Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Strongly correlated structure of axial-symmetric proteins. III. Complexes with DNA/RNA

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

Three cases are considered of protein–DNA (or protein–RNA) complexes with a strongly correlated structure based on symmetry. In the first the symmetry of the nucleic acid is the determinant element, the second contains a dominant protein and an adaptive DNA/RNA and in the third a perturbed symmetry arises from elements of both components. The first situation is exemplified by the filamentous bacteriophage Pf1 in a low- and high-temperature state. The *Pyrococcus abyssi* Sm core and the *trp* RNA-binding attenuation protein are examples of the second situation. Finally, the nucleosome core particle represents the cooperative compromise between histone and DNA. In all the cases, the strong correlation in the structure is based on polygrammal scaling relations and on a molecular polygonal form lattice which depends on a single parameter.

1. Introduction

Helical nucleic acids and axial-symmetric proteins both have an architecture which can be characterized in terms of the symmetry properties of a molecular form lattice. In all the proteins considered in parts I (Janner, 2005*a*) and II (Janner, 2005*b*), this lattice depends on a single parameter only. For nucleic acids the situation is analogous. This implies that these structures are strongly correlated. Fibrous nucleic acids and globular proteins have very different geometries and one may wonder how the formation of a nucleoprotein complex is compatible with the building principles of the two subsystems.

In nature, symmetry is always subsidiary with respect to biological function. A symmetric geometry is also important when cooperativity is required between the symmetry-related components. In the present case, two situations are likely to occur: one where the protein adapts its structure to that of the nucleic acid and one where it is the symmetry of the protein that dominates an adaptive nucleic acid. This does not imply that the structure of the protein in the free state is the same as in the bound state, but usually the overall architecture is conserved and it is the nucleic acid which changes its symmetry. A third situation is more ambiguous from the symmetry point of view, because of the requirement of great flexibility, realised by a kind of symbiosis between protein and nucleic acids, with an approximate symmetry mutually induced. The first case occurs for the coat protein of a filamentous bacteriophage, illustrated here by the Pf1 inovirus. The second case is presented for two binding proteins which have a similar double-wheel architecture: the *Pyrococcus abyssi* Sm core (PA-Sm1) and the *trp* attenuation protein (TRAP), whose free states have already been discussed in part II. An example of the third situation is given by the nucleo-

© 2005 International Union of Crystallography
Printed in Denmark – all rights reserved

some core particle (NCP), a nucleoprotein complex of about half DNA and half histone protein.

2. Dominant DNA and adaptive coat protein

The superhelical structure of major coat proteins in a whole strain of filamentous bacteriophages (*Inoviridae*) is an indication of the dominance of the DNA symmetry with respect to that of the binding protein. This is particularly evident in the Pf1 virus structure determined by Liu & Day (1994) (PDB code 1pfi), where the DNA has the same pitch as the surrounding helix of the protein subunits. These are curved helices, largely α-helices, with 46 amino-acid residues. The peculiarity of the DNA model of Liu and Day is an inverted (or paraxial) structure with the phosphates in and the bases out, requiring a charge compensation.

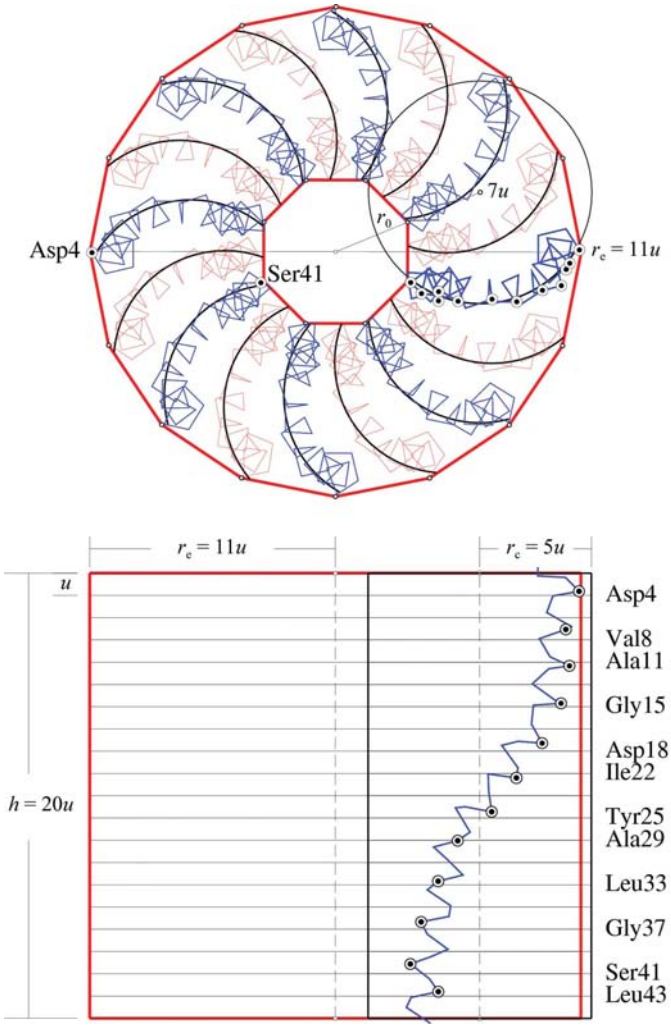

Figure 1
The low-temperature structure of the major coat protein of the filamentous bacteriophage Pf1 is shown in a view along the axis and perpendicular to it. In the asymmetric unit there are eight sets of two curved α-helices with 46 residues each and a unit rise u. Each chain is enclosed in a cylinder with height $H = 20u$, radius $r_c = 5u$ and centre at a distance of $7u$ from the axis. The radius r_e of the 16-fold basal polygon of the envelope is $11u$ and the central hole is octagonal, with a radius r_0 scaled from r_e by two the two star octagons {8/2} and {8/3} (see Fig. 2). The same scaling transformation relates the residues Asp4 and Ser41 at extremal radial distance from the axis. Shown are the observed C^{α} positions of the residues at the boundary of the cylinder.

In more recent work, a structural transition of the same Pf1 capsid at about 283 K has been found, leading in the high-temperature state $Pf1^{H}$ to a slightly different helix form for the major coat protein (Welsh *et al.*, 2000; PDB codes 1ql1 and 1ql2). In the same paper, in a comparison between different models (PDB code 2ifm, 4ifm and 1pfi) for the low-temperature $Pf1^{L}$, the authors express a clear preference for the models 2ifm and 4ifm with respect to that of 1pfi. The competence of these authors is beyond question and in particular D. A. Marvin is an expert on the statics and dynamics of molecular models of inoviruses (Marvin, 1989, 1990; Marvin *et al.*, 1994; Gonzalez *et al.*, 1995).

