A unified convention for biological assemblies with helical symmetry

Chung-Jung Tsai[a]* and Ruth Nussinov[a,b]

[a]Basic Science Program, SAIC-Frederick Inc., Center for Cancer Research Nanobiology Program, NCI-Frederick, Frederick, MD 21702, USA, and [b]Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Correspondence e-mail: tsaic@mail.nih.gov

Assemblies with helical symmetry can be conveniently formulated in many distinct ways. Here, a new convention is presented which unifies the two most commonly used helical systems for generating helical assemblies from asymmetric units determined by X-ray fibre diffraction and EM imaging. A helical assembly is viewed as being composed of identical repetitive units in a one- or two-dimensional lattice, named 1-D and 2-D helical systems, respectively. The unification suggests that a new helical description with only four parameters $[n_1, n_2, \text{twist}, \text{rise}]$, which is called the augmented 1-D helical system, can generate the complete set of helical arrangements, including coverage of helical discontinuities (seams). A unified four-parameter characterization implies similar parameters for similar assemblies, can eliminate errors in reproducing structures of helical assemblies and facilitates the generation of polymorphic ensembles from helical atomic models or EM density maps. Further, guidelines are provided for such a unique description that reflects the structural signature of an assembly, as well as rules for manipulating the helical symmetry presentation.

1. Introduction

Under physiological conditions, many biomolecules are either organized in functional tubular forms or aggregated in disease-related filaments. Tubular and filamentous structures grow with a helical symmetry. The determination of helical structures is important because it provides clues to functional regulation and to the mechanisms of polymerization and depolymerization, and can help in figuring out how to prevent unwanted disease-related fibril aggregation. The most famous example is the determination of the Watson–Crick double-helical DNA structure in 1953 (Watson & Crick, 1953), which created a new era in the history of molecular biology. In proteins, even a slight difference in the interactions between molecules is sufficient to create similar filamentous or tubular structures with distinct helical symmetries. For this reason, structural polymorphism is a common characteristic of tubular or fibril entities. Depending on the specificity and rigidity of the interacting molecules, some, such as the amyloidogenic peptide $A\beta_{1-40}$ (Sachse *et al.*, 2008; Schmidt *et al.*, 2009), can exhibit a broad spectrum of polymorphic assemblies, whereas others only show limited variability, as in the case of microtubules (Sui & Downing, 2010). This underscores the importance of revealing the structural characteristics of helical assemblies directly from a simple helical symmetry description.

Helical symmetry can be formulated in many different ways. Helical transformations can be classified into two categories: one-dimensional (1-D) helical systems and two-dimensional (2-D) helical systems. In structural determination by X-ray fibre diffraction (Klug *et al.*, 1958), a helical structure is described as a set of *n* 1-D molecular helices related by an *n*-fold axial symmetry. However, in both systems, if the helically repeating motif has C_2 symmetry, the helical structure has an additional dyad symmetry (Klug *et al.*, 1958). Given the asymmetric units, the helical assembly can be constructed by rotohelical transformations which are defined by a specified line group (Damnjanović *et al.*, 2007). Because of the prevalence of structural polymorphism (DeRosier *et al.*, 1999) and helical discontinuities (seams; Kikkawa, 2004) in electron-microscopy (EM) images, a description of helical assemblies by a rolled planar 2-D lattice sheet was devised to solve the EM structure in reciprocal space. In this representation, the framework of a helical structure is viewed as a helical net; that is, a set of equivalent points wrapped around a cylindrical surface. Various 2-D lattice wrappings were defined by a circumference vector $\mathbf{c} = n_1\mathbf{a} + n_2\mathbf{b}$, where n_1 and n_2 are two integer constants and $\mathbf{a}$ and $\mathbf{b}$ are the 2-D lattice vectors. On the other hand, the reconstruction of helical structures in real space is typically based on rotohelical transformations which are applied iteratively using the single-particle method (Sachse *et al.*, 2007; Egelman, 2007, 2010).

A simple convention for defining the helical symmetry of biological assemblies has been suggested in the remediated Protein Data Bank (PDB; Lawson *et al.*, 2008) and EM Data Bank (Heymann *et al.*, 2005) archives. Both have used a definition of rotohelical transformation that does not fully capture the underlying symmetry properties of helical assemblies. It is therefore not surprising that two very similar tubular structures might be described by very different helical parameters that provide no clue to the fact that they are actually quite similar. This is the case for the two bacteriophage major coat protein helical tubes determined by X-ray fibre diffraction [PDB entries 1hgv (Pederson *et al.*, 2001) and 1ifd (Marvin, 1990)]; the first is presented as a one-start and the second as a five-start helical tube with each helix related by a fivefold rotational symmetry. The discrepancy is understandable because the two PDB structures were the outcome of structure-determination procedures in which the helical symmetry was preset in the minimization procedure. This shortcoming underscores the importance of a standard system that would report helical structures and provide parameters that reflect their structural characteristics. It appears that to date an unambiguous, simple and systematic standard for defining a unique helical specification for constructing helical assemblies from asymmetric units is lacking.

In this paper, we present a new unified convention for the construction of helical assemblies from asymmetric units determined by X-ray fibre diffraction and EM imaging. The unification is made possible by an augmented 1-D helical system (described below) that extends the traditional 1-D helical scheme to adopt the helical symmetry descriptor $[n_1, n_2]$ which is used in the 2-D helical system. A helical

structure can be prepared by rolling a planar sheet composed of identical 2-D unit cells (Stewart, 1988). In order to create a seamless 2-D lattice tube, two integer constants $[n_1, n_2]$ define the wrapping process: n_1 refers to the number of cells that are needed to complete a full round of cylinder wrapping and n_2 to the number of cells sliding along the cell edge after the wrapping. The helical symmetry of the tubular structure is explicitly determined by $[n_1, n_2]$ and the corresponding 2-D wrapping transformations can be found in the literature (Tsai *et al.*, 2006; Kikkawa, 2004).

In a traditional 1-D helical system, a helical structure is depicted as either a one-start or an *n*-start helical structure (Egelman, 2007; Klug *et al.*, 1958). For a one-start helical structure, the assembly consists of only a single helix with two helical parameters, twist (φ) and rise (δ); these denote the transformation of the 1-D unit cell which is used to build the entire structure. In fibre diffraction, a one-start helix is formulated as *u* units in *v* turns with a helical repeat distance of *c*, which straightforwardly gives $\varphi = 2\pi v/u$ and $\delta = c/(uv)$. An *n*-start helical structure has *n* helices related by an *n*-fold axial symmetry (C_n), with the axis coinciding with the helical axis. In the augmented 1-D helical system described below, in addition to the rotational operation of the C_n symmetry there is an extra translational operation along the helical axis. In the 2-D helical system this extra translational operation is implicitly included in the 2-D wrapping transformation; however, it is ignored in the traditional 1-D helical scheme. This prevents the 1-D and 2-D systems from being unified in a common helical symmetry description. In contrast, the helical symmetry in the augmented 1-D helical system with the four parameters $[n_1, n_2, \text{twist}, \text{rise}]$ is defined by two consecutive helical (screw) operations: the first helical operation is specified by two helical parameters [twist, rise] exactly as in the traditional 1-D helical transformation and the second screw operation is defined by two $[n_1, n_2]$ constants. n_1 refers to the n_1-fold rotational symmetry exactly as in the traditional 1-D helical transformation and n_2 specifies the translation part of the second screw operation. We will illustrate the operational transformations as well as the interconversion between the augmented 1-D helical system and the 2-D helical system below.

2. Methods

In our definition, a helical structure is composed of *repetitive identical* units, similar to a single crystal which is built from a three-dimensional (3-D) lattice. The definition of repetitive relates to the entire helical structures, which are built from a unit cell with a specified helical symmetry. The unit cell is either a 2-D lattice or a 1-D line segment. The definition of identical implies that each repetitive unit in the construct has exactly the same environment; that is, the independent variables of a helical structure include only parameters involved in helical transformation and coordinates of asymmetric units within a unit cell. Identical implies that when evaluating the energy of the assembly there is no need to include the interactions between all units but just the interactions between one unit cell and its surrounding cells.

Given a 3-D unit cell, it is straightforward to generate the entire single crystal with fractional coordinates. The newly generated fractional coordinates are the number of cells (n_a, n_b, n_c) away from the origin $(0, 0, 0)$ along each edge. The Cartesian coordinates of a new cell can be converted from the fractional coordinates by an orthogonal matrix (Evans, 2001) computed from 3-D lattice constants, a, b, c, α, β and γ. However, unlike the 3-D crystal system, a specified helical transformation is needed to generate the helical assembly from a given 1-D or 2-D unit cell. In the following, the equations for the 2-D wrapping and the augmented 1-D helical transformation will be derived and explained in detail.

2.1. 2-D helical system

As stated in §1, a planar sheet of 2-D lattice can be wrapped into a tube. If all sheet units are identical, the two constant integers n_1 and n_2 are sufficient to define all possible distinct tubes obtained by rolling it. Here, n_1 is the number of cells along one edge of the 2-D lattice (a) which are required to make a full round of wrapping and n_2 refers to the number of cells sliding along the other edge of the lattice (b) after wrapping.

