Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Strongly correlated structure of axial-symmetric proteins. II. Pentagonal, heptagonal, octagonal, nonagonal and ondecagonal symmetries

A. Janner

Institute for Theoretical Physics, Radboud University, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Correspondence e-mail: alo@sci.kun.nl

Received 23 March 2004
Accepted 9 December 2004

The investigation of the geometry of the molecular envelope and channel in the proteins discussed in part I [Janner (2005*a*), *Acta Cryst.* D**61**, 247–255] is extended to axial-symmetric proteins with orders of rotation N = 5, 7, 8, 9 and 11, non-crystallographic in dimension 3. In these cases also, the vertices of the molecular form which encapsulate the C^{α} backbone have integral coordinates (indices) in a symmetry-adapted basis which generates a polygonal lattice. As in the crystallographic case of part I, a characteristic rational axial ratio squared is observed that reduces to one the number of free lattice parameters and enhances the symmetry. Furthermore, there is a crystallographic scaling relation between the envelope and the channel which depends on the order of the axial symmetry and is expressible in terms of star polygons. Possible biological implications are suggested within a more general context.

1. Introduction

The remarkable geometric properties of the molecular forms discussed in part I (Janner, 2005*a*) also occur in the case of axial-symmetric proteins having a non-crystallographic point group, as in quasi-crystals. As is known from the theory of aperiodic crystals, crystallographic characterization is possible in a higher-dimensional space, the superspace. A corresponding three-dimensional description is based on so-called polygonal lattices (Yamamoto, 1996). The extension needed of the basic crystallographic concepts is commented upon and the notation introduced.

Symmetry-adapted basis. The rational indices of a point are the integral coordinates expressed in a symmetry-adapted basis. For a protein with a rotational symmetry of order N, this basis is defined by the n vectors

$$a_k = a[\cos(k2\pi/N), \sin(k2\pi/N), 0], \quad k = 1, 2, ..., \varphi(N) = n - 1$$
$$a_n = c(0, 0, 1), \qquad (1)$$

with components (x, y, z) indicated with respect to an orthonormal coordinate system and $\varphi(N)$ the number-theoretic Euler φ-function. This basis is indicated by the two real parameters (a, c).

Polygonal lattices. The integral linear combinations of the (a, c) basis vectors form a polygonal lattice. This term has been introduced by Yamamoto for the crystallography of quasi-crystals (Yamamoto, 1996). For N = 2, 3, 4 and 6 the three-dimensional monoclinic or orthorhombic, trigonal, tetragonal and hexagonal lattices are obtained, respectively. For the other N values the points of a lattice-plane perpendicular to the rotational z axis are dense, but only a discrete set of points has a structural meaning.

Rational indices. The indices are the coordinates of a lattice point with respect to a basis of the lattice, which in the present case is a symmetry-adapted basis. A lattice point $P = \sum_{i=1}^{n} n_i a_i$ has indices $n_1, n_2 \ldots n_n$. Rational indices indicate that the n_i are rational numbers. Only those expressible in terms of small integers are structurally relevant.

Higher-dimensional lattices. The same set of indices n_i also corresponds to points P_s of an n-dimensional lattice (in the so-called superspace), with basis vectors a_{si} having the a_i of the symmetry-adapted basis as a three-dimensional component in the physical space. Accordingly, the two bases coincide for $N = 2, 3, 4, 6$. For the other values $N = 5, 7, 8, 9, \ldots$, the polygonal lattice is the one-to-one projection of a higher-dimensional lattice in the superspace.

Star polygons. Consider a regular polygon of order N for $N > 4$. By joining the vertices with their Mth nearest ones, respectively, one obtains a star polygon (also called polygram) with Schläfli symbol $\{N/M\}$. Examples can be found in Coxeter (1961). The intersection points define new regular polygons scaled (or scale-rotated) with respect to the original one. Conversely, a larger star polygon can be obtained by prolongation of the M nearest edges of the starting polygon. The corresponding scaling factors depend in a characteristic way from the order N of the rotation symmetry. For example, for $N = 5$ one obtains a pentagram, the star polygon $\{5/2\}$, a magic symbol from time immemorial (one finds it in Goethe's Faust) used by the Pythagoreans as a distinctive mark. The scaling factor of the two pentagons (in reverse orientation) of a pentagram is $-\tau^2$, with $\tau = (1 + 5^{1/2})/2$, the golden mean. The pentagram plays a structural role in a variety of pentamers and decamers.

Planar and linear scaling. A planar scaling S_μ with scaling factor μ is defined in two and three dimensions by

$$S_\mu(x, y) = (\mu x, \mu y), \quad S_\mu(x, y, z) = (\mu x, \mu y, z), \qquad (2)$$

and linear scalings X_μ and Y_μ in the x and y directions, respectively, by

$$X_\mu(x, y) = (\mu x, y), \quad Y_\mu(x, y) = (x, \mu y), \qquad (3)$$

so that their product is a planar scaling: $S_\mu = X_\mu Y_\mu$. These scalings are crystallographic if they transform lattice points into lattice points (as already explained in part I).

The polygrammal scalings considered are crystallographic. Expressed with respect to a polygonal lattice basis a [or (a, c) in three dimensions] they are represented by invertible integral matrices. This is not always the case for the linear scalings with the same scaling factor.

These concepts are illustrated in Fig. 1 in the two-dimensional case for $N = 5$ and in Fig. 2 for $N = 8$.

For $N = 5$, one has $\varphi(5) = 4$. As basis of the pentagonal lattice Λ one can choose the four vectors a_1, a_2, a_3, a_4 pointing from the centre to the four vertices of a regular pentagon. The lattice points, labelled accordingly by four integral indices $n_1n_2n_3n_4 = n_1a_1 + n_2a_2 + n_3a_3 + n_4a_4$, are dense in the plane, so that only positions with small indices have a structural meaning. Of particular importance are lattice points which are in a pentagrammal relation, or in other words, at the vertices

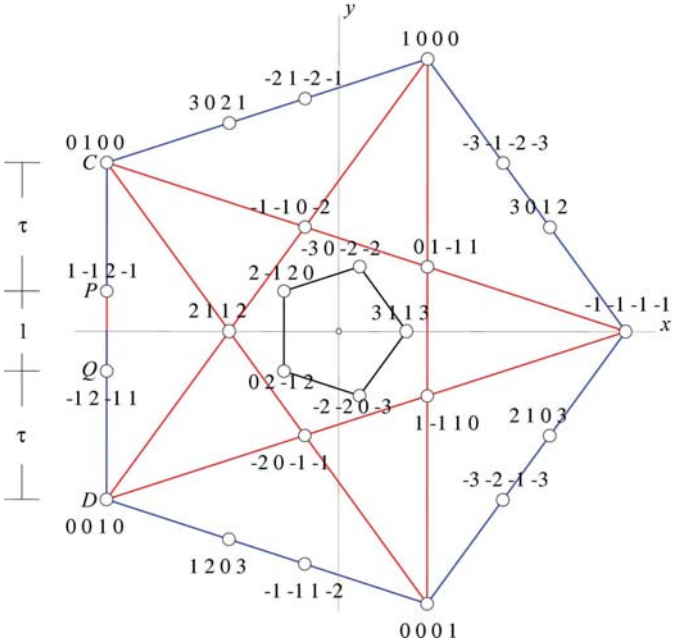

Figure 1
Indexed pentagram with vertices at points of a pentagonal lattice. A smaller pentagon follows from the external one by a planar scaling with scaling factor $-1/\tau^2 = -0.3819\ldots$, where $\tau = (1 + 5^{1/2})/2$ is the golden mean. The smallest pentagon added to the pentagram is scaled by the factor $1/\tau^3 = 0.236\ldots$ from the external one. The integral linear scaling Y_{1/τ^3} transforms C into P and D into Q and is revealed by the golden mean partition of the pentagonal edge $CP{:}PQ{:}QD = \tau{:}1{:}\tau$, as in the star pentagon. The matrices of all these scaling transformations are given in the text.

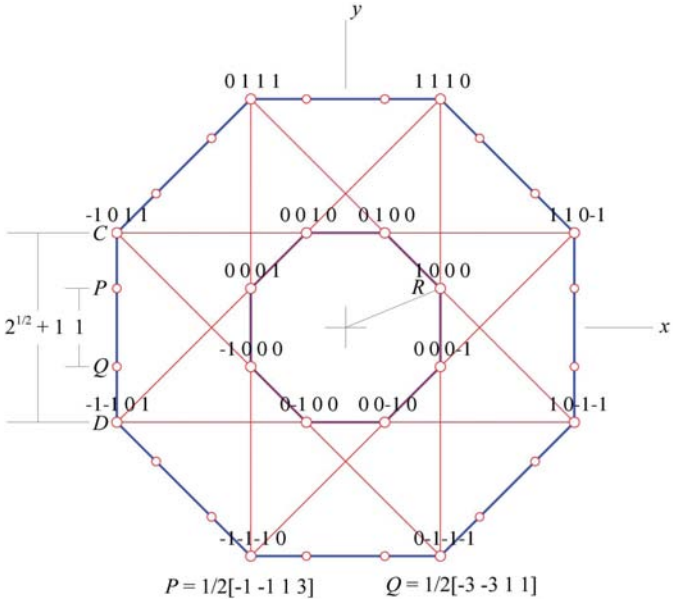

Figure 2
Polygrammal scaling and indexed star polygon for $N = 8$. The basis vectors of the octagonal lattice are defined from the central octagon (with radius R), which is related by the star octagon $\{8/3\}$ to the external octagon [with radius $(2^{1/2} + 1)R$] by the integral scaling transformation $S_{2^{1/2}+1}$ (given in the text). This planar scaling is the product of the two linear scalings $X_{2^{1/2}+1}$ and $Y_{2^{1/2}+1}$ in the x and y directions, respectively, which relate vertices with the edge points indicated. In particular, $Y_{2^{1/2}+1}P = C$ and $Y_{2^{1/2}+1}Q = D$ yield the ratio $PQ{:}CD = 2^{1/2} - 1$ observed in Pf3 (Fig. 6).