Several years ago, a preliminary investigation of symmetries in biomacromolecules showed that in the Pf1 model of Liu and Day the coat protein and the DNA share the same scaling symmetry, a property certainly not considered by the authors in their model building. Later, it became desirable to check whether their model is also consistent with a form lattice analysis, as presented in part II for the Pf3 inovirus. This is indeed the case, as discussed later. Moreover, the comparison between the low-temperature $Pf1^{L}$ of Liu and Day with the high-temperature $Pf1^{H}$ of Welsh and coworkers shows that the two structures are built according to the same architectural principles. This does not imply a preference for the 1pfi model with respect to the other two, 2ifm and 4ifm, which have not yet been analyzed in the same way. It is however a clear indication of the quality of the structural models derived by both groups. Symmetry is not required, but its presence in a model, while not used as input, strengthens the validity of the model and of its interpretation.

2.1. Low-temperature filamentous bacteriophage $Pf1^{L}$

The structure considered is that of the major coat protein bound to a circular single-stranded DNA (ssDNA), with one up and one down strand and paraxial phosphates (at 2.5 Å from the helical axis) according to the model of Liu & Day (1994). The original symmetry of the coat protein is that of an α-helix; that of the ssDNA is a helix with a 6.1 Å rise per nucleotide and a rotation of 131.84°. Each protein subunit binds to one nucleotide with alternating up and down DNA strands. Accordingly, the rise per protein subunit is 3.05 Å, with a rotation of 65.92°. Clearly, the coat protein adopts a superhelical structure which fits exactly to the helix of the ssDNA. This has been known for about 10 y.

Here, two additional types of structural correlations are taken into account, expressible in terms of a molecular form lattice: crystallographic scaling and rational axial ratio squared (see parts I and II). An ideal 3_{16} helical symmetry is adopted with the same unit rise u of 3.05 Å as above and a rotation of 67.5° for identical subunits (instead of the observed rotation of 65.92°) and a corresponding 3_8 symmetry for eight sets of two chains in three full turns with a rise of $2u$ instead of u. In this approximation, the whole architecture depends only on the

single parameter u, which defines an isometric 16-fold polygonal form lattice with parameters $a = c = u$. Indeed, the height of the protein subunit is $h = 20u$ and the radius of the 16-gon envelope is $r_e = 11u$, as shown in Fig. 1. The α-helix has a radius of curvature given by $r_c = 5u$, defining a cylinder with centre at a distance $7u$ from the axis of the DNA. The coat protein therefore builds a superhelix around the DNA with the same pitch.

The octagonal central channel is related to alternate vertices of the regular 16-gon of the envelope by the two successive {8/2} and {8/3} star-polygon transformations, so that the radius r_0 of the central hole is scaled from that of the envelope r_e by a factor $\mu = \mu_{\{8/2\}}\mu_{\{8/3\}} = 0.3170\ldots$ (see Fig. 2). Moreover, the residues at the surface of the curvature cylinder with radius r_c tend to have a height given by a multiple of the unit rise u, as indicated in Fig. 1. In axial projection, two of these residues, Asp4 and Ser41, which have an extremal distance from the centre, are related by the octagrammal scaling $S_\mu = S_{\{8/2\}}S_{\{8/3\}}$, so that S_μAsp4 = Ser41. As indicated in Fig. 2, starting from the tubular channel of the capsid, successive $S_{\{8/3\}}$ scaling transformations give rise to a partition of the DNA nucleotides into regions delimiting the bases, the sugar–phosphates and the DNA octagonal hole, respectively. In particular, the radius of this hole is related to that of the capsid by an $S^3_{\{8/3\}}$ scaling transformation. In the notation adopted $S_{\{M/N\}}$ indicates a planar scaling, with scaling factor $\mu_{\{M/N\}}$, relating the external regular polygon to the internal one in a $\{M/N\}$ star polygon. A similar partition (in an inverted order) has been shown to occur in the double helix of A-DNA and of B-DNA (see Fig. 6 in Janner, 2001). This confirms that in the Pf1^{L} model of Liu and Day the major coat protein adopts the symmetry of the ssDNA.

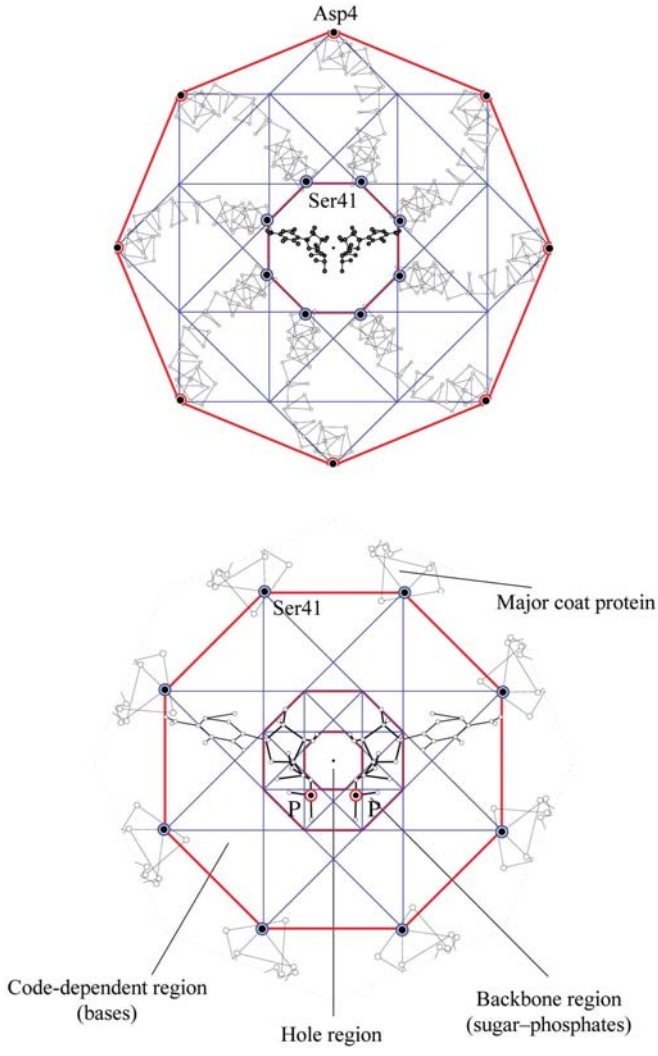

Figure 2
The circular single-stranded DNA has the same pitch as the superhelix of the major coat protein and the same octagrammal scaling symmetry, with nucleotides in a 1:1 ratio with the chains of the capsid. Only eight chains and two nucleotides in a inverse DNA conformations are shown, with the phosphates in and the bases out. The octagrams relating the hole of the capsid to the hole of the DNA (in a $\mu^3_{\{8/3\}}$ scaling relation) define a partition of space according to the molecular components of the nucleotides.

Table 1
Structural correlations in the filamentous bacteriophage Pf1 (Figs. 1, 2 and 3).

