Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 1 February 2008
Accepted 25 April 2008

Comparative architecture of octahedral protein cages. II. Interplay between structural elements

Aloysio Janner

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

In paper I [Janner (2008). *Acta Cryst.* A**64**, 494–502], the enclosing forms of the monomers of four octahedral holoenzymes (bacterio and mitochondrial ferritins, small heat-shock protein and sulfur oxygenase reductase) were derived, with vertices at points of a cubic lattice and indexed accordingly. The correspondence between vertices and neighboring residues allows a sequential ordering of the vertices within the polyline defined by the C^{α} atoms of the primary structure. The alignment of these sequences shows that the form vertices denoted as turning points delimit the elements of the secondary structure (α-helices, β-strands and loops). This relation is analyzed further in a plot of angular changes in orientations of the polyline segments and planes as a function of the residue numbers and of the form vertices, respectively, leading to an alternative characterization of the ternary structure. Finally, two simple connectivity models of monomers, oriented according to the symmetry axes of the octahedral point group 432, suggest possible patterns in the self-assembly process from clusters of monomers to the quaternary structure of the cubic cage.

1. Introduction

In paper I (Janner, 2008), polyhedral enclosing forms with vertices at points of a form lattice have been derived for four protein cages with octahedral symmetry: bacterio ferritin Bfr, PDB 1nf4 (Coelho *et al.*, 2001; Macedo *et al.*, 2003), recombinant human mitochondrial ferritin rMtF, PDB 1r03 (Langlois d'Estaintot *et al.*, 2004), sulfur oxygenase reductase SOR, PDB 2cb2 (Urich *et al.*, 2006) and small heat-shock protein sHSP, PDB 1shs (Kim *et al.*, 1998). In corresponding views along the symmetry axes, the chains contained in the various forms have been plotted without any further analysis, leaving open the problem of relating form and content, which in the present context is represented by the C^{α} backbone of the polypeptide chain.

To begin with, one can make a distinction between residues well inside and those near the form boundaries, devoting special attention to vertices. One expects that C^{α}'s near form vertices are closely related to the folding of the chain into its tertiary structure. Indeed, a chain segment reaching the boundary of the enclosing form, and vertices in particular, has either to fold or to stop. For reasonably simple polyhedral forms, this is a general rule only because of the observed possibility of protruding elements.

If the vertices of the monomeric form are known, it is not difficult to recognize a C^{α} atom near to a vertex and to a folding point. In a few cases, one has to make a choice between more than one possibility. This is of no further consequence because, in any case as already pointed out in paper I, indexed enclosing forms require a certain degree of approximation, which is the price to pay for the possibility of extracting the relevant architectural elements from the structural data. In the present case, the identification of C^{α}'s near indexed vertices reduces by an order of magnitude the backbone elements to be considered, as one can see from the tables collected in Appendix *A* (Tables 2, 3, 4 and 5). In order to render graphically the relation between the two sets, the tetrameric cap of bacterio ferritin, presented in paper I with vertices at points of the form lattice, is shown in Fig. 1 with vertices at the corresponding C^{α} positions.

The bottom-to-top path followed in the present paper II is based on this relation between indexed and occupied positions. It allows, first of all, the vertices of the monomeric form to be ordered according to the primary structure and chain subunits running from a vertex to the next one to be compared with typical elements of the secondary structure (α-helices, β-strands and loops). To the topological folds of the tertiary structure, one can then add the metric of the indexed vertices. Finally, the self-assembly process can be analyzed on the basis of monomeric forms and of their projection (the tiles) along the cubic symmetry axes, only requiring a reduced number of C^{α} atoms without the need to consider the full set of residues.

The geometrical approach has a consequence because the chemical and physical interactions are only taken into account through the symmetry of the biologically active macromolecule: a possible understanding of the 'why' is reduced to a phenomenological characterization of the 'how'.

2. Secondary structures and form vertices

Monomeric form vertices, ordered according to the subsequent residues of the primary structure by the assignments indicated in Tables 2, 3, 4 and 5 of Appendix *A*, subdivide the monomer into chain segments which correspond to secondary elements.

This interesting property is shown in Fig. 2 for a number of chain segments delimited by vertices and the corresponding secondary elements (α-helices and final loop) of chain subunits in a few enclosing forms of bacterio ferritin viewed along the fourfold axis. The general case is summarized in Table 1 for the four different monomers of Bfr, SOR, rMtF and sHSP.

From the data reported, one sees that a given secondary structure involves one or more adjacent segments delimited by vertices and may require a shift in the residue number at the boundaries of the corresponding subunits. It is remarkable that this relation appears to be independent of the fold involved. In the case of the bacterio and mitochondrial ferritins, the fold is a four α-helix bundle (Macedo *et al.*, 2003; Langlois d'Estaintot *et al.*, 2004). For sulfur oxygenase reductase, it is an $\alpha\beta$ motif with an internal β barrel surrounded by α-helices, recalling the ferredoxin $(\beta\alpha\beta)_2$ topology (Urich *et al.*, 2006). In the case of small heat-shock protein, it is a sandwich of β-strands (Kim *et al.*, 1998).

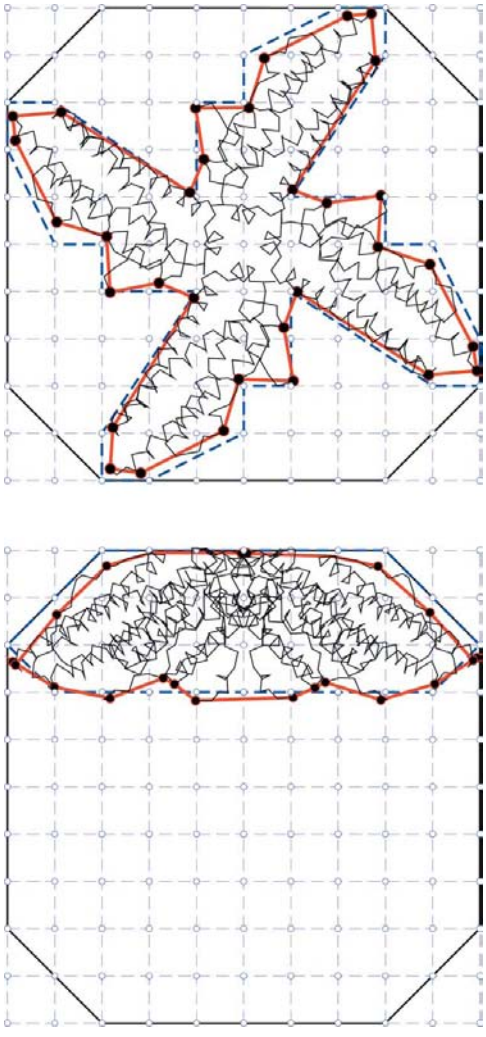

Figure 1
Comparison of a tetrameric cap element of bacterio ferritin with vertices of the enclosing form at cubic lattice points (empty circles and dashed line) and at the corresponding C^{α} positions (filled circles and thick line) in a view along and perpendicular to the [001] axis, respectively (to be compared with Fig. 3*a* of paper I).

