Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 1 February 2008
Accepted 25 April 2008

Comparative architecture of octahedral protein cages. I. Indexed enclosing forms

Aloysio Janner

Theoretical Physics, Radboud University, Toernooiveld, NL-6525 ED Nijmegen, The Netherlands.
Correspondence e-mail: a.janner@science.ru.nl

The architectural elements of four protein cages (bacterio ferritin, human mitochondrial ferritin, sulfur oxygenase reductase and small heat-shock protein) are compared top-to-bottom. The starting points are polyhedra with octahedral symmetry 432 enclosing the cage and delimiting the central cavity, respectively, which have vertices at points of a species-dependent cubic form lattice. The approach is extended from the whole cage to axial-symmetric clusters down to polyhedral forms of single monomers viewed along the fourfold, the threefold and the twofold axes, respectively. The corresponding projected monomeric forms can be approximated by two-dimensional tiles, 13 enantiomorphic pairs in total. The determination of the cubic indices of the monomeric vertices opens the possibility of the way back analysis, bottom-to-top, as discussed in paper II [Janner (2008). *Acta Cryst.* A**64**, 503–512].

1. Introduction

Protein cages like viral capsids, heat-shock proteins and ferritins play an important biological role and are increasingly interesting as nanocontainers in medical and technical applications. Much attention is devoted to their architecture and/or active sites (the last aspect is disregarded here for reasons which should become clear further on). By comparison, one tries to deduce phenomenological laws or, at least, common trends. This is usually done by amino-acid-sequence alignment of primary structures by recognizing the same topological motifs in ternary structures based on elements of the secondary structure and by considering the architecture of the quaternary structure of monomeric chains organized as multimers and as complexes in a biologically active macromolecule.

Here, as in previous similar publications, the approach is reversed. The starting point is a biologically active macromolecule with a given point-group symmetry enclosed in a polyhedral form with vertices at points of a lattice with the same or higher symmetry, the *form lattice*. This morphological characterization fits surprisingly well for the whole and downwards from multimeric clusters to enclosing forms of the monomeric chains. It becomes more approximate at the level of ternary domains and of secondary structural elements (α-helices, β-strands and loops) and it is only meaningfully related to C^{α}'s with positions slightly deviating from the ideal ones at vertices of the enclosing forms indexed according to the form lattice. In a given monomer, the number of these selected C^{α}'s is typically about one tenth of the number of residues in the monomer. Despite the somewhat arbitrary assignment of an ideal position to the selected C^{α}'s, it allows a way back from the indexed residues to a chain segmentation into secondary structural elements up to metrical ternary folds and indexed quaternary structures, assembled in symmetry-adapted orientations to build subunits leading eventually to the morphology of the whole.

The feasibility of this program is tested on four different oligoproteins self-assembled to form a cage with octahedral symmetry. The choice has been dictated by the strength combined with geometrical simplicity of the cubic point group 432 which generates 24 identical copies of the given monomer. Looking at structural data available from the Brookhaven Protein Data Bank (PDB), the more or less arbitrarily chosen representatives are: a bacterio ferritin Bfr from *Desulfovibrio desulfuricans* (PDB entry 1nf4) (Coelho *et al.*, 2001; Macedo *et al.*, 2003), a ferritin, the recombinant human mitochondrial ferritin rMtF (PDB entry 1r03) (Langlois d'Estaintot *et al.*, 2004), a sulfur oxygenase reductase SOR from the thermoacidophilic archeon *Acidianus ambivalens* (PDB entry 2cb2) (Urich *et al.*, 2006) and a small heat-shock protein sHSP from *M. jannaschii* (PDB entry 1shs) (Kim *et al.*, 1998).

The top-to-bottom crystallographic approach is presented in this first paper (I) devoted to indexed enclosing forms and the way back bottom-to-top in a second paper (II) (Janner, 2008), where the interplay between the various structural elements is discussed.

2. The overall architecture

The architectural elements of the protein cages considered are typically the external boundaries, the central hole of the cage cavity, the axial holes of the funnels, the protruding chain segments and the interface of symmetry-adapted subunits viewed along the rotation axes (fourfold, threefold and

twofold) of the octahedral point group 432. To begin with, we consider the complete structure formed by the 24 chains of the cage.

2.1. The form lattice

The protein cages have octahedral symmetry, therefore also their polyhedral enclosing forms and, in particular, the enclosing cube. The vertices of a cube are at points of a cubic lattice with lattice parameter L, half-edge of the cube. Because the cages are nearly spherical, other polyhedra with octahedral symmetry fit better than the cube to the external surface. The vertices of each polyhedron are partitioned by the point group 432 into disjoint orbits of equivalent positions and the polyhedron is generated by applying 432 to the representatives of the orbits. As 432 is crystallographic, each orbit has points on its own cubic lattice, in a similar way as for the cube.

The present approach is based on the property that for a given protein cage all these lattices have their points in common with a single cubic lattice, the *form lattice*, with lattice parameter a. One finds, furthermore, that L and a are related by an integral factor z:

$$L = za, \quad z \in \mathbb{Z}. \tag{1}$$

Only in simple cases does this relation follow from the fact that cube and polyhedron enclose the same protein cage. In general, this relation expresses a phenomenological property.

In order to find the appropriate value z for the four cases listed in the *Introduction* (Bfr, SOR, sHSP and rMtF), one considers first a simple enclosing polyhedron with one or two generating vertices, the *basic polyhedron*. One then checks the validity of the result by moving to refined forms more directly related to the specific architectural elements of the cage. This is done in the projected views along the symmetry axes. Along and perpendicular to the threefold axis, the cubic lattice appears as hexagonal and in a similar way for a twofold axis as orthorhombic, whereas it remains cubic when viewed along the fourfold axes. The mutual relations between these lattices, their bases and the corresponding indices of the lattice points are presented in Appendix *A*. The indices indicated in this article are the cubic ones of the form lattice:

$$[n_1 n_2 n_3] = \sum_{i=1}^{3} n_i a_i, \quad n_i \in \mathbb{Z} \tag{2}$$

with $a_i = a e_i$ and e_1, e_2, e_3 orthonormal vectors.

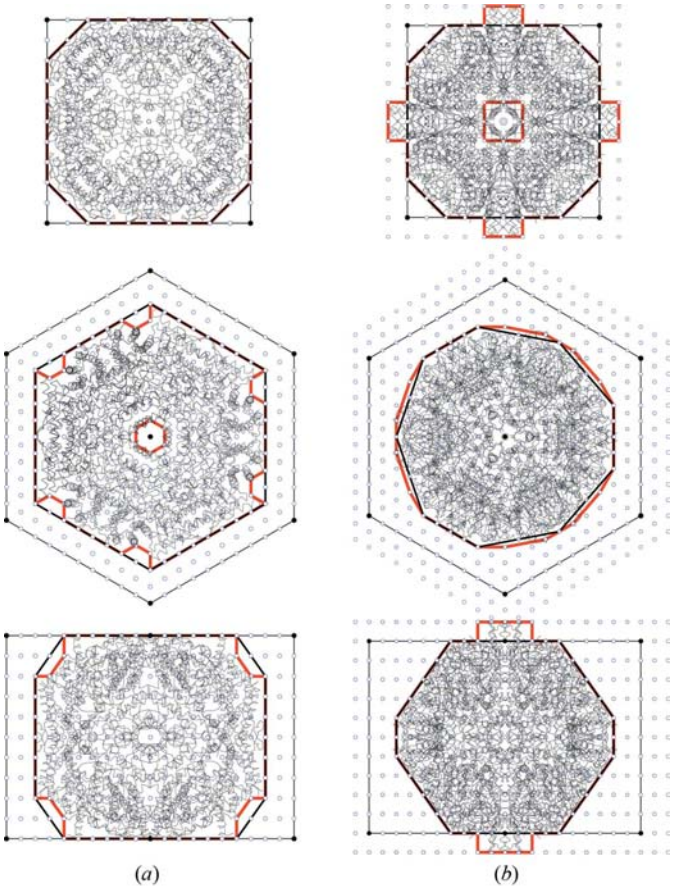

Figure 1
The enclosing cube (with filled vertices), the enclosing basic polyhedron (black thick lines) and the refined one (red thick lines), with vertices at points of the form lattice (empty circles), viewed along the fourfold, the threefold and the twofold axes, respectively, are shown for (*a*) bacterio ferritin Bfr and (*b*) sulfur oxygenase reductase SOR.