If we place edge **a** along the x axis of the Cartesian coordinate and a 2-D lattice (a, b, γ) is placed on the xy plane, the wrapping equations for an $[n_1, n_2]$ tube are

$$x_{\mathrm{w}} = -(z_{\mathrm{c}} + r)t_y \sin \alpha + (x_{\mathrm{c}}t_x + y_{\mathrm{c}}t_y)t_x,$$
$$y_{\mathrm{w}} = (z_{\mathrm{c}} + r)t_x \sin \alpha + (x_{\mathrm{c}}t_x + y_{\mathrm{c}}t_y)t_y,$$
$$z_{\mathrm{w}} = (z_{\mathrm{c}} + r) \cos \alpha,$$

where

$$r = [(n_1 a + n_2 b \cos \gamma)^2 + (n_2 b \cos \gamma)^2]^{1/2},$$
$$t_x = n_2 b \sin \gamma / (2\pi r),$$
$$t_y = -(n_1 a + n_2 b \cos \gamma) / (2\pi r),$$
$$\alpha = (-x_{\mathrm{c}}t_y + y_{\mathrm{c}}t_x) / r,$$

$(x_{\mathrm{w}}, y_{\mathrm{w}}, z_{\mathrm{w}})$ are the wrapped Cartesian coordinates of the tube and $(x_{\mathrm{c}}, y_{\mathrm{c}}, z_{\mathrm{c}})$ are the Cartesian coordinates of the associated 2-D sheet. In the wrapping equations, the 2-D lattice sheet is at a distance of the tube radius (r) from the tube axis $(t_x, t_y, 0)$. The helical transformation is implicitly specified by the helical twist α and the helical rise $x_{\mathrm{c}}t_x + y_{\mathrm{c}}t_y$. Fig. 1 provides a graphical summary of the 2-D helical transformation. A more detailed description has been given previously (Tsai et al., 2006). Given asymmetric units in a 2-D lattice and a helical symmetry specified by the 2-D helical system in five parameters $[n_1, n_2, a, b, \gamma]$, one can build a complete helical construct based on the 2-D helical transformation equations formulated above.

2.2. Augmented 1-D helical system

Instead of rolling a planar 2-D sheet, a helical structure can also be expressed by a single helix or n helices, with the n helices related by an n-fold screw axis instead of just a rotational axis. Because the helical assembly must consist of

identical subunits, the rotational part of the screw axis must display a C_n rotational symmetry and the translational part should be limited by some discrete numbers. In the augmented 1-D helical system, the four parameters $[n_1, n_2, \varphi, \delta]$ indicate that there are n_1 helices in the assembly, with each individual helix characterized by a unit twist (φ) and a unit rise (δ). Because the helices are also related by an n_1-fold screw axis, each helix denoted by $m_1 = 0, 1, 2, \ldots, n_1 - 1$ has an additional twist of $m_1(2\pi/n_1)$ and a rise of $m_1(n_2/n_1)\delta$. Note that the rise, which is specified by n_2 with a quantity of $n_2/n_1\delta$, was not included in the traditional 1-D helical system. Note also that $m_1 = n_1$ refers back to the first helix as specified by (φ, δ), which will give $n_2\delta$ rise after a complete round of n_1 rotations. In the 2-D helical system, this corresponds to the number of cells involved in the helix sliding after a complete wrapping.

A helical structure in the symmetrical construct is identified by the cell coordinates $[m_1, m_2]$. The asymmetric units are given in cell $[0, 0]$ and an $[m_1, m_2]$ cell is located in the m_1 helix m_2 units away from the cell $[m_1, 0]$ along the helix. If the helical axis is parallel to the y axis and passes through the origin $(0, 0, 0)$, the helical transformation equations for an $[m_1, m_2]$ cell in an $[n_1, n_2]$ helical construct are

$$x_{\mathrm{w}} = x_{\mathrm{c}} \cos \alpha + z_{\mathrm{c}} \sin \alpha,$$
$$y_{\mathrm{w}} = y_{\mathrm{c}} + h_\varphi \delta,$$
$$z_{\mathrm{w}} = -x_{\mathrm{c}} \sin \alpha + z_{\mathrm{c}} \cos \alpha,$$

where

$$h_\varphi = m_2 - m_1(n_2/n_1),$$
$$\alpha = h_\varphi \varphi + m_1(2\pi/n_1),$$

$(x_{\mathrm{w}}, y_{\mathrm{w}}, z_{\mathrm{w}})$ are the transformed Cartesian coordinates for the cell $[m_1, m_2]$ and $(x_{\mathrm{c}}, y_{\mathrm{c}}, z_{\mathrm{c}})$ are the Cartesian coordinates of asymmetric units in cell $[0, 0]$. h_φ and α specify the overall rise and twist for the $[m_1, m_2]$ cell as specified by an n_1-fold screw axis with n_2 unit shift. In the case of a helical structure with a single helix, in which $n_1 = 1$ and $m_1 = 0$, the helical transformation above reduces to a simple helical operation defined by $[\varphi, \delta]$ only. A graphical summary of the augmented 1-D helical transformation is given in Fig. 2.

2.3. 2-D helical system → augmented 1-D helical system

There are four ways to convert a helical system from 2-D to 1-D: view the continuation of lattice edge **b** as a helix, view the continuation of lattice edge **a** as a helix or view the continuation along the vector of **a** + **b** or along the vector of **a** − **b**. The first is the most convenient choice. By selecting the vector **b** as an individual helix, the new 1-D helical system retains the same symmetry notation as the 2-D helical system $[n_1, n_2]$. The unit twist φ (in unit of radians) and rise δ of the n_1-start helices are calculated as

$$\varphi = (-x_{\mathrm{c}}t_y + y_{\mathrm{c}}t_x) / r,$$
$$\delta = x_{\mathrm{c}}t_x + y_{\mathrm{c}}t_y,$$

where t_x, t_y are the helical axes of the 2-D helical system and x_{c}, y_{c} are the planar Cartesian coordinates at the cell origin

$$\mathbf{a} = (a, 0, 0)$$
$$\mathbf{b} = (b\cos\gamma, b\sin\gamma, 0)$$

$$\mathbf{w} = n_1\mathbf{a} + n_2\mathbf{b} = (n_1 a + n_2 b\cos\gamma, n_2 b\sin\gamma, 0)$$
$$|\mathbf{w}| = [(n_1 a + n_2 b\cos\gamma)^2 + (n_2 b\sin\gamma)^2]^{1/2} = 2\pi r$$
$$r = |\mathbf{w}|/2\pi$$
$$\mathbf{w}_u = (w_x, w_y, w_z) = \mathbf{w}/|\mathbf{w}|$$

$$\mathbf{x}_w = \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ z_c + r \end{pmatrix} + h\begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

$$w_z = 0;\ t_x = w_y;\ t_y = -w_x;\ t_z = 0$$

$$\alpha = (x_c w_x + y_c w_y)/r$$
$$h = x_c t_x + y_c t_y$$
$$x_w = -t_y\sin\alpha(z_c + r) + h t_x$$
$$y_w = t_x\sin\alpha(z_c + r) + h t_y$$
$$z_w = \cos\alpha(z_c + r)$$
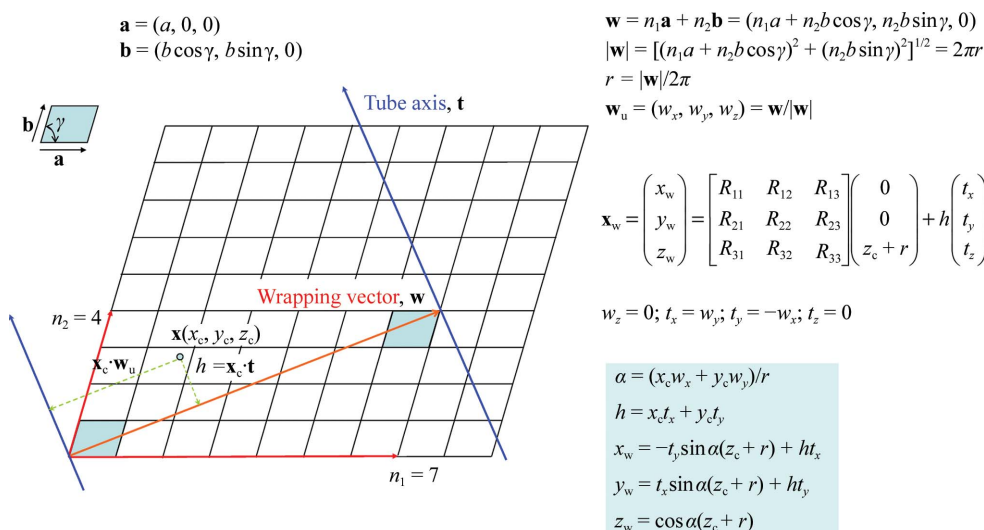
Figure 1
A brief graphic summary of the 2-D helical system. On the left, a 2-D lattice wrapping specified by a circumference vector, $\mathbf{w} = n_1\mathbf{a} + n_2\mathbf{b}$, is sketched with an example of $n_1 = 7$ and $n_2 = 4$. The 2-D lattice is highlighted in color with the angle γ between the two axes $\mathbf{a}$ and $\mathbf{b}$. The lattice lies on the Cartesian xy plane and the axis $\mathbf{a}$ lies along the Cartesian x axis. Thus, the axis $\mathbf{a}$ in Cartesian coordinates is $(a, 0, 0)$ and $\mathbf{b}$ is $(b\cos\gamma, b\sin\gamma, 0)$. The wrapped helical coordinates (x_w, y_w, z_w) of the Cartesian coordinates (x_c, y_c, z_c) in the 2-D planar system can then be calculated by the helical transformation in terms of the helical axis $\mathbf{t}$ (which is perpendicular to $\mathbf{w}$) with the two parameters α and h, twist angle and rise distance, respectively. It is then straightforward to determine the circumferential unit vector $\mathbf{w}_u$, the helical radius r and the helical axis $\mathbf{t}$ as formulated on the right-hand side of the sketch with the summarized wrapping equations highlighted in color at the bottom. Note that vector (t_x, t_y, t_z) as calculated from $\mathbf{w}_u$ is also a unit vector.