3. Tertiary structure and indexed folds

The fold of a monomer into a tertiary structure is a fundamental biological problem investigated from so many different point of views and on so many species that one can hardly hope to add a new relevant element. Nevertheless, vertices of monomeric enclosing forms (at indexed form lattice points or at corresponding C^{α} positions) have not yet been considered in this context, despite the fact that these vertices imply a fold of the enclosed polypeptide chain, are connected with symmetry and have metrical properties, all this however in an indirect way only. For elucidating this last aspect, consider a monomeric form, as derived in paper I. It has been obtained from the quaternary structure and its point-group symmetry, but the form by itself does not have any particular symmetry, despite the symmetry-based indexing of the vertices. From the previous section, one already knows that part of the vertices occurs at (or near to) transition points between secondary elements (see Table 1) and can be regarded as *turning points* of the the chain. The remaining vertices are expected to represent *bending points* of the secondary elements involved.

The simplest way to express graphically these expectations is by an alignment plot, as used for comparing residue sequences for different species. In the spirit of the present

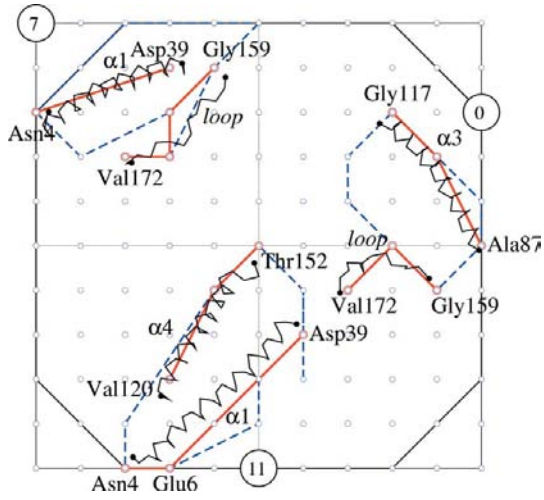

Figure 2
Secondary elements of monomers viewed along the [001] axis: α_1 and α_3 together with the final loop of the monomers 0 and 7, respectively, and α_1 and α_4 of the monomer 11. The polyline segments between monomeric form vertices (double circles) at lattice points and the delimiting residues (filled circles) show the correspondence. The dashed line gives the projected boundary of the monomeric forms involved.

Table 1
Correspondence between secondary elements and chain segments delimited by monomeric form vertices in the bacterio ferritin Bfr, sulfur oxygenase reductase SOR, human mitochondrial ferritin rMtF and the small heat-shock protein sHSP.

Secondary structure		Chain segments		Secondary structure		Chain segments	
Element	Residues	Vertices	Residues	Element	Residues	Vertices	Residues
Bacterio ferritin Bfr							
α_1	4–41	V_0–V_1	4–41	α_4	118–150	V_{10}–V_{13}	116–150
α_2	42–70	V_3–V_5	42–70	α_5	150–158	V_{13}–V_{15}	150–158
loop_1	70–87	V_5–V_8	70–87	loop_2	158–172	V_{15}–V_{18}	158–172
α_3	87–116	V_8-V_{10}	87–116				
Sulfur oxygenase reductase SOR							
β_1	5–14	V_0-V_1	2–16	α_5	173–187	V_{18}–V_{19}	174–186
α_1	17–34	V_1-V_2	16–38	β_6	103–204	V_{20}–V_{21}	191–204
β_2	40–51	V_2-V_3	38–57	loop_4	204–214	V_{21}–V_{22}	204–215
loop_1	51–71	V_3–V_5	57–70	α_6	214–221	V_{22}–V_{23}	215–224
β_3	71–80	V_5–V_6	70–80	loop_5	221–231	V_{23}–V_{24}	224–228
α_2	83–92	V_6–V_7	80–92	α'	231–233	V_{24}–	228–
α_3	94–106	V_7–V_9	92–106	α''	239–241	V_{25}	241
β_4	108–123	V_9–V_{11}	106–126	β_7	246–250	V_{26}–V_{27}	244–250
loop_2	123–133	V_{11}–V_{12}	126–131	α_7	257–269	V_{28}–V_{29}	255–269
α_4	133–142	V_{12}–V_{13}	131–142	α_8	271–281	V_{30}–V_{31}	271–281
α	146–148	V_{14}–V_{15}	144–151	β_8	281–299	V_{31}–V_{33}	281–299
loop_3	148–160	V_{15}–V_{16}	151–156	α_9	302–306	V_{33}–V_{34}	299–308
β_5	160–168	V_{16}–V_{17}	156–171				
Human mitochondrial ferritin rMtF							
loop_1	6–13	V_0–V_2	6–15	α_3	95–124	V_{11}–V_{13}	97–125
α_1	13–41	V_2–V_3	15–44	α_4	126–159	V_{13}–V_{18}	125–158
α_2	48–75	V_4–V_7	49–78	α_5	165–174	V_{19}–V_{22}	162–176
loop_2	75–95	V_7–V_{11}	78–97				
Small heat-shock protein sHSP							
β_1	36–40	V_0–V_1	33–42	β_7	104–110	V_9–V_{10}	102–118
β_2	45–49	V_1–V_2	42–51	α_1	117–119	V_{10}–	118–
β_3	53–59	V_2–V_3	51–62	β_8	121–125	V_{11}	126
β_4	68–73	V_4–V_5	66–74	β_9	129–134	V_{11}–V_{12}	126–137
β_5	76–82	V_5–V_6	74–90	α_2	137–139	V_{12}–	137–
β_6	90–97	V_6–V_8	90–98	β_{10}	142–144	V_{13}	147

investigation, the variation in the side chains of the residues is disregarded and only the successive C^α positions in space are considered. Geometrically, these positions define a polyline. A folding of the chain implies points where the polyline changes direction. It is, therefore, natural to consider angular properties of the polyline. In particular, three successive points P_{n-1}, P_n, P_{n+1} define an angle γ_n and a plane π_n, and two subsequent planes π_{n-1}, π_n form an angle δ_n between them. More details are given in Appendix *A*, together with the γ and δ dependencies of the C^α backbone polyline of the monomers of the four proteins considered in this paper and shown in Fig. 3 for the bacterio ferritin, in Fig. 4 for the mitochondrial ferritin, in Fig. 5 for the small heat-shock protein and in Figs. 6 and 7 for sulfur oxygenase reductase.

The evidence gathered so far of the relation between vertices and folding does not ensure that form vertices are the way to group the folds of different species according to motifs with similar properties as usually done with topological plots of successive secondary elements. As presented in Fig. 8, with added indexed vertices at folding positions, the bacterio and mitochondrial ferritins have the same motif of a four α-helix bundle. But no essentially new insight is gained. Possibly a better way is to consider the angular and distance variations in the polyline defined by vertices at turning points only. This is shown for Bfr and rMtF in Fig. 9 of Appendix *A*, where the secondary elements are indicated along the C^α polyline. To fit the two linear representations, the total length of both polylines have been plotted as equal. In fact, they have approximately equal length. This supports the idea that the vertices at folding positions could be used in a translation–libration–screw analysis of the dynamical properties of monomers (Schomaker & Trueblood, 1968), treating the intervertices segments as rigid.