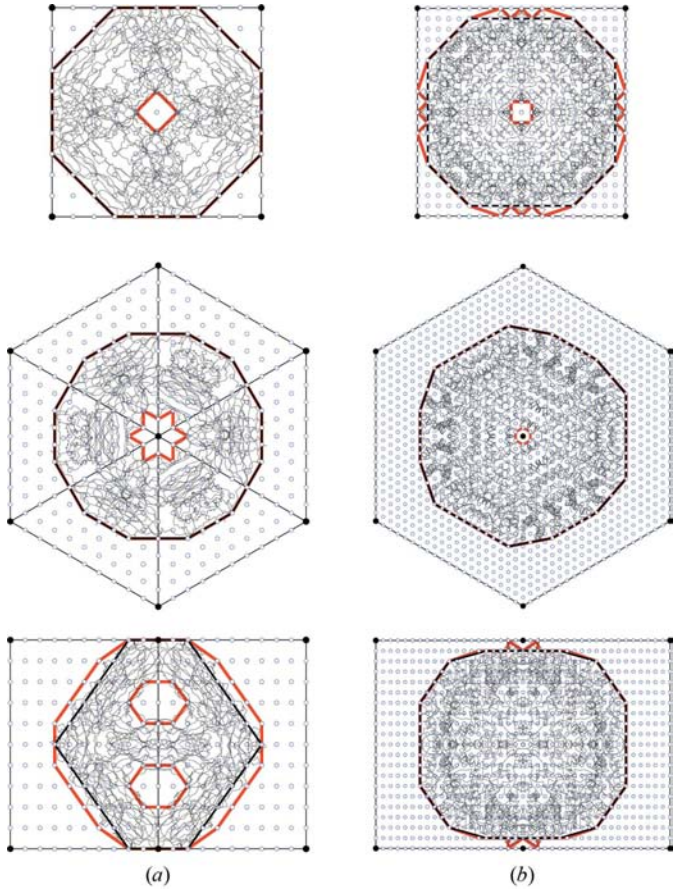

Figure 2
In a way similar to Fig. 1, the enclosing forms (cube, basic and refined polyhedra) are shown for (*a*) small heat-shock protein sHSP and (*b*) human mitochondrial ferritin rMtF.

2.2. The enclosing polyhedra

Bacterio ferritin Bfr (1nf4) is the simplest case. One finds $L = 56$ Å for the enclosing cube, $z = 5$ and $a = 11.2$ Å. The basic enclosing polyhedron is generated by 432 from the vertex [335]. It has 24 vertices (V), 48 edges (E) and 26 faces (F): 6 squares, 8 equilateral triangles and 12 rectangles, so that the Euler relation $V + F = E + 2$ is satisfied. It is indicated in Fig. 1(*a*) in the projected views along [001], [111] and [110] together with the enclosing cube and a refined form, all having vertices at points of the form lattice. In the refined form, one recognizes that the cubic lattice parameter a is equal to the radius of the hexagonal channel along the threefold axis. This non-trivial relation between external and channel surfaces further corroborates the validity of the crystallographic characterization and the value $z = 5$.

The overall architecture of the holoenzyme *sulfur oxygenase reductase* SOR (2cb2) is similar. If we disregard the six chimney-like protrusions, the enclosing cube has $L = 63.5$ Å. One finds again $z = 5$ so that $a = 12.7$ Å. The basic enclosing polyhedron is generated from [513]. It has 24 vertices, 60 edges and 38 faces (6 squares, 8 equilateral and 24 isosceles triangles), so that (VEF) = (246038) (Fig. 1*b*). The lattice parameter a appears to be related to the characteristic morphological elements of SOR, the six chimneys, which have height a and a $2a$ square basis enclosing a circular axial hole along the fourfold axis.

The *small heat-shock protein* sHSP (1shs) has an enclosing cube with $L = 56$ Å and again $z = 5$ implying $a = 11.2$ Å, as for the bacterio ferritin Bfr. The basic polyhedral form which encloses the whole, up to small protruding chain elements, is generated from [520]. It has 24 vertices, 36 edges and 14 faces: 6 squares and 8 hexagonal polygons (VEF) = (243614) (Fig. 2*a*). This time the parameter a appears as the radial distance from the fourfold axis of the vertices of a square delimiting the axial hole. The same parameter value also fits with the smallest axial distance of a hexagonal star-like funnel along the threefold axis. Hexagonal holes approximately delimited by lattice points are observed in the view along a twofold axis.

Finally, the *human mitochondrial ferritin* rMtF (1r03) has a morphology between that of the bacterio ferritin Bfr and the SOR. The enclosing cube has $L = 59$ Å but now the z value is 10, so that the form lattice has parameter $a = 5.9$ Å, which corresponds to the smallest axial distance of a square hole along the fourfold axis (Fig. 2*b*). The basic enclosing polyhedron is obtained disregarding small protruding chain elements delimiting the funnel and rising with height a above a face of the basic polyhedron, in a way similar to the chimneys of SOR. This polyhedron requires two generators, [905] and [6$\bar{4}$8], ensuring a better spherical shape for the cage than in the previous examples. It has 48 vertices, 84 edges of three different lengths and 38 faces (6 squares, 8 equilateral triangles, 24 pentagonal polygons) so that (VEF) = (488438).