$[\varphi, \delta]$ helical transformation for $(0, m_2)$ unit

$$\alpha = m_2 \cdot \varphi$$

$$\begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} = \begin{bmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix}\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} + \begin{pmatrix} 0 \\ m_2\delta \\ 0 \end{pmatrix}$$

$[n_1, n_2, \varphi, \delta]$ helical transformation for (m_1, m_2) unit

$$h = m_2 - m_1(n_2/n_1)$$
$$\alpha = h\varphi - m_1(2\pi/n_1)$$

$$\begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} = \begin{bmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix}\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} + \begin{pmatrix} 0 \\ h\delta \\ 0 \end{pmatrix}$$
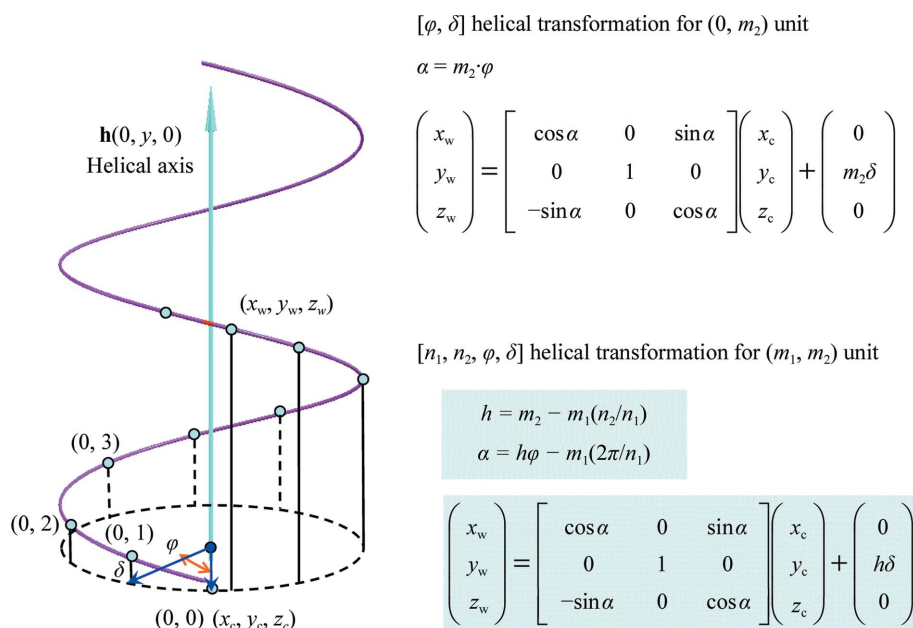
Figure 2
A brief graphic summary of the 1-D helical system. For simplicity, on the left-hand side of the figure, only a single helix is drawn to illustrate the 1-D helical transformation. The helical axis $\mathbf{t}$ is along the Cartesian y axis and the 1-D cell (asymmetric subunits) is sitting at (x_c, y_c, z_c) which is labeled as the $(0, 0)$ cell. The transformed Cartesian coordinates (x_w, y_w, z_w) labeled as $(0, m_2)$ accordingly are calculated by the matrix operations with a rotational angle of $m_2\varphi$ and a translational rise of $m_2\delta$ in the upper right-hand side of the figure. For the augmented 1-D helical system with four parameters $[n_1, n_2, \varphi, \delta]$, the helical transformation equation expressed by matrix operations is given in the lower right corner highlighted in color. α and $h\delta$ respectively specify the overall helical twist and helical rise for the (m_1, m_2) subunits with respect to the $(0, 0)$ asymmetric subunits, with m_1 referring to the n_1-order helix and m_2 referring to the subunits along the denoted helix.

$(0, 1)$. Because the tube axis of the 1-D helical tube (along the y axis) is different from the 2-D helical tube (on the xy plane), the Cartesian coordinates referenced in the 2-D system require some transformations in order to correspond to the new 1-D system. This can be performed either directly in wrapped Cartesian coordinates or in native fractional coordinates. In Cartesian coordinates, the transformation is equivalent to aligning the 2-D helical axis of the xy plane back onto the y axis with the z axis $(0, 0, 1)$ as the rotational axis. Under the right-handed rotational system, the angle θ between the old tube axis and the y axis is calculated as $\mathrm{atan}(-t_x/t_y)$ and the transformations are

$$x_{w1} = x_{w2}\cos\theta + y_{w2}\sin\theta,$$
$$y_{w1} = -x_{w2}\sin\theta + y_{w2}\cos\theta,$$

where x_{w1}, y_{w1} are the wrapped Cartesian coordinates of the new 1-D helical system and x_{w2}, y_{w2} are the wrapped Cartesian coordinates of the old 2-D helical system.

2.4. Augmented 1-D helical system → 2-D helical system

The conversion from a 1-D to a 2-D helical system is not as straightforward as the opposite conversion. This is partly because a 2-D lattice is loosely defined by a single helix structure and partly because of the necessity to revert from the wrapped tube coordinates back to 2-D planar Cartesian coordinates. Given a 1-D helical structure, we first calculate the implicit helical radius from the center of mass of the representative units by assuming that the center of the 1-D helical assembly is located at the origin $(0, 0, 0)$. Secondly, we either use the original n_1 of the 1-D system or determine a new n_1 for the 2-D helical system and define accordingly two wrapped coordinates, (x_{w1}, y_{w1}, z_{w1}) and (x_{w2}, y_{w2}, z_{w2}),

from the 1-D helical system to serve as the origin of cells $(1, 0)$ and $(0, 1)$ of the new 2-D helical system, respectively. Thirdly, we reverse the two wrapped coordinates back to unwrapped planar Cartesian coordinates, (x_{c1}, y_{c1}, z_{c1}) and (x_{c2}, y_{c2}, z_{c2}). With the new calculated planar coordinates, it is straightforward to calculate the 2-D lattice constants as follows:

$$a = (x_{c1}^2 + y_{c1}^2)^{1/2},$$
$$b = (x_{c2}^2 + y_{c2}^2)^{1/2},$$
$$\gamma = (x_{c1}x_{c2} + y_{c1}y_{c2})/(ab).$$

Fourthly, given the newly determined r, a, b, γ and n_1, n_2, the new 2-D helical tube can be determined by solving the quadratic equation $b^2x^2 + 2n_1ab \cos \gamma \, x + (n_1a)^2 - (2\pi r)^2 = 0$.

Finally, all Cartesian coordinates of the old 1-D system are reversed back to the new 2-D planar coordinates.

2.5. Properties of $[n_1, n_2]$ helical system

We have shown that both the 1-D and 2-D helical systems can be represented by two integers, $[n_1, n_2]$, and that the helical assembly can be built through the helical transformation with the associated parameters. However, the helical symmetry specified by these two integers can also be interpreted in a way different from the helical systems' definitions. In the traditional helical description, an assembly with $[n_1, n_2]$ symmetry can be viewed as two sets of n-start helices in which either the arrangement of the n_1-start helices is specified by n_2 or, *vice versa*, that of the n_2-start helices is specified by n_1. The best way to illustrate $[n_1, n_2]$ helical symmetry is by using a helical net: an unwrapped flattened 2-D net bound by the circumference in one direction and extended to infinity parallel to the helical axis. Figs. 3(a) and 3(b) illustrate an example of wrapping and unwrapping of the helical net with the EM structure of a microtubule with $[11, 3]$ symmetry (Sui & Downing, 2010). The colored circular dots in the helical net represent asymmetric units and a line passing through a set of dots is a helix. The number of intersections (n) between the set of parallel lines with the circumference is exactly the number n of helices that are required to fill the helical assembly. This is the origin of the n-start helices definition. In terms of a helical net description, the helical symmetry can be specified by picking a particular set of two intersecting lines (helices) corresponding to n_1-start and n_2-start lines. The intersections define the locations of repeating asymmetric units in the
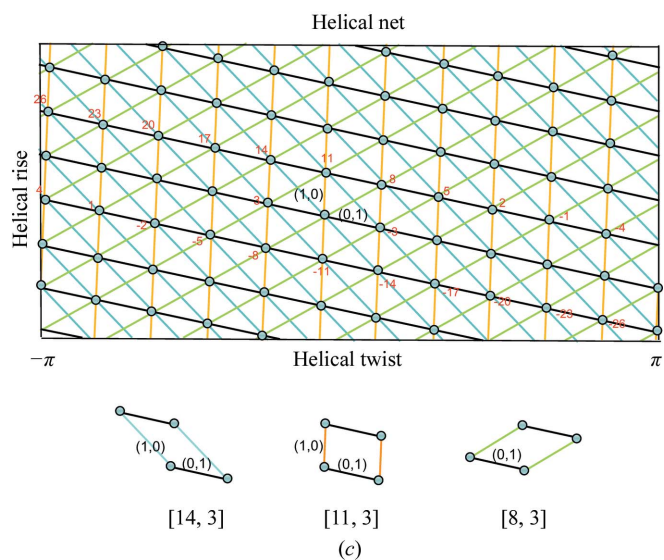

(a)

(b)

(c)
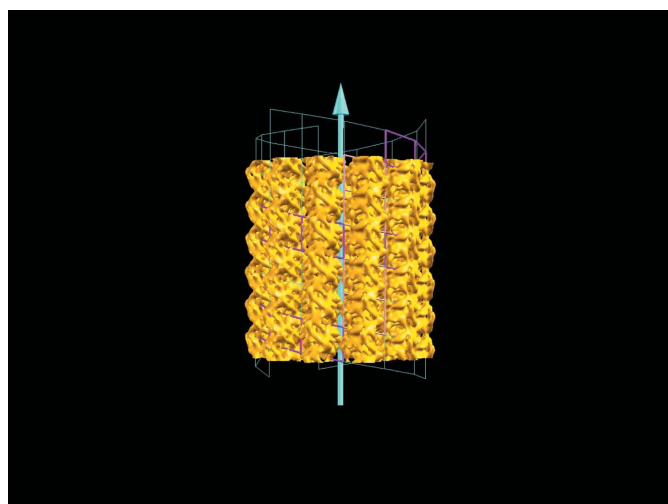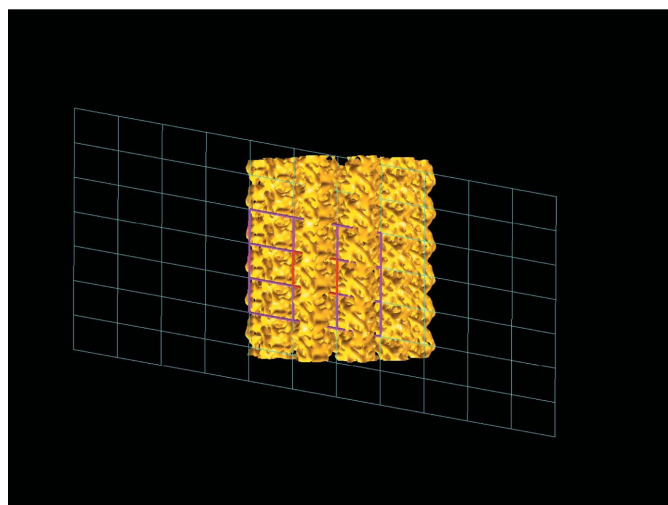
Figure 3
Illustrations of $[n_1, n_2]$ helical symmetry with respect to helical nets. In (a), an EM segment of the microtubule structure is shown in a wrapped helical net with $[11, 3]$ symmetry. The corresponding flattened unwrapped helical net is demonstrated in (b). A section of the corresponding $[11, 3]$ helical net is drawn in (c), with the x axis covering the helical circumference, a twist range of 2π and the y axis parallel to the helical axis, corresponding to the helical rise. The colored circular dots in the net are the asymmetric subunits and a solid line that passes through a set of dots is a helix. A helical net can be redefined by any two sets of lines with their intersections covering all dots. With n_2 fixed at 3, there are ten additional sets of helices which can be used to define the same helical structure. See text for an explanation of why only limited sets of helices are feasible with n_2 fixed at 3. In (c), feasible sets of helices are marked beside the dots with the value of n_1 colored red. Two of the new helical nets with helical symmetry $[8, 3]$ and $[14, 3]$ are superimposed in (c). The individual lattice drawn under the helical nets is to help in the visualization of individual helical nets.