It appears difficult to extract from these plots features common to the two similar motifs. One clearly needs more. A statistical analysis of angular correlations in the C^α polylines is a possible issue.

4. Quaternary structure and self-assembly process patterns

To get the quaternary structure of the biomolecule from the monomers and the point-group symmetry of the whole, or to get the enclosing form of the cage from the indexed monomeric forms, is straightforward. One simply has to apply the point group 432 to the initial element.

The fundamental problem is to obtain the quaternary structure and its point-group symmetry from the coordinates

of the monomer, even with the knowledge of the tertiary structure, the folding motifs and the primary structure. In the present case, this would correspond to reconstructing the cage from the vertices of the monomeric forms, even if indexed by the cubic form lattice, without knowing the center of rotation. No attempt is made to solve this problem, although simpler than the fundamental one mentioned above.

Attention is focused on the aspect of self-assembly on the basis of all that has been derived. Two models are explored: the connecting monomer forms (CMF) model based on vertices shared by two or more monomeric forms, and the connected tiles model (CTM) where connection is expressed in terms of form vertices projected along the cubic axes.

4.1. Connected monomeric forms (CMF) model

We consider the full set of form vertices obtained by applying the octahedral group 432 to the set of M vertices of the monomeric form labeled by the point-group identity as can be found in Table 4 of paper I, indicated as Table 4I. The other tables of paper I are quoted in a similar way.

Two monomers are said to be connected if their enclosing forms have two or more vertices in common. In the fully connected case, all the 24 monomers are pairwise connected. The number of shared vertices provides the rules according to which the self-assembly process may occur. The results presented for the four cage proteins should clarify the approach.

Bacterio ferritin (Bfr). Each of the 24 monomers ($k = 0, \ldots, 23$) has an enclosing form with 19 vertices ($i = 0, \ldots, 18$). Some of these have the same set of indices. So, for example, starting from $V_{14}(0) = [500]$ (see Table 4I), one finds

$$V_{14}(8) = V_{14}(21) = [005] \qquad (1)$$

(see Table 5I). Accordingly, the monomers 8 and 21 are connected. As [005] is invariant with respect to the fourfold rotation around [001], the full tetramer (8, 11, 21, 22) is connected and by symmetry this is the case for all tetramers indicated in Table 2I. The trimers are disconnected, whereas the dimer (4, 16) has monomers with three vertices in common:

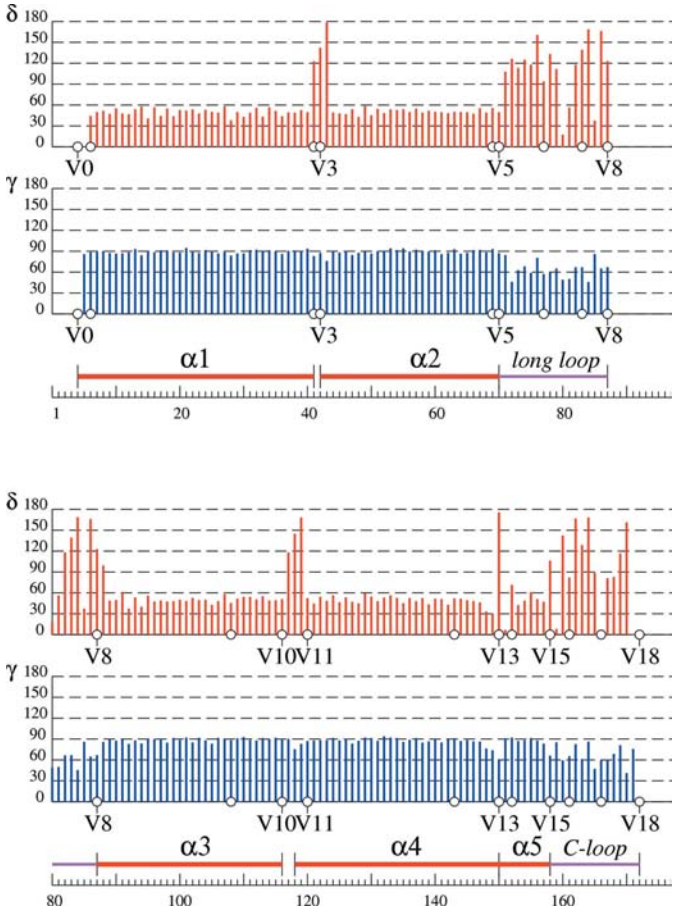

Figure 3
Angular plot of the C^{α} polyline as a function of the residue number n of the bacterio ferritin monomer Bfr (with n from 1 to 172). Plotted are the angles γ_n and δ_n (in degrees) associated with a change in direction and of plane orientation. The position of the vertices is indicated by empty circles. The turning points, labeled by Vk, delimit the secondary elements in the way indicated in the PDB file (1nf4). The regions of the α-helices are characterized by nearly constant γ and δ values.

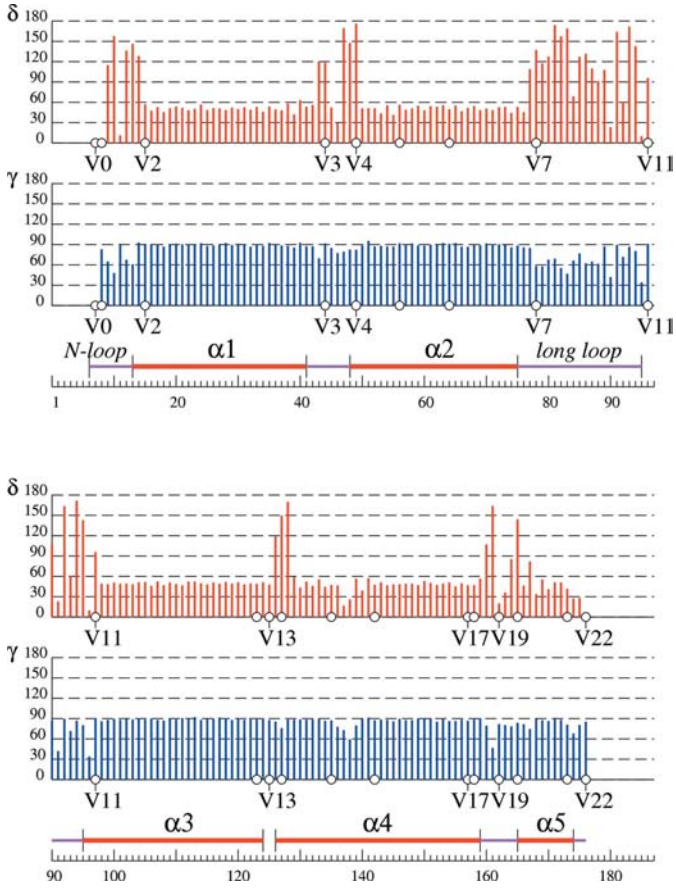

Figure 4
Same angular plot as in Fig. 3 for the mitochondrial ferritin rMtF, which has a similar four α-helix fold as bacterio ferritin.

$$V_6(4) = V_6(16) = [440],$$
$$V_2(4) = V_4(16) = [24\bar{1}], \quad (2)$$
$$V_2(16) = V_4(4) = [421].$$

This set of vertices is left invariant by the twofold rotation around [110], all dimers listed in Table 2I are connected through three vertices and, because the tetramers are also connected by one vertex, the monomers are fully connected.

