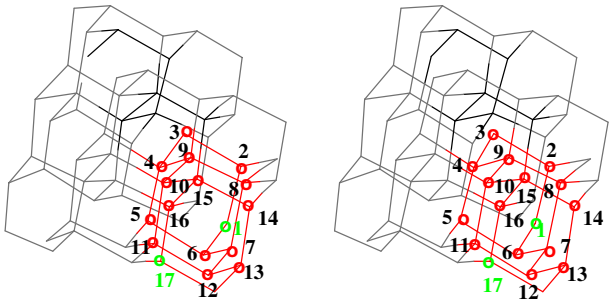


**Fig. 11** Stereo-view of the helix **SRRSRRSRRSRRSRR** on the diamond network, beginning with 1 (green) and ending with 17 (green). It forms a hexagon from overlapping bonds when viewed along the helix axis

**Table 3** Point coordinates of the helix on the diamond network **SRRSRRSRR** . . .

| Vertex | **SRRSRRSRRS** | | |
|---|---|---|---|
| | $x$ | $y$ | $z$ |
| 1 | −1 | 3 | 1 |
| 2 | 0 | 2 | 0 |
| 3 | 1 | 1 | 1 |
| 4 | 0 | 0 | 2 |
| 5 | −1 | 1 | 3 |
| 6 | −2 | 2 | 2 |
| 7 | −3 | 1 | 1 |
| 8 | −2 | 0 | 0 |
| 9 | −1 | −1 | 1 |
| 10 | −2 | −2 | 2 |
| 11 | −3 | −1 | 3 |
| 12 | −4 | 0 | 2 |
| 13 | −5 | −1 | 1 |
| 14 | −4 | −2 | 0 |
| 15 | −3 | −3 | 1 |
| 16 | −4 | −4 | 0 |
| 17 | −5 | −3 | 2 |

It is also possible to have helices with the repeating sequence formed by more than one diamond-3-path, as we saw in Fig. 9, for instance with repeating sequence SRS as presented in Fig. 11. The coordinates for the 17 points of Fig. 11 are displayed in Table 3. It can be seen that the $z$-coordinates become repeated with a period of 6, in agreement with the fact that the helix appears as a hexagon formed by overlapping bonds when viewed along the axis of the helix.

### 1.5 Dihedral angles for 3-paths

The three lines of a 3-path define two intersecting planes. One can see in Fig. 12 the four red points and the three red lines of the ortho-3-path, and also the three broken red lines that with the full red lines define an irregular tetrahedron. It is easy to see that in this case the dihedral angle is a right angle, because the two planes are isosceles triangles (half-faces of the cube).

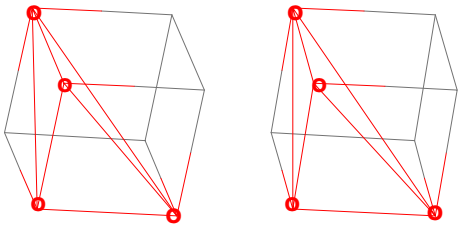

**Fig. 12** Stereo-view of an ortho-3-path (full red lines) on a cube

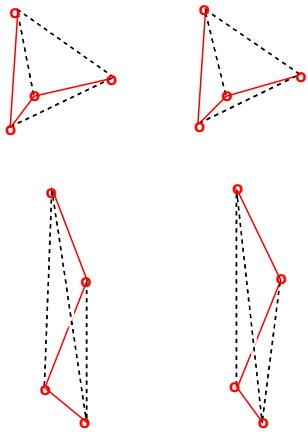

**Fig. 13** Stereo-views of a tight-tetrahedron-3-path (upper part) and a diamond-3-path (lower part)

For the 3-path on a regular tetrahedron, the upper part of Fig. 13 indicates similarly to Fig. 12 by four red points and three red lines the tight-tetrahedron-3-path. Black broken lines complete the regular tetrahedron. The lower part of Fig. 13 illustrates similarly the red-colored diamond-3-path, but in this case the broken black lines and the full red lines define an irregular tetrahedron. Interestingly, for both 3-paths in Fig. 13, the dihedral angle is the same ($\arcsin 3^{-1/2} = 70.5°$).

## 2 Paths in the diamond net

We follow Balaban et al. [13] to describe a method of assigning integer coordinates to the points on the diamond net. Let $A = (1, 1, 1), B = (1, -1, -1), D = (-1, 1, -1)$ and $E = (-1, -1, 1)$ be the four basic vectors. For convenience let us also use a redundant notation $a = -A = (-1, -1, -1)$, $b = -B = (-1, 1, 1)$, $d = -D = (1, -1, 1)$, and $e = -E = (1, 1, -1)$. We have deliberately avoided the use of the letter C to reduce the confusion between the capital and small letters. The alternating sums of the form $\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 + \cdots + (-1)^{n-1}\alpha_n$ where $\alpha_i$'s are from the set {A, B, D, E} for $n > 0$, give the coordinates of the points of the diamond network. Let $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n$ be the points of a directed path P with distinct edges in the diamond network. The letter sequence of vectors $\boldsymbol{r}_2 - \boldsymbol{r}_1, \boldsymbol{r}_3 - \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n - \boldsymbol{r}_{n-1}$, form an alternating sequence of capital and small letters. We remark that the geometry of the diamond net does not allow repetition of the same letter (no distinction between capital and small letters is made in this remark). For instance none of AA, aA, Aa, or aa are allowed and similar comments hold for the other three letters. The path P is made of $n - 3$ overlapping

3-paths corresponding to the sets $\{\boldsymbol{r}_2-\boldsymbol{r}_1, \boldsymbol{r}_3-\boldsymbol{r}_2, \boldsymbol{r}_4-\boldsymbol{r}_3\}$, $\{\boldsymbol{r}_3-\boldsymbol{r}_2, \boldsymbol{r}_4-\boldsymbol{r}_3, \boldsymbol{r}_5-\boldsymbol{r}_4\}$, . . ., respectively. Each of these 3-paths has a unique assignment of a protochiron **S, R** or **Z**. These protochirons form a sequence $S_1, S_2, \ldots, S_{n-3}$ providing a standard type of code for the path P. Thus it is easy to see that given a path we can construct a unique code. Clearly the code depends on the first four points, $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{r}_4$ of the path P. Conversely, given the first four points a path P and the code of the path P we can evaluate the coordinates of the rest of the points of the path P. To achieve this goal we first list all possible directed paths of length 3 with the tail end of the path at the origin. This list essentially captures all paths of length 3 since any directed 3-path $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{r}_4$ in the diamond network can be translated so that $\boldsymbol{r}_1$ coincides with the origin. In the diamond network only half of these paths, 36 in number, appear at the origin. These paths are denoted by a string of three letters with the first and the last from {A, B, D, E} and the middle from {a, b, d, e}. The remaining paths are denoted by a string of three letters with the first and last from {a, b, d, e} and the middle from {A, B, D, E}. In both types of 3-paths the same letter is not repeated. We will use the word "3-path designations" or "3pd" for these strings. In the next example we consider a 3-path and it's 3pd. The *overlap operation* can now be explained as follows: a 3pd, say "xyz" can be combined with another succeeding 3pd "pqr" if and only if y = p and z = q. The result is essentially the standard code used in polymer chemistry for conformations of polymethine chains—see, e.g., Flory's books [10,11].