Structural parameters	Low-temperature form 1pfi (two chains)	High-temperature form 1ql2 (three chains)
Helical symmetry	13_{71}, approx. 3_{16} (3_8)	5_{27} (5_9)
Rise unit u (Å)	3.05 ($2u$ = 6.1)	2.9 ($3u$ = 8.7)
Rotation φ (°)	65.9, approx. 67.5 (2φ = 135)	66.67 (3φ = 200)
Subunit height h	$20u$	$24u$
Envelope radius r_e	$11u$	$10u$
Curvature radius r_c	$5u$	$6u$
Axial distance of the centres	$7u$	$5u$
Planar scaling	Octagrammal	Nonagrammal
Protein hole radius r_0	$S_{\{8/2\}}S_{\{8/3\}}r_e$	$S^2_{\{9/2\}}S_{\{9/3\}}r_e$
DNA hole radius	$S^3_{\{8/3\}}r_0$	—
Molecular form lattice	N = 16 ($c = a = u$)	N = 27 ($c = a = u$)

2.2. High-temperature filamentous bacteriophage Pf1^{H}

A similar analysis for the capsid of Pf1 in the high-temperature conformation based on the model derived by Welsh *et al.* (2001) reveals that the high- and low-temperature forms of Pf1 are built according to the same architectural principles as 1pfi, despite a difference in axial symmetry and thus of polygrammal scaling relations. The missing structural data of the DNA atomic positions in the PDB files indicated does not allow a full comparison of the two states of the virion, but the relevant facts already follow from the properties of the capsid.

The authors consider two models: ^{R}Pf1^{H} with a 5_{27} superhelix of 27 identical units in five turns (PDB code 1ql1) and ^{3R}Pf1^{H} with a 5_9 superhelix of nine sets of three subunits, *A*, *B* and *C*, in five turns (PDB code 1ql2). For the present analysis, the differences between 1ql1 and 1ql2 are marginal. Here, the three-subunit model is adopted because it better suits comparison with 1pfi presented above in the approximation of eight sets of two subunits in three full turns. For clarity, the structural relations shown in Fig. 3 and listed in Table 1 refer to the *C* subunit; those of the other two subunits are similar.

The axial repeat of the model is 78.3 Å for 27 units, so that the unit rise u is 2.9 Å. For the C chains it is three times larger, $3u$, with a rotation of 200°. The height h of the curved α-helix of this subunit is $24u$. The radius r_e of the envelope of the whole capsid (a prism with a 27-fold polygonal basis) is $10u$, so that the corresponding molecular form has an axial ratio of 24/10, and the structure is characterized by an isometric polygonal lattice with axial ratio one: $c = a = u$. The radius r_c of curvature of the α-helix is $6u$. The centres of the curvature cylinders are at a distance $5u$ from the centre, about 2/3 of the corresponding low-temperature value. The central channel is delimited by a nonagon with a radius r_0 scaled by a factor $\mu = \mu^2_{\{9/2\}}\mu_{\{9/3\}} = 0.353\ldots$ from the radius r_e of the envelope. The same nonagrammal scaling transformation S_μ relates two residues of neighbouring chains having an extremal distance from the central axis: S_μSer6 = Lys45. The corresponding positions in a same chain are indicated in Fig. 3. As in the low-temperature case, the residues at the boundary of the curvature cylinder with radius $6u$ have a tendency to adopt heights which are a multiple of u. A comparison of the structural parameters for the high- and the low-temperature structures of the two models presented is given in Table 1.

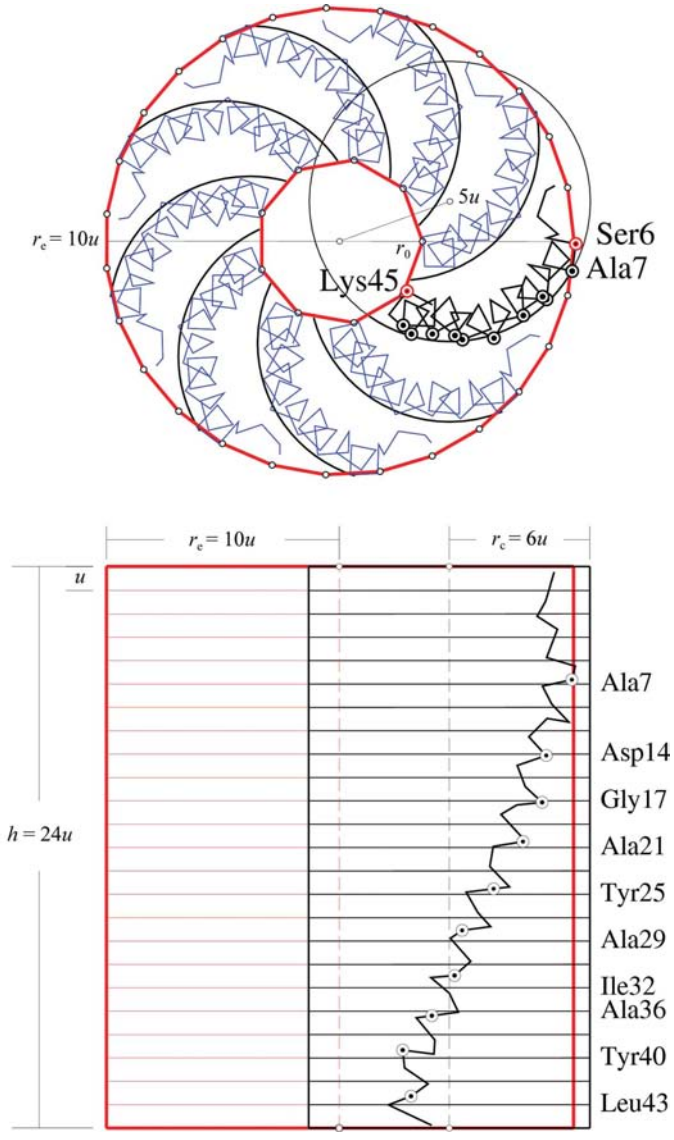

Figure 3
The high-temperature conformation of the capsid of Pf1 differs from that at low temperature. There are now nine sets of three α-helices (A, B and C, respectively) in five full turns. Only the C chains are shown. The cylinder defining the curvature of an α-chain has a height h of $24u$, a radius r_c of $6u$ and a centre at a $5u$ distance from the axis, where u is the unit rise of the 27 chains. The radius r_e of the 27-gon basis of the envelope is $10u$. The central hole is nonagonal with a radius r_0 scaled by the combination of the two star polygons {9/2} and {9/3}. Ser6 and Lys45 are mutually related by the same nonagrammal transformation. The C^α atoms of the residues at the surface of the curvature cylinder are indicated by a dot in a circle.