The model suggests a self-assembly pattern based on a dimerization followed by a clustering of the dimers around the fourfold axes in the same orientation as the tetramers. Relevant in this process is the mutual orientation of the monomers with respect to the rotational axes of 432.

Mitochondrial ferritin (rMtF). In this case, the monomeric form has 23 vertices, as indicated in Table 4I. The tetramers and the trimers are disconnected. The monomers 4 and 16 share their V_9 vertex (indicated in Table 4I with the indices [707]). Indeed,

$$V_9(4) = V_9(16) = [770]. \quad (3)$$

Each dimer of Table 2I is therefore connected but two different dimers are not, so this model only predicts a dimerization and not a self-assembly pattern in the cubic cage.

Sulfur oxygenase reductase (SOR). The monomer of SOR has about twice as many form vertices (35) as the bacterio ferritin (19). This reflects the correspondingly larger number of residues involved (308 in SOR and 172 in Bfr). One expects accordingly for SOR more connecting vertices. One finds indeed four common vertices in the tetramers, one in the trimers and no less than six in the dimers. In particular, the monomers 8 and 21 of the tetramer (8, 11, 21, 22) share the following vertices:

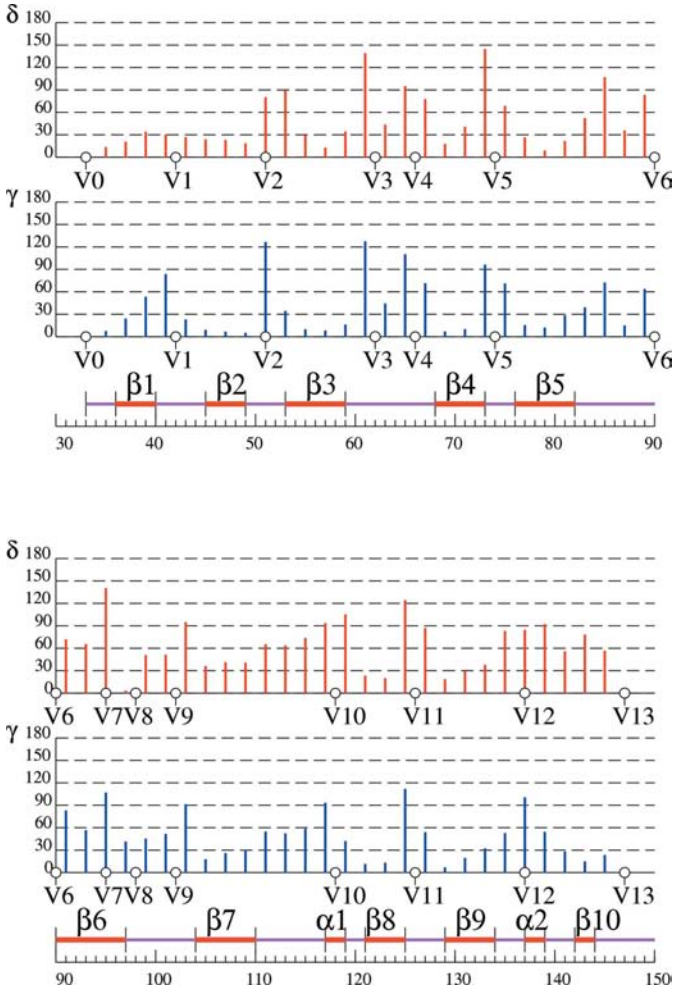

Figure 5
In the (γ, δ) angular plot of the small heat-shock protein sHSP, the C^{α} polyline considered is at even residue numbers only for taking the zigzag character of the β-strands into account, which now appear associated with fairly small γ and δ values. Nearly all monomeric vertices represent turning points delimiting (more or less well) the secondary elements, which are short.

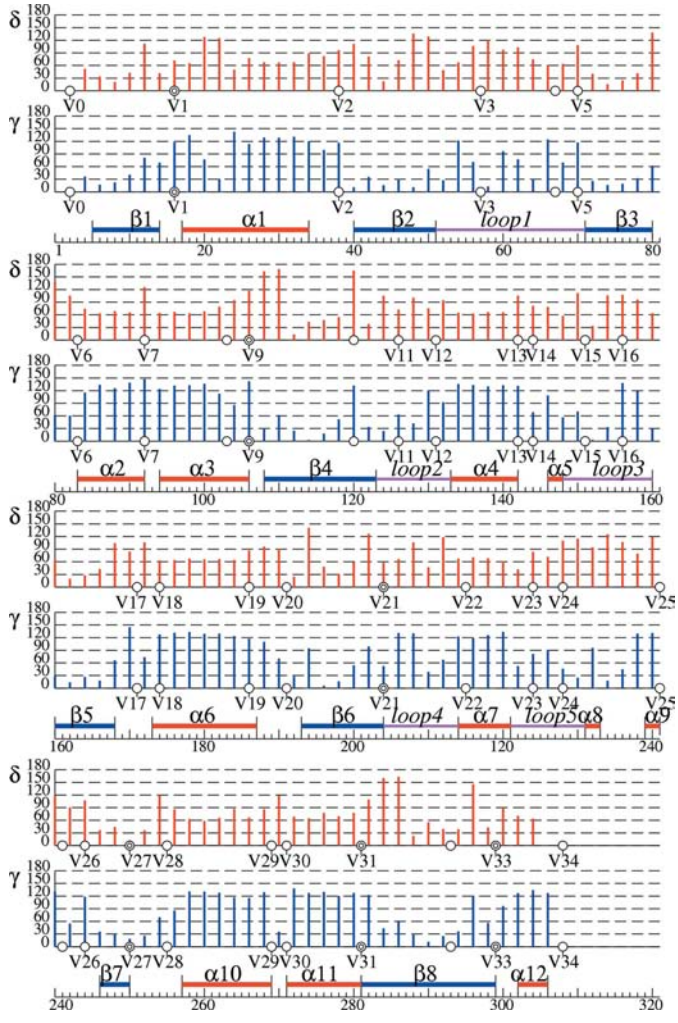

Figure 6
Same even-C^{α} angular plot as in the previous figure for the sulfur oxygenase reductase SOR. The general behavior of the β-strands is similar to that in Fig. 5. That of the α-helices is more regular, as expected. Here also, the turning-point vertices fit fairly well with the boundaries of the secondary elements.

$$V_2(8) = V_{30}(21) = [225], \quad V_{12}(8) = V_{28}(21) = [104],$$
$$V_{16}(8) = V_{20}(21) = [\bar{1}14], \quad V_{13}(8) = V_{13}(21) = [006]. \tag{4}$$

The trimer (0, 4, 8) is connected through the vertex V_8:

$$V_8(0) = V_8(4) = V_8(8) = [222], \tag{5}$$

and for the connected vertices of the dimer (4, 16) one finds:

$$V_3(4) = V_{23}(16) = [23\bar{2}], \quad V_5(4) = V_5(16) = [220],$$
$$V_{16}(4) = V_{17}(16) = [441], \quad V_{17}(4) = V_{16}(16) = [44\bar{1}], \tag{6}$$
$$V_{23}(4) = V_3(16) = [322], \quad V_{26}(4) = V_{26}(16) = [440].$$

SOR is clearly fully connected. According to this model, the dimerization dominates the process of self-assembly in competition with the formation of tetramers, whereas the role of the trimers is negligible.