*Example 2.1* Consider the directed 3-path with vertices (0, 0, 0), (1, 1, 1), (0, 2, 2), and (−1, 3, 1). Here the vector (1, 1, 1) − (0, 0, 0) = (1, 1, 1) corresponds to the directed edge A, (0, 2, 2) − (1, 1, 1) = (−1, 1, 1) corresponds to the directed edge b, and (−1, 3, 1) − (0, 2, 2) = (−1, 1, −1) corresponds to the directed edge D. The 3pd of this 3-path is AbD. Conversely, given AbD, the same 3pd, the letter A corresponds to the directed edge with tail end at (0, 0, 0) and head at (1, 1, 1). Thus the first two points of the directed path are (0, 0, 0) and (1, 1, 1). The letter b corresponds to the vector (−1, 1, 1) and the addition of this vector to (1, 1, 1) results in the third point (0, 2, 2) of the path. The vector (−1, 1, −1) corresponds to the last letter D and adding (0, 2, 2) and (−1, 1, −1) gives the fourth point (−1, 3, 1).

### 2.1 The 3-path designations

All of the required 3-paths can be constructed by adding the vectors {A, B, D, E} and {a, b, d, e} alternatively while ensuring the same letter is not immediately repeated. There are a total of 72 different paths. Half of these 3pd's begin with a capital letter and can be realized at the origin in the diamond net, while the rest of the paths, whose 3pds begin with small letter are not realizable in the same diamond net centered at the origin. We divide the 3pds into three classes based on the type of protochiron each 3-path designation represents.

We may note that every 3pd beginning with a capital letter has a corresponding 3pd beginning with a small letter and is obtained by reversing the order and interchanging the capital letters and small letters. Given a path, we first find the sequence of letters corresponding to the directed edges along the path, then using a sliding window of three letters, and the list of 3pds we have derived, we arrive at the protochiron code

**List of 3pds**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **R**-type | AbE | BaD | DaE | EaB | eBa | dAb | eAd | bAe |
| | AdB | BdE | DbA | EbD | bDa | eDb | aBd | dBe |
| | AeD | BeA | DeB | EdA | dEa | aEb | bEd | aDe |
| **S**-type | AbD | BaE | DaB | EaD | dBa | eAb | bAd | dAe |
| | AdE | BdA | DbE | EbA | eDa | aDb | eBd | aBe |
| | AeB | BeD | DeA | EdB | bEa | dEb | aEd | bDe |
| **Z**-type | AbA | BaB | DaD | EaE | aBa | bAb | dAd | eAe |
| | AdA | BdB | DbD | EbE | aDa | bDb | dBd | eBe |
| | AeA | BeB | DeD | EdE | eAe | bEb | dEd | eDe |

of the path. For instance if the sequence of letters is AbAdBeAdB then the first three letters AbA corresponds to a 3pd of **Z**-type, sliding to the next window bAd gives a 3pd of **S**-type, and next AdB gives an **R**-type and so on leading to the code **ZSRRRRR**. Next we have some examples.

*Example 2.2* The first three columns contain the coordinates of the points of a path. The next three columns contain the coordinates of the vector obtained by subtracting the coordinates of the points (first three columns) in upper row from the coordinates of the point immediately following lower row. Next column contains the label of the directed edge. The last column shows the protochiron of the sliding window of the row.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4 | 0 | 2 | | | | | |
| 3 | 1 | 3 | 1 | 1 | 1 | A | |
| 2 | 0 | 4 | 1 | 1 | 1 | d | |
| 1 | 1 | 3 | 1 | 1 | 1 | B | **R** |
| 0 | 0 | 2 | 1 | 1 | 1 | e | **R** |
| 1 | 1 | 3 | 1 | 1 | 1 | A | **R** |
| 2 | 0 | 4 | 1 | 1 | 1 | d | **R** |
| 3 | 1 | 3 | 1 | 1 | 1 | B | **R** |
| 4 | 0 | 2 | 1 | 1 | 1 | e | **R** |

*Example 2.3* In this example we chose a self intersecting path with distinct edges to evaluate the code. The edges are distinct if no consecutive points are repeated.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | | | | | |
| 0 | 0 | 0 | 1 | 1 | 1 | b | |
| 1 | 1 | 1 | 1 | 1 | 1 | A | |
| 2 | 0 | 2 | 1 | 1 | 1 | d | **S** |
| 1 | 1 | 3 | 1 | 1 | 1 | E | **S** |
| 0 | 2 | 2 | 1 | 1 | 1 | a | **R** |
| 1 | 1 | 1 | 1 | 1 | 1 | D | **S** |
| 0 | 0 | 0 | 1 | 1 | 1 | e | **R** |
| 1 | 1 | 1 | 1 | 1 | 1 | D | **Z** |
| 2 | 0 | 2 | 1 | 1 | 1 | a | **S** |

Next we consider examples illustrating the use of the list of 3pds in evaluating the coordinates of the path, given the first four points of the path and the code. Here we are given the first four rows and the last column containing the code. We have the first two letters of the second 3pd and the code letter (**R**, **S**, or **Z**). From this information we can find the third letter of the second 3pd by consulting the list of the 3pds corresponding to the code (**R**, **S** or **Z**) and this uniquely determines the second 3pd. Adding the coordinates of the vector corresponding to the new letter to the coordinates of the fourth point yields the fifth point. Next we move on to evaluating the third 3pd and so on. We can repeat this process as long as the code is available. We may note that we do not really need the first four points of the path, as the first three will suffice. The fourth point can be evaluated since the first code letter is known. We will summarize these observations in the following Lemma 2.4, which is stated without proof.

**Lemma 2.4** *Given a code* $S_0, S_1, \ldots$ *of a path* P *in the diamond net and the coordinates of four points* $\boldsymbol{r}_{i+1}, \boldsymbol{r}_{i+2}, \boldsymbol{r}_{i+3}, \boldsymbol{r}_{i+4}$ *corresponding to any single* $S_i$*, the rest of the points of the path P can be determined uniquely.*

*Example 2.5* Let us reconstruct the points of the path in Example 2.3 given the first four points and the code **SSRSRZS**.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | −1 | −1 | | | | | |
| 0 | 0 | 0 | −1 | 1 | 1 | b | |
| 1 | 1 | 1 | 1 | 1 | 1 | A | |
| 2 | 0 | 2 | 1 | −1 | 1 | d | **S** |

The second protochiron is **S** and first two letters are A and d. There is exactly one 3pd namely "AdE" in the list for **S** with Ad as the first two letters. Thus E is the fourth directed edge and E=(−1, −1, 1). With this information our table grows a little.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | −1 | −1 | | | | | |
| 0 | 0 | 0 | −1 | 1 | 1 | b | |
| 1 | 1 | 1 | 1 | 1 | 1 | A | |
| 2 | 0 | 2 | 1 | −1 | 1 | d | **S** |
| | | | −1 | −1 | 1 | E | **S** |

Addition of (−1, −1, 1) and (2, 0, 2) yields the fifth point (1, −1, 3). The next protochiron is **R** and d,E are the first two letters pointing to the unique 3pd namely dEa and the directed edge a=(−1, −1, −1). Again the addition of (1, −1, 3) and (−1, −1, −1) yields the sixth point.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | −1 | −1 | | | | | |
| 0 | 0 | 0 | −1 | 1 | 1 | b | |
| 1 | 1 | 1 | 1 | 1 | 1 | A | |
| 2 | 0 | 2 | 1 | −1 | 1 | d | **S** |
| 1 | −1 | 3 | −1 | −1 | 1 | E | **S** |
| 0 | −2 | 2 | −1 | −1 | −1 | a | **R** |

This process can be continued as long as the code is available.