of a {5/2} star pentagon. By starting from the pentagonal vertices with indices 1000, 0100, 0010, 0001 and $\bar{1}\bar{1}\bar{1}\bar{1}$, by joining next-nearest vertices one obtains a pentagram with an internal pentagon having vertices at the lattice points $\bar{2}0\bar{1}\bar{1}$, $1\bar{1}10$, $01\bar{1}\bar{1}$, $\bar{1}\bar{1}0\bar{2}$ and 2112 (see Fig. 1). These vertices follow from those of the external pentagon by the integral matrix transformation

$$S_{-1/\tau^2}(a) = \begin{pmatrix} \bar{2} & 1 & 0 & \bar{1} \\ 0 & \bar{1} & 1 & \bar{1} \\ \bar{1} & 1 & \bar{1} & 0 \\ \bar{1} & 0 & 1 & \bar{2} \end{pmatrix}, \qquad (4)$$

with scaling factor $\mu = -1/\tau^2 = -0.3820\ldots$, where $\tau = (1 + 5^{1/2})/2 = 1.618\ldots$ is the golden mean. This is equivalent to a scaling by $1/\tau^2$ combined with a π rotation. As $\det S_{-1/\tau^2} = 1$, the inverse transformation $S_{-\tau^2}$ is also integral. Both scalings leave the pentagonal lattice Λ invariant. The same lattice is also invariant with respect to linear scaling transformations X_{1/τ^3} and Y_{1/τ^3}. In the a basis oriented as in (1), they are represented by the matrices:

$$X_{1/\tau^3}(a) = \begin{pmatrix} \bar{2} & 1 & 1 & \bar{3} \\ \bar{1} & 1 & 0 & \bar{1} \\ \bar{1} & 0 & 1 & \bar{1} \\ \bar{3} & 1 & 1 & \bar{2} \end{pmatrix}, \quad Y_{1/\tau^3}(a) = \begin{pmatrix} 0 & 1 & \bar{1} & 1 \\ 1 & \bar{1} & 2 & \bar{1} \\ \bar{1} & 2 & \bar{1} & 1 \\ 1 & \bar{1} & 1 & 0 \end{pmatrix} \qquad (5),$$

with $1/\tau^3 = 0.2361\ldots$. Their product defines the planar scaling S_{1/τ^3},

$$S_{1/\tau^3}(a) = \begin{pmatrix} \bar{3} & 2 & 0 & \bar{2} \\ 0 & \bar{1} & 2 & \bar{2} \\ \bar{2} & 2 & \bar{1} & 0 \\ \bar{2} & 0 & 2 & \bar{3} \end{pmatrix}, \qquad (6)$$

leading to the smallest pentagon of Fig. 1. In particular, applying Y_{1/τ^3} to the pentagrammal point $C = 0100$ of Fig. 1 one finds the lattice point $P = 1\bar{1}2\bar{1}$ and in a similar way $Y_{1/\tau^3}D = Q = \bar{1}2\bar{1}1$. By fivefold rotations the additional indexed positions shown are correspondingly obtained. These positions have the remarkable property of subdividing the pentagonal edge in the golden-mean ratio

$$CP : PQ : QD = \tau : 1 : \tau. \qquad (7)$$

In a molecular structure it is difficult to see whether linear scalings are relevant or not. On the contrary, it is fairly easy to recognize pentagonal boundaries subdivided according to the golden-mean ratios. It occurs in particular for the (projected) positions of the residue Glu15 on the pentagonal boundaries of the decameric cyclophilin (see Fig. 3 and the next section). This property does not follow from the 52 point-group symmetry of the decamer and suggests the larger point group K generated by the fivefold rotation R_5, the twofold perpendicular rotation R_2, the linear scaling Y_{τ^3} and planar scaling S_{1/τ^3}, which relates the external boundary to the internal one. Therefore, the effect of this scaling transformation has been included in Fig. 1.

For $N = 8$, as $\varphi(8) = 4$, the octagonal lattice is generated by four radial vectors of a regular octagon. The corresponding vertices have the indices 1000, 0100, 0010 and 0001, respectively. The prolongation of the octagonal edges defines an {8/3} star octagon and accordingly a larger regular octagon with vertices having the integral indices indicated. All these points are points of the octagonal lattice projection on the plane of a four-dimensional lattice. The scaling transformation from the central to the external octagon is represented by the matrix

$$S_{2^{1/2}+1}(a) = \begin{pmatrix} 1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ -1 & 0 & 1 & 1 \end{pmatrix}, \quad \det S_{2^{1/2}+1} = 1, \qquad (8)$$

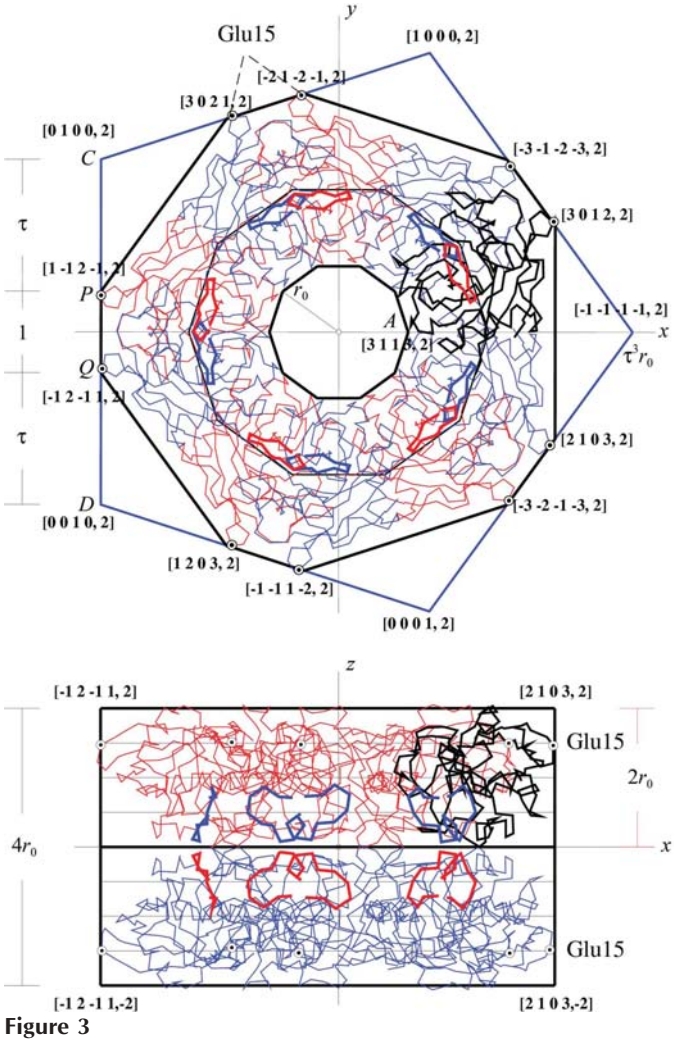

Figure 3
The decameric complex cyclophilin A–cyclosporin A (CyPA–CsA), with point group 52, is formed by two adjacent pentamers which are dyadically related. A pair of corresponding vertices of the pentamers (labelled P and Q, respectively, at the projected positions of Glu15) subdivides the pentagonal edge of the envelope according to the pentagrammal ratios τ:1:τ, where τ is the golden mean. The radius of the pentagonal envelope is τ^3 times the radius r_0 of the decagonal channel. The height of the decamer is $4r_0$ and that of the pentamer $2r_0$. The molecular form lattice of CyPA–CsA is pentagonal isometric c:a = 1, with in this case $a = r_0$. The axial indices of Glu15 are $\pm 3/2$, whereas those of the external form are ± 2.

with scaling factor $\lambda\{8/3\} = 2^{1/2} + 1$. As $\det S_{2^{1/2}+1} = 1$ the inverse transformation is also integral and has scaling factor $2^{1/2} - 1$.

The planar scalings $S_{2^{1/2}\pm 1}$ can be factorized into the linear scalings $X_{2^{1/2}\pm 1}$ and $Y_{2^{1/2}\pm 1}$ along the x axis and the y axis, respectively. In the orientation adopted in Fig. 2, with the x axis along the basis vectors $a_1 - a_4$ [which is not the orientation of (1)] these linear scalings have half-integral entries,

$$X_{2^{1/2}-1} = X_{2^{1/2}-1}(a_1 - a_4) = \frac{1}{2}\begin{pmatrix} 1 & 1 & \bar{1} & 1 \\ 1 & \bar{1} & 3 & \bar{1} \\ \bar{1} & 3 & \bar{1} & 1 \\ 1 & \bar{1} & 1 & 1 \end{pmatrix},$$

$$Y_{2^{1/2}-1} = Y_{2^{1/2}-1}(a_1 - a_4) = \frac{1}{2}\begin{pmatrix} \bar{1} & 1 & 1 & 3 \\ 1 & 1 & \bar{1} & 1 \\ 1 & \bar{1} & 1 & 1 \\ \bar{3} & 1 & 1 & \bar{1} \end{pmatrix}. \quad (9)$$

In particular (see Fig. 2),

$$Y_{2^{1/2}-1}C = P = \left(\frac{\bar{1}\bar{1}}{2}\frac{1}{2}\frac{3}{2}\right), \quad Y_{2^{1/2}-1}D = Q = \left(\frac{\bar{3}\bar{1}}{2}\frac{1}{2}\frac{1}{2}\right), \quad (10)$$

so that the octagonal edge is partitioned according to

$$PQ : CD = 1 : (2^{1/2} + 1) = 2^{1/2} - 1$$

as indicated in Fig. 2.

2. Proteins with pentagonal symmetry

In the pentagonal case, the symmetry-adapted basis (a, c) is spanned by five vectors: four in the xy plane and the fifth along the fivefold rotational axis taken in the z direction. The vertices of a molecular form with integral coordinates in the basis (a, c) are labelled by five indices $[n_1n_2n_3n_4,n_5]$ where, in order to enhance clarity, the axial index n_5 is separated by a comma from the four planar indices. These vertices are at points of a pentagonal lattice generated by integral linear combination from the vectors of the symmetry-adapted basis, as already stated. Only points with small (planar) indices have a structural meaning, because large enough indices allow an approximation with arbitrary precision of any point of the xy plane.