3. Dominant binding protein and adaptive RNA

The two RNA-binding proteins, the *P. abyssi* Sm core (PA-Sm1) and the *trp* attenuation protein TRAP, have been analyzed in part I in the free state. Here, their complexes with RNA are considered. Despite the differences in axial

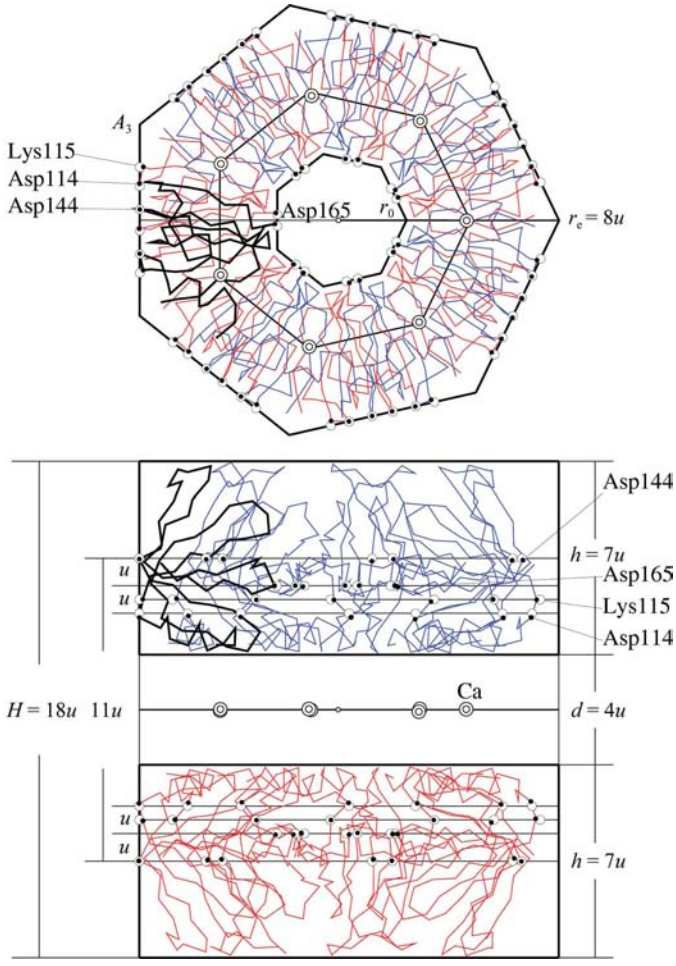

Figure 4
The PA-Sm1 protein bound to RNA has the same pair of heptamers as in the free state, but at a doubled distance $d = 4u$ (compare with Fig. 5 of part II). The remaining parameters indicated (r_e, r_0, h, H) have the same values as in the free state. The residues at the lateral boundaries, which differ in the free and in the bound states, all have integer or half-integer indices and are related by heptagrammal scaling symmetries. This is also the case for the Ca^{2+} ions.

symmetry (sevenfold for PA-Sm1 and 11-fold for TRAP) and in their binding with RNA (two different types of binding sites for PA-Sm1 and one for TRAP) the two have many common features: the binding modifies the structure of the protein, but it is the RNA which fits the geometry of the protein and not the other way around.

The symmetry of the whole and of each subsystem indicates a predominance of the geometry of the protein, even with respect to the constraints of the binding to the RNA. This is particularly evident for PA-Sm1, where the internal pocket required for one of the RNA-binding sites is simply twice the space between the two free heptamers. In this way, the symmetry conditions imposed by a one-parameter form lattice, which ensures rigidity by a structural strong correlation, remain satisfied.

Although the RNA adapts its conformation to the geometry of the protein, it has its own one-parameter molecular form lattice. As a whole, however, the complex remains strongly correlated, because the unit-of-length parameters of the protein and of the RNA are symmetry-related. It is, however, convenient to treat the protein and RNA as separate subsystems.

Table 2
One-parameter u structural correlations in the *P. abyssi* Sm core and in the *trp* RNA-binding attenuation protein (Figs. 4–8).

A negative distance indicates a corresponding overlap.

Structural parameters	PA-Sm1		TRAP	
Scaling symmetry	Heptagrammal		Ondecagrammal	
Molecular form lattice	Isometric heptagonal		Isometric ondecagonal	
	u' (RNA) = $S_{\{0.862\}}u$ (protein)		u (protein) = $S\{11/2\}u'$ (RNA)	
RNA-binding sites	External ($'$) and internal ($''$)		Wrapped around ($'$)	
Subsystems	Protein	RNA	Protein	RNA
Envelope	$r_e = 8u$	$r'_e = 8u'$	$r_e = 16$ u	$r'_e = 16u'$
Total height	$H = 18u$	$H' = 5u'$ $H'' = 9u'$	$H = 26u$	$H' = 20u'$
Subunit height	$h = 7u$	$h' = 3u'$ $h'' = 2u'$	$h = 11u$	$h' = 6u$ $h' = 5u'$
In-between distance	$d = 4u$	$d' = -u'$ $d'' = 5u'$	$d = 4u$	$d' = 10u'$
Hole–envelope scaling	$r_0 = S_{0.3079}r_e$	$r'_0 = S_{0.413}r_e$ $r''_0 = S_{0.1098}r_e$	$r_0 = S_{0.379}r_e$	$r'_0 = S_{0.768}r'_e$

3.1. *P. abyssi* Sm core in complex with RNA

The structure of the *P. abyssi* Sm core in complex with RNA is given in PDB entry 1m8v (Thore *et al.*, 2003). The biologically active molecule is a double ring of PA-Sm–U_7 heptamers.

The protein subsystem of this complex is characterized by the same isometric heptagonal form lattice ($c = a$) and the same parameter u value as in the free state. The main difference is the doubling of the distance d between the two heptamers, as already mentioned ($d = 4u$ instead of $2u$). This implies an increase of the total height H by $2u$: $H = 18u$. The other parameters are unchanged: $r_e = 8u$, $h = 7u$ and $r_0 = S_{0.3079}r_e$. The notation is the same as in part II and the meaning should be clear by looking at Fig. 4.

Minor changes occur in the symmetry relations of the residues at the molecular form boundaries: Asp114, Lys115, Asp144 and Asp165. Note that, for example, Asp114 and Asp14 denote the same residue, the different labelling having been adopted by the authors in order to distinguish between free and bound states. As in the free state, these boundary residues have heights and are scaled with respect to the vertices of the envelope in a way which ensures integral or half-integral indices for the corresponding ideal C^α positions. This is also the case for the positions of the Ca^{2+} ions. With indices expressed with respect to the symmetry-adapted basis ($c = u$, $a = 10u$) and in the asymmetric unit one has

$$\begin{aligned}
\text{Lys115} &= Y_{0.555}A_3 = \tfrac{1}{2}(1\bar{1}201\bar{1}, 8)\\
\text{Asp114} &= Y_{0.356}A_3 = \tfrac{1}{2}(2\bar{1}021\bar{2}, 7)\\
\text{Asp144} &= Y_{0.109}A_3 = (1\bar{1}101\bar{1}, \tfrac{11}{2})\\
\text{Asp165} &= Y_{0.198}S_{0.3079}A_3 = \tfrac{1}{2}(2\bar{3}6\bar{4}30, 9)\\
\text{Ca}^{2+} &= S_{0.597}A_3 = (\bar{7}6\,\bar{1}0\,6\bar{7}0, 0),
\end{aligned} \tag{1}$$

where S_μ and Y_μ denote planar and linear scaling, respectively, with scaling factor μ, as explained in the introduction of part II. Planar and axial indices are separated by a comma.