helical structure. In addition to the [11, 3] symmetry, two feasible helical nets with symmetries [8, 3] and [14, 3] are also depicted in Fig. 3(c) for the same helical structure. In the special case of a one-start helix, the entire assembly is built from a single helix instead of a set of helices (n-start helices). Although n_2 is not required in helical symmetry denoted as a one-start helix, it is still represented in the [1, n_2] notation.

Based on the augmented 1-D helical system, it is not difficult to realise that a helical symmetry with [n_1, n_2, twist, rise] is equivalent to [$-n_1$, $-n_2$, twist, rise], [$-n_1$, n_2, $-$twist, $-$rise] and [n_1, $-n_2$, $-$twist, $-$rise]. To reduce the redundancy in the helical symmetry representation, we set several simple rules. Firstly, n_1 is always positive. For consistency in the inter-conversion between the 1-D and 2-D helical systems (with the sign of n_2 kept unchanged), we choose the rise to also be positive. Secondly, the value of n_2 is always smaller than that of n_1. In this way, the handedness of the n_1 helices is determined by the sign of twist: if positive the n_1 helix is right-handed, otherwise it is left-handed. The sign of n_2 gives the handedness of the n_2 helices: if negative it is a right-handed helix, otherwise it is left-handed. To calculate the [twist, rise] of the n_2-start helices the helical symmetry can be swapped from [n_1, n_2] to [n_2, n_1].

2.6. Local C_2 (dyad) symmetry

Above, the helical symmetry operation has been applied to asymmetric subunits without first assigning a plausible local symmetry. In order to generate all symmetric subunits in a planar 2-D lattice, the local symmetry can be specified by one of the 17 wallpaper groups. However, the local symmetry in the planar 2-D lattice is largely lost by the [n_1, n_2] helical transformation; therefore, we only describe pseudo-local symmetry. A local C_2 symmetry operation is an exception: not only is it maintained between asymmetric subunits within the unit cell, but also in the entire helical construct. In terms of the 1-D helical system, the C_2 symmetry is defined as an additional dyad symmetry (with axial C_2 along the z axis and the helical axis along the y axis). To include a local C_2 symmetry operation, we can assign a wallpaper group $p2$ before applying the helical transformation.

2.7. Manipulation of [n_1, n_2] symmetry

A helical symmetry can be described by many [n_1, n_2] combinations. For a particular preset n_2 there are a limited number of n_1; similarly, a preset n_1 will have a limited selection of n_2 for the same helical assembly. To understand the manipulation and limitations of changing from one [n_1, n_2] to another, the helical net is the best reference. For illustration, we use the [11, 3] helical symmetry of the polymorphic helical structure of the microtubule. In terms of the (h, k; n) notation (Toyoshima & Unwin, 1990; Toyoshima, 2000), the set of n-start helices can be specified by the equation

$$n = hn_{10} - kn_{01},$$

where $n_{10} = 11$ and $n_{01} = 3$ for the [11, 3] symmetry. Note that the (h, k) index has to be confined within the circumference

range. Starting with the [11, 3] symmetry and a fixed $n_2 = 3$, the redundant helical symmetry [n_1, 3] can have $n_1 = h\,11 - k\,3$, with $h = \pm1$. On the other hand, given a fixed $n_1 = 11$, we can have many redundant [11, n_2] helical symmetries with $n_2 = h\,11 - k\,3$ and $k = \pm1$. In the case of ($h = 1$, $k = \pm1$), we have new redundant helical symmetries of [8, 3] and [14, 3] (Fig. 3c), respectively. The complete list of redundant [n_1, 3] symmetries with $h = 1$ are [26, 3], [23, 3], [20, 3], [17,3], [14,3], [11,3], [8, 3], [5, 3], [2, 3], [-1, 3] and [-4, 3]. For [n_1, 3] there is an infinite number of helical symmetries in a planar 2-D lattice rather than just the 11 sets restricted by the circumference. In Fig. 3(c), the corresponding redundant sets of helical symmetries are noted next to the helical dots. For the redundant symmetry [-1, 3], we have created an equivalence between 11-start helices [11, 3] and one-start helix [-1, 3] (or [1, -3]) symmetry.

In order to check whether the (h, k) index is within the circumference, we convert the 1-D system to its equivalent planar 2-D helical net. It is then straightforward to calculate the new helical parameters φ for the new n-start helix. If $|\varphi|$ is less than π then it falls within the circumference range.

2.8. Relevance to X-ray fibre diffraction and the EM method

In helical structure determination by X-ray fibre diffraction or EM based on the Fourier–Bessel method (in reciprocal space), the first step is indexing the layer-line diffraction pattern to a specified helical symmetry. There are two possible systems for indexing a diffraction pattern. In the first, assuming that a helical assembly can be described by a single (one-start) helix, the 'selection rule' $l = tn + um$ can be utilized to assign (n, l) pairs to layer-lines in which each layer-line is associated with a set of n-start helices. A more general formalism using (n, Z_l) instead of (n, l), which removes the requirement for t/u to be a rational number, is more appropriate for fibre diffraction. However, for simplicity, we prefer to use the (n, l) system here. A successful layer-line indexing then gives the helical organization as the selection rule implies: u units require t turns of the one-start helix to complete a true repeat with a rise distance of c. A second more systematic (h, k; n) indexing system (Toyoshima & Unwin, 1990; Toyoshima, 2000) interprets diffraction patterns based on the helical surface lattice. In a planar 2-D lattice, diffraction by a set of lines gives a row of dots in reciprocal space. Therefore, it is straightforward to determine a 2-D planar symmetry from an ideal diffraction pattern. For a helical structure which is obtained by wrapping of a 2-D lattice, a set of lines now becomes a set of helices and the corresponding diffraction dots become layer-lines. To define a surface lattice, two indices, (1, 0; n_{10}) and (0, 1; n_{01}), are first assigned where n is the start number of the associated helices, which can be estimated from its peak position in the layer-line diffraction (Toyoshima & Unwin, 1990; Toyoshima, 2000). If the remaining layer-lines can be indexed and related by the equation $n = hn_{10} - kn_{01}$ then the helical symmetry is determined. Fig. 4 illustrates the relationship between the new [n_1, n_2] helical scheme and the symmetry in both indexing

systems. The nodes in Fig. 4 represent (i) asymmetric units based on a simple helical structure which is described by a one-start helix ($t = 4$ and $u = 13$), *i.e.* with 13 units and four turns completing a true repeat of distance c, and (ii) a simplified diffraction pattern of the same helical structure. However, instead of the layer-line pattern for n-order Bessel diffraction, each dot gives the position of (n, l) diffraction where the layer-line pattern has a maximum diffraction peak at the $\sim n + 2$ position (Diaz *et al.*, 2010). In the figure, $[n_1, n_2]$ of the new helical scheme correspond to the choice of n_{10} and n_{01} in the $(h, k; n)$ indexing system. The same ($t = 4$ and $u = 13$)
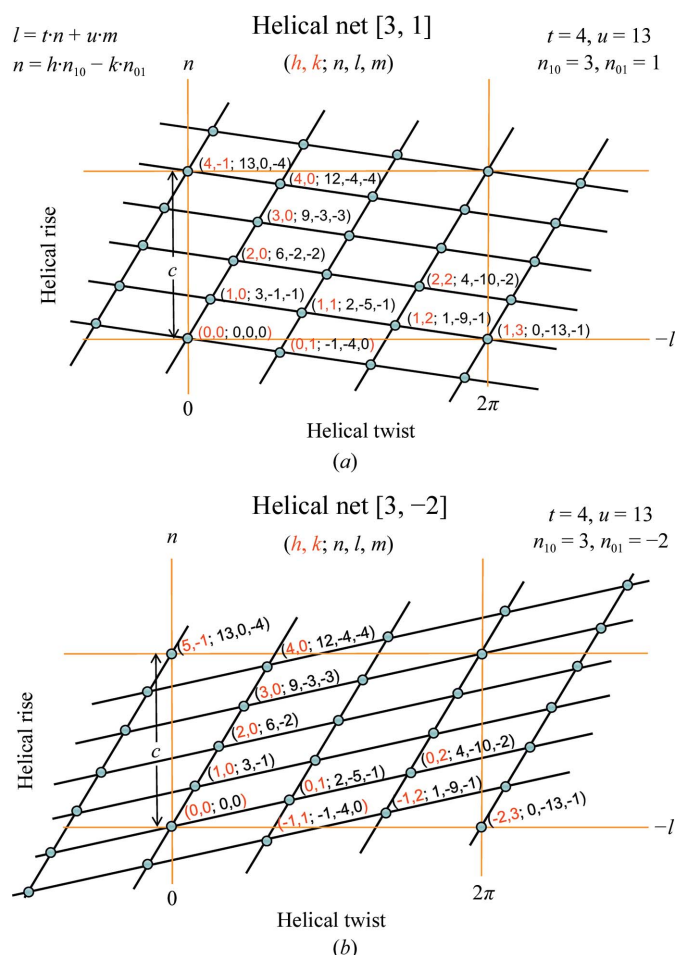