*Example 2.6* Let $(-1, -1, 1)$, $(0, -2, 2)$, $(1, -1, 3)$ and $(2, 0, 2)$ be the first four points of a path with code **SRZSSRZZSZR**. Find the rest of the points on this path. Note that we can solve this problem with just $(-1, -1, 1)$, $(0, -2, 2)$, $(1, -1, 3)$, the first three points.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| −1 | −1 | 1 | | | | | |
| 0 | −2 | 2 | 1 | −1 | 1 | d | |
| 1 | −1 | 3 | 1 | 1 | 1 | A | |
| 2 | 0 | 2 | 1 | 1 | −1 | e | **S** |
| 1 | 1 | 1 | −1 | 1 | −1 | D | **R** |
| 2 | 2 | 0 | 1 | 1 | −1 | e | **Z** |
| 3 | 3 | 1 | 1 | 1 | 1 | A | **S** |
| 2 | 4 | 2 | −1 | 1 | 1 | b | **S** |
| 1 | 3 | 3 | −1 | −1 | 1 | E | **R** |
| 0 | 4 | 4 | −1 | 1 | 1 | b | **Z** |
| −1 | 3 | 5 | −1 | −1 | 1 | E | **Z** |
| −2 | 2 | 4 | −1 | −1 | −1 | a | **S** |
| −3 | 1 | 5 | −1 | −1 | 1 | E | **Z** |
| −4 | 2 | 6 | −1 | 1 | 1 | b | **R** |

## 2.2 The overlap operation and the De Bruijn sequences

The overlap operation is a classic combinatorial operation occurring in a problem of practical origin: the so-called *Rotating drum problem*. This problem asks if it is possible to construct a circular binary sequence of length $2^n$ so that $2^n$ possible $n$-tuples of consecutive 0's and 1's arise, see [14,15] for further details. Such a circular sequence is called a De Bruijn sequence. The concepts of overlap operation and line graph are used in the construction of De Bruijn sequences. These concepts can be used to obtain the 3pd's and describe how they can be combined by the overlap operation. We now briefly outline the ideas. First let us recall that a *directed line graph* $L(G)$ of a directed graph $G$ is a graph whose vertex set is the set of directed arcs of $G$ and two vertices $v = (x, y)$ and $u = (a, b)$ of $L(G)$ form an arc of $L(G)$ if and only if the head of the directed arc $v$ of $G$ is same as the tail of the directed arc $u$ of $G$, i.e., $y = a$.

We may first explore the diamond network or the hexagonal network in two dimensions. Let $\mathbf{X} = (1/2, \sqrt{3}/2)$, $\mathbf{Y} = (1/2, -\sqrt{3}/2)$, $\mathbf{Z} = (0, -1)$, $\mathbf{x} = -\mathbf{X}$, $\mathbf{y} = -\mathbf{Y}$, and $\mathbf{z} = -\mathbf{Z}$ and the alternating sums of the form $\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 + \cdots + (-1)^{n-1}\alpha_n$ where $\alpha_i$ 's are from the set $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ for $n > 0$, give the coordinates of the points of the hexagonal network. Let G be the directed graph with the vertex set $V(G) = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{x}, \mathbf{y}, \mathbf{z}\}$ and the arc set $A(G) = \{(\mathbf{X}, \mathbf{y}),(\mathbf{y}, \mathbf{X}),(\mathbf{X}, \mathbf{z}),(\mathbf{z}, \mathbf{X}),(\mathbf{Y}, \mathbf{x}),(\mathbf{x}, \mathbf{Y}),(\mathbf{Y}, \mathbf{z}),(\mathbf{z}, \mathbf{Y}),(\mathbf{Z}, \mathbf{x}),(\mathbf{x}, \mathbf{Z}),(\mathbf{Z}, \mathbf{y}),(\mathbf{y}, \mathbf{Z})\}$. In 2 dimensions there are two chiral protochirons, namely **R** and **S**:**Xy**, **Yz**, **Zx**, **xY**, **yZ**, **zX** are of **R**-type and **Yx**, **Zy**, **Xz**, **yX**, **zY**, **xZ** are of **S**-type. We may note that these protochirons correspond to the arcs of $G$. The protochirons are exactly the vertices of the line graph $L(G)$ and the arcs of $L$(G) show the possible combinations of protochirons.

Now, for the three-dimensional diamond lattice let the set of eight vectors {A, B, D, E, a, b, d, e}, be the vertex set of a directed graph $H$. Given two vertices $x$ and $y$ there is a directed arc from $x$ to $y$ if and only if (i) $x$ and $y$ do not represent the same letter, (ii) exactly one of $x$ and $y$ is a small letter. The vertices of $L(L(H))$, i.e., the line graph of the line graph of $H$, are 72 in number and are exactly the 3pd's listed in Sect. 2.1. The acrs of $L(L(H))$ show the possible overlap operations.

## 3 Periodic paths and periodic codes in the diamond network

Many results in this section can be generalized to arbitrary networks, however in this section we focus exclusively on the diamond network. Presumably a periodic path should give rise to a periodic code. In the diamond network the converse is also true. We remark to the reader that this result is not immediate since each code letter (**R**, **S** and **Z**) corresponds to 24 distinctly positioned (non-congruent) 3-path designations (3pds). Thus it is necessary to provide precise definitions of periodic path and periodic code to aid in the proof of this claim.

**Definition 3.1** An infinite path with vertices $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots$, is periodic if for some positive integer $n$, for $\boldsymbol{\tau} = (\boldsymbol{r}_n - \boldsymbol{r}_0)$, one has $\boldsymbol{r}_{n+1} = \boldsymbol{r}_1 + \boldsymbol{\tau}, \boldsymbol{r}_{n+2} = \boldsymbol{r}_2 + \boldsymbol{\tau}, \ldots, \boldsymbol{r}_{2n-1} = \boldsymbol{r}_{n-1} + \boldsymbol{\tau}$, $\boldsymbol{r}_{2n} = \boldsymbol{r}_0 + 2\boldsymbol{\tau}, \boldsymbol{r}_{2n+1} = \boldsymbol{r}_1 + 2\boldsymbol{\tau}, \ldots, \boldsymbol{r}_{3n-1} = \boldsymbol{r}_{n-1} + 2\boldsymbol{\tau}, \boldsymbol{r}_{3n} = \boldsymbol{r}_0 + 3\boldsymbol{\tau}, \boldsymbol{r}_{3n+1} = \boldsymbol{r}_1 + 3\boldsymbol{\tau}$, $r_{3n+2} = r_2 + 3(r_n - r_0), \ldots$ and so on. In a more compact form the sequence of $\boldsymbol{r}_t$ is periodic if $\boldsymbol{r}_t = \boldsymbol{r}_s + q(\boldsymbol{r}_n - \boldsymbol{r}_0)$, $t = qn + s$ with $0 \leq s < n, q > 0$. The smallest integer $n$ for which the sequence is periodic is the period of the path with the periodic block $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_n$.

**Definition 3.2** The code $S_0, S_1, \ldots$, of an infinite path is periodic if for some positive integer $m$, $S_m = S_0, S_{m+1} = S_1, \ldots, S_{2m-1} = S_{m-1}, S_{2m} = S_0, S_{2m+1} = S_1$, and so on. In a compact form if $S_t = S_s$, where $t = qm + s$ with $0 \leq s < m, q > 0$. The smallest integer m for which the sequence is periodic is the period of the code with the periodic block $S_0, S_1, \ldots, S_{m-1}$.