2.3. The cage

It is not easy to determine the polyhedral shape delimiting the cavity of the protein cage. Here this is done in terms of a

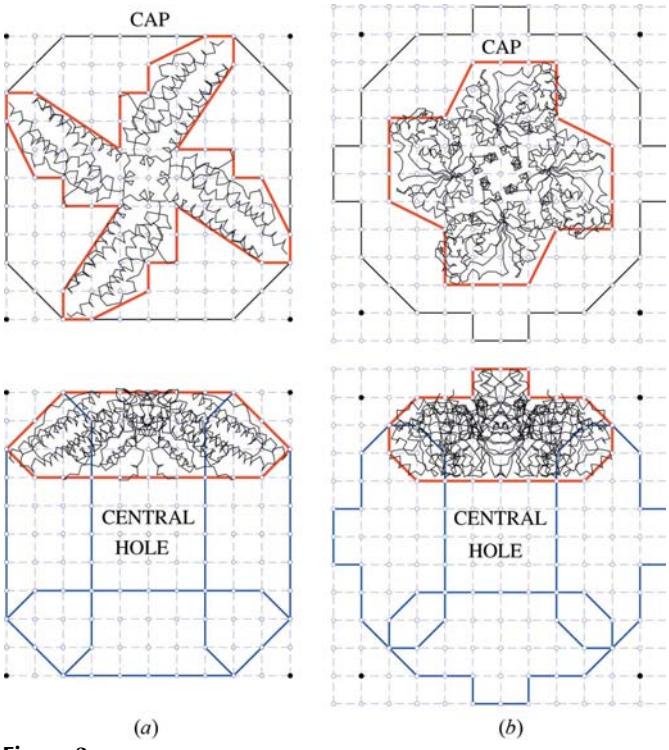

Figure 3
The tetrameric caps around the fourfold axes delimit a central polyhedral hole. Only one cap is shown filled with the corresponding chains, together with the boundaries of the refined polyhedron enclosing the cage and vertices at form lattice points in views along [001] and [010], respectively. (*a*) Bacterio ferritin Bfr. (*b*) Sulfur oxygenase reductase SOR.

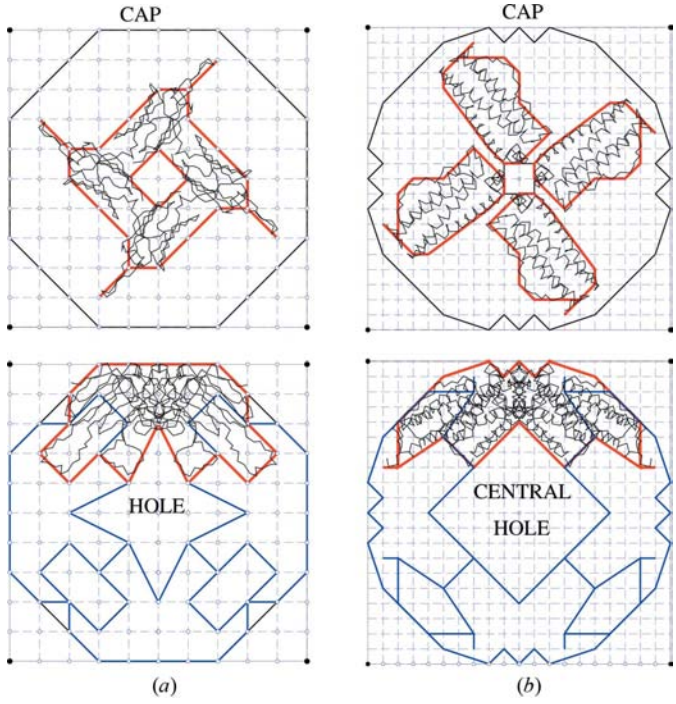

Figure 4
In a way similar to Fig. 3, tetrameric cap and central hole are shown. (*a*) Small heat-shock protein sHSP. (*b*) Human mitochondrial ferritin rMtF.

Table 1
The basic polyhedra delimiting the external envelope and the central cavity are generated by applying the octahedral group 432 to the sets of indices indicated in square brackets.

Sq = squares, ET = equilateral triangles, IT = isosceles triangles, Re = rectangles, He = hexagonal polygons, Pe = pentagonal polygons, Oct = octahedra, V = vertices, E = edges, F = faces.

	Bacterio ferritin Bfr	Sulfur oxygenase reductase SOR	Heat-shock protein sHSP	Mitochondrial ferritin rMtF
Envelope	[335]	[513]	[520]	[905], [6$\bar{4}$8]
(V E F)	(24 48 26)	(24 60 38)	(24 36 14)	(48 84 38)
Faces	6Sq + 8ET + 12Re	6Sq + 8ET + 24IT	6Sq + 8He	6Sq + 8ET + 24Pe
Cavity	[221]	[202]	[131]	[604]
(V E F)	(24 36 14)	(12 24 14)	(24 44 22)	(24 36 14)
Faces	8ET + 6Oct	6Sq + 8ET	6Sq + 8Re + 8ET	6Sq + 8He

basic internal polyhedron by looking at the tetrameric clusters forming a cap around a fourfold axis and disregarding the protruding chains mentioned above. Viewed from another fourfold axis and together with the corresponding three tetramers belonging to the same upper hemisphere, one sees a projected image of the central hole. From this image, one identifies the indexed vertices of a polyhedron which approximates the cavity. The result is shown in Figs. 3(*a*), (*b*) and Figs. 4(*a*), (*b*).

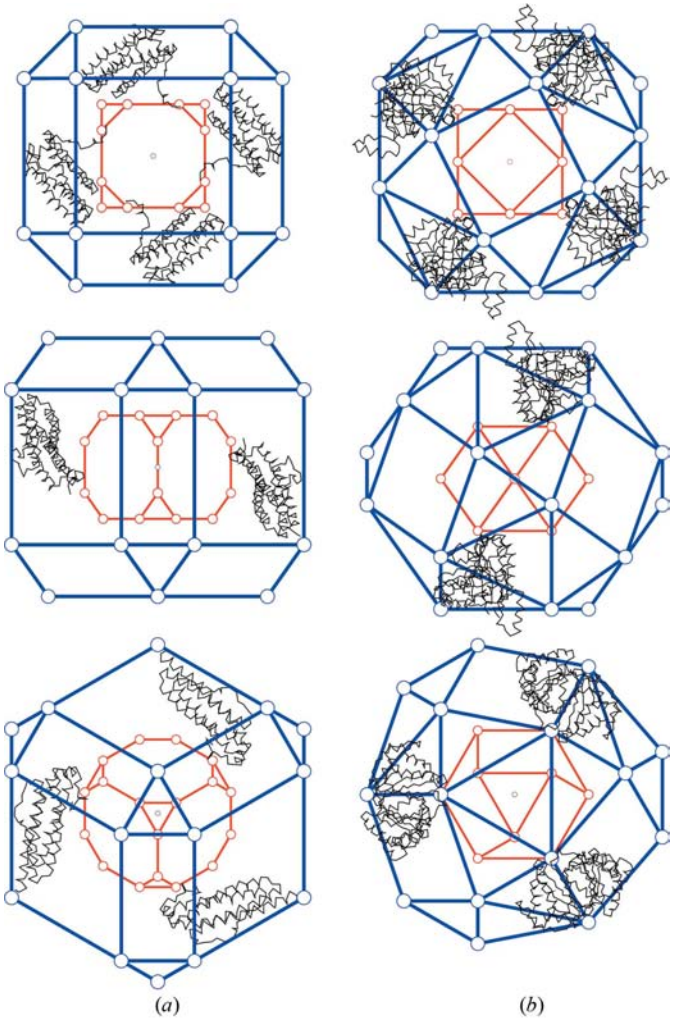

Figure 5
Cage model in terms of external and internal basic polyhedra delimiting the external and the internal surfaces, respectively, of the protein cage viewed along [001], [110] and [111] together with the chains in the corresponding equatorial regions and running between the surfaces of the two polyhedra (see Table 1 for more details). (*a*) Bacterio ferritin Bfr. (*b*) Sulfur oxygenase reductase SOR.