Figure 4
The relationship between the $[n_1, n_2]$ helical scheme and the helical symmetry utilized in two common indexing systems. The dots in the figure represent two properties. Firstly, they represent asymmetric units based on a simple helical structure which is described by a one-start helix ($t = 4$ and $u = 13$) with 13 subunits and four turns completing a repeat of distance c. Secondly, they describe a simplified diffraction pattern of the same helical structure. Thus, instead of showing the layer-line pattern for n-order Bessel diffraction, each dot gives the position of (n, l) diffraction where the layer-line pattern has a maximum diffraction peak at the $\sim n + 2$ position (Diaz *et al.*, 2010). The diffraction pattern in terms of $-l$ and n is related to the helical net description with 13 subunits enclosed by orange lines as a repeating unit. The two implicit helical symmetries, $l = tn + um$ and $n = h\,n_{10} - k\,n_{01}$, are then related by the [3, 1] and [3, −2] helical symmetry in (*a*) and (*b*), respectively, with $n_{10} = n_1$ and $n_{01} = n_2$. In the figure, each dot is labeled with an $(h, k; n, l, m)$ index and the helical lines are in terms of the $[n_1, n_2]$ helical symmetry.

structure is related by two different helical systems [3, 1] and [3, −2] in Figs. 4(*a*) and 4(*b*), respectively, for a simplified diffraction pattern in terms of n and $-l$. The figure only shows one fourth of the diffraction pattern. For example, the $n = 4$, $l = 3$ diffraction is at the left-upper corner of the figure without a label of $(h, k; n, l, m)$.

2.9. General guidelines for presenting a helical structure

There is no clear-cut advantage in treating a helical assembly as a 1-D or a 2-D system; both systems have pluses and minuses. However, we believe that the augmented 1-D helical system is more suitable than the 2-D system for describing a helical structure, even though the two systems are equivalent and interchangeable. There are two reasons for favoring the 1-D helical system for describing a helical symmetry. Firstly, the 1-D helical system is simpler than the 2-D system, with one fewer parameter. Secondly, the 1-D helical scheme is independent of the helical radius, while the surface lattice parameters (a, b, γ) in the 2-D system will change with different radii. Here, based on a 1-D helical system we suggest general guidelines for helical structure representation. With our guidelines, if the assembly units can be unambiguously defined and follow the helical paths, each helical assembly is expected to provide a unique symmetry $[n_1, n_2, \varphi, \delta]$ that also explicitly reflects the helical structural characteristics.

In terms of a helical net, a helical structure is composed of a set of n helices in which the individual helix is named an n-start helix. If a helical structure can be expressed by just a single helix, it is a one-start helical structure. In the augmented 1-D helical system, the $[n_1, n_2]$ representation implies that the organization of the n_1 helices is specified by n_2, with the individual n_1-start helix defined by $[\varphi, \delta]$. We can also swap the representation to say that there are n_2 helices in the structure related by n_1. Since there are many $[n_1, n_2]$ combinations for a particular helical structure, the first and the most important guideline is to define the rule for choosing a unique $[n_1, n_2]$ specification. In order to reflect the helical structural characteristics, the rule states that only protofilaments will be candidates for the $[n_1, n_2]$ selection. In our definition, if adjacent asymmetric subunits in an assigned helix are in physical contact, this helix is a protofilament. Therefore, we first sort protofilaments according to the extent of contacts between adjacent asymmetric subunits. Of the best four protofilaments, the one with the twist angle closest to zero is set as the primary protofilament n_1 and the next best protofilament is selected as the secondary protofilament n_2. Note that n_1 is always larger than $|n_2|$ under this guideline.

To ensure a unique helical symmetry representation for a helical assembly, redundancy needs to be reduced to singular $[n_1, n_2, \varphi, \delta]$. The reduction guideline requires that $n_1, \delta > 0$ and $n_1 > |n_2|$. In the case of $n_1 < 0$, one can apply the equivalent rule that the new $[n_1, n_2] = [-n_1, -n_2]$. If δ is negative, one can simply apply the equivalent rule $[n_1, n_2, \varphi, \delta] = [n_1, -n_2, -\varphi, -\delta]$ to make it a positive value.

Table 1
Helical parameters for helical structures solved by X-ray fibre diffraction and EM imaging.

The first and second columns provide the PDB code and molecular name of the helical structure. The third column indicates whether the helical assembly was solved by X-ray fibre diffraction (XFD) or cryo-EM imaging (CryoEM). If the structure was solved by cryo-EM imaging, the PDB file records atomic models which have been docked into the corresponding EM map. If the EM map was deposited in EMDB, the entry gives its EMDB ID code. The next three columns (C_n, φ, δ) report the helical symmetry if specified in the PDB or EMDB. The following two columns give the unified helical symmetry of $[n_1, n_2]$ by following the symmetry-determination guideline (provided in the text). Next are the newly determined helical parameters $[\varphi, \delta]$ of the 1-D helical system and the lattice constants (a, b, γ) of the 2-D helical system.

| PDB code	Molecule	XFD/cryoEM	C_n	Twist φ (PDB) (°)	Rise δ (PDB) (Å)	n_1	n_2	Twist φ (1-D) (°)	Rise δ (1-D) (Å)	a	b	γ
1ifd	Inovirus	XFD	C_5	−33.23	16.00	10	−5	5.54	32.00	21.16	32.06	37.35
1hgv	PH75 inovirus	XFD	C_1	66.67	2.90	11	−6	13.33	31.90	23.05	32.29	32.01
1cgm	Mosaic virus	XFD	C_1	22.04	1.44	16	−1	−7.34	23.11	22.59	24.31	104.34
2zwh	F-actin model	XFD	C_1	−166.40	27.59	2	1	27.20	55.18	69.06	56.14	77.07
3a5x	L-type flagellar	CryoEM	C_1	65.30	4.79	11	5	−1.70	52.65	42.96	52.68	125.82
1mwk	ParM filament open	EMD-5128	C_1	165.20	24.30	2	−1	−29.50	48.60	62.30	49.67	78.94
1mwk	ParM filament closed	EMD-5129	C_1	165.00	24.20	2	−1	−30.10	48.60	62.75	49.73	79.47
2hi5	Bacteriophage	EMD-1240	C_5	−34.62	17.40	10	−5	2.60	34.80	22.50	34.82	138.88
3b5u	Acrosomal actin	EMD-1088	C_1	−164.90	27.75	2	1	23.34	54.60	67.74	55.27	75.18
3dik	HIV-1 CA	EMD-5136				13	2	−11.00	89.80	96.64	96.68	119.96
3a69	Flagellar hook	EMD-1647	C_1	64.79	4.12	11	−6	−7.31	45.32	38.37	45.93	52.25
3dco	Microtubule	EMD-5038				13	3/2	0.00	80.00	51.94	80.00	100.24

3. Results

There are many helical filaments and tubular structures in the PDB which have been solved either directly by X-ray fibre diffraction or by fitting individual crystal structures into cryo-EM density maps. Similar to X-ray crystal structures where only the coordinates of the asymmetric units are included in the PDB file, most of the helical structures deposited in the PDB also contain only asymmetric units. Therefore, in principle, the entire helical structures should be constructed from the deposited asymmetric units by a specified helical symmetry. In the case of crystal structures, a space group and six lattice constants are defined in the keyword 'CRYST1' in the PDB for calculating all symmetric units in the unit cell. However, owing to the lack of a simple, complete and widely accepted system for helical symmetry, no keyword has been set to define the helical symmetry and helical parameters are implicitly stated in the comments. Furthermore, the creation of the entire helical structure relies on a set of translational and rotational matrices which are hard-coded in the PDB.

We have applied the augmented 1-D helical scheme along with the suggested guidelines to all helical structures deposited in the PDB. A small portion of the results are given in Table 1 and a complete list is available on the web at http://protein3d.ncifcrf.gov/helicalSymmetry/table1.html. The newly determined helical parameters [n_1, n_2, twist, rise] not only directly reflect the helical characteristics but also provide sufficient information for constructing an entire helical structure from given asymmetric units. Here, we propose four helical parameters in a new keyword named HELSYM in the PDB for the specified helical symmetry to avoid using matrices and comments when specifying a helical symmetry.