Let P be an infinite path with points $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots$, and code $S_0, S_1, \ldots,$ . It is intuitively clear that a periodic infinite path must have a periodic code. It should be remarked that in general the period of a periodic path is not the same as the period of its code and the code itself does not readily yield the translation vector $\boldsymbol{\tau}$.

**Proposition 3.3** *Every infinite periodic path has a periodic code.*

*Proof* Let P be an infinite path $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots$ with period n. Let t be any positive integer not less than n. Suppose $t = qn + s$ with $0 \leq s < n$. The code $S_t$ corresponds a protochiron depending on the directed edges $\boldsymbol{r}_{t+1} - \boldsymbol{r}_t, \boldsymbol{r}_{t+2} - \boldsymbol{r}_{t+1}, \boldsymbol{r}_{t+3} - \boldsymbol{r}_{t+2}$. Since the path P is periodic we can express these four points $\boldsymbol{r}_t, \boldsymbol{r}_{t+1}, \boldsymbol{r}_{t+2}, \boldsymbol{r}_{t+3}$ in terms of points of the periodic block: $\boldsymbol{r}_t = \boldsymbol{r}_s + q(\boldsymbol{r}_n - \boldsymbol{r}_0), \boldsymbol{r}_{t+1} = \boldsymbol{r}_{s+1} + q(\boldsymbol{r}_n - \boldsymbol{r}_0)$, $\boldsymbol{r}_{t+2} = \boldsymbol{r}_{s+2} + q(\boldsymbol{r}_n - \boldsymbol{r}_0)$, and $\boldsymbol{r}_{t+3} = \boldsymbol{r}_{s+3} + q(\boldsymbol{r}_n - \boldsymbol{r}_0)$. It is clear that $S_t = S_s$ since $\boldsymbol{r}_{t+1} - \boldsymbol{r}_t = \boldsymbol{r}_{s+1} - \boldsymbol{r}_s, \boldsymbol{r}_{t+2} - \boldsymbol{r}_{t+1} = \boldsymbol{r}_{s+2} - \boldsymbol{r}_{s+1}$, and $\boldsymbol{r}_{t+3} - \boldsymbol{r}_{t+2} = \boldsymbol{r}_{s+3} - \boldsymbol{r}_{s+2}$.

**Proposition 3.4** *Let* P *be an infinite path in the diamond network with points* $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots$ *If the path* P *has a periodic code then* P *is a periodic path and period of the path is not greater than 24* m*, where* m *is the period of the code.*

*Proof* Let $S_0, S_1, \ldots$, be the periodic code of the path P with period m. Consider a sequence $C_0, C_1, \ldots$ of subsequences of the points of P as follows:

$$\begin{aligned}
&C_0 : \boldsymbol{r}_0, \boldsymbol{r}_1, \ldots \\
&C_1 : \boldsymbol{r}_m, \boldsymbol{r}_{m+1}, \ldots \\
&C_2 : \boldsymbol{r}_{2m}, \boldsymbol{r}_{2m+1}, \ldots \quad \text{and in general } C_t : \boldsymbol{r}_{tm}, \boldsymbol{r}_{tm+1}, \ldots \quad \text{for } t = 0, 1, 2, \ldots
\end{aligned}$$

We note that each of the sequences $C_0, C_1, \ldots$ all have the same code due to the periodicity of the original code. Translations of a path by a vector does not alter the code, thus the following sequences all have the same code as the original path P,

$$\begin{aligned}
&D_0 : (0,0,0), \boldsymbol{r}_1 - \boldsymbol{r}_0, \boldsymbol{r}_2 - \boldsymbol{r}_0, \boldsymbol{r}_3 - \boldsymbol{r}_0, \ldots \\
&D_1 : (0,0,0), \boldsymbol{r}_{m+1} - \boldsymbol{r}_m, \boldsymbol{r}_{m+2} - \boldsymbol{r}_m, \boldsymbol{r}_{m+3} - \boldsymbol{r}_m, \ldots \\
&D_2 : (0,0,0), \boldsymbol{r}_{2m+1} - \boldsymbol{r}_{2m}, \boldsymbol{r}_{2m+2} - \boldsymbol{r}_2, \boldsymbol{r}_{2m+3} - \boldsymbol{r}_{2m}, \ldots \quad \text{and in general} \\
&D_t : (0,0,0), \boldsymbol{r}_{tm+1} - \boldsymbol{r}_{tm}, \boldsymbol{r}_{tm+2} - \boldsymbol{r}_{tm}, \boldsymbol{r}_{tm+3} - \boldsymbol{r}_{tm}, \ldots \quad \text{for } t = 0, 1, 2, \ldots
\end{aligned}$$

All of the initial 3pds given by the first four points in $D_0, D_1, D_2, \ldots$ belong to the same type of protochiron since they all have the same code letter. There are only 24 distinct 3pds for each of the code letters **R**, **S** and **Z**. Then by the Dirichlet box principle (more commonly known as Pigeonhole principle), there must exist two non-negative integers i and j, $i < j \leq 24$ for which the initial four points in $D_i$ and $D_j$ are identical. Since $D_i$ and $D_j$ have exactly the same code and the same first four points it follows that the sequences $D_i$ and $D_j$ are identical. Thus we have,

$$\boldsymbol{r}_{jm+t} - \boldsymbol{r}_{jm} = \boldsymbol{r}_{im+t} - \boldsymbol{r}_{im} \quad \text{for } t = 1, 2, \ldots$$

Letting $\mathbf{A} = \boldsymbol{r}_{jm} - \boldsymbol{r}_{im}, \boldsymbol{r}_{jm+t} = \boldsymbol{r}_{im+t} + \mathbf{A}$ for $t = 1, 2, \ldots$ We now show that the path corresponding to the sequence $D_i$ is periodic. Let $t = q\,(jm - im) + s$, $0 \leq s < (jm - im)$ and $q > 0$, then,

$$\begin{aligned}
\boldsymbol{r}_{jm+t} &= \boldsymbol{r}_{im+t} + \mathbf{A} \\
&= \boldsymbol{r}_{jm+(q-1)(jm-im)+s} + \mathbf{A} \\
&= \boldsymbol{r}_{im+(q-1)(jm-im)+s} + 2\mathbf{A} \\
&= \boldsymbol{r}_{jm+s} + q\mathbf{A}.
\end{aligned}$$

If $s = 0$, then $\boldsymbol{r}_{jm+q(jm-im)} = \boldsymbol{r}_{jm} + q\mathbf{A}$, and if $s \neq 0$, $\boldsymbol{r}_{jm+q(jm-im)+s} = \boldsymbol{r}_{im+s} + (q+1)\,\mathbf{A}$.