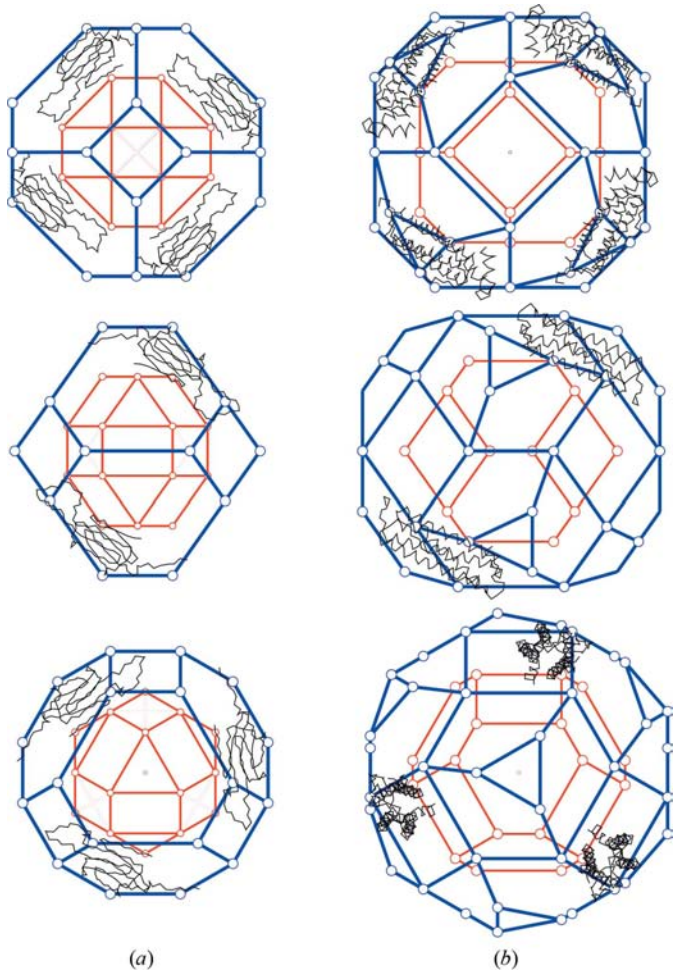

Figure 6
As in Fig. 5, the polyhedral cage model is shown together with equatorial chains. (*a*) Small heat-shock protein sHSP. (*b*) Human mitochondrial ferritin rMtF.

Table 2
Sets of chains forming tetramers, trimers and dimers with symmetry 4, 3 and 2, respectively, lying in planes perpendicular to the symmetry axes [001], [111] and [110], respectively.

	Tetramers	Trimers		Dimers	
Chains	(0, 1, 14, 15)	(0, 4, 8)	(12, 19, 22)	(0, 12)	(6, 18)
	(2, 3, 12, 13)	(1, 5, 9)	(13, 18, 23)	(1, 13)	(7, 19)
	(4, 6, 17, 19)	(2, 6, 10)	(14, 17, 20)	(2, 14)	(8, 20)
	(5, 7, 16, 18)	(3, 7, 11)	(15, 16, 21)	(3, 15)	(9, 21)
	(8, 11, 21, 22)			(4, 16)	(10, 22)
	(9, 10, 20, 23)			(5, 17)	(11, 23)

A refinement is then possible by repeating the procedure along other axial directions. For the four protein cages considered, this appears not to be necessary. In any case, the finite thickness of the chains limits the precision obtained from the projected views. More important for allowing comparison of the architecture is to adopt the same recipe in each case.

Combining the external enclosing polyhedron with the internal one of the cavity into a polyhedral pair, one then gets an indexed model of the protein cage. A check of the validity of the model derived is obtained by including in the axial views of these pairs the multimeric clusters belonging to the corresponding equatorial regions, as presented in Figs. 5(*a*), (*b*) and Figs. 6(*a*), (*b*).

The geometrical properties of the polyhedral cage models are summarized in Table 1. Note that, in the case of the small heat-shock protein, one deduces from the central hole shown in Fig. 4(*a*) that the cavity polyhedron has two generating vertices: one at [131] and one at [230]. This polyhedron has 48 vertices, 108 edges and 62 faces: 6 squares, 8 equilateral triangles, 24 isosceles triangles and 24 trapeziums. In Fig. 6(*a*) and in Table 1, the simpler [131] polyhedron is indicated.

3. Cluster and monomeric forms

Previous investigations on the chaperonin complex GroEL–GroES (Janner, 2003*a*,*b*) and on various serotypes of the human rhinovirus (Janner, 2006) have shown that the property of enclosing forms to have vertices at points of the form lattice obtained from the morphology of the whole is also observed for axial-symmetric clusters of subunits. In all these cases, the identification of the indexed vertices in three dimensions is done on projected views along and perpendicular to symmetry directions, in the way illustrated in the previous section for the four protein cages. The aim is to extend the verification of these properties to their axial-symmetric multimers down to the enclosing forms of single monomers.

3.1. Comparative labeling of the monomeric chains

All 24 chains building the cage are by symmetry identical (in principle at least), not, however, in the projected views used for the determination of the vertices of their enclosing form. To allow a comparison, one has to adopt a uniform labeling scheme which bridges the gap between the different descriptions used by the authors of the PDB files. As already

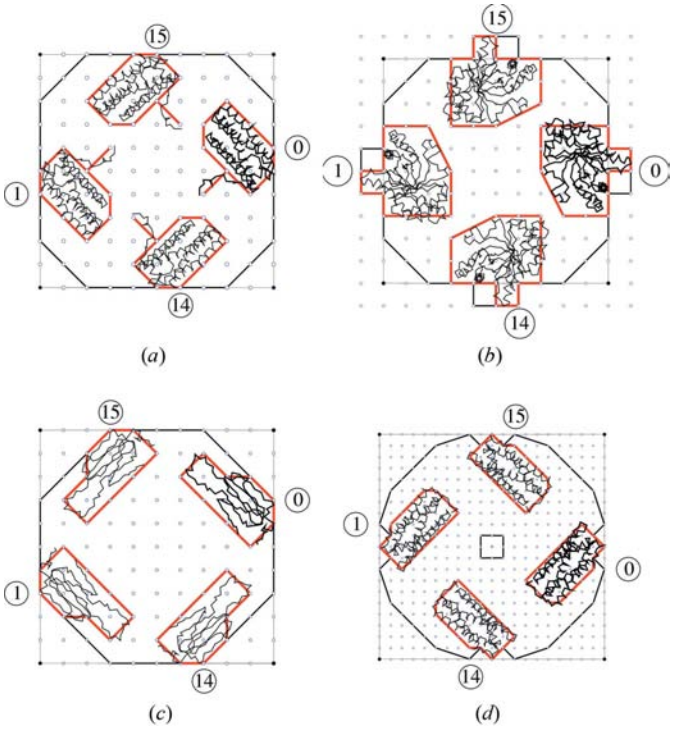

Figure 7
Enclosing forms with vertices at form lattice points of the tetrameric cluster (0, 1, 14, 15) around [001]. (*a*) Bfr, (*b*) SOR, (*c*) sHSP, (*d*) rMtF.