Small heat-shock protein (sHSP). One is back to a relatively small number of monomeric form vertices (14), again related to the length of the monomer (with 147 residues, for 33 of which the position could not be determined). As in the case of the mitochondrial ferritin, the tetramers and the trimers are disconnected and only single dimers are connected. For example, in the dimer (0, 12) one finds three common vertices:

$$V_0(0) = V_7(12) = [31\bar{1}],$$
$$V_7(0) = V_0(12) = [131], \tag{7}$$
$$V_{11}(0) = V_{11}(12) = [330].$$

Again no self-assembly rules can be obtained from this model. Therefore, as an alternative, the connected tiles model is considered where connectivity is associated with the set of form vertices projected along the cubic axes.

4.2. Connected tiles model (CTM)

The concept of tile as projected monomeric form along one of the axes [001], [111] and [110] was introduced in paper I. The indices of the projected vertices have to be expressed with respect to the two-dimensional lattices square (s), hexagonal (h) and rectangular (r), which are sublattices of the cubic, hexagonal and orthorhombic lattices discussed in Appendix *A* of paper I. The corresponding planar indices of a cubic lattice point given by $[n_1, n_2, n_3]$ are

$[n_1, n_2]_s$	for the square lattice,
$[n_1 - n_3, n_2 - n_3]_h$	for the hexagonal lattice and
$[-n_1 + n_2, n_3]_r$	for the rectangular lattice.

Considered as connected are tiles in the same upper hemisphere (in Table 3I marked by +) sharing projected vertices

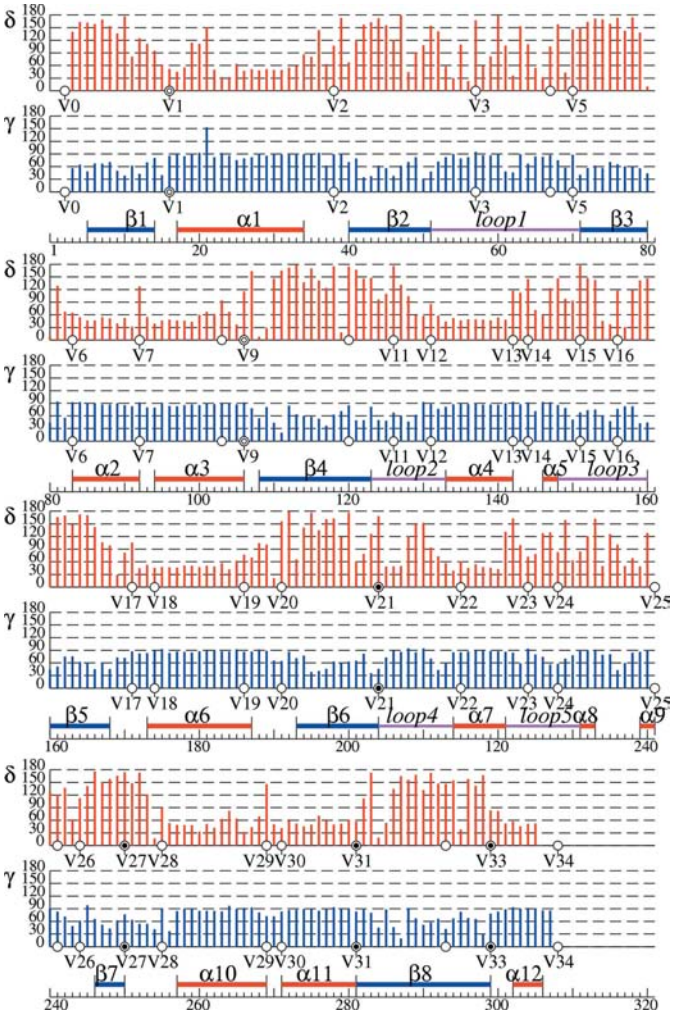

Figure 7
The angular plot of the full C^{α} polyline of SOR is shown, for a comparison with the even-*n* plot of the previous figure.

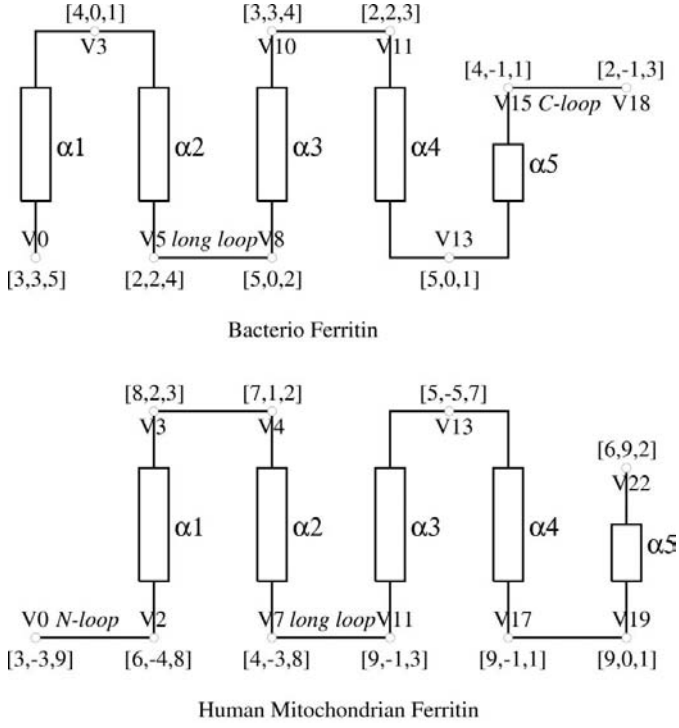

Figure 8
Comparison of the four α-helix bundle motif in bacterio ferritin and in mitochondrial ferritin, with the indices of the turning points.

and therefore having the same planar indices. The two cases for which the previous CMF model failed to make a prediction are discussed. In the remaining two cases, only the results are quoted.