Taking these peculiarities into account, the following proteins with a pentagonal axial symmetry can be treated in the same way as the examples of part I: cyclophilin A–cyclosporin A complex (CyA–CsA; PDB code 2rma; Ke *et al.*, 1994), D-aminopeptidase (DppA; PDB code 1hi9; Remaut *et al.*, 2001) and pentameric human serum amyloid P component (SAP; PDB code 1sac; Emsley *et al.*, 1994).

The pentagonal lattice of the cyclophilin is isometric (c:a = 1). That of the DppA appears at first to have an axial ratio c:$a = \tau^4$ for the decamer and c:$a = \tau^3$ for the pentamer, where τ denotes the golden mean $(1 + 5^{1/2})/2 = 1.681\ldots$. In SAP the axial ratio is τ^2, as has already been shown in Janner (2002*a*), so that only the first two cases will be discussed here. In fact all these lattices are isometric, as can be seen by converting to an equivalent symmetry-adapted basis.

2.1. Pentagonal isometric form lattice c:a = 1

2.1.1. Cyclophilin A–cyclosporin A complex (CyPA–CsA). Disregarding scaling, the decameric CyPA–CsA complex has 52 point-group symmetry and is formed by two pentamers of CyPA–CsA molecules arranged face-to-face (Ke *et al.*, 1994; PDB code 2rma). The central channel is a decagonal prism with a height which is four times the radius for the decamer ($H = 4r_0$) and twice the radius for the pentamer ($h = 2r_0$). From this observation, one immediately deduces that both symmetry-adapted bases and the corresponding pentagonal form lattices are isometric: $r_0 = c = a$. The envelope is a pentagonal prism with a same height as the channel, but the pentamer differs from the decamer both in size and orientation (Fig. 3).

The envelope of the decameric cyclophilin has a radius scaled by a factor τ^3 with respect to the channel ($r_e = \tau^3 r_0$) and has an orientation fixed by the 52 symmetry. In Fig. 3, the x axis has been chosen along a dyadic axis. This same dyad transforms one pentamer into the other one, without fixing either the orientation or the size of the single pentamer. It is the scale-rotational symmetry of the pentagonal lattice which restricts these degrees of freedom in such a way that the vertices of the envelope have small integral indices. Indeed, the dyadically related vertices of the pentamers, like the two vertices labelled P and Q in Fig. 3, at the projected positions of Glu15 (see also Fig. 1), subdivide the edge of the decameric pentagon according to the pentagrammal ratio CP:PQ:QD = τ:1:τ. As explained in the previous section, this pentagrammal ratio arises from an integral linear-scaling transformation Y_{τ^3} relating (as in Figs. 1 and 3) the y coordinate of P to that of C by a factor τ^3 (and correspondingly Q with the subsequent vertex D). Remarkably enough, the same partition also occurs, in axial projection, for the position of the residue Asp215 in the case of the protein DppA (see Fig. 4) and for the P-atomic positions in the double helix of B-DNA (Janner, 2001). Finally, the envelope of the cyclosporin is a decagonal prism with radius and height having half the corresponding lengths in the pentagonal envelope of the cyclophilin.

Summarizing, in the 52 asymmetric unit of the decamer, the channel has two decagonal vertices with indices A = [3113,2] and $[\bar{3}\bar{1}\bar{1}\bar{3},2]$, the pentagonal envelope one vertex C = [0100,2] and the refined one two vertices at $P = Y_{1/\tau^3}C = [1\bar{1}2\bar{1},2]$ and $Q = [\bar{1}2\bar{1}1,2]$. The envelope of one of the two pentamers (the upper one), up to fivefold rotations, has two vertices P and P' with the same planar indices $P = [1\bar{1}2\bar{1},2]$ and $P' = [1\bar{1}2\bar{1},0]$. The pentameric channel is only half the height of that of the decamer (see Fig. 3).

2.1.2. D-Aminopeptidase (DppA). The decamer DppA has 52 point-group symmetry and consists of two face-to-face interpenetrating pentamers which are dyadically related (Remaut *et al.*, 2001; PDB code 1hi9).

The channel is a pentagonal prism with radius r_0. In the decamer the height is a factor τ^4 larger than the radius

($H = \tau^4 r_0$), whereas in the pentamer this factor is τ^3 ($h = \tau^3 r_0$). As above, τ is the golden mean. These morphological properties lead to isometric pentagonal form lattices by taking $c = a = \tau^4 r_0$ and $c = a = \tau^3 r_0$, respectively. It is, however, more convenient to convert to an equivalent symmetry-adapted basis by choosing $a = r_0$, as in the case of cyclophilin, while keeping $c = \tau^4 r_0$ for the decamer and $c = \tau^3 r_0$ for the pentamer. This changes the integral value of the planar indices. The axial indices of the decameric molecular form are the half-integers $\pm 1/2$, mutually related by the dyadic transformation, and after a shift of the origin become 1 and 0, respectively. The situation is similar for the pentamer with respect to its own lattice parameters. Scaling the radius r_0 of the channel by a factor $2\tau^2$, the large pentagonal envelope of the decamer is obtained:

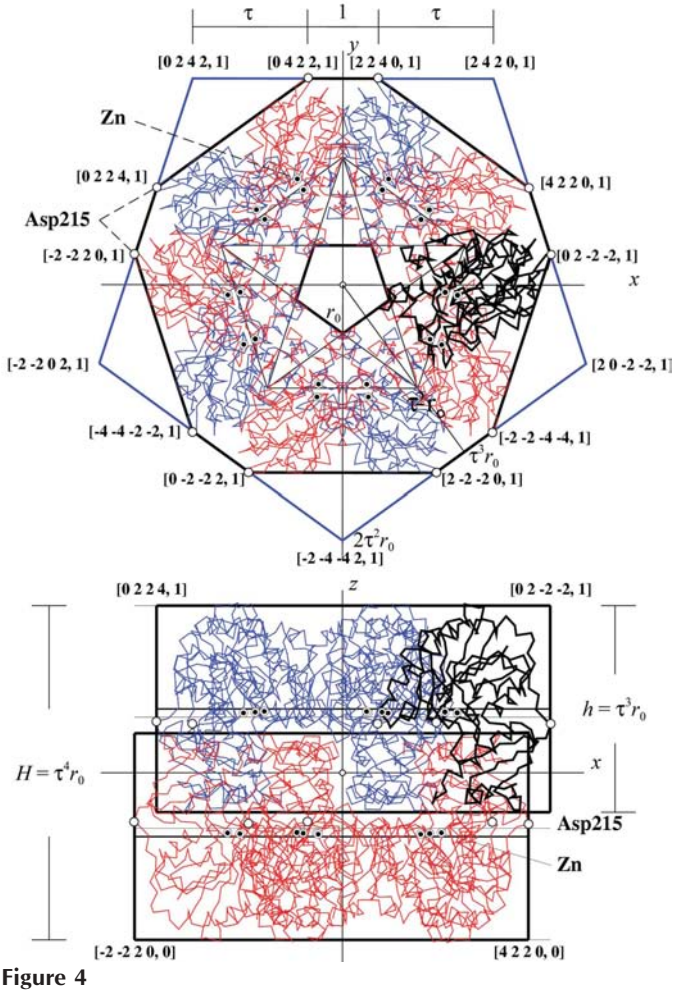

Figure 4
The decameric D-aminopeptidase (DppA), with point group 52, consists of two dyadically related interpenetrating pentamers. The height of the decamer is τ^4 times the radius r_0 of the pentagonal channel; that of the pentamer is $\tau^4 r_0$, where $\tau = (1 + 5^{1/2})/2$ is the golden mean. The radius of the pentagonal envelope is $2\tau^2 r_0$. Pairs of vertices of the two pentamers at the projected positions of Asp215 (empty circles) subdivide the pentagonal edge of the envelope according to τ:1:τ and have integral planar indices and fractional axial indices (1/3, 2/3) with respect to the pentagonal isometric lattice $c = a = \tau^4 r_0$. The indices of the molecular form are given with respect to the equivalent basis $(c, a) = (\tau^4 r_0, r_0)$, turned by 90° with respect to (1). The Zn^{2+} ions (dotted circles) are in a pentagrammal relation with the channel and approximatively at half-height of the pentamer.

$r_e = 2\tau^2 r_0$. As in the case of cyclophilin, the pentameric vertices of the envelope, given by the projected positions of Asp215, subdivide the pentagonal edge of the decamer according to the pentagrammal proportions τ:1:τ. The axial distance $\tau^3 r_0$ defines a truncated pentagon enclosing the decamer (Fig. 4). In the symmetry-adapted bases chosen and in the 52 asymmetric unit of the decamer, the indices of the channel are $[\bar{1}\bar{1}\bar{1}\bar{1},1]$ and those of the larger pentagonal envelope $[\bar{2}\bar{4}\bar{4}2,1] = 2S_{\tau^2}[\bar{1}\bar{1}\bar{1}\bar{1},1]$. For the truncated pentagon (outlined in Fig. 4), taking into account the pentagrammal ratio τ:1:τ and the corresponding linear scalings from the pentagonal vertices of the envelope, one finds two vertices at $[2240,1] = X_{1/\tau^3}[2420,1]$ and $[0422,1] = X_{1/\tau^3}[0242,1]$, as indicated in Fig. 4. The other vertices follow by using the 52 point-group symmetry. The vertices of the pentameric molecular form have correspondingly the same indices, but now referred to the pentagonal lattice parameters $c = \tau^3 r_0$ and $a = r_0$. Note that the matrices expressing Y_{1/τ^3} and X_{1/τ^3} given in (8) have to be interchanged because of the 90°-rotated orientation adopted in Fig. 4 for graphical reasons.

Finally, the Zn^{2+} ions are symmetrically arranged at mid-height of each pentamer on both sides of an intermediate boundary in a pentagrammal τ^2-scaling relation with the central channel (see Fig. 4).