In the PDB, the axial symmetry of a helical structure is conventionally along the z axis and passes through the origin (0, 0, 0). The helical symmetry specified in the PDB usually follows the rotohelical description, which provides the helical parameters (φ, δ) for a single helix or n helices related by a C_n rotational axis. The manually extracted data, the helical twist φ and the helical rise δ, are first verified against the helical transformation matrices if also given in the PDB file. The corresponding augmented 1-D helical parameters will then be either [1, 0, φ, δ] or [n, 0, φ, δ] for one-start helices or n-start helices, respectively. We then use our graphics tool named *PNAS* (*Protein Nanoscale Architecture by Symmetry*), in-house software running both under Linux and Windows, to search for the first four protofilaments with the largest contact between the asymmetric units and determine their n_1, n_2, φ, δ values accordingly. Next, we select among them the helical protofilament with the lowest absolute value of twist angle as
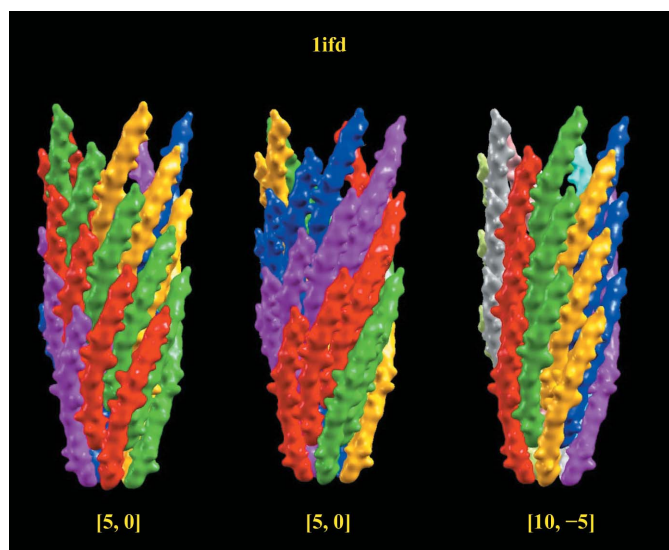
Figure 5
A helical structure with various distinct descriptions of helical assembly. The helical structure of bacteriophage major coat protein (PDB entry 1ifd) is used as an example here to illustrate the variations. In the figure, an individual asymmetric subunit is presented by a color isosurface entity and each color represents a helix specified by the helical symmetry. The first [5, 0] helical symmetry given in the PDB file has a corresponding 1-D helical symmetry of [5, 0, −33.2, 16.0]. The other two helical symmetry assignments, which are more meaningful in terms of reflecting structural characteristics of the helical assembly, are given as [5, 0, 38.8, 16.0] and [10, −5, 5.5, 32.0].

the primary n_1 helix and its associated twist and rise are set as the 1-D helical parameters $[\varphi, \delta]$. Finally, the highest contact protofilament other than the chosen primary helix is assigned as the secondary n_2 helix to complete the determination of $[n_1, n_2, \varphi, \delta]$.

In Fig. 5, the helical assembly of the bacteriophage major coat protein (PDB entry 1ifd; Marvin, 1990) is assigned into three different 1-D helical symmetries: [5, 0, −33.2, 16.0], [5, 0, 38.8, 16.0] and [10, −5, 5.5, 32.0]. The first symmetry [5, 0] corresponds to the rotohelical assignment in the PDB. Apparently, each individual helix is not a protofilament since no contact between helical subunits (shown in the same color) is observed. Therefore, the helical structural characteristics will not be conveyed clearly from its helical parameters. The second [5 ,0] symmetry is based on the first protofilament with the largest number of contacts between helical subunits. However, the guideline suggests using the [10, −5] helical symmetry to represent this structure. This symmetry is advantageous for three reasons: firstly, the [10, −5] symmetry corresponding to the second and first protofilaments in the structure presents the best structural characteristics, unlike the second [5, 0] assignment which only contains information for the first protofilament; secondly, by looking down the helical axis the structure is composed of ten helices, not just five; and thirdly, another inovirus coat protein (PDB entry 1hgv; Pederson *et al.*, 2001) also gives a similar structure with [11, −6] helical symmetry. Here, the primary (11-start) helix is the first protofilament and the secondary (six-start) helix is the second protofilament. These two examples show that the augmented 1-D helical representation not only describes similar helical structures by similar parameters but at the same time also differentiates between similar helical organizations. In Fig. 6, three additional helical structures are depicted in 1-D symmetry. Pictorial descriptions with 1-D helical symmetry for the complete list of known helical structures can be accessed from links on the webpage http://protein3d.ncifcrf.gov/helicalSymmetry/table1.html.

For helical structures deposited in the EMDB (Lawson *et al.*, 2011) only the helical classification is indicated but no helical symmetry is explicitly given in the data bank. However, it is not difficult to deduce the 1-D helical parameters if the helical axis can be determined from the EM density map. The graphics tool *PNAS* can be utilized to assign 1-D helical symmetry to the EM structure. Firstly, we determine the location of the primary protofilament by visual inspection of the density map when shown in various isosurface presentations. Looking down the EM map along the helical axis, the number of assigned protofilaments can be counted to give the n_1 helical parameter. *PNAS* then determines the position of the helical axis using either the given map center or the calculated coordinates of the center of density. Next, we determine the $[\varphi, \delta]$ pair for the visually assigned primary protofilament by calculating the correlation coefficient (Grubisic *et al.*, 2010) between the origin map density and the helical transformed density specified by a pair of manually adjustable parameters $[\varphi, \delta]$. In this procedure, we follow the guideline to keep the helical twist as close to zero as possible and at the same time change the twist and rise to reach the
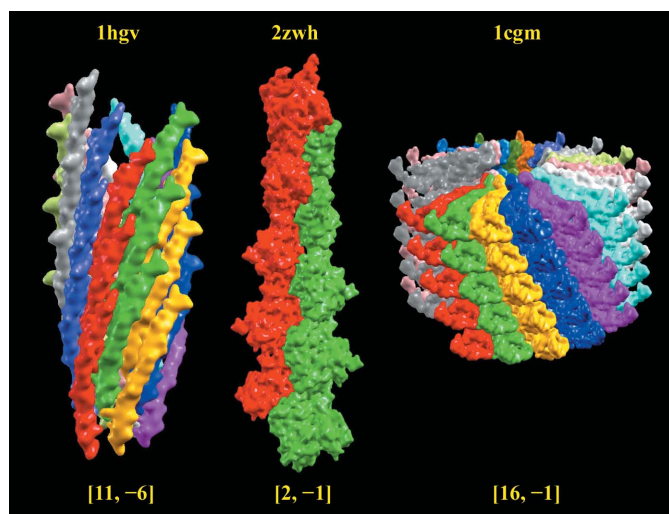

Figure 6
Pictorial 1-D helical symmetry description of three helical structures determined by X-ray fibre diffraction. The three depicted helical structures are the filamentous bacteriophage ph75 (PDB entry 1hgv), F-actin (PDB entry 2zwh) and the cucumber green mottle mosaic virus (PDB entry 1cgm). The same isosurface and color definition described in Fig. 5 is used for the three helical assemblies. The pictorial presentation directly indicates the helical signatures of the three helical structures in 11, two and 16 colored protofilaments, which are respectively implied by the specified [11, −5], [2, −1] and [16, −1] helical symmetry.
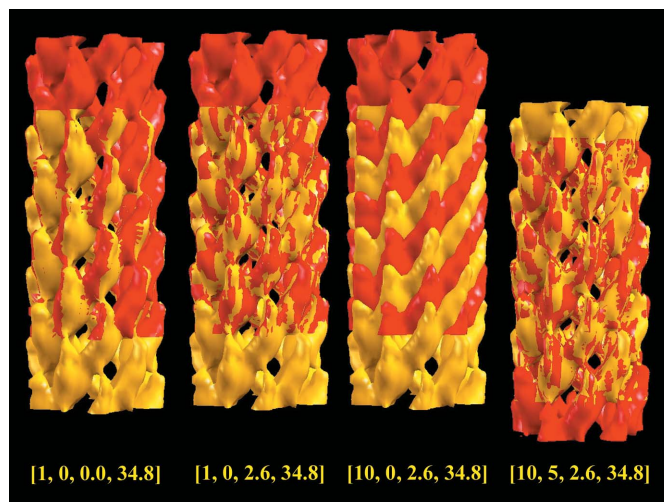

Figure 7
The procedure of serial density-map superimpositions illustrates the determination of the 1-D helical symmetry from a helical EM structure. The bateriophage fd coat protein (EMD-1240) is used here for demonstration. The original EM structure is displayed as a yellow isosurface and the symmetry-transformed density map is shown as a red isosurface. A first superimposition, denoted by the 1-D helical symmetry operation [1, 0, 0.0, 34.8], gives the result of a 34.8 Å translation only. Following an additional rotation of 2.6° denoted by [1, 0, 2.6, 34.8], the outcome of perfect superposition determines the $[\varphi, \delta]$ of 1-D helical symmetry for the set of ten-start primary protofilaments. Applying a tenfold rotation [10, 0, 2.6, 34.8] then gives the third superimposition. Finally, an extra translation defined by n_2 = 5 completes the determination of 1-D helical symmetry [10, 5, 2.6, 34.8] for the EMD-1240 structure.

Table 2
Comparison of helical parameters between the reported symmetries determined in the EM reconstruction and the new helical symmetries for six polymorphic microtubule structures.

See Table 1 for column name description. The last column gives the helical radius where the 2-D lattice was defined.

CryoEM	C_n	Twist φ	Rise δ	n_1	n_2	Twist φ (1-D)	Rise δ (1-D)	a	b	γ	Radius (COD)
EMD-5191	C_1	−32.47	11.08	11	3	0.95	40.6	52.58	40.63	100.03	90.71
EMD-5192	C_1	−29.88	10.16	12	3	0.50	40.6	52.35	40.61	99.97	98.49
EMD-5193	C_1	−27.69	9.39	13	3	0.00	40.6	52.59	40.60	100.26	107.07
EMD-5194	C_1	−25.77	8.72	14	3	−0.25	40.6	51.92	40.60	100.35	113.81
EMD-5195	C_1	−23.83	10.81	15	4	0.65	40.6	51.00	40.62	100.34	119.85
EMD-5196	C_1	−22.40	10.18	16	4	0.40	40.6	51.12	40.61	100.19	128.16

best match as guided by visual superimposition between the origin and the transformed density map in isosurface presentation. The optimal correlation coefficient should be very close to 1.0. Now, we can use the newly determined helical parameters n_1, φ and δ to determine n_2: simply try integer numbers between $-n_1$ and n_1 and perform the 1-D helical transformation to determine n_2 from the result of the superimposition as stated above.
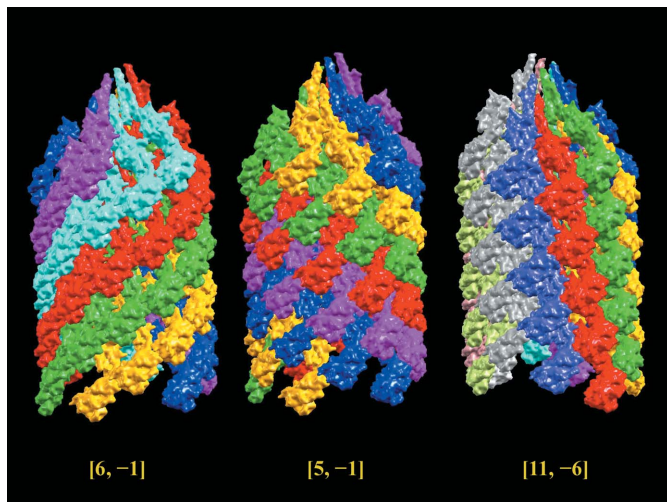
To illustrate the procedure of 1-D helical symmetry determination, four superimpositions between the original EM (EMD-1240; bateriophage fd coat protein B; Wang *et al.*, 2006) and transformed density maps relating to the four stages are given in Fig. 7. The first two superimpositions illustrate the determination of $[\varphi, \delta]$ for the assigned primary protofilament. The last two superimpositions illustrate the determination