Mitochondrial ferritin (rMtF). The tetramers remain disconnected in the tile approximation also, whereas the tiles VI of the trimer (15, 16, 21) (see Table 3I) are connected by the [111] projection of two vertices. In particular, for the tiles VI(15) and VI(16), one finds

$$\left.\begin{array}{r} V_{21}(15) = [171] \\ V_{0}(16) = [393] \end{array}\right\} \stackrel{[111]}{\rightarrow} [06]_{h}. \qquad (8)$$

The tiles X of the dimer (4, 16) have their origin in common:

$$V_{9}(4) = V_{9}(16) = [770] \stackrel{[110]}{\rightarrow} [00]_{\rm r}. \qquad (9)$$

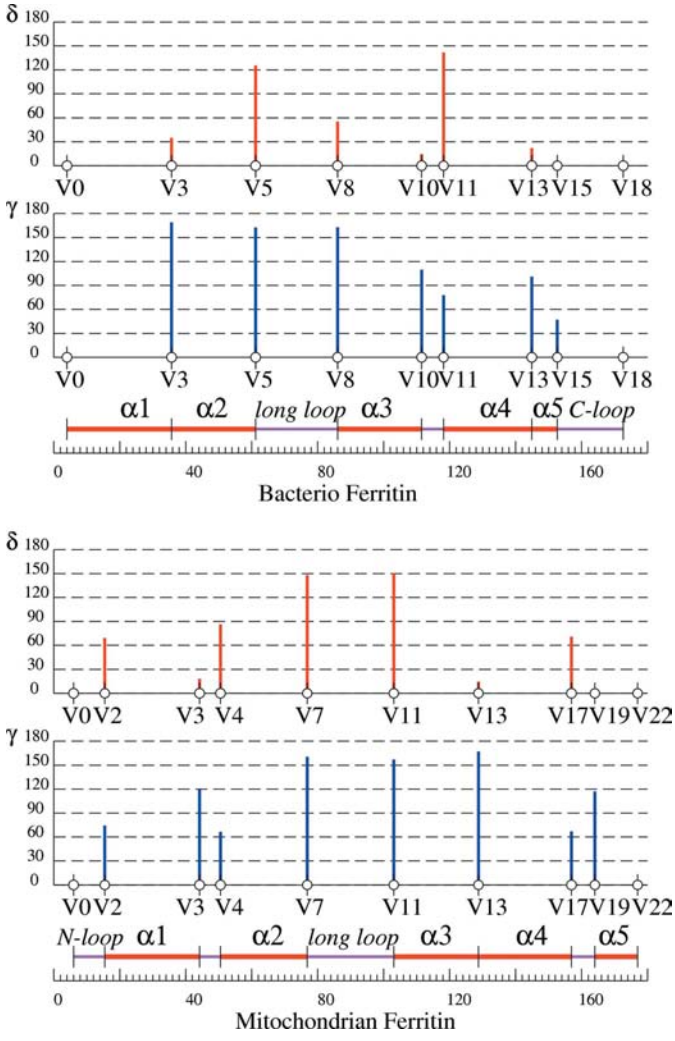

Figure 9
(γ, δ) plot for the polyline defined by the turning points indicated (V0 to V18 and V0 to V22, respectively) as a function of the distance between two successive vertices. The secondary elements are indicated as a function of the residue number n, adopting the same total length as for the polyline of the turning points. The correspondence between vertices and residue numbers delimiting secondary elements is a characteristic feature.

In the tile model, rMtF is fully connected and the model predicts the formation of trimers which then cluster in the dimeric orientations.

Small heat-shock protein (sHSP). The tiles III of the tetramers share one vertex. In (8, 11, 21, 22), one finds for the pair III(8), III(21)

$$\left.\begin{array}{r} V_{13}(8) = [0\bar{3}4] \\ V_{11}(21) = [0\bar{3}3] \end{array}\right\} \stackrel{[001]}{\rightarrow} [0\bar{3}]_{\rm s}. \qquad (10)$$

The tiles V of the trimer (12, 19, 22) also have a common vertex. For the pair V(12) and V(19), one has

$$\left.\begin{array}{r} V_{8}(12) = [32\bar{1}] \\ V_{13}(19) = [430] \end{array}\right\} \stackrel{[111]}{\rightarrow} [43]_{h}. \qquad (11)$$

Finally, the tiles VIII(0), VIII(12) of the dimer (0, 12) have three projected vertices in common:

$$\begin{aligned} V_{11}(0) = V_{11}(12) &= [330] \stackrel{[110]}{\rightarrow} [00]_{\rm r} \\ V_{7}(0) = V_{0}(12) &= [131] \stackrel{[110]}{\rightarrow} [21]_{\rm r} \\ V_{0}(0) = V_{7}(12) &= [31\bar{1}] \stackrel{[110]}{\rightarrow} [\bar{2}\bar{1}]_{\rm r}. \end{aligned} \qquad (12)$$

In this case, the connected tiles model predicts the formation of dimers which then cluster around the fourfold and threefold axes according to the orientations implied by the tetramers and the trimers, respectively.

The predictions of the CTM for the bacterio ferritin and for sulfur oxygenase reductase are consistent with those of the CMF model. In bacterio ferritin, the tiles III(8) and III(21) of the tetramer (8, 11, 21, 22) share the two projected vertices $[00]_{\rm s}$ from V_{14} and $[11]_{\rm s}$ from $V_{12}(8)$ and $V_{15}(21)$. The tiles $X(4)$ and $X(16)$ of the dimer (4, 16) are connected by the three vertices V_2, V_4 and V_6 which in [110] projections give the lattice points $[00]_{\rm r}$, $[2\bar{1}]_{\rm r}$ and $[\bar{2}1]_{\rm r}$, respectively. The trimeric tile clusters remain disconnected and one is back to the predictions of the CMF model. The situation is similar for SOR where the common vertices of the tiles are the same as derived above according to the CMF model.

5. Concluding remarks

There are two main results of this second part, where a bottom-to-top approach has been applied to what has been obtained in paper I in a top-to-bottom analysis.

The first new insight is that the vertices of the monomeric form denoted as turning points delimit the elements of the secondary structure.

The second result is represented by the possibility of suggesting, by means of very simple models of connectivity (CTM and CMF), suitable building blocks for the self-assembly of single monomers to the quaternary structure of the protein cage. Moreover, these models underline the importance of the mutual symmetry-adapted orientation of the monomers in the process of cluster formation. The two models are very coarse. Even if a refinement is possible, one has to be aware that a purely geometrical description does not replace a physico-chemical computation of the attachment

Table 2
Vertices of the monomeric form of bacterio ferritin Bfr.