3. Proteins with heptagonal symmetry

The planar heptagrammal scaling symmetry has already been demonstrated for three proteins: α-hemolysin, GroEL and GroES (Janner, 2002*b*, 2003*a*,*b*,*c*). The peculiar isometric axial ratio c:$a = 1$ of GroEL in the symmetric free state has been the motivation for extending the investigation to proteins with other rotational symmetries, as presented in this paper. Moreover, these features are retained even in the asymmetric GroEL–GroES complex. The free *Pyrococcus abyssi* SM core (PA-Sm1) considered here is a symmetric double heptameric ring with a central cavity (PDB code 1h64; Thore *et al.*, 2003), as in GroEL, which keeps its symmetric state even when complexed with RNA. The discussion of the Sm-core complex is postponed to part III (Janner 2005*b*), which is devoted to complexes with DNA/RNA. In the heptagonal case, as $\varphi(7) = 6$, seven indices are needed for labelling the lattice points $[n_1 n_2 \ldots n_6, n_7]$: six planar indices and one axial index.

3.1. Heptagonal isometric form lattice c:$a = 1$

3.1.1. *P. abyssi* Sm Core (PA-Sm1). The double heptameric ring of PA-Sm1 has point-group symmetry 72 (Thore *et al.*, 2003; PDB code 1h64). The overall height H of the dimer is twice the radius r_e of the envelope and the distance d between the heptamers is one quarter of r_e. If one takes the dyadic axis into account, one immediately sees that the natural unit of length u of the molecular form is half the distance d. Accordingly, each heptamer has height $h = 7u$ and radius $r_e = 8u$. The symmetry-adapted basis and the heptagonal form lattice are isometric with $c = a = u$. The central cavity is related to the envelope by an heptagrammal planar scaling transfor-

mation S_μ with scaling factor $\mu = 0.3079\ldots$, expressible in terms of the three integers $t_0 = 2$, $t_1 = -1$, $t_2 = 1$, according to (in two dimensions)

$$S_{\mu(t_0,t_1,t_2)} = \begin{pmatrix} t_0 - t_1 & t_1 - t_2 & t_2 & 0 & -t_2 & -t_1 + t_2 \\ 0 & t_0 - t_2 & t_1 & t_2 & -t_2 & -t_1 \\ -t_1 + t_2 & t_1 - t_2 & t_0 & t_1 & 0 & -t_1 \\ -t_1 & 0 & t_1 & t_0 & t_1 - t_2 & -t_1 + t_2 \\ -t_1 & -t_2 & t_2 & t_1 & t_0 - t_2 & 0 \\ -t_1 + t_2 & -t_2 & 0 & t_2 & t_1 - t_2 & t_0 - t_1 \end{pmatrix}, \tag{11}$$

where the scaling factor is given by

$$\mu(t_0, t_1, t_2) = t_0 + 2\,t_1 \cos(2\pi/7) + 2\,t_2 \cos(4\pi/7). \tag{12}$$

The condition $\det S_{\mu(t_0,t_1,t_2)} = 1$ ensures that the inverse scaling matrix also has integral entries (Janner, 2003*c*).

The structural relevance of the chosen unit distance u is confirmed by looking at the C^α positions of the residues on the surface of the heptagonal rings: Asp14, Gly58, Asp44 on the envelope and Asp65 on the central channel (Fig. 5). While Asp44 and Gly58 occur at $h/2$, the half-height of the heptamer, Asp65 is at $(h/2) - u$ and Asp14 at $(h/2) - 2u$, so that all have half-integer axial indices with respect to the isometric basis $(c, a) = (u, u)$.

This is also the case for the corresponding planar indices, which follow from linear scaling from vertices of the larger and smaller regular heptagon, respectively. So, for example, in the orientation adopted in Fig. 5 according to (1), by linear scaling $Y_{\mu(t_0,t_1,t_2)}$ of the y coordinate by a factor $\mu(t_0, t_1, t_2)$, in the same parameterization as (11), is given by (Janner, 2003*c*),

$$Y_{\mu(t_0,t_1,t_2)} = \frac{1}{2}\begin{pmatrix} 1 + t_0 - t_2 & t_1 & t_2 & -t_2 & -t_1 & 1 - t_0 + t_2 \\ t_1 & 1 + t_0 & t_1 - t_2 & -t_1 + t_2 & 1 - t_0 & -t_1 \\ t_2 & t_1 - t_2 & 1 + t_0 - t_1 & 1 - t_0 + t_1 & -t_1 + t_2 & -t_2 \\ -t_2 & -t_1 + t_2 & 1 - t_0 + t_1 & 1 + t_0 - t_1 & t_1 - t_2 & t_2 \\ -t_1 & 1 - t_0 & -t_1 + t_2 & t_1 - t_2 & 1 + t_0 & t_1 \\ 1 - t_0 + t_2 & -t_1 & -t_2 & t_2 & t_1 & 1 + t_0 - t_2 \end{pmatrix}. \tag{13}$$

Starting from $A_3 = 001000$ one obtains the planar indices $\frac{1}{2}(t_2,\ t_1 - t_2,\ 1 + t_0 - t_1,\ 1 - t_0 + t_1,\ -t_1 + t_2,\ -t_2)$. The ideal (projected) position Asp14 indicated is obtained with $\mu(1, 0, 1) = 0.5549\ldots$, that of Gly58 with $\mu(1, -1, -1) = 0.1980\ldots$ and that of Asp44 with $\mu(1, 0, 2) = 0.1099\ldots$. In the same way from the corresponding vertex of the central cavity with $\mu(1, 0, 1)$ one finds Asp65. The full set of indices for these representatives of point-group equivalent positions is given by

$$\begin{aligned} \text{Asp14} &= \tfrac{1}{2}[1\bar{1}201\bar{1}, 5], \quad & \text{Gly58} &= \tfrac{1}{2}[\bar{1}03\bar{1}01, 9], \\ \text{Asp44} &= \tfrac{1}{2}[2\bar{2}202\bar{2}, 9], \quad & \text{Asp65} &= \tfrac{1}{2}[45\bar{7}5\bar{5}\bar{2}, 7]. \end{aligned} \tag{14}$$

All these indices become integers if one chooses the alternative isometric lattice basis $(a, c) = (u/2, u/2)$.

4. Proteins with octagonal symmetry

The eightfold case occurs for proteins investigated by NMR spectroscopy at the Laboratory of Biophysical Chemistry of the University of Nijmegen. These proteins, encoded by the bacteriophages Pf3 and M13, bind as homodimers to single-stranded DNAs (ssDNA) and form a superhelix with 822 point-group symmetry (Folmer, 1997). A full helical turn is confined within an octagonal prism with a central hole, similar to the channels of the examples discussed. One can take such a prism as a unit cell of a molecular form which, potentially at least, is infinitely long. The interest of including the new situation in the present analysis is evident, in particular because it is tightly connected to the helical structure of ssDNA. The example chosen here is the helical left-handed configuration of homodimeric single-stranded DNA-binding protein encoded by the filamentous *Pseudomonas* phage Pf3 (Folmer *et al.*, 1995; Folmer, 1997).

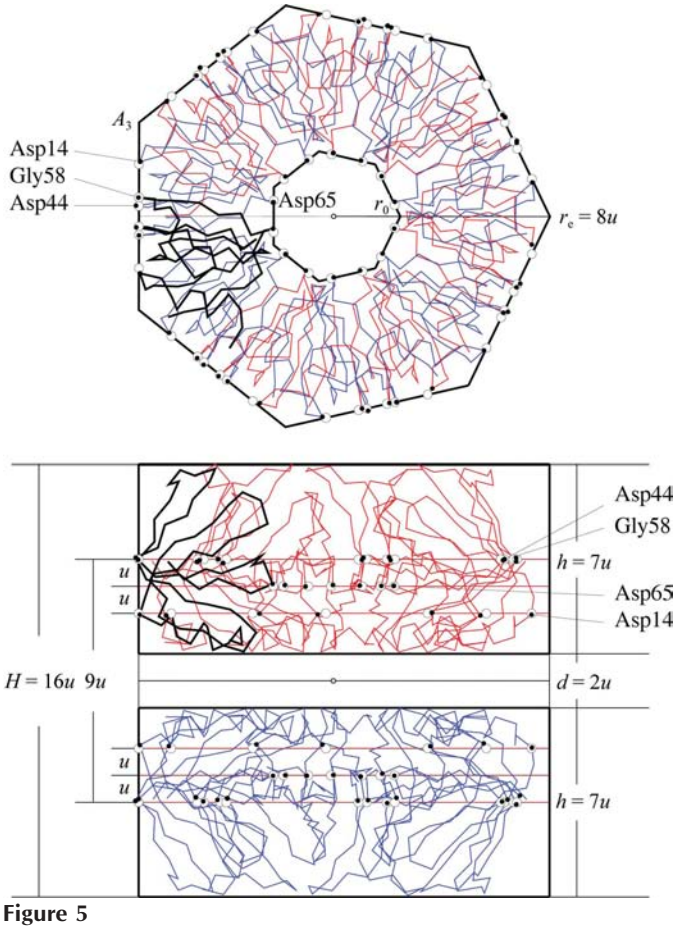

Figure 5
The architecture of the dyadically related double heptameric ring of the Sm core of *P. abyssi* is expressible in terms of the single unit-length parameter u given by the distance of each heptamer from the dyad. The radius r_e of the envelope is then $8u$. That of the channel r_0 follows from the envelope by a planar scaling S_μ with scaling factor $\mu = 0.4450\ldots$ and integral entries in the isometric lattice basis $(a, c) = (u, u)$. The height of the dimer is twice r_e and the height of each heptamer $7u$. In the same basis, all the residues at the lateral boundaries of the double ring (Asp14, Gly58, Asp44 and Asp65) have C^α atoms with half-integer indices. Empty and filled circles indicate the ideal positions and the observed positions, respectively.

4.1. Octagonal isometric form lattice c:a = 1

4.1.1. Superhelix of the Pf3-encoded ssDNA-binding protein (Pf3 ssDBP). The structural data of the major coat superhelix of Pf3 ssDBP, which have not been published, have been made available by R. Folmer, whereas those of the dimer have been deposited as PDB code 1pfs (Folmer *et al.*, 1995).