[6, −1]   [5, −1]   [11, −6]

Figure 8
1-D helical symmetry determination for a complicated helical structure. This example illustrates that the guideline defined for 1-D helical symmetry determination is capable of giving a unique symmetry assignment $[n_1, n_2, \varphi, \delta]$ to an intricate helical structure. The asymmetric subunit of the bacterial flagellar hook (Fujii *et al.*, 2009; PDB entry 3a69; EMD-1647) contains three protein domains spanning the inner, middle and outer layers of the helical structure. Based on individual domains, the best protofilament in each domain forms a set of six-start, five-start and 11-start helices, respectively, from the outer to the inner layer of the helical structure. The pictorial helical descriptions for three different symmetry assignments are given under [6, −1], [5, −1] and [11, −6] symmetry. The guideline prefers the [11, −6] symmetry assignment simply because the 11-start helix has a twist angle closest to zero and the six-start helix is the protofilament with the largest number of contacts between the asymmetric units along the protofilament.

of the secondary protofilament n_2, giving the 1-D helical symmetry [10, 5, 2.6, 34.8]. The 1-D helical symmetry for each helical EM map deposited in the EMDB has been determined with the graphics tool *PNAS*. Some of the results are listed in Table 1 and a complete list is reported on the webpage http://protein3d.ncifcrf.gov/helicalSymmetry/table1.html. To highlight the importance of a comprehensive helical scheme, Table 2 provides a comparison between the reported helical symmetries determined in the EM reconstruction and the new helical symmetries for six polymorphic helical structures of the microtubule. The results clearly show that the inherent structural characteristics of the microtubule obtained by the new helical scheme can directly discover polymorphic ensembles. The very similar surface lattice within different helical symmetries implies very similar subunit–subunit interactions which the microtubule uses to assemble into divergent helical organizations.

4. Discussion

4.1. Will the guidelines give a unique [n_1, n_2, φ, δ] helical system?

A helical description using four parameters [n_1, n_2, φ, δ], determined according to the augmented 1-D helical symmetry guidelines (in §2) provides the helical signature of the structure. This is because the two sets of defined helices, the n_1-start and the n_2-start, correspond to the two sets of protofilaments. However, will the guidelines also always give a unique [n_1, n_2] combination for a given helical structure? The answer is yes, as illustrated by the example below. The docked atomic model of the bacterial flagellar hook (Fujii *et al.*, 2009; PDB entry 3a69) contains an asymmetric subunit with three protein domains spanning the inner, middle and outer layers of the helical cryo-EM map. In terms of individual protein domains, the best protofilament of each domain yields an 11-start, five-start and six-start helix, respectively, from the inner to the outer layers. Even though different helical descriptions of different layers are observed, the guidelines still give an unambiguous helical symmetry of [11, −6, −7.31, 45.32] for this structure. The assignment is based on two clear elements in the structural data: the 11-start helix (the third protofilament in the protein) has a twist angle closest to zero and the six-start helix is the first protofilament. In Fig. 8, each color presents an assigned helix and the assembly illustrates six, five and 11 helices, respectively, for the [6, −1], [5, −1] and [11, −6] symmetries.

Not all helical structures have unambiguous primary protofilaments, especially when the growth mechanism does not follow a helical path. The tubular structure of the HIV-1 capsid protein (CA; Byeon *et al.*, 2009) is such an example. In solution, CA forms a dimer *via* the association of its C-terminal domain (CTD). The cryo-EM tubular structure (EMD-5136) reveals that the basic unit is a trimer of CA dimers with a pseudo-threefold at the CTD–CTD interfaces and the CA dimer is shared between two trimers. Following our guideline, we obtain a helical symmetry of [24, 13, 7.39, 165.78] for the CA tubular structure. The unit cell depicted by the [24,13] symmetry does not correspond to the observed CA hexamer; however, after applying the symmetry-manipulation rules ($n_1 = n_1 - n_2$, $[n_1, n_2]$ swapping and $n_2 = n_2 + n_1$) the new helical symmetry of [13, 2, −11.00, 89.80] gives the cell dimensions of the hexamer. The surface lattices for both helical symmetries are highlighted in red in Fig. 9. The fact that the assigned asymmetric units in both helical symmetries ([24, 13] and [13, 2]) do not correspond to the assembly unit implies that the path of a trimer of CA dimers is not helical. In this case, our guidelines will fail to offer an unambiguous helical specification.

The guidelines have two limitations in fulfilling the aim that every helical structure would have a unique $[n_1, n_2, \varphi, \delta]$ helical symmetry. The first arises when the primary protofilament is ambiguous, as discussed above, and the second is encountered when there is a continuous helical density along a protofilament in the cryo-EM structure rather than a clear boundary between asymmetric units. Under such circumstances, for a determined $[n_1, n_2]$ symmetry the helical structure can be described by an infinite number of $[\varphi, \delta]$ pairs, which are always related by a constant. The helical structure of the tubular $A\beta_{1-42}$ amyloid with a hollow core (Miller *et al.*, 2010; Zhang *et al.*, 2009) is an example of this limitation. The cryo-EM structure gives a [2, 0] (or [2, 1]) helical symmetry and the two helical parameters $[\varphi, \delta] = [−3.75c, 4.8c]$, where c is a constant.

4.2. Relevance to experimental diffraction patterns

The diffraction patterns of helical structures consist of a series of layer-lines. Assuming that the layer-lines do not overlap, each layer-line is the result of diffraction by a set of *n*-start helices. The position of the peak with the maximum diffraction intensity in each layer-line can be indexed to correspond to a node in the helical net. The relationship between the 1-D helical system $[n_1, n_2]$ symmetry and the diffraction pattern is detailed in Figs. 4(*a*) and 4(*b*). The two figures illustrate the different assignments of n_{10} and n_{01}, which give different helical symmetries, [3, 1] and [3, −2], for the same structure that has a simple helical symmetry of $t = 4$ and $u = 13$. The assignment of n_{10} with the first peak close to the equator (*n*) of the diffraction pattern is consistent with the guideline for selecting the n_1-start helices with a twist angle closest to zero. The assignment of n_{01} to the position of the diffraction which is close to the origin is also likely to constitute a main protofilament of the given helical structure.

4.3. The minimal number of helices needed for a complete helical structure description

From the rotohelical transformation, we learnt that a single (one-start) helix description is not always sufficient to generate the entire helical structure from given asymmetric units. The helix may need to be related by a C_n rotational symmetry, which implies that a minimum of *n* helices are required to cover the entire helical assembly. Given an $[n_1, n_2]$ symmetry, there are n_1 or n_2 assigned helices for the entire helical description. To determine the minimal number of helices for complete structural description (or to be correlated with the rotohelical transformation), the $[n_1, n_2]$ symmetry is reduced to an equivalent symmetry with $n_2 = 1$ or 0. In the case of a reduced $[n_1, 0]$ symmetry, the new n_1 is the minimal number of helices.

It is straightforward to deduce the minimal number of helices for a given $[n_1, n_2]$ helical symmetry. If the numbers n_1 and n_2 do not have a common factor, the symmetry can always be reduced to a one-start helix description by using a combination of the swap and the equivalence rules of $n = h\,n_{10} - k\,n_{01}$ as described in §2. For example, [7, 3] can be reduced to [1, 3] with $h = 0$, $k = 2$. On the other hand, the largest common factor between n_1 and n_2 is the minimal number of helices for a complete helical structure description. For example, [8, 4] can be reduced to [4, 0] with the number 4, the largest common factor of 4 and 8.

4.4. A new description of helical symmetry

Despite the fact that so many helical structures have been determined, a universal formulation for representing helical symmetry is still lacking. The absence of agreement in the community has been attributed to three main reasons. The first apparent reason is a consequence of the fact that helical
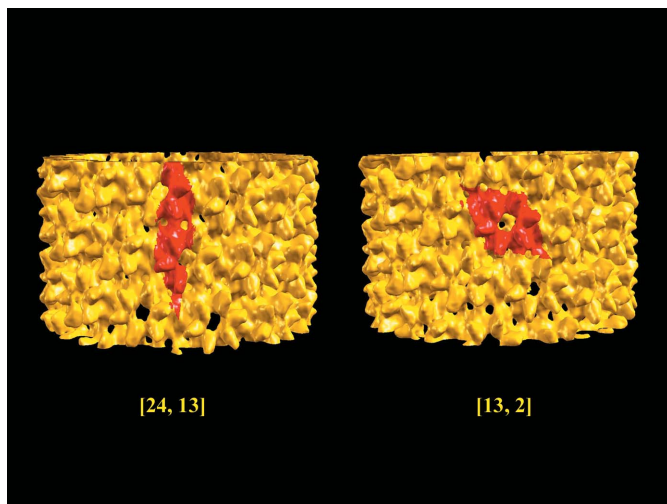

Figure 9
Different surface lattices with different helical symmetry assignment. The span of the surface lattice of the tubular structure of HIV-1 capsid protein (EMD-5136) is highlighted in red for two helical symmetries of [24, 13] and [13, 2].

symmetries have been formulated in distinct ways to fulfill a particular requirement or convenience in different structure-determination methods. The diversified helical representations can be classified into two commonly adopted helical schemes named the 1-D and 2-D helical systems. In this study, the two helical schemes were unified into a single helical specification by two constants $[n_1, n_2]$ and we have shown that the two systems are interchangeable and complementary to each other. Because of the simplicity of using one less parameter and the lack of involvement of the axial radius, we suggest using the augmented 1-D helical system with four parameters $[n_1, n_2, \text{twist, rise}]$ for representing a helical structure.