Vertices	Indices	Residues	C^{α} coordinates
V_0	[335]	Asn4	(2.7, 2.7, 4.9)
V_1	[325]	Glu6	(2.6, 2.2, 4.9)
V_2	[4$\bar{1}$2]	Asp41	(4.0, −0.7, 1.4)
V_3	[401]	Tyr42	(4.0, −0.3, 1.3)
V_4	[214]	Glu69	(2.1, 1.6, 4.0)
V_5	[224]	Leu70	(2.3, 1.7, 4.2)
V_6	[404]	Gln77	(3.6, 0.5, 4.0)
V_7	[503]	Val83	(4.6, 0.2, 2.9)
V_8	[502]	Ala87	(4.9, 0.0, 2.0)
V_9	[423]	Gln108	(3.8, 2.3, 3.4)
V_{10}	[334]	Gln116	(3.0, 2.7, 4.1)
V_{11}	[223]	Val120	(2.3, 2.3, 3.4)
V_{12}	[511]	Gly143	(4.6, 1.1, 1.2)
V_{13}	[501]	Gly150	(5.0, 0.4, 0.8)
V_{14}	[500]	Thr152	(4.7, 0.1, 0.5)
V_{15}	[4$\bar{1}$1]	Ala158	(4.1, −0.4, 0.9)
V_{16}	[301]	Pro161	(3.3, −0.4, 1.1)
V_{17}	[2$\bar{1}$2]	Thr166	(2.3, −0.3, 1.8)
V_{18}	[2$\bar{1}$3]	Val172	(1.8, −1.0, 2.9)

Table 3
Vertices of the monomeric form of human mitochondrial ferritin rMtF.

Vertices	Indices	Residues	C^{α} coordinates
V_0	[3$\bar{3}$9]	Ser6	(3.6, −3.0, 8.7)
V_1	[3$\bar{3}$8]	Val8	(3.2, −3.4, 8.0)
V_2	[6$\bar{4}$8]	Asp15	(6.1, −3.7, 8.0)
V_3	[823]	Asp44	(7.9, 2.2, 3.0)
V_4	[712]	Asn49	(6.8, 0.9, 2.5)
V_5	[604]	Leu56	(6.2, 0.0, 3.9)
V_6	[5$\bar{1}$5]	Glu64	(5.0, −1.3, 5.1)
V_7	[4$\bar{3}$8]	Gly78	(4.0, −2.9, 7.9)
V_8	[6$\bar{1}$8]	Leu82	(5.8, −1.0, 7.6)
V_9	[707]	Asp84	(6.8, −0.4, 7.3)
V_{10}	[905]	Gln90	(8.6, 0.2, 5.2)
V_{11}	[9$\bar{1}$3]	Leu97	(8.6, −0.7, 2.9)
V_{12}	[6$\bar{5}$7]	Asp123	(6.2, −5.2, 7.1)
V_{13}	[5$\bar{5}$7]	Gly125	(5.3, −5.3, 7.0)
V_{14}	[5$\bar{5}$6]	Pro127	(4.8, −4.9, 6.3)
V_{15}	[5$\bar{4}$5]	Thr135	(4.9, −3.7, 4.7)
V_{16}	[7$\bar{3}$4]	Val142	(6.7, −3.3, 3.8)
V_{17}	[9$\bar{2}$2]	Lys157	(9.1, −1.6, 1.4)
V_{18}	[9$\bar{1}$1]	Met158	(8.9, −1.0, 1.2)
V_{19}	[10,0,2]	Asp162	(9.8, 0.1, 1.9)
V_{20}	[901]	Leu165	(8.7, 0.3, 1.0)
V_{21}	[7$\bar{1}$1]	His173	(6.9, −0.5, 1.0)
V_{22}	[602]	Gly176	(6.3, −0.2, 1.6)

Table 4
Vertices of the monomeric form of small heat-shock protein sHSP.

Vertices	Indices	Residues	C^{α} coordinates
V_0	[31$\bar{1}$]	Thr33	(3.5, 0.5, −0.8)
V_1	[321]	Phe42	(2.6, 2.0, 0.4)
V_2	[41$\bar{1}$]	Asp51	(4.3, 1.4, −1.4)
V_3	[331]	Gly62	(2.9, 3.0, 0.5)
V_4	[431]	Glu66	(3.9, 2.9, 0.8)
V_5	[510]	Gly74	(4.5, 1.0, 0.0)
V_6	[242]	Glu90	(1.9, 3.8, 1.9)
V_7	[131]	Ile95	(1.2, 3.3, 1.0)
V_8	[231]	Glu98	(1.7, 2.7, 0.5)
V_9	[221]	Glu102	(2.5, 2.3, 1.1)
V_{10}	[520]	Glu118	(5.0, 1.8, 0.0)
V_{11}	[330]	Asn126	(3.2, 3.2, −0.2)
V_{12}	[51$\bar{1}$]	Glu137	(5.0, 1.0, −0.9)
V_{13}	[40$\bar{3}$]	Glu147	(4.3, 0.1, −3.0)

energy between the monomers responsible for the process of self-assembly. They represent, however, an important preliminary stage in the selection of the relevant orientations.

The angular plots introduced in the analysis of the tertiary structure can also be refined by appropriate statistical methods. The surprising element revealed by the simple plots adopted is that the changes in chain orientation related to the presence of form vertices, expected to be relevant for the fold of the polypeptide chain, are of course there, but often with an angular variation of about the same value as that occurring between neighboring C^{α} inside a given secondary structure.

In the views of the author, the meaning of this work is not limited to the results obtained in the four examples considered. Its relevance is given by the derivation of a methodology applicable to many other biomacromolecules, viral capsids in particular, permitting predictions which, in principle at least, can be tested and compared.

APPENDIX *A*

A1. Assignment of C^{α} positions to indexed vertices V_n of monomeric forms

Near the boundaries of a polyhedral enclosing form (possibly having additional protruding arms), one finds subunits of the monomer and near form vertices one expects a residue where the chain is obliged to stop or to fold or, at least, to have a bend. This allows each vertex V_n to be assigned to a given C^{α} position and to the corresponding residue the indices of V_n. These indices are the integral coordinates of the vertex with respect to a basis of the form lattice, as indicated in paper I. On the same basis, the coordinates of the selected C^{α} are in general not integral because they are not at lattice point positions. Their values are indicated in the fourth column of Tables 2, 3, 4 and 5, respectively.

The deviation between actual and fitted C^{α} positions can be expressed in terms of the distance d_n between the indexed vertex V_n and the corresponding C^{α} atom, as given in Tables 2, 3, 4 and 5. In units of the cubic lattice parameter a, one finds in all four cases similar values for the mean distance Δ and for the standard deviation σ, as indicated in Table 6.

A2. Angular properties of C^{α} polylines

The monomeric form vertices V_n and the C^{α} of the polypeptide chain correspond to a set of N points $P_0, P_1, \ldots, P_{n-1}, P_n, P_{n+1}, \ldots, P_{N-1}$ in space, which defines a polyline and a sequence of vectors $\mathbf{v}_n$ pointing from P_n to P_{n+1}:

$$\mathbf{v}_n = P_{n+1} - P_n, \tag{13}$$

after identification of positions with position vectors. The total length L of the polyline is defined as the sum of the lengths $|\mathbf{v}_n|$ of the vectors $\mathbf{v}_n$:

Table 5
Vertices of the monomeric form of sulfur oxygenase reductase SOR.