Looking along the helical axis, a two-step octagrammal scaling relation can be recognized. Starting from the envelope with radius r_e by repeating twice the construction of an {8/3} star polygon, the central channel with radius r_0 is obtained, so that $r_e = (2^{1/2} + 1)^2 r_0$ (see Fig. 6). Note that in the octagonal case the two reciprocal scaling factors $2^{1/2} \pm 1$ play a similar role as the golden-mean factors $(5^{1/2} \pm 1)/2$ in the pentagonal case, as known from the atomic positions in octagonal and decagonal quasi-crystals. In the molecular case, similar relations are also found in dyadically related proteins which do not have a generic allowed position, but subdivide (in projection) the edge of the envelope in a characteristic way, as already described in the pentagonal case for cyclophilin A (Fig. 3) and for D-aminopeptidase (Fig. 4). In the octagonal superhelix of Pf3 ssDBP, neighbouring projected C-terminal positions Ala78 appear on the external octagon at a distance PQ scaled by a factor $(2^{1/2} - 1)$ with respect to the length of the octagonal edge l_0, so that $l_0 = (2^{1/2} + 1)PQ$, as already discussed in §1. The same ratio also occurs between the height h of the molecular form and the radius r_e of the envelope: $h = (2^{1/2} + 1)r_e$ (Fig. 6). In the orientation of the x axis along $(a_1 - a_4)$, as in Fig. 6, the P and Q positions indicated follow from C and D, respectively, (as in Fig. 2) by the linear scaling $Y_{2^{1/2}-1}$,

$$Y_{2^{1/2}-1}[\bar{2}023, 1] = \left[\tfrac{\bar{5}\bar{1}}{22}\tfrac{3}{2}\tfrac{5}{2}, 1\right],$$
$$Y_{2^{1/2}-1}[\bar{3}\bar{2}02, 1] = \left[\tfrac{\bar{5}\bar{3}}{22}\tfrac{1}{2}\tfrac{5}{2}, 1\right]. \quad (15)$$

Successive full turns, each consisting of eight dimers bound to two antiparallel DNA strands interpenetrate in a similar way to pentamers in the decamer of DppA (see Fig. 4). In the present case, the penetration depth is $r_e/2^{1/2}$, so that the pitch p of the ssDNA helix is scaled by the same factor with respect to the total height of the 16-mers: $p = h/2^{1/2} = [(1 + 2^{1/2})r_e]/2^{1/2}$, as indicated in Fig. 6.

A first natural choice of a symmetry-adapted basis (a_0, c_0) of the molecular lattice of this form is $a_0 = r_0$ and $c_0 = h_0 = (2^{1/2} + 1)^3 r_0$. With respect to this basis, the indices of the representative vertices of the channel are [1000,0], [1000,1] and by applying to these positions the the planar scaling $S^2_{2^{1/2}+1}$, one obtains for the envelope [320$\bar{2}$,0], [320$\bar{2}$,1], respectively.

Actually, there is an equivalent basis (a, c) for the same form lattice which is isometric. One can indeed also choose $a = c = (2^{1/2} + 1)^3 r_0 = (2^{1/2} + 1)r_e$. This choice only changes the planar indices. So, for example, $1000 \rightarrow \bar{7}50\bar{5}$ for the channel and $320\bar{2} \rightarrow \bar{1}10\bar{1}$ for the envelope.

Concluding, the overall geometry of the Pf3 ssDBP superhelix nucleoprotein complex and the underlying octagonal molecular form lattice can both be expressed in terms of one single parameter, for example the radius r_e of the envelope or the pitch p of the helix.

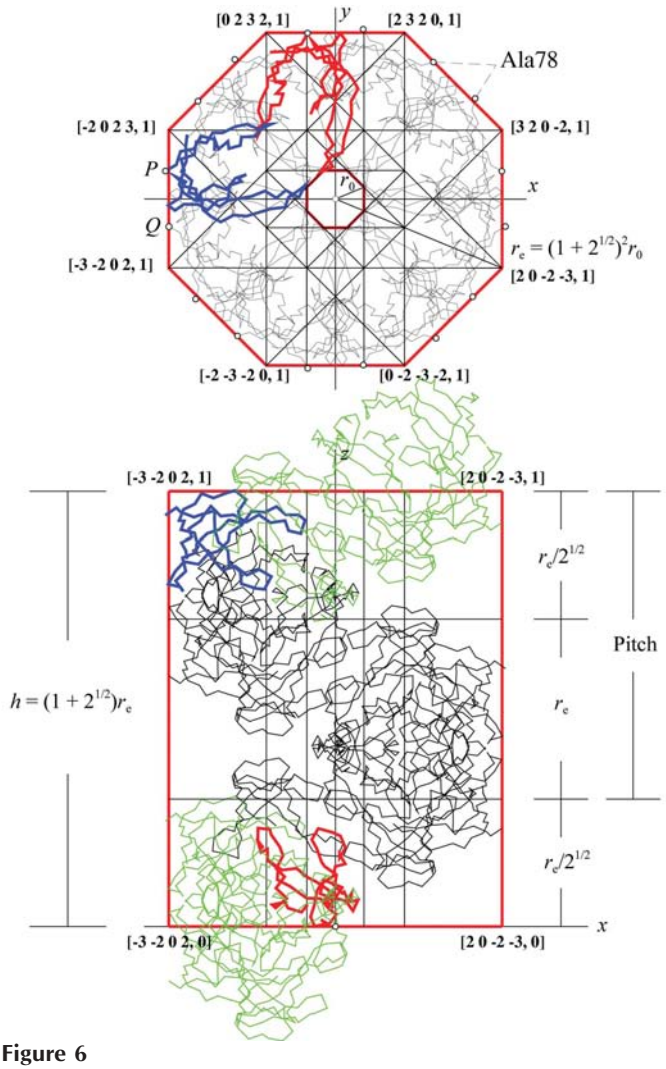

Figure 6
A full turn of the Pf3 ssDBP superhelix consists of eight dimers bound to each of two antiparallel helices of a single-stranded DNA (not shown) and is enclosed in a double octagonal prism with envelope (radius r_e) and channel (radius r_0) related by two {8/3} star octagons: $r_e = (2^{1/2} + 1)^2 r_0$. The height h is given by $h = (2^{1/2} + 1)r_e$. For a better visualization, more dimers have been added in a light colour. The pitch p of the superhelix and the inter penetration depth d of two full turns have the scaling relation $p = (2^{1/2} + 1)r_e/2^{1/2}$ and $d = r_e/2^{1/2}$, so that $h = 2^{1/2}p$. The lowest and highest monomer of this molecular form (enhanced in the figure) belong to different DNA helices. The C-terminal Ala78 residues of two adjacent molecules at P and Q, respectively, appear in axial projection on the octagonal edge l_0 of the envelope at a planar distance $(2^{1/2} - 1)l_0$ and are related by linear scalings to the neighbouring vertices (see Fig. 2).

5. Proteins with nonagonal symmetry

The peripheral LH2 complex from *Rhodopseudomonas acidophila* strain 10050 (PDB code 1kzu; Prince *et al.*, 1997) belongs to a family of light-harvesting proteins, like the R-phycoerythrin discussed in §3.1 of part I (Janner, 2005*a*). Both have a similar arrangement of apoproteins α and β, but a different point-group symmetry. While in R-phycoerythrin they are assembled as a hexamer $(\alpha\beta)_6$, in the present LH2

complex their arrangement is a nonamer $(\alpha\beta)_9$. In both cases, the corresponding polygonal form lattice is isometric.

5.1. Nonagonal isometric form lattice $c{:}a = 1$

5.1.1. LH2 light-harvesting protein (*R. acidophila*). The point group of this complex is 9. Here, only the subsystem of the α, β apoproteins is considered, disregarding the chromophores. The attention is focused on the scaling relations between envelope and channel and on the height-to-width axial ratio.

The external envelope encloses the larger ring formed by the nine β-apoproteins, whereas the smaller ring formed by the other nine α-apoproteins delimits the central channel of this double-ring protein. As shown in Fig. 7, the prismatic channel and envelope are in a nonagrammal scaling relation,

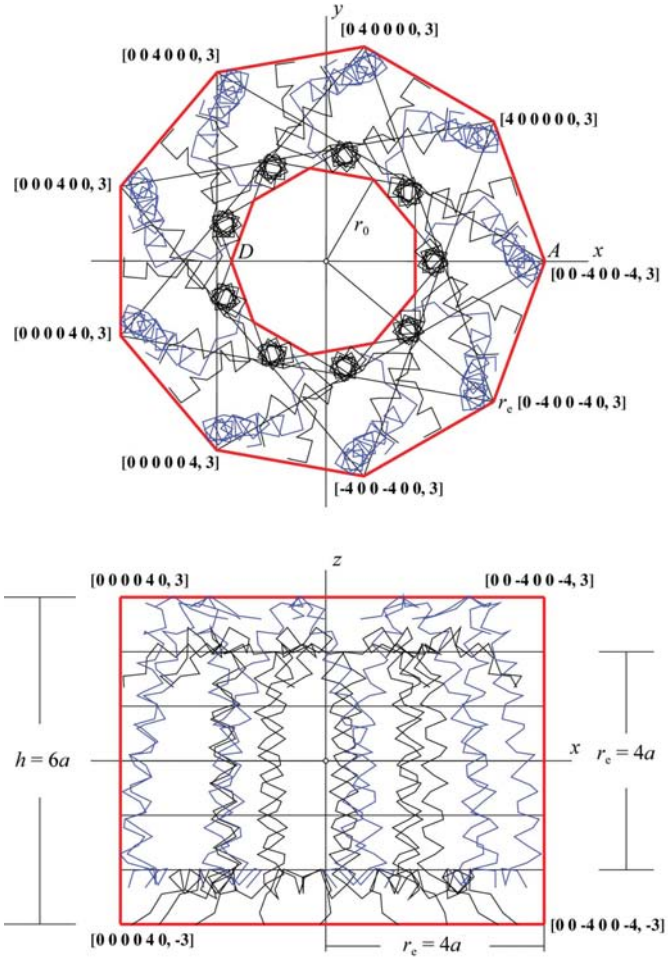

Figure 7
The peripheral light-harvesting protein LH2 from *R. acidophila* is an $(\alpha\beta)_9$ nonamer enclosed in a ring-like regular nonagonal prism. The wall of the central channel is formed by nine α-helices of the α-apoproteins, whereas the outer nine α-helices of the β-apoproteins form the wall of the envelope. The nonagons of the channel (with radius r_0) and of the envelope (with radius r_e) are scaled according to the two star polygons {9/3} and {9/2} indicated: this implies the scaling relation $r_0 = 0.4337\ldots r_e$. The height of the prism is $3/2r_e$, so that the molecular form of LH2 has vertices with integral indices with respect to a nonagonal isometric lattice ($c{:}a = 1$) with $h = 6a$ and $r_e = 4a$. The upper and lower D vertices of the channel have indices [20 $\bar{2}$4 24 0 $\bar{8}$ 16,$\pm$3], respectively.