The RNA subsystem is split into two parts: one bound at the internal site of the protein and one in the external region located between the two heptamers. Both subsystems obey the same structural principles of heptagrammal scaling and isometric heptagonal form lattice, now with the single parameter u', scaled by a factor $\mu = 0.8629\ldots$ with respect to the free value u. One can then verify that the total height of the complex is a multiple of both lengths (within 0.5%): $H = 18u \simeq 21u'$. Holes and envelopes of the various subsystems are related by heptagrammal integral scalings

$$r''_0 = S_{0.1098}r_e, \quad r''_e = r'_0 = S_{0.4134}r_e, \tag{2}$$

$$r'_e = S_{0.8629}r_e = 8u', \tag{3}$$

where the prime labels the parameters of the external RNA subsystem and the double prime those of the internal one and $r_e = 8u$ is the radius of the envelope of the protein. The remaining structural parameters of the RNA subsystems are multiples of u', as indicated in Fig. 5 and listed in Table 2.

3.2. *Trp* RNA-binding attenuation protein in complex with RNA

The structure of the *trp* RNA-binding attenuation protein TRAP has been determined by Antson *et al.* (1999) and deposited as PDB entry 1c9s. The packing of the TRAP–RNA complex in the crystal consists of four 11-mers, two of which bind to RNA. As these two form a dyadically related double-wheel structure as in the free state (Antson *et al.*, 1995; PDB

code 1wap) similar to that of PA-Sm1, only these two 11-mers (with chains numbered from 12 to 22 in 1c9s) are considered here, quite independently of the question of the biologically active molecule.

Let us consider first the TRAP subsystem of the complex. The orientation of the two neighbouring 11-mers is different from that in the free state presented in part II: it is face-to face in 1wap and back-to-back in 1c9s. The geometrical shape of the central hole is also slightly different owing to the changed conformation of the chains (see Fig. 6 and compare with Fig. 8 of part II). The star polygon {11/2} of the free state is reduced in the bound state to a regular 11-gon with same the radius r_0 as the larger radius of the free-state star polygon. In both cases the scaling factor with respect to the envelope is the same: $r_0 = \mu r_e$, with $\mu = \mu_{\{11/2\}}\mu_{\{11/4\}} = 0.3796\ldots$, $\mu_{\{11/2\}} = 0.8767\ldots$ and $\mu_{\{11/4\}} = 0.4329\ldots$ (Table 2). Moreover, the L-tryptophan molecules take a symmetric position, with the C^α atoms at the vertices of a 11-gon with radius r_{trp} scaled by a factor $\mu_{trp} = 0.6476\ldots$ from the radius r_e of the envelope.

The architecture remains the same, with equal value of the unit parameter u of the isometric ondecagonal form lattice ($c = a = u$). In the bound state, there a doubling of the distance d between the 11-mers with respect to the free case: $d = 4u$ and $d = 2u$, respectively, exactly as observed for the PA-Sm1 complex, even if for TRAP this doubling is not required by a RNA-binding pocket. The total height H is increased accordingly: from $24u$ (free) to $26u$ (bound). The remaining structural parameters do not change: $r_e = 16u$ and $h = 11u$.

The residues at the lateral boundaries of the envelope and channel are different from those in the free state, but are also related by planar or linear scaling with the vertices of the envelope and have heights which approximate multiples of u. Their ideal C^α-atom positions have integral or half-integral

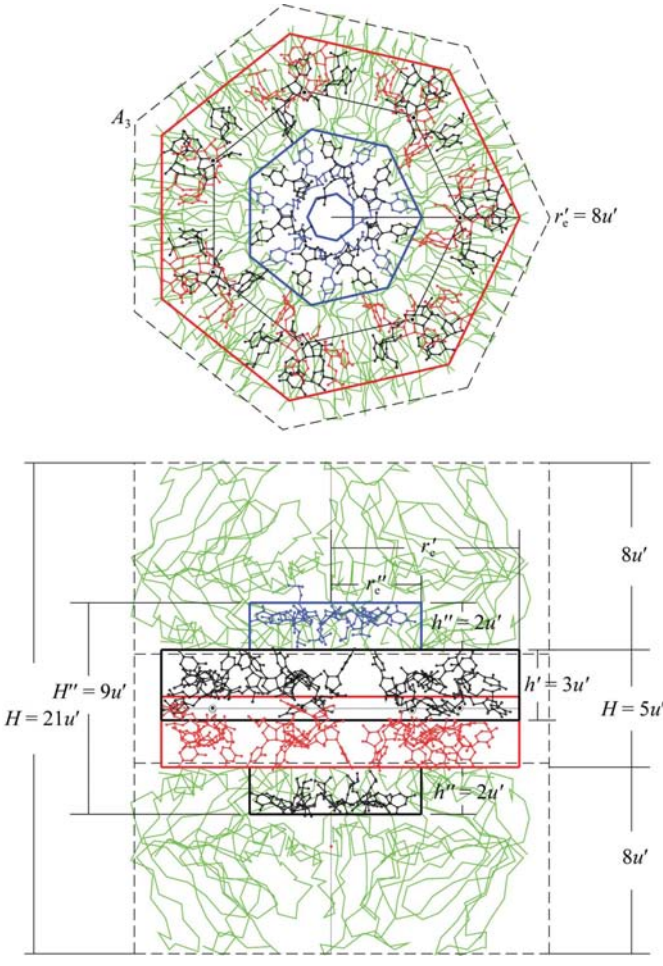

Figure 5
The RNA subsystem of the PA-Sm1 complex has the same architecture as the binding protein. The RNA envelope has a radius r'_e scaled by the heptagrammal factor $\mu = 0.8629$ with respect to r_e, as is the unit length u' with respect to u. There are two binding sites. The outer one, with radius $r'_e = 8u'$, occupies the enlarged space between the two heptameric rings. The inner binding site has radius r''_e and a central hole in a heptagrammal scaling relation with the envelope. The remaining structural parameters are multiples of u', as indicated.