Letting $\mu = (jm - im)$, we have shown that the sequence $C_i$ has the following structure: $\boldsymbol{r}_{im}, \boldsymbol{r}_{im+1}, \boldsymbol{r}_{im+2}, \ldots, \boldsymbol{r}_{jm}$,

$$\begin{aligned}
&\boldsymbol{r}_{jm+1} = \boldsymbol{r}_{im+1} + \mathbf{A}, \boldsymbol{r}_{jm+2} = \boldsymbol{r}_{im+2} + \mathbf{A}, \ldots, \boldsymbol{r}_{jm+\mu} = \boldsymbol{r}_{jm} + \mathbf{A} \\
&\boldsymbol{r}_{jm+1+\mu} = \boldsymbol{r}_{im+1} + 2\mathbf{A}, \boldsymbol{r}_{jm+2+\mu} = \boldsymbol{r}_{im+2} + 2\mathbf{A}, \ldots, \boldsymbol{r}_{jm+2\mu} = a_{jm} + 2\mathbf{A} \\
&\boldsymbol{r}_{jm+1+2\mu} = \boldsymbol{r}_{im+1} + 3\mathbf{A}, \ldots \quad \text{and so on.}
\end{aligned}$$

Thus, we have shown that the sequence $C_i$ is periodic with the periodic block $\boldsymbol{r}_{im}$, $\boldsymbol{r}_{im+1}$, $\boldsymbol{r}_{im+2}, \ldots, \boldsymbol{r}_{jm}$. Each row in the above structure shows the periodic blocks. Now we will adjoin $\boldsymbol{r}_{im-1}$ to the sequence $C_i$ and show the resulting sequence is also periodic. First we note that $\boldsymbol{r}_{jm} = \boldsymbol{r}_{im} + (\boldsymbol{r}_{jm} - \boldsymbol{r}_{im})$ or $\boldsymbol{r}_{jm} = \boldsymbol{r}_{im} + \mathbf{A}$. The code letter $S_{im-1}$ and the last three points $\boldsymbol{r}_{im}$, $\boldsymbol{r}_{im+1}$, and $\boldsymbol{r}_{im+2}$ uniquely determine the first point $\boldsymbol{r}_{im-1}$. So, we can conclude that the code letter $S_{im-1}$ and the points $\boldsymbol{r}_{im} + \mathbf{A}, \boldsymbol{r}_{im+1} + \mathbf{A}$, $\boldsymbol{r}_{im+2} + \mathbf{A}$ uniquely determine $\boldsymbol{r}_{im-1} + \mathbf{A}$. Similarly, the code letter $S_{jm-1}$ and the last three points $\boldsymbol{r}_{jm}$, $\boldsymbol{r}_{jm+1}$, and $\boldsymbol{r}_{jm+2}$ uniquely determine the first point $\boldsymbol{r}_{jm-1}$. Because the code is periodic the code letters of $S_{im-1}$ and $S_{jm-1}$ are same and since $\boldsymbol{r}_{jm} = \boldsymbol{r}_{im} + \mathbf{A}$, $\boldsymbol{r}_{jm+1} = \boldsymbol{r}_{im+1} + \mathbf{A}$, and $\boldsymbol{r}_{jm+2} = \boldsymbol{r}_{im+2} + \mathbf{A}$, it follows that $\boldsymbol{r}_{jm-1} = \boldsymbol{r}_{im-1} + \mathbf{A}$. We may note that $\mathbf{A} = (\boldsymbol{r}_{jm-1} - \boldsymbol{r}_{im-1})$. Using the equation $\boldsymbol{r}_{jm} = \boldsymbol{r}_{im} + \mathbf{A}$, we see that $\boldsymbol{r}_{jm+\mu} = \boldsymbol{r}_{jm} + \mathbf{A} = \boldsymbol{r}_{im} + 2\mathbf{A}, \boldsymbol{r}_{jm+2\mu} = \boldsymbol{r}_{jm} + 2\mathbf{A} = \boldsymbol{r}_{im} + 3\mathbf{A}$ and so on. We will add the new element $\boldsymbol{r}_{im-1}$ to the first row and remove the last element $\boldsymbol{r}_{jm}$ and add it as the first element in the second row. Again remove the last element from the second row and make it the first element of the third row and so on. Rewriting $\mathbf{A}$ as $\boldsymbol{r}_{jm-1} - \boldsymbol{r}_{im-1}$ the sequence $\boldsymbol{r}_{im-1}, \boldsymbol{r}_{im}, \boldsymbol{r}_{im+1,} \ldots$, has the following structure:

$$
\begin{aligned}
&\boldsymbol{r}_{im-1}, \boldsymbol{r}_{im}, \boldsymbol{r}_{im+1}, \ldots, \boldsymbol{r}_{im+\mu-1},\\
&\boldsymbol{r}_{jm} = \boldsymbol{r}_{im} + \mathbf{A}, \boldsymbol{r}_{jm+1} = \boldsymbol{r}_{im+1} + \mathbf{A}, \boldsymbol{r}_{jm+2} = \boldsymbol{r}_{im+2} + \mathbf{A}, \ldots, \boldsymbol{r}_{jm+\mu-1} = \boldsymbol{r}_{jm-1} + \mathbf{A}\\
&\boldsymbol{r}_{jm+\mu} = \boldsymbol{r}_{im} + 2\mathbf{A}, \boldsymbol{r}_{jm+1+\mu} = \boldsymbol{r}_{im+1} + 2\mathbf{A}, \boldsymbol{r}_{jm+2+\mu} = \boldsymbol{r}_{im+2} + 2\mathbf{A}, \ldots, \boldsymbol{r}_{jm+2\mu-1}\\
&\qquad = \boldsymbol{r}_{jm-1} + 2\mathbf{A}\\
&\boldsymbol{r}_{jm+2\mu} = \boldsymbol{r}_{jm} + 3\mathbf{A}, \boldsymbol{r}_{jm+1+2\mu} = \boldsymbol{r}_{im+1} + 3\mathbf{A}, \ldots \quad \text{and so on.}
\end{aligned}
$$

We can repeat this until we reach the first point. This shows that the path P is periodic with the periodic block $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_\mu$, and $\mathbf{A} = (\boldsymbol{r}_{jm-im} - \boldsymbol{r}_0)$. Since $\mu = (j - i)\, m$ and $(j - i) \leq 24$ it follows that an infinite path P with periodic code of period m has a period not more than 24 times the period of the periodic code. This completes the proof of this theorem.

*Remark 3.5* Theorem 3.4 says a bit more than the statement indicates. The theorem implies that the first 3pd that appears in an infinite path with periodic code must reappear making a periodic block. This is not a consequence of the Dirichlet box principle. This is the consequence of Lemma 2.4.

*Remark 3.6* Results from this section ensure that an infinite path with periodic code is a periodic path. Thus if we are interested in periodic paths such as a helices then it is enough to restrict our attention to paths with periodic codes.

## 4 Helices in the diamond network

Our interest is in helices in the diamond net. By a *helix* in a given overall direction we understand a curve which overall proceeds along that direction while also writhing around that direction, with the net number of writhings (signed + and − for opposite writhings) overall diverging ever farther from 0 as one continues farther and farther along the helix. Admittedly this characterization is somewhat ambiguous, though it

can be made precise at least in the periodic case, as we now do. That is, attention is limited to periodic paths infinitely extending in both directions. In Sect. 3, we have characterized the infinite periodic paths in the diamond network: *An infinite path in the diamond network is a periodic path if and only if it has a periodic code.* Thus it is enough to search for the helices among the paths with infinite periodic codes. It is clear that if the periodic code corresponds to a helix, then the periodic block is to correspond to a net turning or writhing in the helix. If a path P is of period $n$ with periodic block $\boldsymbol{r}_{i+1}, \boldsymbol{r}_{i+2}, \ldots, \boldsymbol{r}_{i+n}$ and $\tau = \boldsymbol{r}_n - \boldsymbol{r}_0$, then the projection of the points of the periodic block projected onto a plane $\pi_\perp$ normal to the vector $\boldsymbol{\tau}$ are seen to form a self-returning walk in $\pi_\perp$. And if this walk in $\pi_\perp$ makes some non-zero number of full 360° turns, then the parent path corresponds to a *periodic helix*. In more detail, one imagines a walk along the regular path between two sites separated by the repeat vector $\boldsymbol{\tau}$, and this constrains the corresponding walk $w_\perp$ projected into the normal plane. The different steps of the projection $w_\perp$ are in general of different lengths, and from one step to the next step in the walk $w_\perp$, there is often a well-defined change of direction $\Delta\vartheta$, measured as + or − as the (minimum-angle) turn is in the counter-clockwise or clockwise direction. If one of the projected steps is of length 0 (as when the step of the unprojected path is parallel to $\boldsymbol{\tau}$), then $\Delta\vartheta = 0$. If the sum of these $\Delta\vartheta$ over a full period (of $n$ steps in the parent path on the diamond net) gives a non-zero multiple of 360°, then the path is a regular helix.