as expected: a combination of the two star polygons {9/3} and {9/2}. Accordingly, the ratio $r_0{:}r_e$ of the radius r_0 of the channel and the radius r_e of the envelope is the product of the scaling factors $0.532\ldots$ and $0.815\ldots$ observed in {9/3} and {9/2}, respectively. One has $r_0{:}r_e = 0.4337\ldots$. The internal regular nonagon is in a reverse orientation with respect to the external one. Two scaling-related vertices are labelled D and A, respectively. The height of the prismatic molecular form is 3/2 times the radius of the envelope: $h = 3r_e/2$. Therefore, the molecular form lattice, leading to an integral indexing of the vertices of the form, is isometric, $c = a$, with $r_e = 4a$ and $h = 6a$. As the Euler φ-function yields $\varphi(9) = 6$, a point of the monagonal lattice has seven indices, indicated by $[n_1 n_2 \ldots n_6, n_7]$, with n_1–n_6 the planar indices and n_7 the axial index of the point. In the isometric basis (a, a) given above, the vertices of the envelope have indices [400000,$\pm$3], [040000,$\pm$3], ..., [000004,$\pm$3], [$\bar{4}00\bar{4}00$,$\pm$3], ..., [00$\bar{4}$00$\bar{4}$,$\pm$3] (see Fig. 7). Those of the channel are given by D = [20 $\bar{2}$4 24 0 $\bar{8}$ 16,$\pm$3], ..., [$\bar{2}$4 24 $\bar{2}$0 $\bar{8}$ 16 $\bar{2}$0,$\pm$3]. The missing sets follow from those indicated by applying the ninefold rotations.

6. Proteins with ondecagonal symmetry

The ondecagonal case can be analyzed in the same way as the heptagonal one, because their crystallographic scale-rotational symmetries are based on arithmetic properties, which for prime numbers (here 7 and 11) are similar. In §3, for $N = 7$ an RNA-binding protein was taken as an example: the *P. abyssi* Sm core (PA-Sm1). Another RNA-binding protein is now chosen for $N = 11$, the *trp* attenuation protein (TRAP). In the free state, both have a dyadically related double-ring structure with a large central cavity. Similar scale-rotational properties of their architectural elements are therefore expected. This is indeed the case, as shown below for the free TRAP, according to the PDB data file 1wap (Antson *et al.*, 1995). The discussion of the bound state is postponed to part III, which is devoted to protein–DNA/RNA complexes, as previously mentioned.

6.1. Ondecagonal isometric form lattice $c{:}a = 1$

6.1.1. *Trp* RNA-binding attenuation protein (TRAP) of *Bacillus subtilis*. The RNA-binding *trp* attenuation protein in the free state consists of 22 subunits forming a dyadically related double β-ring structure with a large central channel. It has a nearly exact 11-fold rotational axis combined with twofold perpendicular rotations (Antson *et al.*, 1995). The chains labelled *A*–*V* in PDB file 1wap have been oriented with the 11-fold axis in the z direction and one twofold axis along x. The origin has been chosen at their intersection point (see Fig. 8). The 22 L-tryptophan molecules bound to adjacent chains are also indicated by their C^α positions. Owing to the point-group symmetry, corresponding atoms in the 11 subunits are situated (on average) in the same plane. As remarked in the paper referenced above (Antson *et al.*, 1995, p. 699) this is

generally true for oligomers with a prime number of subunits (for example pentamers and heptamers).

In view of what has been said above about the arithmetic properties, this is an interesting confirmation coming from people working in the field.

As in PA-Sm1, there is a natural unit of length u, which is the distance of the two 11-mers from the origin. In the present case, the overall height H of the dimer is $24u$, three halves of the radius $r_e = 16u$ of the envelope, and each 11-mer has a height h of $11u$. The corresponding values for PA-Sm1 are $H = 16u$, $r_e = 8u$, $h = 7u$, with u having the same geometrical meaning. This is clear evidence that the two RNA-binding proteins obey the same building principles. While in PA-Sm1 one residue (Asp65) is responsible for the heptagonal central hole, in TRAP two residues (Tyr67 and Ala66) give rise to an {11/2} star polygonal hole (compare Fig. 5 with Fig. 8). The two

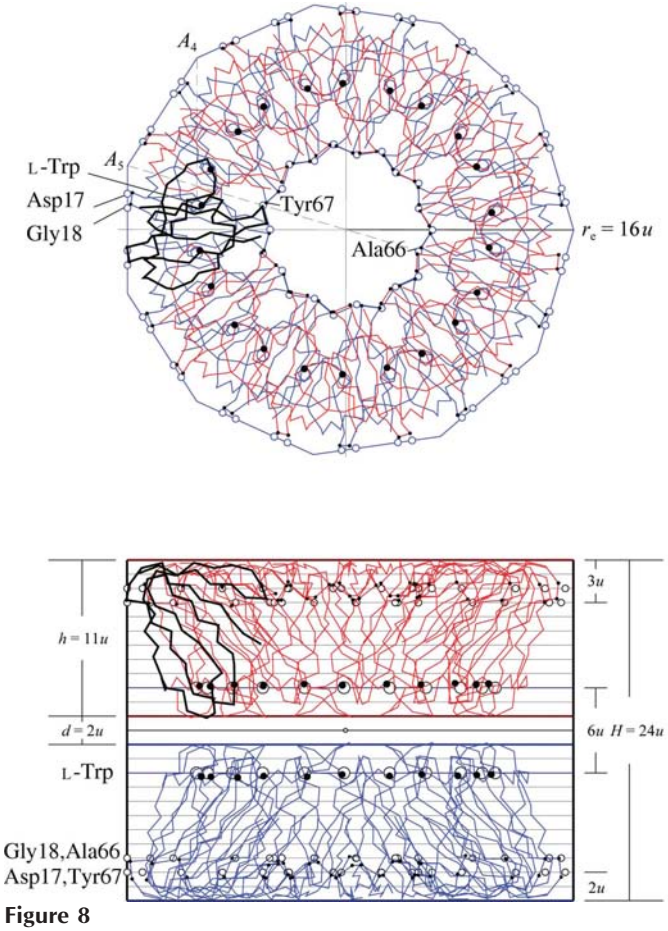

Figure 8
The structure of the double ondecagonal ring of the *trp* RNA-binding attenuation protein is expressible in the unit of length u given by the distance from the origin of the dyadically related 11-mers, as for PA-Sm1 (see Fig. 5). The radius of the envelope r_e is $16u$ and the central channel has as basis an {11/2} star polygon in ondecagrammal scaling relation with the envelope. The height h of the 11-mer is $11u$; the overall height H is $24u$. The ideal positions of the residues at the lateral boundaries (Tyr67, Ala66, Asp17, Gly18) and of the L-tryptophan molecules (L-Trp) are related by planar or by linear scalings and have integral and half-integral indices in the symmetry-adapted basis $(a, c) = (u, u)$, as given in the text. Empty and filled circles indicate the ideal position and the observed position, respectively.

radial distances r_{01} and r_{02} of this star polygon are related to the external 11-fold polygon by an {11/4} scale transformation, followed by one and two {11/2} scalings, respectively,

$$r_{01} = \mu_{\{11/2\}}\mu_{\{11/4\}} r_e = 0.3796\, r_e$$
$$r_{02} = \mu^2_{\{11/2\}}\mu_{\{11/4\}} r_e = 0.3328\, r_e. \qquad (16)$$

Let us look, as in the PA-Sm1 case, at the residues at the lateral boundaries of the molecular form: Asp17 and Gly18 in addition to Ala66 and Tyr67, which have already been considered. It is found that they all have (on average) an integral height when expressed in units of length u: $\pm 9u$ for Gly18 and Ala66 and $\pm 10u$ for Asp17 and Tyr67. The same is true for L-tryptophan at $\pm 3u$ (Fig. 8). The positions of the corresponding C^α atoms are related by ondecagrammal scale-rotational transformations that are either planar or linear. The simplest way is to indicate these scalings in projection from vertices A_k of the external ondecagon. For the residues labelled in Fig. 8 one has

$$\begin{aligned}
\text{Tyr67} &= S_{\mu_1} A_5, \quad \mu_1 = \mu_{\{11/2\}}\mu_{\{11/4\}} = 0.3796\ldots \\
\text{Ala66} &= S_{-\mu_2} A_5, \quad \mu_2 = \mu^2_{\{11/2\}}\mu_{\{11/4\}} = 0.3328\ldots \\
\text{Asp17} &= Y_{\mu_3} A_5, \quad \mu_3 = \mu(3, \bar{2}, 2, \bar{2}, 1) = 0.5561\ldots \\
\text{Gly18} &= Y_{\mu_4} A_5, \quad \mu_4 = \mu(\bar{1}, 1, \bar{2}, 0, \bar{1}) = 0.3305\ldots \\
\text{L-Trp} &= Y_{\mu_5} A_4, \quad \mu_5 = \mu(\bar{1}, 2, \bar{2}, 2, 0) = 0.13409\ldots, \qquad (17)
\end{aligned}$$

where S_μ indicates a planar scaling by a factor μ of the (x, y) coordinates and Y_μ a corresponding linear one of the y coordinate. The parameterization of the μ factor is given by

$$\mu(t_0, t_1, t_2, t_3, t_4) = t_0 + 2t_1 \cos\varphi + 2t_2 \cos 2\varphi + 2t_3 \cos 3\varphi + 2t_4 \cos 4\varphi, \qquad (19)$$

with $\varphi = 2\pi/11$. For the star polygons {11/2} and {11/4} the scaling factors are

$$\mu_{\{11/2\}} = \mu(\bar{1}, 1, \bar{1}, 1, \bar{1}) = 0.8767\ldots$$
$$\mu_{\{11/4\}} = \mu(1, \bar{1}, 1, \bar{1}, 0) = 0.4329\ldots \qquad (19)$$