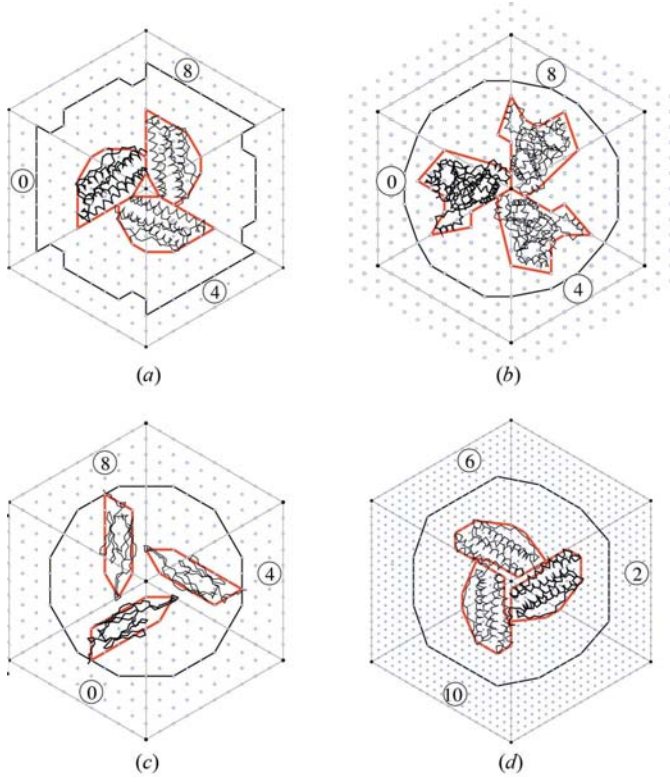

Figure 8
Enclosing forms around [111] with vertices at form lattice points of the trimeric clusters (0, 4, 8) and (2, 6, 10), respectively. (*a*) Bfr, (*b*) SOR, (*c*) sHSP and (*d*) rMtF.

Table 3
Correspondence between chains and tiles belonging to the upper (+) and lower (−) hemisphere, respectively, of the corresponding monomeric enclosing forms.

Axes	Tiles	Chains		Tiles	Chains	
001	I	+ 0, 1, 14, 15	− 2, 3, 12, 13	II	+ 4, 6, 17, 19	− 5, 7, 16, 18
	III	+ 8, 11, 21, 22	− 9, 10, 20, 23			
111	IV	+ 0, 4, 8	− 13, 18, 23	V	+ 12, 19, 22	− 1, 5, 9
	VI	+ 15, 16, 21	− 2, 6, 10	VII	+ 3, 7, 11	− 14, 17, 20
110	VIII	+ 0, 12	− 1, 13	IX	+ 3, 15	− 2, 14
	X	+ 4, 16	− 6, 18	XI	+ 7, 19	− 5, 17
	XII	+ 8, 20	− 11, 23	XIII	+ 9, 21	− 10, 22

done for the various serotypes of the human rhinovirus (Janner, 2006), one starts from a labeling of the elements of the symmetry point group. In the follow up, the monomer obtained by applying a group element to a chosen initial chain gets the same label. The procedure possibly requires a change in the coordinate system of the PDB data to a symmetry-adapted one. It is an important but purely technical detail not mentioned further. The choice of the starting monomer labeled by the group identity has consequences for the comparison between different biomolecules. How to reach an optimal choice is not self-evident when the systems show large differences in the details of their structure. The choice made here can possibly be improved, but yields reasonable results. The details of the comparative labeling adopted can be found in Appendix *A* and in Table 5, in particular.

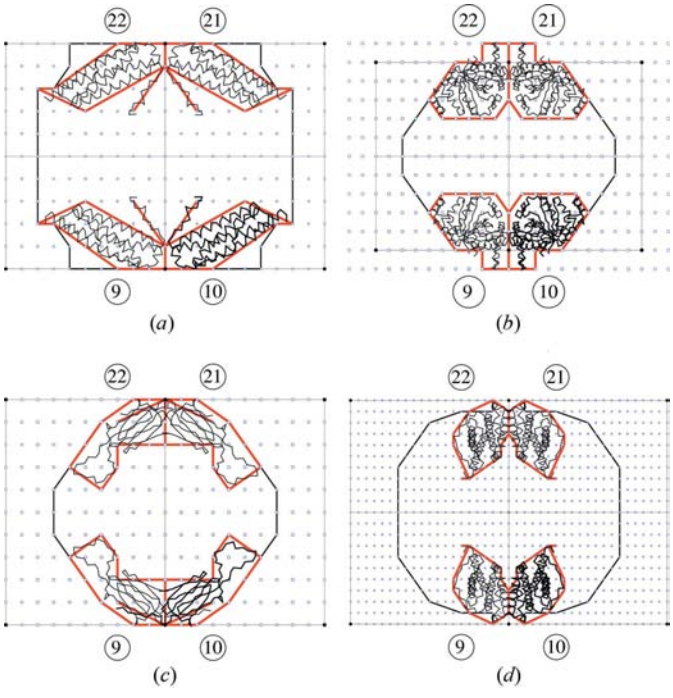

Figure 9
Enclosing forms around [110] with vertices at form lattice points of the dimeric clusters (9, 21) and (10, 22), respectively. (*a*) Bfr, (*b*) SOR, (*c*) sHSP and (*d*) rMtF.

3.2. Axial-symmetric clusters

There are six tetrameric clusters with symmetry 4 around the axis [001] grouping chains as indicated in Table 2. The enclosing forms of the cluster (0, 11, 14, 15) for the four oligoproteins Bfr, SOR, sHSP and rMtF are shown in Figs. 7(*a*), (*b*), (*c*), (*d*) as a comparative illustration. The cluster (8, 11, 21, 22) builds the fourfold cap of the corresponding cages, as plotted in projected views in Figs. 3(*a*), (*b*) and 4(*a*), (*b*).

In a similar way, each of the eight sets of chains around the [111] axis indicated in the same Table 2 forms a trimer with symmetry 3. As shown in Figs. 8(*a*), (*b*), (*c*), the trimeric clusters (0, 4, 8) of Bfr, SOR and sHSP are arranged in a similar way, but for rMtF one has to consider instead the cluster (2, 6, 10), shown as Fig. 8(*d*). From the point of view of fold, this is surprising because, as discussed in more detail in paper II, while the folds of Bfr, SOR and sHSP are very different, the motif of rMtF is essentially the same as in Bfr. Apparently, for the projected enclosing form, the relative

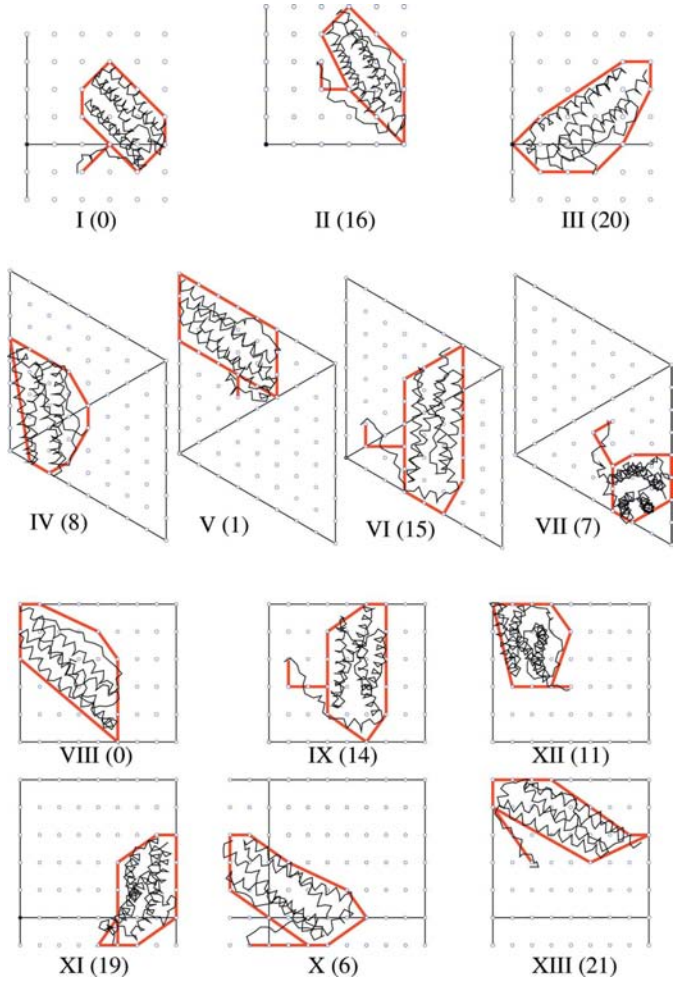

Figure 10
Projected monomeric enclosing forms of bacterio ferritin Bfr defining different indexed tiles perpendicular to [001] (first row), [111] (second row) and [110] (last two rows). Given in brackets is the label of the chain selected (see Table 3).