There remains an ambiguity in this definition of a (regular) helix. For if there are steps in the path which appear as reversals in the normal plane $\pi_\perp$, then there is an ambiguity as to whether to count the turn angle as $\Delta\vartheta = +180°$ or $\Delta\vartheta = -180°$. In such a circumstance, one approach is to choose both possibilities, and use both as candidates in making separate candidate summations for the net turn angle for an $n$-path. The different possible candidate answers for the net turn angle correspond to different results obtainable upon infinitesimal distortions from the ideal with "well-signed" angles that deviate but slightly from strict reversals. Now granted several such net sums for the turn angle, if there is no net sum = 0, then the result is still a (regular) helix. If there is a net sum = 0, but the average of these net sums is not = 0, then we might call the periodic path a regular *quasi-helix*—in the sense that most infinitesimal distortions lead to something which is a helix. And if the average of these sums is 0, then the result is in some sense *quasi-a helical*. Again, if there is only one candidate sum, the decision is clear, being a helix or non-helix respectively as the sum is $\neq 0$ or $= 0$.

Many mathematics books identify helices as curves on a surface such as a cylinder and such as to have constant value for the *pitch* (which is the rate of change of the candidate curve in the overall direction $\boldsymbol{\tau}$ of the helix). But these here are viewed just as examples of helices, of which we consider a more general type. If helices are restricted to have constant local pitch, then this is a strong constraint on the subset of helices which occur in the diamond network—and indeed there are only a few possible choices for the translation vector $\boldsymbol{\tau}$ which allow such constant pitch helices. But more especially most of the molecular helices found in chemistry and molecular biology are not such constant-pitch helices—if the pitches are computed for the individual bonds along the backbone of the polymeric structures. The definition of a periodic helix may be used to search for such helices in the diamond net—as indicated in Sect. 5.2.

### 4.1 Testing for helicity

Given a periodic code we may test if it corresponds to a helix. First we evaluate the points of a periodic path corresponding to the code by starting at the origin and choose a 3pd corresponding to the first code letter from the list of 24 possible choices of 3pd's. The period of the path is found by extending the path until the first 3pd (of our choosing) makes a second appearance. Let $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots$, denote the points of the path evaluated one at time. The points $\boldsymbol{r}_0 = \mathbf{0}, \boldsymbol{r}_1, \boldsymbol{r}_2$ correspond to the first 3pd and suppose $\boldsymbol{r}_{n+1}, \boldsymbol{r}_{n+2}, \boldsymbol{r}_{n+3}$ correspond to its second appearance, then the period of the path is n, the translation vector $\boldsymbol{\tau}$ is $\boldsymbol{r}_n - \boldsymbol{r}_0$ and the periodic block is $\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1}$. Theorem 3.4 assures us that the period of $n$ of the periodic path is not greater than $24\chi$, where $\chi$ is the period of the code. The next step proceeds with the test for helicity, utilizing the step projections into the plane normal to $\boldsymbol{\tau}$, the projection of a unit step $\boldsymbol{r}_{i+1} - \boldsymbol{r}_i \equiv \boldsymbol{\Delta}$ along the direction of $\boldsymbol{\tau}$ being expressed in terms of the unit vector $\hat{\boldsymbol{\tau}} \equiv \boldsymbol{\tau}/|\boldsymbol{\tau}|$ as $(\boldsymbol{\Delta} \cdot \hat{\boldsymbol{\tau}})\hat{\boldsymbol{\tau}} \equiv \boldsymbol{\Delta}_\parallel$, so that the normal projection of $\boldsymbol{\Delta}$ is $\boldsymbol{\Delta}_\perp \equiv \boldsymbol{\Delta} - \boldsymbol{\Delta}_\parallel$. And if $\boldsymbol{\Delta}_\perp$ and $\boldsymbol{\Delta}'_\perp$ are two subsequent non-zero projections, then the accompanying turn angle (for the path projected in $\pi_\perp$) is

$$\vartheta = \cos^{-1}\left\{(\boldsymbol{\Delta}_\perp \cdot \boldsymbol{\Delta}'_\perp)/(\left|\boldsymbol{\Delta}_\perp\right|\left|\boldsymbol{\Delta}'_\perp\right|)\right\}$$

with the sum of such turn angles checked to see if it is nonzero. The particular non-zero multiple of 360° gives the number of turns per unit of translation—this might also be called the *helicity* or *writhe number* (per translational unit) for the helix. The sign of this number might also be said to be the sign of the helicity and chirality. The *mean pitch* of the helix evidently is $|\boldsymbol{\tau}|/n$. And local pitches may be identified as $\left|\boldsymbol{\Delta}_\parallel\right|/|\boldsymbol{\Delta}_\perp|$. It may be noted that the regular helices (on the diamond network) are necessarily of even period, because the translation $\boldsymbol{\tau}$ must carry sites into those which are translationally equivalent, such as must occur in the same part of the two separate parts of the bipartite decomposition of the diamond network. For a regular helix, given $\boldsymbol{\Delta}_\perp$ and $\boldsymbol{\Delta}'_\perp$( two subsequent non-zero vectors projections), the angle between $\boldsymbol{\Delta}_\perp \times \boldsymbol{\Delta}'_\perp$ and $\boldsymbol{\tau}$ is either 0° or 180°, if this angle is 0° then the helix is right handed otherwise it is left handed.

*Example 4.1* Consider the periodic code **SRSSRSSRS**... and test if this code corresponds to a helix. The periodic block for the code is **SRS**.

From the list we choose AbD for the first 3-path and begin at the origin. The following shows the points and the code.

| Points on the path | | | | vector $\Delta_{i+1} = \boldsymbol{r}_{i+1} - \boldsymbol{r}_i$ | | | 3pd | Code |
|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{r}_0$ | 0 | 0 | 0 | | | | | |
| $\boldsymbol{r}_1$ | 1 | 1 | 1 | 1 | 1 | 1 | A | |
| $\boldsymbol{r}_2$ | 0 | 2 | 2 | −1 | 1 | 1 | b | |
| $\boldsymbol{r}_3$ | −1 | 3 | 1 | −1 | 1 | −1 | D | **S** |
| $\boldsymbol{r}_4$ | −2 | 2 | 0 | −1 | −1 | −1 | a | **R** |
| $\boldsymbol{r}_5$ | −1 | 1 | −1 | 1 | −1 | −1 | B | **S** |
| $\boldsymbol{r}_6$ | 0 | 2 | −2 | 1 | 1 | −1 | e | **S** |
| $\boldsymbol{r}_7$ | 1 | 3 | −1 | 1 | 1 | 1 | A | **R** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{r}_8$ | 0 | 4 | 0 | −1 | 1 | 1 | b | **S** |
| $\boldsymbol{r}_9$ | −1 | 5 | −1 | −1 | 1 | −1 | D | **S** |
| ... | | ... | | | ... | | ... | ... |