With respect to a symmetry-adapted basis with $a = 16u$ and $c = u$, one finds integer indices for the ideal positions of Tyr67, Ala66 and L-Trp and half-integer indices for Asp17 and Gly18. In particular, for the planar indices of the positions labelled by residue and by the L-tryptophan in Fig. 8 one has

$$\begin{aligned}
\text{Asp17} &= \tfrac{1}{2}[1\bar{3}4\bar{4}6\bar{4}4\bar{4}3\bar{1}], \\
\text{Gly18} &= \tfrac{1}{2}[\bar{1}1\bar{2}3\bar{1}3\bar{3}2\bar{1}1] \\
\text{Tyr67} &= [\bar{4}4\bar{7}6\bar{8}6\bar{7}4\bar{4}0], \\
\text{L-Trp} &= [1\bar{1}1\bar{1}2\bar{2}2\bar{1}1\bar{1}] \\
\text{Ala66} &= [\bar{2}7\ 25\ \bar{4}8\ 40\ \bar{5}6\ 40\ \bar{4}8\ 25\ \bar{2}7\ 0]. \qquad (20)
\end{aligned}$$

The alternative basis $(a, c) = (u, u)$ shows that the molecular ondecagonal lattice of the TRAP protein is isometric.

7. Possible biological implications

The 14 proteins discussed in parts I and II, together with the few others investigated previously and those presented in part

Table 1
Examples of residues at the boundary of protein enclosing forms having ideal positions with rational indices.

Protein	Form lattice	Residue	Indices	Boundary	Figs.†
CK	Cubic	Gly365	[614]	Vertex	I-3/12
CyPA	Pentagonal isometric	Glu15	$[3021, \frac{3}{2}]$	Lateral edge	II-3
DppA	Pentagonal isometric	Asp15	$[02\bar{2}4, \frac{1}{3}]$	Lateral edge	II-4
PA-Sm1	Heptagonal isometric	Asp14	$\frac{1}{2}[11\bar{2}011,5]$	Lateral edge	II-5
		Asp65	$\frac{1}{2}[45\bar{7}5\bar{5}2,7]$		
		Gly58	$\frac{1}{2}[\bar{1}03\bar{1}01,9]$		
		Asp44	$\frac{1}{2}[\bar{2}\bar{2}202\bar{2},9]$		
Pf3 ssDBP	Octagonal isometric	Ala78	$\frac{1}{2}[\bar{5}135,z]$	Lateral face	II-6
TRAP	Ondecagonal isometric	Asp17	$\frac{1}{2}[\bar{1}34\ldots43\bar{1}]$	Lateral edge	II-8
		Gly18	$\frac{1}{2}[\bar{1}1\bar{2}\ldots2\bar{1}1]$		
		Tyr67	$[44\bar{7}\ldots4\bar{4}0]$		
		Ala66	[high indices]		

† I-3/12 refers to Figs. 3 and 12 of part I, II-3 is Fig. 3 of part II, II-4 is Fig. 4 of part II, II-5 is Fig. 5 of part II, II-6 id Fig. 6 of part II, II-8 id Fig. 8 of part II.

III, are too different to allow a consideration of the biological implications of their morphology at the individual level of each single case. Their number is too small to deduce laws that are valid for axial-symmetric proteins in general. Nevertheless, the properties of the geometrical models presented are so precise and valid in all cases examined that one can safely claim that they are not accidental. The only remaining possibility is to conjecture on how the geometric symmetry principles observed in the proteins presented could allow these proteins to realise their function.

Before doing so, it is important to make clear what it is that is being discussed. Only some of the geometric properties of proteins as they occur in crystals are taken into account, without considering essential physico-chemical aspects such as chemical binding, charge distribution, molecular potentials, thermodynamics and *in vivo* conditions and without an appropriate statistical analysis of the fitting adopted.

The symmetry properties leading, in particular, to the one-parameter characterization, which gives rise to the strong correlations in the structure, are those of an encapsulating form and not necessarily of the atomic positions. Typical in this respect is the polygrammal scaling, which relates the external boundary to the internal channel but does not apply to the individual polypeptide chains. Indeed, the geometric elements considered are not those of the detailed protein structure reported in the PDB, not even in the C^{α}-backbone approximation, because only a few residues are possibly relevant for the molecular form. These few residues are those at the boundary and their positions therefore play a crucial role in the geometry of the forms. The biological properties, however, certainly depend on the other residues as well.

The molecular forms considered are delimited by planes. Owing to the axial symmetry of the protein, these forms are prismatic, with two basal faces and a number of lateral faces. Their fundamental property is that their vertices are at points of a lattice having at least the point-group symmetry of the protein. Accordingly, vertices, edges and faces obey the Law of Rational Indices and are expressible in terms of rational coordinates (integral in particular) with respect to a symmetry-adapted basis.

The structural role of the boundary residues depends on their position with respect to the encaging form: on a vertex, on an edge (lateral or basal) or on a face (also lateral or basal). Indeed, while for a residue on a face shifts in position within the same face do not affect the molecular form, the form is always modified by any shift of a vertex residue, also affecting the integral indices of the ideal position. Other boundary residues may only have basal and/or axial rational indices related to vertex positions by some symmetry element. In saying this, one implicitly accepts the idea that the boundary character of the residues considered is not an accidental one. A number of concrete examples from this work of boundary residues considered to be structurally relevant are listed in Table 1. Their general schematic characterization is shown in Fig. 9.

Under these considerations, to attempt to derive biological implications from the symmetries of the geometric form in a protein at first appears to be unreasonable. Nevertheless, the molecular forms which occur in nature are the result of physico-chemical interactions. Moreover, the encagement of

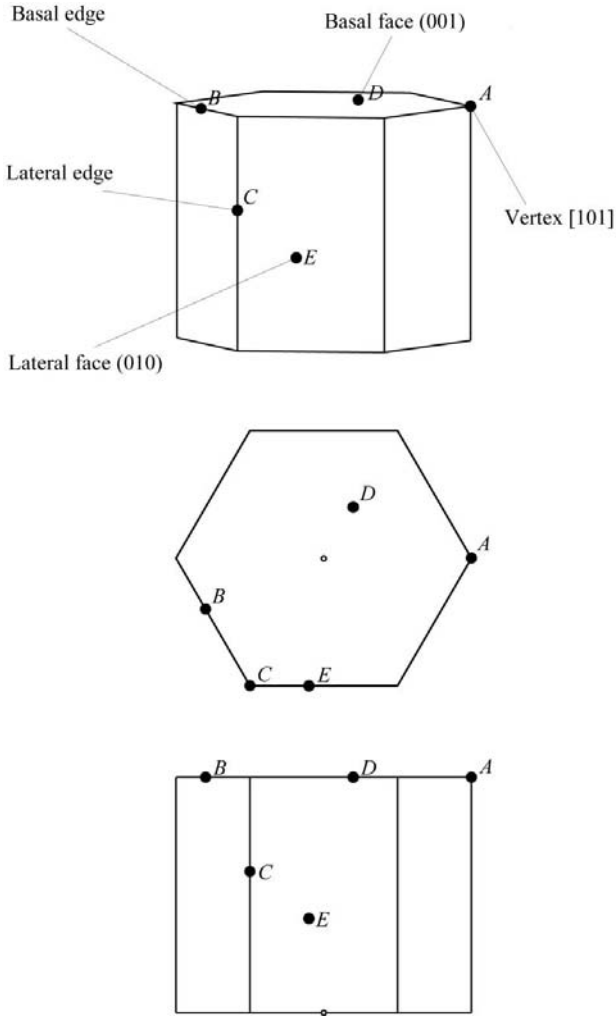

Figure 9
Various types of possible residue positions at the boundary of an indexed hexagonal form are shown in a prespective view along the sixfold axis and perpendicular to it, respectively.

an axial-symmetric protein in the highly symmetrical crystallographic forms derived in this and previous papers reflects similar properties observed in a large number of crystals (compare, for example, Figs. 1 and 2 and Figs. 4 and 5 of part I, respectively). These are facts which cannot be ignored, even if not yet understood and supported by a statistical analysis. Apparently, symmetry in proteins plays a more important role than currently assumed and represents another example of what E. P. Wigner called

> the unreasonable effectiveness of mathematics in the natural sciences

(Wigner, 1967).

In seeking to link the observations discussed above with consideration of protein function, one can consider concepts such as structural stability and rigidity, variability or flexibility, identity and compatibility.

7.1. Structural stability and rigidity

The basic idea is that additional geometric symmetry elements of a given form enhance the molecular stability and its rigidity because they restrict the possible symmetry-conserving deformations. Consider, for example, the axial view of a protein with fivefold symmetry and a pentagonal boundary for both the envelope and the channel. A rigid rotation of the envelope with respect to the central channel which keeps the fivefold symmetry is possible; however, it is not if the envelope and central channel are related by an additional pentagrammal scaling symmetry. The same reasoning applies to the more general deformation illustrated in Fig. 10 and to arbitrary changes in the height-to-width ratio of proteins with an integral form lattices. As shown in part III, there are situations where the rigidity of the form is not a desirable property for the functionality of the protein. The symmetry is then low, even if one still recognizes an underlying badly broken high symmetry. In these cases, properties other than the geometric properties are expected to ensure the required stability.