orientation of the monomeric chains with respect to the symmetry axes of the cage is more important than the fold.

This different behavior does not occur in the dimeric clusters of the four different cages. All clusters are correspondingly arranged in a similar way around the [110] axis as shown in Figs. 9(*a*), (*b*), (*c*), (*d*) for the pair of dimers (9, 21) and (10, 22) of Bfr, SOR, sHSP and rMtF, respectively.

Summarizing the situation: in each of the four protein cages considered, the enclosing forms of the axially symmetric clusters have vertices at points of the form lattice.

3.3. Monomeric enclosing forms

The step from the enclosing forms of axial-symmetric clusters to the forms of the single monomers is straightforward because most of these already appear disconnected in the forms of the cluster involved. The indexing of the vertices, however, is less straightforward due to superposition in projection of different lattice points. The method used for arriving at the correct indexing is based on the property that polyhedral vertices always appear either at the boundary or inside the polygon with vertices at points of one of the projected cubic lattices. One has also to consider a few additional vertices for chain segments giving rise to arms attached to the polyhedral form. For symmetry reasons it is, however, not necessary to perform this check for all the 24 monomers in the cage. Disregarding height difference in the direction of projection, the forms can be regarded as two-dimensional tiles lying in a plane perpendicular to the corresponding axis of symmetry. For the fourfold, threefold and twofold axes, one finds three, four and six tiles, respectively, only differing in orientation from other monomeric tiles lying in the same plane. Chains occurring in a given axial cluster correspond to the same tile, and so do monomers related by twofold rotations of the projected cubic lattices. These rotations possibly change the orientation.

With Roman numbers used for labeling the tiles, the correspondence between chains and tiles is indicated in Table 3. These tiles occur in enantiomorphic pairs (+, −), depending on whether the axial cluster belongs to the upper (+) or lower (−) hemisphere with respect to the equatorial plane perpendicular to the corresponding rotational axis. Disregarding enantiomorphism, there are 13 different tiles shown in Figs. 10, 11, 12 and 13 for Bfr, SOR, sHSP and rMtF, respectively. In

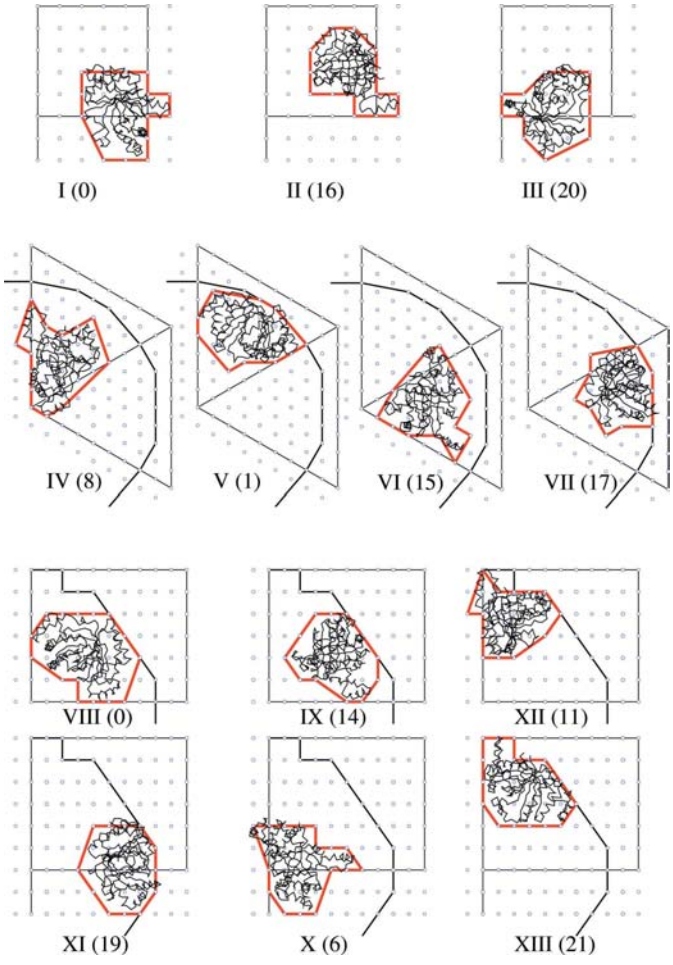

Figure 11
Set of monomeric tiles of sulfur oxygenase reductase SOR shown in a way similar to Fig. 10.

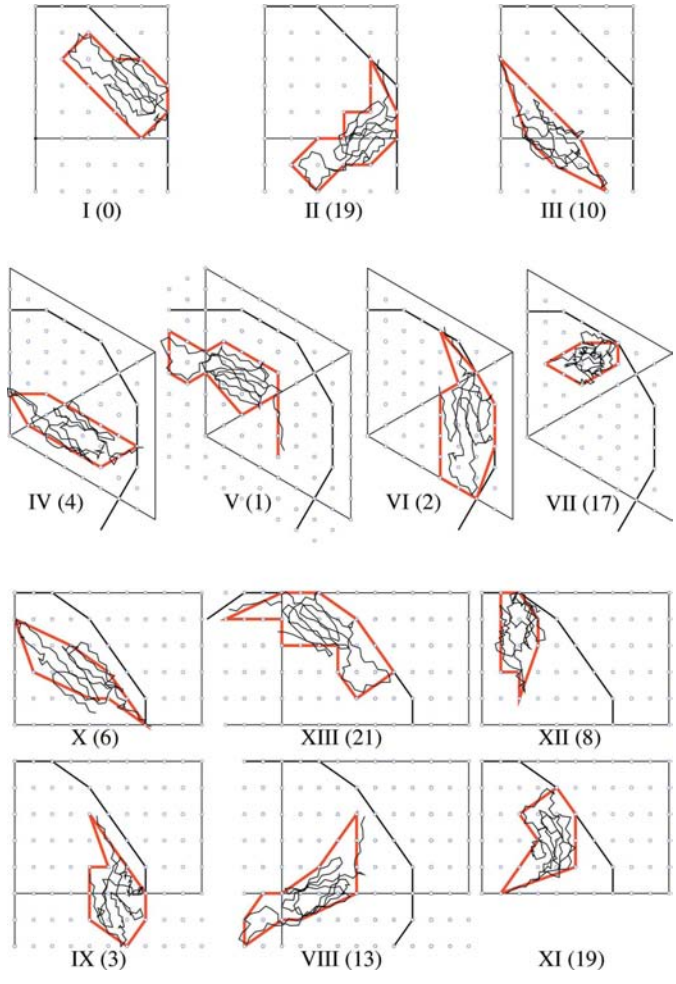

Figure 12
Set of monomeric tiles of small heat-shock protein sHSP shown in a way similar to Fig. 10.