The second hurdle for defining a helical description is that a helical structure can be pictured in many ways, *i.e.* in many $[n_1, n_2]$ combinations as two (n_1-start and n_2-start) sets of helices. However, in principle, the generalized guidelines for describing a helical symmetry are expected to give a unique $[n_1, n_2]$ specification that reflects the characteristics of the structure, although in a limited number of cases a unique specification is impossible.

The fact that no standard helical symmetry has been accepted so far can be attributed to the last obstacle: a complete coverage of helical description includes the capability of handling helical discontinuity (a seam). However, building an entire helical construct with a seam from given asymmetric units requires no additional modification in our formulation of helical transformation. Instead, a helical structure with a seam is simply reflected in the value of n_2. By definition, the helical discontinuity indicates that n_2 is no longer an integer but a rational number.

4.5. Presentation of a structure with a helical discontinuity

An implicit requirement of the 2-D helical system (§2.1) is that in a seamless helical arrangement $[n_1, n_2]$ must be specified by integer numbers. By treating two consecutive asymmetric subunits in the primary protofilament as a new single asymmetric subunit, the new augmented 1-D helical symmetry becomes $[n_1, n_2/2, 2\varphi, 2\delta]$, which is equivalent to the original helical symmetry $[n_1, n_2, \varphi, \delta]$ except that the asymmetric units are doubled in size. When the tubulin subunit is treated not as a dimer of $\alpha\beta$ subunits but as a single subunit by ignoring the small difference between the α and β subunits (Sui & Downing, 2010), we do not encounter the microtubule seam problem. However, when treating the $\alpha\beta$ dimer as an asymmetric subunit in the new 1-D helical symmetry, helical structures with an odd number for the n_2 symmetry (in single subunit representation) create a seam with a new rational n_2.

The microtubule EM structure (Cochran *et al.*, 2009; EMD-5038) presents such a helical discontinuity when treating the dimer of $\alpha\beta$ subunits as the asymmetric unit. The augmented 1-D helical symmetry in four parameters [13, 3/2, 0.0, 80.0] is sufficient to generate the entire helical structure with a seam, based on the helical transformation matrix summarized in Fig. 2. Owing to the helical discontinuity, the repetitive asymmetric unit is no longer an identical unit. Instead, a complete round of n_1 subunits (13 dimers in the microtubule case) now constitutes the identical unit in the helical structure with a seam. Therefore, the subunit coordination index $[m_1, m_2]$ can no longer have an index with $m_1 \geq n_1$ when applying the helical transformation to generate the repetitive subunits for a helical structure with a seam.

A seam in a helical structure can be classified visually with respect to its helical axis into a strictly vertical seam or a seam that wraps around the helix. The microtubule case above is an example of a vertical seam. Under the restriction that only a rational n_2 and integer $n_1 > |n_2|$ are allowed in the augmented 1-D helical representation, the corresponding helical structure always produces a vertical seam and the handedness of the seam is determined by the sign of n_2, with positive indicating a left-handed seam and negative a right-handed seam. In contrast, a seam described by a rational $n_1 > |n_2|$ and integer n_2 should correspond to the type of seam that wraps around the helix.

4.6. Application to polymorphic structural assemblies

Both the 1-D and 2-D helical systems are designed to create helical assemblies from asymmetric subunits with specified helical parameters. The conformational heterogeneity of molecular assemblies is known to set limits on solving cryo-EM structures at high resolution. Polymorphism is particularly problematic in the determination of structures with helical symmetry since even a slight deviation in the interactions between two asymmetric subunits will create distinct structures with different symmetries. We have seen such an example in Table 2 for the microtubule structure. The question is can all such polymorphic structures be generated based on a single helical structure which is given in an atomic model or an EM map? The answer is yes, because the interactions between the asymmetric subunits are preserved in the definition of the 2-D helical system. Thus, to create distinct polymorphic structures with almost the same subunit–subunit interactions we only need to change the specific $[n_1, n_2]$ helical symmetry.

5. Conclusions

In this paper, we give two helical formulations (augmented 1-D and 2-D) to describe a helical structure. Unlike the rotohelical transformation (1-D formulation) with a helical plus an additional rotational operation, a new augmented 1-D formulation with two consecutive helical operations enables unification with the widely adopted 2-D formulation, giving a common helical symmetry descriptor with two integers $[n_1, n_2]$. The new formulation requires only four parameters $[n_1, n_2, \text{twist, rise}]$ for the augmented 1-D helical system and five parameters $[n_1, n_2, a, b, \gamma]$ for a 2-D helical system to generate the entire structural assembly from given asymmetric units. We propose using the augmented 1-D helical system with four parameters to describe a helical structure owing to its simplicity and independence from the helical radius compared with the 2-D helical system.

research papers

In terms of a helical net representation, a helical structure with an $[n_1, n_2]$ symmetry indicates that its organization is specified by two sets of helices (n_1-start and n_2-start). Because many different $[n_1, n_2]$ combinations exist for the same structure, we suggest general guidelines for selecting a unique $[n_1, n_2]$ symmetry which reflects the structural characteristics of a given helical structure. We provide a computational graphics tool for this purpose which can be used for any helical structure determined by X-ray fibre diffraction or EM imaging.

While there are multiple ways to construct equations that generate the same helical structure, an $[n_1, n_2, \text{twist}, \text{rise}]$ description provides the following advantages: firstly, it provides full helical coverage, including a helical discontinuity (seam) which is indicated by a rational n_2; secondly, it reflects the structural characteristics of the assembly (formation mechanism) directly by four helical parameters; that is, similar structures give similar parameters; thirdly, the unnecessary error in reproducing the entire helical structures, such as editing wrong transformation matrices in the PDB or in the deposited EM parameters in the EMDB, will be prevented; and lastly, the new helical symmetry is expected to be useful for maintaining a pre-determined helical symmetry in structural refinement as well as for the generation of all 'meaningful' polymorphic structural assemblies from a given helical atomic model or EM density map.

References

Byeon, I.-J., Meng, X., Jung, J., Zhao, G., Yang, R., Ahn, J., Shi, J., Concel, J., Aiken, C., Zhang, P. & Gronenborn, A. M. (2009). *Cell*, **139**, 780–790.

Cochran, J. C., Sindelar, C. V., Mulko, N. K., Collins, K. A., Kong, S. E., Hawley, R. S. & Kull, F. J. (2009). *Cell*, **136**, 110–122.

Damnjanović, M., Nikolić, B. & Milošević, I. (2007). *Phys. Rev. B*, 033403.

DeRosier, D., Stokes, D. L. & Darst, S. A. (1999). *J. Mol. Biol.* **289**, 159–165.

Diaz, R., Rice, W. J. & Stokes, D. L. (2010). *Methods Enzymol.* **482**, 131–165.

Egelman, E. H. (2007). *J. Struct. Biol.* **157**, 83–94.

Egelman, E. H. (2010). *Methods Enzymol.* **482**, 167–183.

Evans, P. R. (2001). *Acta Cryst.* D**57**, 1355–1359.

Fujii, T., Kato, T. & Namba, K. (2009). *Structure*, **17**, 1485–1493.

Grubisic, I., Shokhirev, M. N., Orzechowski, M., Miyashita, O. & Tama, F. (2010). *J. Struct. Biol.* **169**, 95–105.

Heymann, J. B., Chagoyen, M. & Belnap, D. M. (2005). *J. Struct. Biol.* **151**, 196–207.

Kikkawa, M. (2004). *J. Mol. Biol.* **343**, 943–955.

Klug, A., Crick, F. H. C. & Wyckoff, H. W. (1958). *Acta Cryst.* **11**, 199–213.

Lawson, C. L. *et al.* (2011). *Nucleic Acids Res.* **39**, D456–D464.

Lawson, C. L., Dutta, S., Westbrook, J. D., Henrick, K. & Berman, H. M. (2008). *Acta Cryst.* D**64**, 874–882.

Marvin, D. A. (1990). *Int. J. Biol. Macromol.* **12**, 125–138.

Miller, Y., Ma, B., Tsai, C.-J. & Nussinov, R. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 14128–14133.

Pederson, D. M., Welsh, L. C., Marvin, D. A., Sampson, M., Perham, R. N., Yu, M. & Slater, M. R. (2001). *J. Mol. Biol.* **309**, 401–421.

Sachse, C., Chen, J. Z., Coureux, P. D., Stroupe, M. E., Fändrich, M. & Grigorieff, N. (2007). *J. Mol. Biol.* **371**, 812–835.

Sachse, C., Fändrich, M. & Grigorieff, N. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 7462–7466.

Schmidt, M., Sachse, C., Richter, W., Xu, C., Fändrich, M. & Grigorieff, N. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 19813–19818.

Stewart, M. (1988). *J. Electron Microsc. Tech.* **9**, 325–358.

Sui, H. & Downing, K. H. (2010). *Structure*, **18**, 1022–1031.

Toyoshima, C. (2000). *Ultramicroscopy*, **84**, 1–14.

Toyoshima, C. & Unwin, N. (1990). *J. Cell Biol.* **111**, 2623–2635.

Tsai, C.-J., Zheng, J. & Nussinov, R. (2006). *PLoS Comput. Biol.* **2**, 311–319.

Wang, Y. A., Yu, X., Overman, S., Tsuboi, M., Thomas, G. J. Jr & Egelman, E. H. (2006). *J. Mol. Biol.* **361**, 209–215.

Watson, J. D. & Crick, F. H. (1953). *Nature (London)*, **171**, 737–738.

Zhang, R., Hu, X., Khant, H., Ludtke, S. J., Chiu, W., Schmid, M. F., Frieden, C. & Lee, J.-M. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 4653–4658.