Vertices	Indices	Residues	C^{α} coordinates
V_0	[512]	Pro2	(5.1, 0.8, 2.2)
V_1	[212]	Glu16	(2.1, 0.7, 1.7)
V_2	[5$\bar{2}$2]	Pro38	(4.6, 1.9, 2.2)
V_3	[3$\bar{2}$2]	Arg57	(3.0, −1.3, 2.3)
V_4	[201]	Lys67	(2.3, −0.1, 1.4)
V_5	[202]	Ser70	(2.1, 0.2, 2.0)
V_6	[513]	Asp80	(4.4, 1.1, 2.9)
V_7	[423]	Gln92	(3.9, 2.1, 2.8)
V_8	[222]	Ser103	(2.6, 1.7, 2.3)
V_9	[212]	Ser106	(2.2, 1.3, 2.4)
V_{10}	[503]	Ile120	(4.8, 0.1, 2.6)
V_{11}	[511]	Ile126	(4.7, 0.6, 1.3)
V_{12}	[410]	Thr131	(4.3, 0.6, 0.2)
V_{13}	[600]	Ala142	(5.8, 0.6, 0.3)
V_{14}	[611]	Gly144	(6.0, 0.3, 0.6)
V_{15}	[501]	Val151	(4.7, −0.0, 0.9)
V_{16}	[4$\bar{1}$1]	Tyr156	(3.7, 0.7, 1.1)
V_{17}	[4$\bar{1}$4]	Gly171	(4.5, −0.7, 3.7)
V_{18}	[5$\bar{0}$3]	Lys174	(4.8, −0.6, 3.2)
V_{19}	[5$\bar{1}$2]	Met186	(4.8, −1.0, 1.8)
V_{20}	[4$\bar{1}$1]	Pro191	(4.3, −0.8, 1.0)
V_{21}	[413]	Val204	(3.8, 0.8, 3.2)
V_{22}	[324]	Lys215	(2.8, 1.8, 3.9)
V_{23}	[223]	Pro224	(2.2, 2.1, 2.9)
V_{24}	[323]	Glu228	(3.0, 2.2, 3.1)
V_{25}	[414]	Ala241	(3.8, 0.6, 3.6)
V_{26}	[404]	Thr244	(3.9, −0.1, 3.7)
V_{27}	[402]	Val250	(4.1, −0.3, 2.3)
V_{28}	[401]	Ala255	(3.9, −0.3, 1.0)
V_{29}	[4$\bar{2}$2]	Leu269	(4.0, −1.6, 1.7)
V_{30}	[5$\bar{2}$2]	Pro271	(4.5, −1.8, 1.9)
V_{31}	[4$\bar{1}$3]	Leu281	(4.1, −1.2, 2.8)
V_{32}	[302]	Leu293	(3.4, 0.1, 1.7)
V_{33}	[412]	Gly299	(4.1, 1.1, 1.5)
V_{34}	[512]	Glu308	(5.3, 0.8, 1.8)

$$L = |\mathbf{v}_0| + |\mathbf{v}_1| + \ldots + |\mathbf{v}_{N-2}| \tag{14}$$

and not as the distance between the first and the last point of the polyline. The angles between two successive vectors follow from their scalar product

$$\gamma_n = \arccos \frac{\mathbf{v}_{n-1} \cdot \mathbf{v}_n}{|\mathbf{v}_{n-1}||\mathbf{v}_n|} \tag{15}$$

and form another sequence. Three successive points P_{n-1}, P_n, P_{n+1} define a plane π_n with normal vector $\mathbf{w}_n$ obtained by cross product of the $\mathbf{v}_n$'s:

$$\mathbf{w}_n = \mathbf{v}_{n-1} \times \mathbf{v}_n. \tag{16}$$

The angle between two successive planes equals that of the corresponding normal vectors

$$\delta_n = \arccos \frac{\mathbf{w}_{n-1} \cdot \mathbf{w}_n}{|\mathbf{w}_{n-1}||\mathbf{w}_n|}. \tag{17}$$

One then gets a linear representation of the γ angular distribution in the polyline by plotting the values of γ_n and of δ_n along a straight line, respectively, at distances given by $|\mathbf{v}_n|$. In Figs. 3, 4 and 7, the polyline is that of the C^{α} positions, where the small variation in length of the vectors $\mathbf{v}_n$ has been disregarded. In Figs. 5 and 6, the points of the polyline are restricted to C^{α} with even residue numbers (the restriction to the odd numbers yields a similar plot), whereas in Fig. 9 the polyline is that of the turning-point vertices only.

Table 6
Deviations between actual C^{α} positions and fitted ones at vertices V_n of the monomeric forms.

Enzyme	Lattice parameter a (Å)	No. of vertices V_n	Mean distance Δ	Standard deviation σ
Bfr	11.2	19	0.47	0.163
rMtF	5.9	23	0.42	0.150
sHSP	11.2	14	0.44	0.197
SOR	12.7	35	0.46	0.165

In Fig. 3 (Bfr), one recognizes the regular structure of the α-helices with nearly constant angles, $\gamma_n \approx 90^{\circ}$ and $\delta_n \approx 50^{\circ}$, and, as expected, a more irregular behavior of the γ and δ values in the loop regions. At (or near to) the turning points (9 of the 19 vertices), fairly large δ values are observed (corresponding to change of planar directions) but, surprisingly enough, one does not see any marked change in the γ value (which could imply a sharp change of direction). The same can be said for most of the bending points.

Fig. 4 shows a similar behavior in mitochondrial ferritin, confirming the validity of the analysis. In both cases, the difference between expectation and observation enhances the interest of the angular plot.

An angular characterization of the β-strands, which have a predominantly zigzag structure, is more conveniently based on the polyline formed by even (or odd) C^{α} positions. One then finds, as expected, small values of the γ and δ values ($\gamma_{2n} \approx \delta_{2n} < 30^{\circ}$) in the β-strand regions as shown in Fig. 5 for the small heat-shock protein. Practically all 14 vertices are turning points because the occurring ten β-strands are short, and so also are the two α-helices. In such an even parity plot, the angular dependency of these vertices is in general larger than 90°.

A similar same-parity plot of sulfur oxygenase reductase, which has a basic $\alpha\beta$-fold, is shown in Fig. 6. In the β regions, the behavior is similar to the previous case, and in the α regions one has $\delta_{2n} \approx 60^{\circ}$ as for the polyline of the full C^{α} sequence, whereas now $\gamma_{2n} \approx 120^{\circ}$ instead of the 90° values observed in Figs. 3 and 4 (consider in particular α_6 and α_{10}). Here also as in Fig. 5, most of the vertices can be considered to be folding points. The values for vertices at the odd-n positions do not appear. For comparison, the plot for the full C^{α} polyline is given in Fig. 7.

In Fig. 9, one sees the occurrence of fairly high angular values of about 150°.

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.

Janner, A. (2008). *Acta Cryst.* A**64**, 494–502.
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.
Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
Urich, T., Gomes, C. M., Kletzin, A. & Frazão, C. (2006). *Science*, **311**, 996–1000.