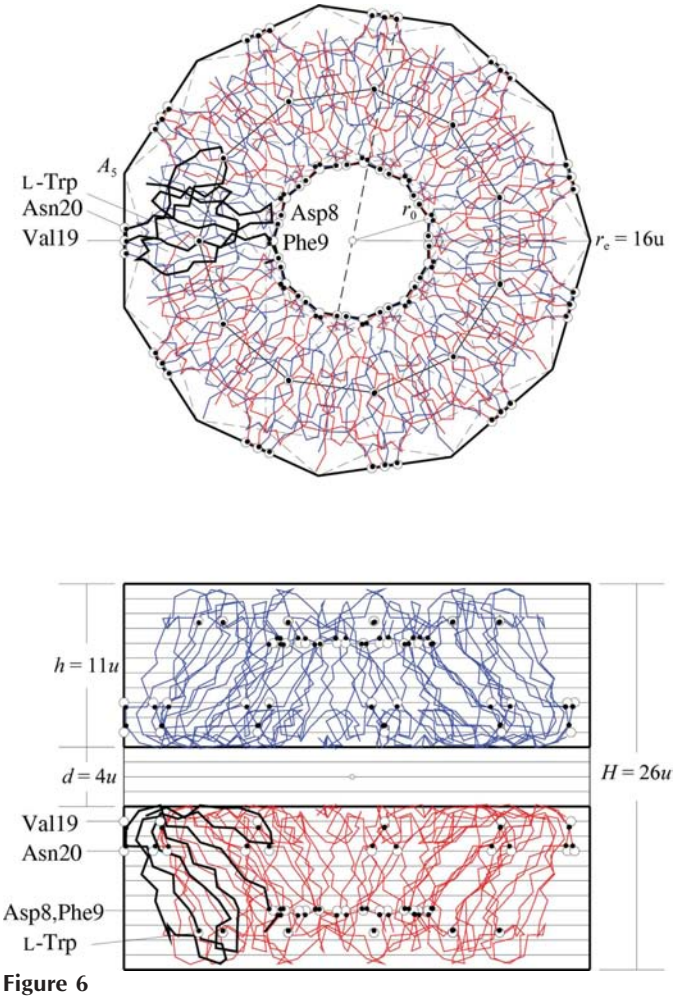

Figure 6
In the crystal of the TRAP complex, four 11-mers appear together (and not two as in the free state) in a different orientation: face-to-face (free) and back-to-back (bound). Only the two which bind to the RNA are shown. There is a doubling of the interface distance d between the two 11-mers (compare with Fig. 8 of part II), even if this space is not a RNA-binding pocket. All the other parameters are correspondingly the same. The chain conformation is changed, but the residues at the lateral boundaries are again related by polygrammal scaling and have C^α-atom positions with integer or half-integer indices. This is also the case for the L-tryptophan molecules.

indices in the symmetry-adapted basis ($c = u$, $a = 16u$). Within the asymmetric units one finds

$$\begin{aligned} \text{Val19} &= S_{0.9594}A_0 \quad z(\text{Val19}) = 3u \\ \text{Asn20} &= Y_{0.1972}A_5 \quad z(\text{Asn20}) = 5u \\ \text{Phe9} &= S_{-0.3796}A_5 \quad z(\text{Phe9}) = 9u \\ \text{Asp8} &= S_{-0.3796}\text{Asn20} \quad z(\text{Asp8}) = 9u \\ \text{L-Trp} &= S_{-0.6467}A_0 \quad z(\text{L-Trp}) = 10\tfrac{1}{2}u, \end{aligned} \tag{4}$$

with S_μ and Y_μ a planar and a linear scaling, respectively, and A_0 the vertex along the x axis (see Fig. 6).

The adaptive character of the RNA single strand consisting of repetitive GAGAU segments (Antson *et al.*, 1999) is evident from Fig. 7. Indeed, the outer and inner lateral boundaries of the RNA are scaled by a factor $1/\mu_{\{11/2\}} = 1.1406\ldots$ and $\mu_{\{11/2\}} = 0.8767\ldots$, respectively, from the TRAP envelope with radius $r_e = 16_u$. Moreover, while the height of the TRAP ondecamer is $11u$, that of the RNA is $6u$.

The architecture of the dyadically related RNAs bound to the TRAP double wheel represented in Fig. 6 requires another

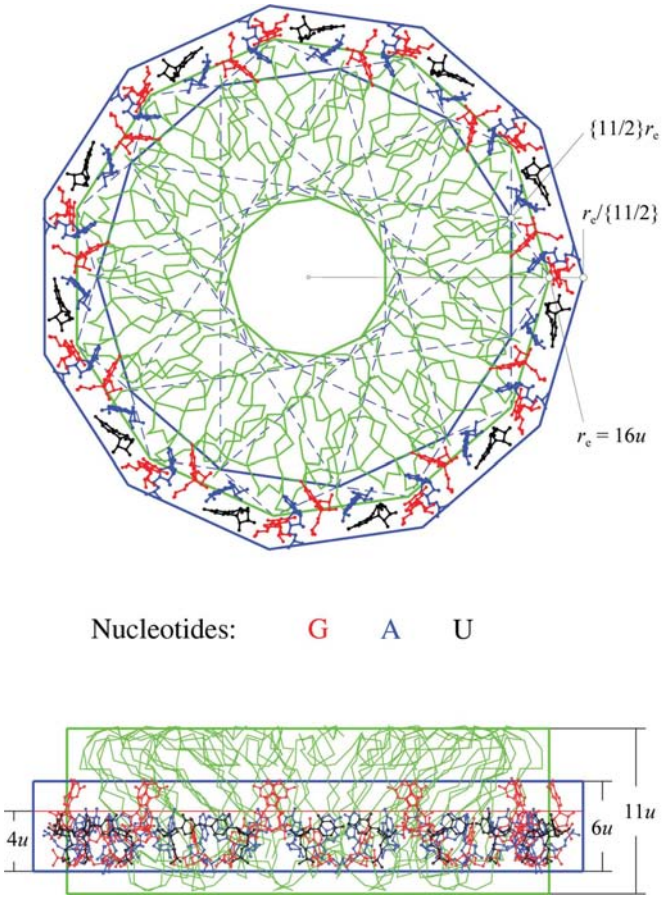

Figure 7
The structural relation between the TRAP protein and the RNA chain consisting of a repetition of G, A and U nucleotides (in different colours) is shown in a view along and perpendicular to the 11-fold axis. Ondecagonal polygrammal scalings {11/2} connect the envelope and hole of the DNA with the TRAP envelope. In the same unit of length u as in Fig. 6, the height of the RNA single strand is $6u$ and that of the 11-mer is $11u$, as indicated in Fig. 6.

structural parameter u', scaled by a factor $1/\mu_{\{11/2\}}$ with respect to the parameter u of the TRAP subsystem, so that one has $r'_e = 16u'$. In projection, the inner boundary appears now as an {11/2} star polygon (Fig. 8). In this new unit the total height H' of the RNA subsystem is $20u'$ and that of one RNA strand $h' = 5u'$; the distance between the two strands is $d' = 10u'$.

The mismatch between the height h' of one RNA strand expressed in the two parameters $h' = 6u \simeq 5u'$ is about 5%, which is a fairly large deviation. As one sees by comparing Fig. 7 with Fig. 8, the parameter u gives a better fitting for h, but not for the whole RNA subsystem.

4. Cooperative nucleoprotein complex

The nucleosome core particle of chromatin (NCP) is presented as an example of a cooperative symmetry interaction between protein and DNA (Luger *et al.*, 1997; Richmond, 2001). The histone protein octamer is not axial-symmetric and the complex with DNA has only some features of a strongly correlated structure. The nucleosome still fits into the context of this work. It represents an intermediate situation with respect to the cases discussed above, where one component of the complex with DNA/RNA dominates the other one. Moreover, the DNA superhelix encircling the histone protein recalls the RNA bound to TRAP. Keeping this last case in mind, similar building principles based on a kind of shadow symmetry can be recognised, allowing an approximate characterization in terms of polygrammal scaling and a one-parameter molecular form lattice.

4.1. Nucleosome core particle: histone in complex with DNA

The DNA binds to the histone octamer as a left-handed superhelix (Luger *et al.*, 1997; PDB code 1aoi; Richmond, 2001). The superhelix configuration can be simplified to a backbone of the P atomic positions paired by bases represented by a line segment, as in the historical diagrammatic representation of the double helix by Watson and Crick.