Table 4
Sets of indices $[n_1n_2n_3]$ of the vertices V_i of the enclosing polyhedral form, possibly with attached arms, of the monomers labeled by the group identity (see Table 5).

	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
SOR	[512]	[212]	[522]	[3$\bar{2}$2]	[201]	[202]	[513]	[423]	[22$\bar{2}$]	[212]
rMtF	[3$\bar{3}$9]	[3$\bar{3}$8]	[6$\bar{4}$8]	[823]	[712]	[604]	[5$\bar{1}$5]	[4$\bar{3}$8]	[6$\bar{1}$8]	[707]
Bfr	[335]	[325]	[4$\bar{1}$2]	[401]	[214]	[224]	[404]	[503]	[502]	[423]
sHSP	[31$\bar{1}$]	[321]	[41$\bar{1}$]	[331]	[431]	[510]	[242]	[131]	[231]	[221]
	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}	V_{16}	V_{17}	V_{18}	V_{19}
SOR	[503]	[511]	[410]	[600]	[611]	[501]	[411]	[4$\bar{1}$4]	[503]	[5$\bar{1}$2]
rMtF	[905]	[9$\bar{1}$3]	[6$\bar{5}$7]	[5$\bar{5}$7]	[5$\bar{5}$6]	[5$\bar{4}$5]	[7$\bar{3}$4]	[9$\bar{2}$2]	[9$\bar{1}$1]	[10,0,2]
Bfr	[334]	[223]	[511]	[501]	[500]	[4$\bar{1}$1]	[301]	[2$\bar{1}$2]	[2$\bar{1}$3]	
sHSP	[520]	[330]	[51$\bar{1}$]	[40$\bar{3}$]						
	V_{20}	V_{21}	V_{22}	V_{23}	V_{24}	V_{25}	V_{26}	V_{27}	V_{28}	V_{29}
SOR	[4$\bar{1}$1]	[413]	[324]	[223]	[323]	[414]	[404]	[402]	[401]	[4$\bar{2}$2]
rMtF	[901]	[7$\bar{1}$1]	[602]							
	V_{30}	V_{31}	V_{32}	V_{33}	V_{34}					
SOR	[5$\bar{2}$2]	[4$\bar{1}$3]	[302]	[412]	[5$\bar{1}$2]					

Table 4, the indexed vertices of the monomeric enclosing forms are listed.

The indexed vertices given in Table 4 for the form of the monomers labeled by the group identity, as indicated in the first column (xyz) of Table 5, represents the final stage of a top-to-bottom approach of indexed enclosing forms. Applying the octahedral group 432 to these coordinates, one gets the vertices of all the other monomeric forms. The enclosing forms of the axial-symmetric clusters and of the cage as a whole are obtained as three-dimensional objects and/or in projection by combination of the monomeric forms involved. In other words, all architectural elements one is possibly interested in can be obtained, analyzed and compared in their metrical characterization from the indexed data derived so far.

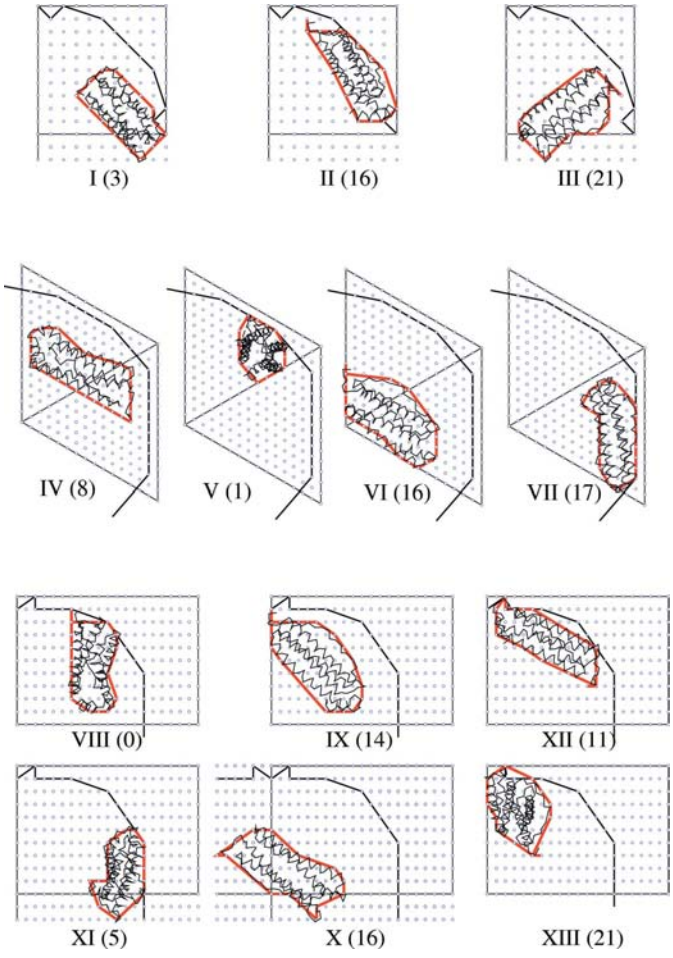

Figure 13
Set of monomeric tiles of human mitochondrial ferritin rMtF shown in a way similar to Fig. 10.

4. Final remarks

Enclosing forms are always an approximation of the atomic structural data, and so also is the indexing of the vertices in terms of a form lattice. From the results presented, one can conclude that, for a comparison between architectural elements of different biological systems sharing the same symmetry group, these approximations are better adapted than the set of all atomic coordinates.

The problem of the relation between form and content (the last one represented by the atomic positions) is analyzed in paper II in a bottom-to-top approach, which is a kind of coarse-grained version of the conventional path from the primary to the quaternary structure of monomers eventually assembled in an active biological unit.

APPENDIX *A*

A1. Axial lattices of a projected cubic lattice

Considered are the axial lattices obtained from the cubic form lattice Λ_c by projection along a rotational axis and perpendicular to it, respectively.

To begin with the fourfold case, the two projections along [001] and [010] define the same cubic lattice.

In the [111] axial projection and in the $[2\bar{1}\bar{1}]$ direction, Λ_c appears as the hexagonal lattice Λ_h with basis $a_h = \{a_{h1}, a_{h2}, a_{h3}\}$:

Table 5
The chains of the four protein cages Bfr, SOR, sHSP and rMtF, indicated by the labeling of their PDB data files, are numbered according to the octahedral group element by which they are obtained by the action on an initial one, indicated in the same column as (xyz).