7.2. Variability or flexibility

A different environment may require a change in the structure while keeping the same symmetry, if this symmetry is relevant for the given function of the protein. Perhaps the best illustration of this idea lies outside biology, in the growth forms of snowflakes. The morphology of many snow crystals is characterized by a macroscopic hexagonal form lattice, as explained in part I and shown in the example of Fig. 1 in part I. Yet, owing to changes in the environmental growing conditions, each snowflake is different. This variability is possible because in a given form only a finite selection of lattice points is involved. In the biological case, to find a similar variability requires a comparison between molecular forms of different proteins. One observes, for example, that the same cubic form lattice allows completely different indexed envelopes: the truncated square pyramid of the *A. niger* acid phosphatase (Fig. 10 of part I), the truncated square prism of the mitochondrial creatine kinase (Fig. 12 of part I) or the much more complex polyhedron of the vanillyl-alcohol oxidase (Fig. 13 of part I). Morphological changes of a protein with the same form lattice symmetry can be found between the pentameric and the decameric conformation of cyclophilin A (Fig. 3) or between the free and the RNA-bound states of the *P. abyssi* Sm core protein discussed in parts II and III (Fig. 5 of part II and Fig. 4 of part III).

7.3. Identity

Essential structural elements cannot be accidental. Also considered not to be accidental is the c/a ratio of the form lattice observed in all the proteins examined. In particular, the $3^{1/2}$ ratio of the hexagonal envelope of the membrane

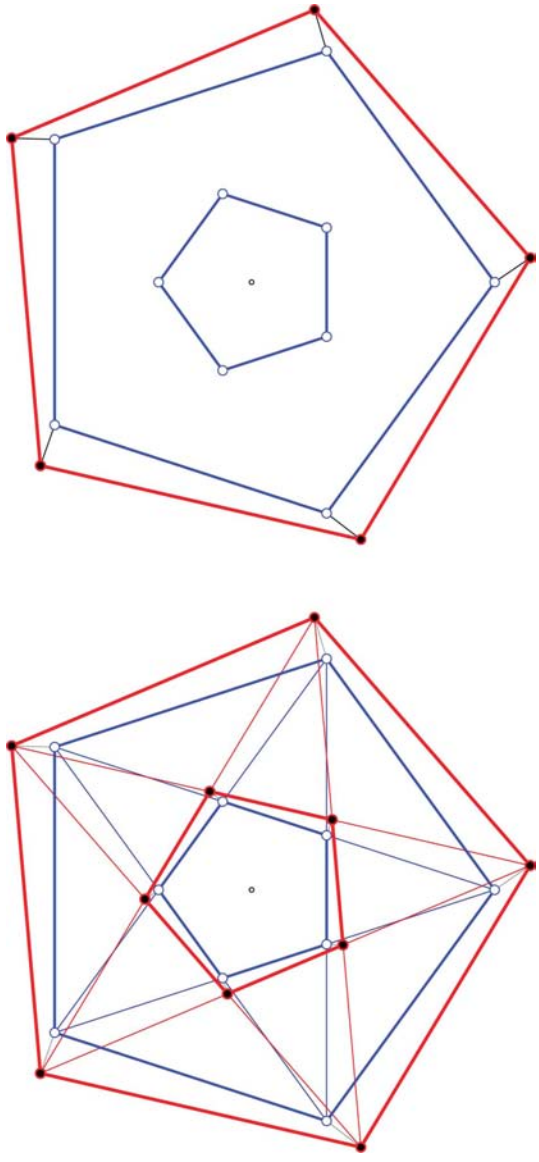

Figure 10
Rotational symmetry allows rigid deformations of the envelope (large pentagons), while leaving the channel (central pentagon) invariant. This is no longer the case if envelope and channel are related by scaling symmetry. For the example shown, one can say that a rigid movement of the envelope is transmitted to the channel by a pentagrammal symmetry.

bacteriorhodopsin ensures compatibility with a cubic environment. In general, however, it is not easy to derive possible biological implications from these ratios. Here, the attention is focused on structural features allowing, at least in principle, a local identification of a given protein. This need is quite general in nature. It is the reason why biometric identifiers such as fingerprints are becoming more and more required in passports. Suitable as biometric protein identifiers are pairs of boundary residues with (small) rational indices related by symmetry. In this case, their metrical properties are anchored to the protein as a whole through the value of the single parameter of the indexed molecular form. Examples of possible biometric identifiers can be derived from Table 1, such as Glu15 of cyclophilin A with indices $[3021,\bar{\frac{2}{3}}]$ and $[\bar{2}1\bar{2}\bar{1},\bar{\frac{2}{3}}]$, respectively (Fig. 3), or similar pairs of Asp215 in DppA (Fig. 4). The combination of the indexed residues found in PA-Sm1 leads to possible biometric identifiers involving more than two residues. A similar situation is found in TRAP.

7.4. Compatibility

The coherent interplay of the various components of a protein, of a protein complex or of a protein with the environment requires compatibility between the different structures involved. Several examples of a symmetry-based compatibility have been shown to occur such as the following.

(i) The hexagonal axial ratio in bacteriorhodopsin, which ensures the compatibility of the envelope with a cubic form and with the cubic lipid phase (Fig. 8 of part I).

(ii) The scaling relation between the envelope of cyclosporin and that of the hosting cyclophilin in the cyclophilin A–cyclosporin A complex (Fig. 3).

(iii) The face-to-face components of quaternary structures which show specific relations in height and/or in orientation, for example between pentamer and decamer or between trimer and hexamer (Figs. 3, 4 and Fig. 9 of part I).

(iv) The axial height of a Pf3 ssDBP and the pitch of the enclosed single-strand DNA (Fig. 6). Additional examples of compatibility are given in part III.

8. Conclusions

The present work is challenging from several points of view.

8.1. From the crystallographic point of view

The geometry of the axial-symmetric proteins analyzed shows unexpected crystallographic properties in terms of molecular form lattices, such as the scaling properties relating the outer to the inner envelope and the integral indices of the vertices. The axial ratio c:a of these lattices is not a generic real number: c^2/a^2 is a rational number. It implies that the corresponding n-dimensional lattice is integral. This means that for a chosen unit of length the metric tensor defined by the scalar product of the basis vectors has integral entries, a property independent of the choice of the lattice basis. If one chooses $a = 1$, the tensor elements are rational numbers.

Integral lattices have a smaller number of free parameters than generic lattices (one instead of two in the axial case) and additional other crystallographic properties: a larger point-group symmetry and scaling relations also involving integral sublattices (Janner, 2004*a*).

The frequency distribution of hexagonal and tetragonal crystals as a function of the c/a ratio shows sharp peaks at values which approximate those of integral lattices (Janner, 2004*b*; de Gelder & Janner, 2004). In fact, this investigation was motivated by the results reported in the present work. It is amazing to see how molecular features can suggest the investigation of new properties of crystals. Apparently, the structural importance of integral lattices has been underestimated until now.

8.2. From the biological point of view

Independently of whether the biological implications suggested are correct or not, one point is clear: symmetry plays a more important role in proteins than commonly assumed. Encapsulation in symmetric geometrical forms with integral indices and crystallographic scaling relations between envelope and channel are a fairly general phenomenon. This leads to strongly correlated structures characterized by an architecture expressible in terms of very few parameters and in the axial case to a single unit of length. It is necessary to understand better the character of the new symmetries, the physico-chemical basis of geometric encapsulation and the possible biological implications.

Thanks are expressed to R. de Gelder, C. Hilbers and B. Souvignier for suggesting improvements, and to A. Fasolino, G. Vriend and R. Hooft for stimulating discussions. R. Folmer is acknowledged for the NMR structural data of the Pf3 nucleoprotein complex.

References

Antson, A. A., Otridge, J., Brzozowski, A. M., Dodson, E. J., Dodson, G. G., Wilson, K. S., Smith, T. M., Yang, M., Kurecki, T. & Gollnick, P. (1995). *Nature (London)*, **374**, 693–700.
Coxeter, H. S. M. (1961). *Introduction to Geometry*. New York: John Wiley.
Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B. & Wood, S. P. (1994). *Nature (London)*, **367**, 338–345.
Folmer, R. (1997). PhD thesis, University of Nijmegen, The Netherlands.
Folmer, R. H. A., Nilges, M., Konings, R. N. H. & Hilbers, C. W. (1995). *EMBO J.* **14**, 4132–4142.
Gelder, R. de & Janner, A. (2004). *Acta Cryst.* A**60** s218.
Janner, A. (2001). *Cryst. Eng.* **4**, 119–129.
Janner, A. (2002*a*). *Z. Kristallogr.* **217**, 408–414.
Janner, A. (2002*b*). *Struct. Chem.*, **13**, 279–289.
Janner, A. (2003*a*). *Proteins Struct. Funct. Genet.* **51**, 126–136.
Janner, A. (2003*b*). *Acta Cryst.* D**59**, 783–794.
Janner, A. (2003*c*). *Acta Cryst.* D**59**, 795–808.
Janner, A. (2004*a*). *Acta Cryst.* A**60**, 611–620.
Janner, A. (2004*b*). *Acta Cryst.* A**60**, 198–200.
Janner, A. (2005*a*). *Acta Cryst.* D**61**, 247–255.

Janner, A. (2005*b*). *Acta Cryst.* D**61**, 269–277.
Ke, H., Mayrose, D., Belshaw, P. J., Alberg, D. G., Schreiber, S. L., Chang, Z. Y., Etzkorn, F. A., Ho, S. & Walsh, C. T. (1994). *Structure*, **2**, 33–44.
Prince, S. M., Papiz, M. Z., Freer, A. A., McDermott, G., Hawthornthwaite-Lawless, A. M., Cogdell, R. J. & Isaacs, N. W. (1997). *J. Mol. Biol.* **268**, 412–423.
Remaut, H., Bompard-Gilles, C., Goffin, C., Frère, J.-M. & Van Beeumen, J. (2001). *Nature Struct. Biol.* **8**, 674–678.
Thore, S., Mayer, C., Sauter, C., Weeks, S. & Suck, D. (2003). *J. Biol. Chem.* **278**, 1239–1247.
Wigner, E. P. (1967). *Symmetries and Reflections*, p. 222. Bloomington/London: Indiana University Press.
Yamamoto, A. (1996). *Acta Cryst.* A**52**, 509–560.