The envelope of the DNA thus appears in a first approximation as a cylinder with radius r_e and height H. In a view perpendicular to the axis, the transition to an upper superhelix can be recognized, followed by a central helix and a lower helix, with an approximate dyadic symmetry overall (see Fig. 9). The vertical distance d between the upper and the lower superhelices is the key for a molecular form description in terms of a single parameter u. If one takes $d = u$, the correlated values $r_e = 10u$, $H = 11u$ and $h = 4u$, $h' = 5u$ are obtained for the height of the corresponding superhelical segments (Fig. 9). This part of the architecture is indeed similar to that found in SA-Sm1 and in TRAP.

The next step is the characterization of a polygrammal scaling relating the envelope of the DNA to its central hole. The P positions in the upper and lower helical part, which have an extremal radial distance from the central axis (indicated by filled dots in Fig. 9), have an approximate 16-fold symmetry (with ideal positions indicated by empty circles). On the basis of this shadow symmetry, one finds two star polygons (drawn

with dash lines) which relate the envelope with radius r_e to the DNA hole with radius r_0,

$$r_0 = S_\mu r_e, \quad \mu = \mu^2_{\{8/2\}} = 0.5857... \tag{5}$$

The hole of the DNA forms the envelope of the histone, which shares the same unit length u and same total height $H = 11u$ and, in the corresponding region, the height $h' = 5u$ of the central part of the DNA superhelix (Fig. 10).

It can be seen that the cooperative symmetry of the nucleoprotein is built from the dyadic axis of the histone and from what remains of the helical DNA periodicity in a wrapped conformation around the globular protein.

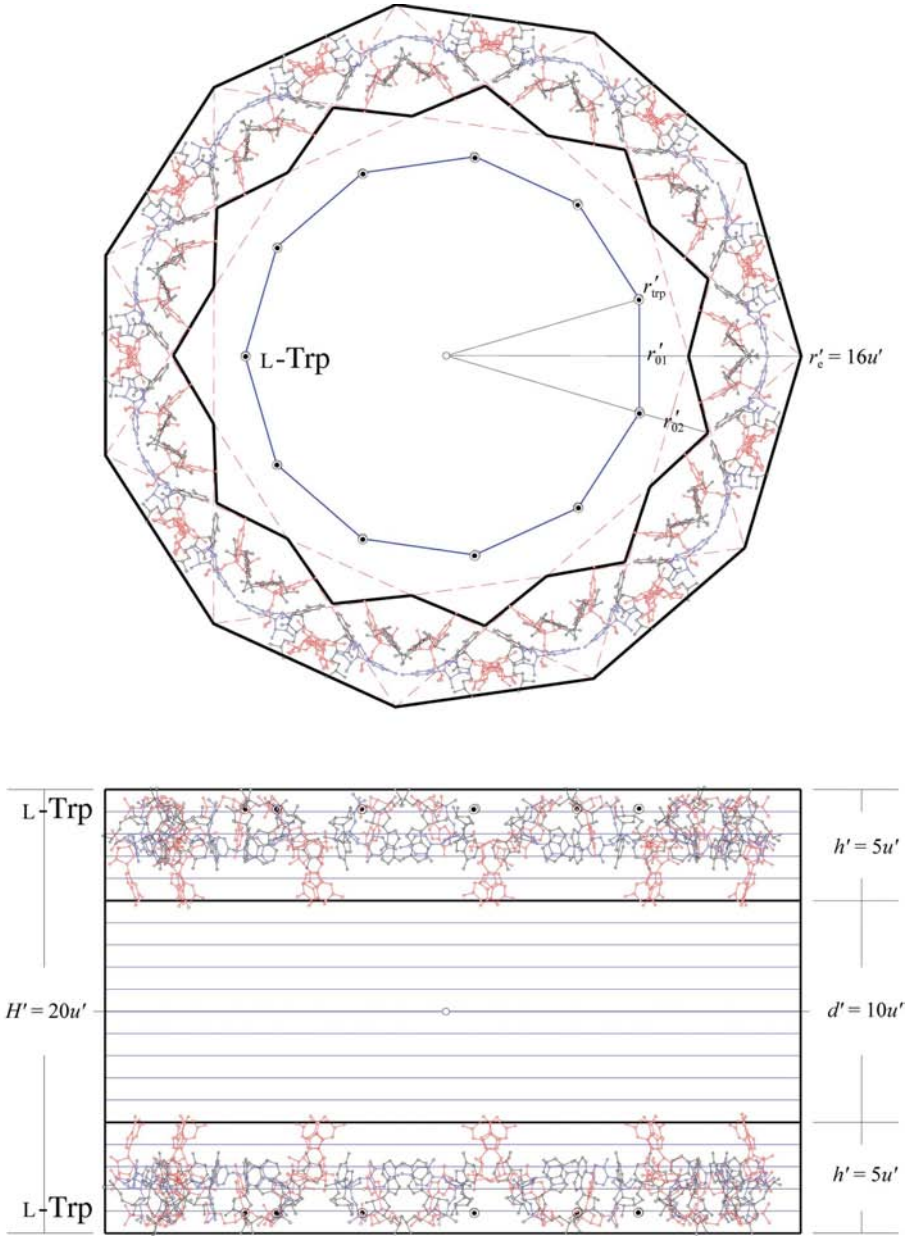

Figure 8
The RNA subsystem of the pair of TRAP 11-mers shown in Fig. 6 has a similar architecture to the protein if one adopts a unit of length u' related to the RNA envelope and scaled by a factor $1/\mu_{\{11/2\}}$ with respect to u in the previous figure. The fitting in height of the new structural parameters expressed as multiples of u' is less good than $h' = 6u$, as indicated in Fig. 7.

Of course, in this cooperative example only part of the molecule fits into a symmetry characterization of the molecular form, but still allows a meaningful description in terms of polygrammal symmetry and of a 16-fold isometric molecular form lattice with $c = a = u$.

5. Conclusions

The proteins presented in these three parts devoted to strongly correlated biomacromolecular structures demonstrate that an incredible variation in form and function can be compatible with a same small set of simple architectural laws based on symmetry. In addition to the normal point-group symmetry observed in a large number of molecules and often considered to be more or less accidental, axial-symmetric proteins reveal the importance of polygrammal scaling which relates the outside with the inside world and involves not only the molecular form but also a number of residues at the boundaries. It is surprising that it is possible to characterize the molecular form in terms of a molecular form lattice that is left invariant by the point group and by the polygrammal scalings. This lattice depends on a single parameter, instead of the two required for a generic polygonal lattice: the parameter a in the plane of rotation and the parameter c along the axis. Under these circumstances, one can hardly speak of an accidental property.