The successive order adopted is the same as the one used in *International Tables for Crystallography* (Hahn, 1992) for the point group 432.

	xyz	$\bar{x}\bar{y}z$	$\bar{x}y\bar{z}$	$x\bar{y}\bar{z}$	zxy	$z\bar{x}\bar{y}$	$\bar{z}\bar{x}y$	$\bar{z}x\bar{y}$	yzx	$\bar{y}z\bar{x}$	$y\bar{z}\bar{x}$	$\bar{y}\bar{z}x$
IT	0	1	2	3	4	5	6	7	8	9	10	11
Bfr	0*A*	2*M*	2*I*	1*H*	1*A*	0*M*	0*I*	2*H*	2*A*	1*M*	1*I*	0*H*
SOR	3*D*	0*D*	0*B*	3*B*	1*E*	3*C*	2*E*	0*C*	1*F*	1*A*	2*A*	2*F*
sHSP	2*E*	1*H*	0*C*	2*B*	0*E*	2*H*	1*C*	0*B*	1*E*	0*H*	2*C*	1*B*
rMtF	0	3	2	1	4	5	7	6	8	10	9	11

	$yx\bar{z}$	$\bar{y}\bar{x}\bar{z}$	$y\bar{x}z$	$\bar{y}xz$	$xz\bar{y}$	$\bar{x}zy$	$\bar{x}\bar{z}\bar{y}$	$x\bar{z}y$	$zy\bar{x}$	$z\bar{y}x$	$\bar{z}yx$	$\bar{z}\bar{y}\bar{x}$
IT	12	13	14	15	16	17	18	19	20	21	22	23
Bfr	1*N*	1*J*	0*G*	2*B*	1*B*	2*G*	0*J*	0*N*	1*G*	0*B*	2*N*	2*J*
SOR	1*B*	2*B*	2*D*	1*D*	1*C*	0*E*	2*C*	3*E*	3*A*	3*F*	0*F*	0*A*
sHSP	0*F*	1*G*	2*D*	0*A*	2*A*	1*D*	0*G*	2*F*	0*D*	1*A*	1*F*	2*G*
rMtF	18	17	19	16	13	15	14	12	23	22	21	20

$$a_{h1} = \tfrac{1}{3}(2a_1 - a_2 - a_3), \quad a_{h2} = \tfrac{1}{3}(-a_1 + 2a_2 - a_3), \quad a_{h3} = \tfrac{1}{3}(a_1 + a_2 + a_3), \tag{3}$$

with a_1, a_2, a_3 the cubic lattice basis given in equation (2). As $|a_{h1}| = |a_{h2}| = \sqrt{2/3}a$ and $|a_{h3}| = \sqrt{1/3}a$, the axial ratio is $|a_{h3}|/|a_{h1}| = \gamma_h = \sqrt{1/2}$, so that Λ_h is integral (Janner, 2004), as expected. The indices $[m_1m_2m_3]_h$ of a point of Λ_h with respect to the basis a_h are related to the indices $[n_1n_2n_3]$ of the same point in the cubic lattice Λ_c by

$$[m_1m_2m_3]_h = \tfrac{1}{3}[2m_1 - m_2 + m_3, -m_1 + 2m_2 + m_3, -m_1 - m_2 + m_3]$$
$$[n_1n_2n_3] = [n_1 - n_3, n_2 - n_3, n_1 + n_2 + n_3]_h. \tag{4}$$

From the projections along the twofold axis [110] and along [001], one obtains an orthorhombic lattice Λ_o. The corresponding basis transformation is

$$a_{o1} = \tfrac{1}{2}(a_1 + a_2), \quad a_{o2} = \tfrac{1}{2}(-a_1 + a_2), \quad a_{o3} = a_3, \tag{5}$$

and the relation between the indices is

$$[m_1m_2m_3]_o = \tfrac{1}{2}[m_1 - m_2, m_1 + m_2, 2m_3],$$
$$[n_1n_2n_3] = [n_1 + n_2, -n_1 + n_2, n_3]_o. \tag{6}$$

As $|a_{o1}| = |a_{o2}| = a/\sqrt{2}$ and $|a_{o3}| = a$, Λ_o is in fact tetragonal and integral with axial ratio $\gamma_t = |a_{o3}|/|a_{o1}| = \sqrt{2}$.

A2. Labeling conventions

The elements of the octahedral group 432 are numbered from 0 to 23 (starting from 0 according to the computer array convention) in the same order as in *International Tables for Crystallography* (*IT*) (Hahn, 1992). For example, a point (xyz) is transformed by R_{14}, indicated in *IT* as (15), to ($y\bar{x}z$).

In the PDB data file (1nf4) of bacterio ferritin Bfr, the coordinates are given for 16 chains labeled $A, B, \ldots, P$. The biologically significant oligomerization yielding the biomolecule follows from the 8 chains A, B, G, H, I, J, M, N and the given threefold rotation, so that the 24 chains get an additional label $k = 0, 1, 2$. In particular, the starting chain A gets the label $0A$ and is transformed by R_1 to the chain $2M$, which is a threefold rotated chain M. The corresponding labeling of all 24 chains is indicated in Table 5.

In the data file (2cb2) of SOR, six chains are given, labeled as A, B, C, D, E, F, together with the rotations around a fourfold axis, so that the 24 chains of the biomolecule are indicated by $kA, kB, \ldots, kF$ with $k = 0, 1, 2, 3$. Here, $3D$ has been chosen as the starting chain because it appears in a projected position similar to chain A of Bfr.

In (1shs) small heat-shock protein sHSP, the coordinates of 8 chains are given, labeled as A, B, C, D, E, F, G, H, which together with the rotations around a threefold axis form the biomolecule with chains labeled $kA, \ldots, kH$ for $k = 0, 1, 2$. In this case, $2E$ is the selected starting chain.

Finally, in (1r03) mitochondrial ferritin rMtF, all 24 chains are obtained from the coordinates of the single chain indicated by applying the transformations given by the octahedral group 432, yielding the corresponding chains $0, 1, 2, \ldots, 23$, but in an order different from that of *IT*. The choice of the single chain of (1r03) as the starting one leads to the correspondence between the two labeling conventions. One finds, for example, $1_{IT} = 3_{\mathrm{rMtF}}$.

The various labeling conventions are summarized in Table 5.

The author thanks the Editor and the referees for their constructive comments and for the suggested text corrections.

References

Coelho, A. V., Macedo, S., Matias, P. M., Thompson, A. W., LeGall, J. & Carrondo, M. A. (2001). *Acta Cryst.* D**57**, 326–329.

Hahn, T. (1992). Editor. *International Tables for Crystallography*, Vol. A. Dordrecht: Kluwer Academic Publishers.

Janner, A. (2003*a*). *Acta Cryst.* D**59**, 783–794.

Janner, A. (2003*b*). *Acta Cryst.* D**59**, 795–808.

Janner, A. (2004). *Acta Cryst.* A**60**, 198–200.

Janner, A. (2006). *Acta Cryst.* A**62**, 270–286.

Janner, A. (2008). *Acta Cryst.* A**64**, 503–512.

Kim, K. K., Kim, R. & Kim, S.-H. (1998). *Nature (London)*, **394**, 595–599.

Langlois d'Estaintot, B., Santambrogio, P., Granier, T., Gallois, B., Chevalier, J. M., Précigoux, G., Levi, S. & Arosio, P. (2004). *J. Mol. Biol.* **340**, 277–293.

Macedo, S., Romão, C. V., Mitchell, E., Matias, P. M., Liu, M. Y., Xavier, A. V., LeGall, J., Teixeira, M., Lindley, P. & Carrondo, M. A. (2003). *Nature Struct. Biol.* **10**, 285–290.

Urich, T., Gomes, C. M., Kletzin, A. & Frazão, C. (2006). *Science*, **311**, 996–1000.