To find the period of the path we look for the second appearance of the first 3pd namely AbD. From the computations we see that $\boldsymbol{r}_7\,\boldsymbol{r}_8\,\boldsymbol{r}_9$ correspond to the second appearance of AbD, thus $\boldsymbol{\tau} = \boldsymbol{r}_6 = (0,\ 2, -2)$ and the period is 6. Let $\theta_{\mathrm{i,i+1}}$ denote the turn angle accompanying the projections of $\Delta_{\mathrm{i}}$ and $\Delta_{\mathrm{i+1}}$. The values of turn angles can be calculated from the formulas provided in this section: $\theta_{1,2} = \theta_{4,5} = \cos^{-1}(1/3)$, and $\theta_{2,3} = \theta_{3,4} = \theta_{5,6} = \theta_{6,7} = \cos^{-1}(\sqrt{3}/3)$, and these six angles add to 360°, also $\boldsymbol{\Delta}_\perp \times \boldsymbol{\Delta}'_\perp = -\boldsymbol{\tau}$. Thus we conclude that the code **SRSSRSSRS**... corresponds to a left handed helix with six points per turn, see Fig. 10. Suppose we use eBd for the first 3pd instead of AbD, then the same result is obtained but the points of the path differ since the 3pd eBd leads to a different path. The points of the path resulting from the choice of eBd are shown below. Also for this choice of the first 3pd we find the same turn angles since the helix is exactly the same.

| Points on the path | | | | vector $\Delta_{i+1} = \boldsymbol{r}_{i+1} - \boldsymbol{r}_i$ | | | 3pd | Code |
|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{r}_0$ | 0 | 0 | 0 | | | | | |
| $\boldsymbol{r}_1$ | 1 | 1 | −1 | 1 | 1 | −1 | e | |
| $\boldsymbol{r}_2$ | 2 | 0 | −2 | 1 | −1 | −1 | B | |
| $\boldsymbol{r}_3$ | 3 | −1 | −1 | 1 | −1 | 1 | d | **S** |
| $\boldsymbol{r}_4$ | 2 | −2 | 0 | −1 | −1 | 1 | E | **R** |
| $\boldsymbol{r}_5$ | 1 | −1 | 1 | −1 | 1 | 1 | b | **S** |
| $\boldsymbol{r}_6$ | 2 | 0 | 2 | 1 | 1 | 1 | A | **S** |
| $\boldsymbol{r}_7$ | 3 | 1 | 1 | 1 | 1 | −1 | e | **R** |
| $\boldsymbol{r}_8$ | 4 | 0 | 0 | 1 | −1 | −1 | B | **S** |
| $\boldsymbol{r}_9$ | 5 | −1 | 1 | 1 | −1 | 1 | d | **S** |
| ... | | ... | | | ... | | ... | ... |

*Example 4.2* Let us consider the two paths given in the Table 2 with codes **RRRRR**... and **SSSSS**... respectively. Ignoring the details, both are seen to be helices of period 4, with a constant turn angle 90°. For the path with code **RRRR**..., we find that $\boldsymbol{\tau} = (4, 0, 0)$, $\boldsymbol{\Delta}_\perp \times \boldsymbol{\Delta}'_\perp = (2, 0, 0)$ and the angle between them is 0°, thus this path is right handed helix. Similarly, for the path with code **SSSSS**..., we find that $\boldsymbol{\tau} = (4, 0, 0)$, $\boldsymbol{\Delta}_\perp \times \boldsymbol{\Delta}'_\perp = (-2, 0, 0)$ and the angle between them is 180°, thus it is a left handed helix.

*Example 4.3* Let us now consider the path described in Table 3.

| Points on the path | | | | Vector $\Delta_{i+1} = \boldsymbol{r}_{i+1} - \boldsymbol{r}_i$ | | | 3pd | Code |
|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{r}_0$ | −1 | 3 | 1 | | | | | |
| $\boldsymbol{r}_1$ | 0 | 2 | 0 | 1 | −1 | −1 | B | |
| $\boldsymbol{r}_2$ | 1 | 1 | 1 | 1 | −1 | 1 | d | |
| $\boldsymbol{r}_3$ | 0 | 0 | 2 | −1 | −1 | 1 | E | **R** |
| $\boldsymbol{r}_4$ | −1 | 1 | 3 | −1 | 1 | 1 | b | **S** |
| $\boldsymbol{r}_5$ | −2 | 2 | 2 | −1 | 1 | −1 | D | **R** |
| $\boldsymbol{r}_6$ | −3 | 1 | 1 | −1 | −1 | −1 | a | **R** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{r}_7$ | $-2$ | 0 | 0 | 1 | $-1$ | $-1$ | B | **S** |
| $\boldsymbol{r}_8$ | $-1$ | $-1$ | 1 | 1 | $-1$ | 1 | d | **R** |

We see that $\boldsymbol{r}_7\ \boldsymbol{r}_8\ \boldsymbol{r}_9$ correspond to the second appearance of BdE, thus $\boldsymbol{\tau} = \boldsymbol{r}_6 - \boldsymbol{r}_0 = (-3, 1, 1) - (-1, 3, 1) = (-2, -2, 0)$ and the period is 6. The turn angles for the periodic block are $\theta_{1,2} = \theta_{4,5} = \cos^{-1}(1/3)$, and $\theta_{2,3} = \theta_{3,4} = \theta_{5,6} = \theta_{6,7} = \cos^{-1}(\sqrt{3}/3)$, and these six angles add to 360°. Since $\boldsymbol{\Delta}_\perp \times \boldsymbol{\Delta}'_\perp = (-2, -2, 0) = \boldsymbol{\tau}$, the angle between $\boldsymbol{\Delta}_\perp \times \boldsymbol{\Delta}'_\perp$, $\boldsymbol{\tau}$ is 0°, the code **RSRRSR**... corresponds to a right handed helix of period 6. Adding $\boldsymbol{\tau}$ to $\boldsymbol{r}_3$ results in the $\boldsymbol{r}_9$ and adding $\boldsymbol{\tau}$ to $\boldsymbol{r}_4$results in the $\boldsymbol{r}_{10}$ and so on.

## 5 Searching for helices in the diamond network

In this section we will present an algorithm to find helices with code of period $n$ ($n \geq 1$) in the diamond network. It may be appropriate to impose an order $<$ on the protochirons as $\mathbf{S} < \mathbf{Z} < \mathbf{R}$ in order to facilitate a lexicographic order on the set of all sequences of protochirons of length $n$. These sequences form the periodic blocks of the code for the infinite periodic paths. We may note that not all periodic blocks of make their first appearance at the $n$th stage, if $d$ is a divisor of $n$, then $n/d$ repeats of a periodic block of length $d$ reappears at the $n$th stage. Also the "cyclic rotations" of a periodic block result in the same infinite periodic path. Thus both types of blocks must be eliminated since they do not create new periodic paths. In the next section we arrive at a formula for counting the number of distinct periodic blocks of size $n$.