The general validity of this integral lattice irreducibility of the biomacromolecules discussed so far allows the discovery of a corresponding tendency in the frequency distribution of a large number of axial-symmetric crystals, hexagonal and tetragonal (de Gelder & Janner, 2004). The observed rational ratio c^2/a^2 eventually leads to the concept of the integral lattice (Janner, 2004) and to conjecture the tendency to reduce the number of free parameters (for a given class of phenomena) as a general principle in nature. In the proteins analyzed in this series of articles, it is the isometric lattice ($c = a$), which occurs most often. This is also true for the proteins in complex with RNA/DNA presented in this last part of the series. The own lattice parameters of the protein and of the RNA/DNA subsystems, respectively, are symmetry-related and in fact the structural para-

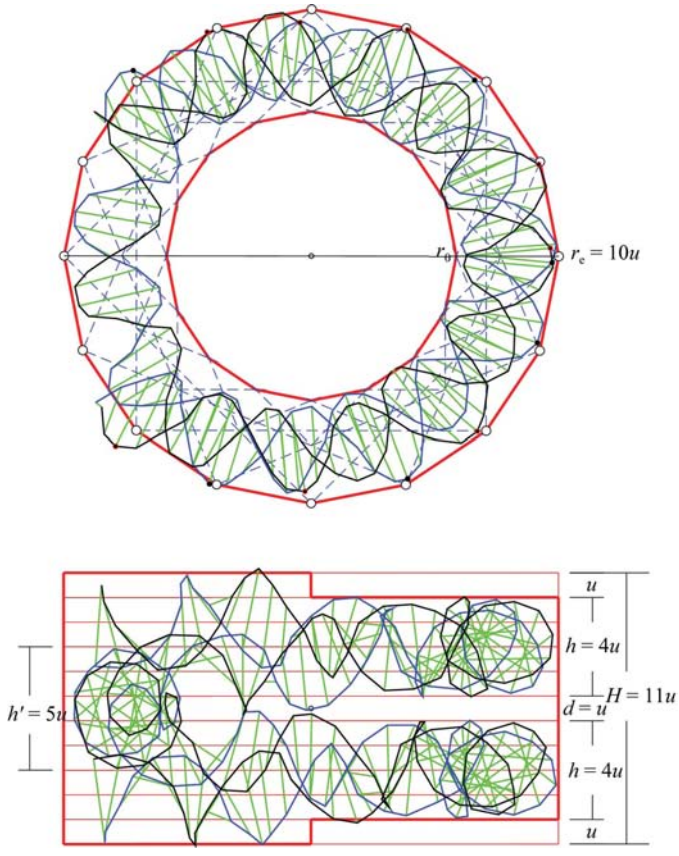

Figure 9
The DNA double helix in the nucleosome core particle is shown by the backbone P positions, with the paired bases represented by a connecting line segment. The periodicity of the superhelix wrapped around the histone protein (see Fig. 10) induces an approximate 16-fold rotational symmetry, indicated by empty circles, the corresponding P positions being marked by black dots. The hole follows from the 16-gon of the envelope by two successive {8/2} star octagons, as indicated by dashed lines. The twofold symmetry of the histone octamer in turn induces a dyadic symmetry for the DNA superhelix, which has an upper, a central and a lower part. The distance between the upper and the lower superhelix defines the unit parameter u, which allows the expression of all the other lengths indicated in multiples of u.

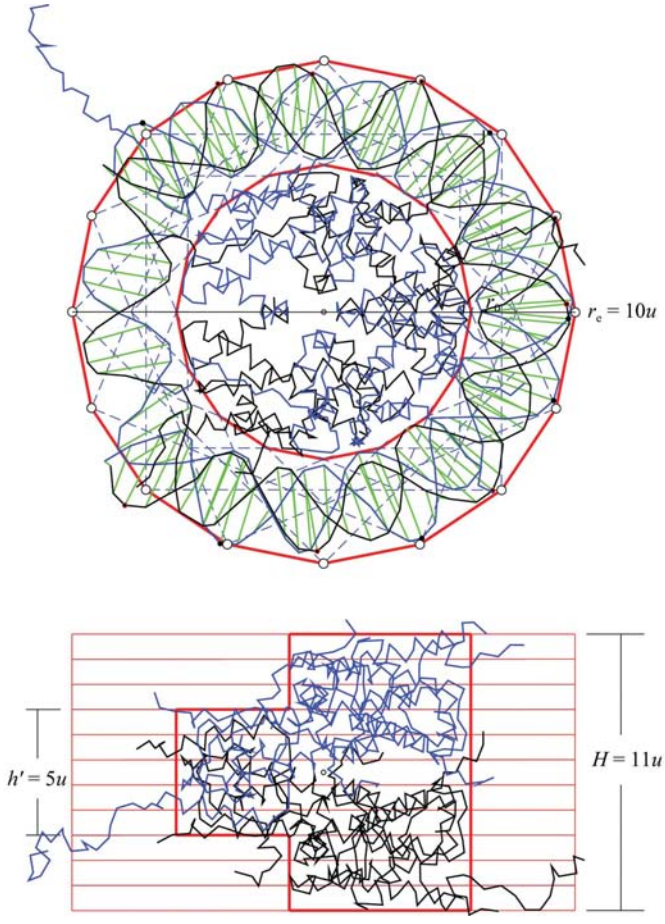

Figure 10
The histone protein is shown in relation to the DNA superhelix as in Fig. 9. The envelope of the histone fits into the hole of the DNA. The total height $H = 11u$ of the histone is equal to that of the DNA. The height $h' = 5u$ is the same as that for the corresponding central part of the superhelix.

meter of the whole is a single one. A similar analysis of the large family of nucleic acids is still missing: only a few cases have been published and the results obtained for the helical conformations have not yet been published.

What is now needed is a proper crystallographic characterization of integral lattices and the application of the new insights to structure determination, together with an appropriate statistical analysis. For biomacromolecules and for crystals the fundamental problem of understanding the physico-chemical basis of the structural features observed still remains.

The careful reading by the Editor and his pertinent suggestions have greatly helped the author to improve the presentation of the three parts of this work.

References

Antson, A. A., Dodson, E. J., Dodson, G., Greaves, R. B., Chen, X. & Gollnick, P. (1999). *Nature (London)*, **401**, 235–242.
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.
Gelder, R. de & Janner, A. (2004). *Acta Cryst.* A**60** s218.
Gonzalez, A., Nave, C. & Marvin, D. A. (1995). *Acta Cryst.* D**51**, 792–804.
Janner, A. (2001). *Cryst. Eng.* **4**, 119–129.
Janner, A. (2004). *Acta Cryst.* A**60**, 198–200.
Janner, A. (2005*a*). *Acta Cryst.* D**61**, 247–255.
Janner, A. (2005*b*). *Acta Cryst.* D**61**, 256–268.
Liu, D. J. & Day, L. A. (1994). *Science*, **265**, 671–674.
Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). *Nature (London)*, **389**, 251–260.
Marvin, D. A. (1989). *Int. J. Biol. Macromol.* **11**, 159–164.
Marvin, D. A. (1990). *Int. J. Biol. Macromol.* **12**, 125–138.
Marvin, D. A., Hale, R. D. & Nave, C. (1994). *J. Mol. Biol.* **235**, 260–286.
Richmond, T. J. (2001). *Chimia*, **55**, 487–492.
Thore, S., Mayer, C., Sauter, C., Weeks, S. & Suck, D. (2003). *J. Biol. Chem.* **278**, 1239–1247.
Welsh, L. C., Symmons, M. F. & Marvin, D. A. (2000). *Acta Cryst.* D**56**, 137–150.