### 5.1 Number of essential periodic blocks of size n

A total of $3^n$ distinct sequences are candidates as periodic blocks of size $n$ for the periodic code, however they do not all lead to distinct codes as mentioned earlier, while further some excluded because of self-intersections. To count the number of essential periodic blocks leading to distinct codes we need to eliminate the cyclic rotations of the sequences. We use a consequence of Burnside's Lemma to achieve this goal [13]. Let $A = \{1, 2, \ldots, n\}$ and $B = \{\mathbf{R}, \mathbf{Z}, \mathbf{S}\}$. Let $D$ be the set of all functions from $A$ to $B$. Let $G$ be the cyclic group of order $n$ acting on A. then $G$ acts on the set of mappings $D$ as follows: $\sigma \in G$, $f \in D$, $(\sigma(f))(x) = f(\sigma^{-1}(x))$ for all $x \in A$ and $\sigma(f) \in D$. The number of orbits of $G$ on $D$ is exactly the number of essential periodic blocks of size $n$. Letting $C_k(G)$ denote the number of elements in $G$ that have exactly $k$ cycles in their cyclic decomposition on A, the number of orbits $\lambda_n$ is given by:

$$\lambda_n = \frac{1}{|G|}\sum_{k=1}^{\infty} C_k(G)\,|B|^k = \frac{1}{n}\sum_{k=1}^{n} C_k(G)3^k.$$

We may note that $C_n(G) = 1$, for k $<$ n, $C_k(G) = 1$ if k divides n, otherwise $C_k(G) = 0$, and $C_n(G) = \varphi(n)$ the Euler's function, the number of relatively prime positive integers less than n, thus

$$\lambda_n = \frac{1}{n}\left(3\varphi(n) + \sum_{k>1,k|n}^{n} 3^k\right)$$

For n = 1, the group is just the identity element, $C_1(G) = 1$, there are three sequences namely, **R**, **Z**, **S**.

For n = 2, $\varphi(2) = 1$, and $\lambda_2 = \frac{1}{2}\left(3 + 3^2\right) = 6$, three of the orbits are from n = 1, and the representatives of the three new orbits are **RS**, **RZ**, and **ZS**.

For n = 3, $\varphi(3) = 2$, and $\lambda_3 = \frac{1}{3}\left((2 \times 3) + 3^3\right) = 11$, three orbits are from n = 1, and the rest of the eight orbits are represented by **SRR**, **SZZ**, **ZSS**, **RSS**, **RZZ**, **ZRR**, **ZSR**,**ZRS**.

For n = 4, $\varphi(4) = 2$, and $\lambda_4 = \frac{1}{4}\left((2 \times 3) + 3^2 + 3^4\right) = 24$, since 1 and 2 divide 4, six of these orbits are from earlier stages, there are 18 new orbits.

### 5.2 Algorithm to search helices of period n in the diamond network

We start with an initial $n$-member sequence (**S**, **S**, **S**, . . ., **S**). The algorithm to generate $n$-step periodic polymers in broad outline is as follows:

(A) From the current sequence $P_{\text{current}}$ generate the next sequence $(c_1, c_2, \ldots, c_n)$ which then is identified as $P_{\text{current}}$. If the sequence is (**R**, **R**, **R**, . . ., **R**) end the generation, and otherwise proceed to step B.
(B) Cycle the sequence $P_{\text{current}}$ to its lexicographic lowest form $P'$, and proceed to C.
(C) Check to see if either $P' < P_{\text{current}}$ or $P'$ consists of $n/d$ repeats of an earlier periodic block of size $d$ for some divisor of $n$. If so, then return to A, and otherwise go to the next step D.
(D) From $P'$ construct the path $\boldsymbol{r}_0, \boldsymbol{r}_1, \boldsymbol{r}_2, \ldots$ and test for intersection, either with itself or a later block. If it intersects return to A, and otherwise go to E.
(E) From the path of D test for helicity as explained in Sect. 4.1. If a helix, place $P'$ in a list of helices, and return to step A.

Here in (B) the generation of the next sequence is lexicographically next—and is a simple operation like turning an odometer 1 step further along. Step (C) utilizes a list of earlier shorter sequences.

## 6 Conclusion

A systematic identification of different possible helices is developed from overlapping of a small set of basic protochirons. Special focus has been directed to the structures so realized on the tetrahedral network, though several of the ideas are more general. A different approach to discussing molecular helices (with mathematical and biological hints concerning helical protein chains) may be found in the scientific literature [16–19]. For non-peptide mimetics of $\alpha$-helices, a recent review was published [20]. The findings of the present communication may be useful in the context of chemicals with helical structure: molecules and macromolecules such as inorganic compounds (sulfur or selenium allotropes, CsSb and NaP), or organic substances (polyethylene,

dendronized polymers, biopolymers such as proteins and DNA). From the rich literature we cite only a few references that in turn can provide an ample bibliography [21–25].

**Acknowledgement** DJK acknowledges the support (through grant BD-0894) from the Welch Foundation of Houston, Texas.

## References

1. V. Prelog, in *Foreword to Chemical Applications of Graph Theory*, ed. by A.T. Balaban (Academic Press, London, 1976)
2. V. Prelog, Chirality in Chemistry. Les Prix Nobel 1975 p. 137, The Nobel Foundation, Almquist and Wiskell Internat., Stockholm, 1976; reprinted in Science **193**, 17 (1976), and in J. Mol. Catal. **1**, 159 (1976)
3. V. Prelog, G. Helmchen, Angew. Chem. Int. Ed. **21**, 567 (1982)
4. A.T. Balaban, J. Chem. Inf. Comput. Sci. **37**, 645 (1997)
5. A.T. Balaban, C. Rücker, J. Chem. Inf. Comput. Sci. **41**, 1145 (2001)
6. R.S. Cahn, C.K. Ingold, V. Prelog, Angew. Chem. Int. Ed. **5**, 385 (1966)
7. E. Estrada, J. Phys. Chem. A **107**, 7482 (2003)
8. K. Mislow, *Introduction to Stereochemistry* (W. A. Benjamin, New York, 1965)
9. E.L. Eliel, S.H. Wilen, L.N. Mander, *Stereochemistry of Organic Compounds* (Wiley, New York, 1994)
10. P.J. Flory, *Principles of Polymer Chemistry* (Cornell Univ. Press, Ithaca, 1953)
11. P.J. Flory, *Statistical Mechanics of Chain Molecules* (Wiley-Interscience, New York, 1969)
12. C. Zheng, R. Hoffmann, D.R. Nelson, J. Am. Chem. Soc. **112**, 3784 (1990)
13. A.T. Balaban, D.J. Klein, J.E. Dahl, R.M.K. Carlson, The Open Org. Chem. J. **1**, 13 (2007)
14. J.H. Van Lint, R.M. Wilson, *A Course in Combinatorics* (Cambridge University Press, 1993)
15. M. Hall, Jr., *Combinatorial Theory* (Blaisdell Publishing Company, 1967)
16. H.S.M. Coxeter, Can. Math. Bull. **28**, 385 (1985)
17. J.F. Sadoc, N. Rivier, Eur. J. Phys. B **12**, 309 (1999)
18. E.A. Lord, Struct. Chem. **13**, 305 (2002)
19. E.A. Lord, S. Ranganathan, Eur. J. Phys. D **15**, 335 (2001)
20. J.M. Davis, L.K. Tsou, A.D. Hamilton, Chem. Soc. Rev. **36**, 326 (2007)
21. C. Zheng, R. Hoffmann, D.R. Nelson, J. Am. Chem. Soc. **112**, 3784 (1990)
22. R. Hoffmann, C. Janiak, C. Kollmar, Macromolecules **24**, 3725 (1991)
23. C. Kollmar, R. Hoffmann, J. Am. Chem. Soc. **112**, 8230 (1990)
24. V. Percec, M. Peterca, J.G. Rudik, W. Aqad, M.R. Imam, P.A. Heiney, Chem. Eur. J. **13**, 9572 (2007) and previous papers by V. Percec et al.
25. T.V. Jones, M.M. Slutsky, G.N. Tew, New J. Chem. **32** 676 (2008)