What Curves α-Solenoids?

EVIDENCE FOR AN α-HELICAL TOROID STRUCTURE OF Rpn1 AND Rpn2 PROTEINS OF THE 26 S PROTEASOME *

Andrey V. Kajava ‡

From the Center for Molecular Modeling, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892-5626

The α-helical solenoid proteins adopt a variety of elongated curved structures. They have been examined to identify the interactions that determine their curvature. A sequence pattern characteristic for strongly curved α-helical solenoids has been constructed and was found to match protein sequences containing the proteasome/cyclosome repeats. Based on this, a structural model of the repeat-containing domains of the Rpn1/S2 and Rpn2/S1 proteins, which represent the largest subunits of the 26 S proteasome, has been proposed. The model has a novel architecture resembling an α-helical toroid. Molecular modeling shows that these toroids have a central pore that would allow passage of an unfolded protein substrate through it. This implies that the Rpn1 and Rpn2 toroids are aligned along the common axial pores of the ATPase hexamer and form an “antechamber” of the 26 S proteasome. The proposed quaternary structure agrees with the available experimental data. It is suggested that the function of this antechamber is assistance to the ATPases in the unfolding of protein substrates prior to proteolysis. An evolutionary link between the PC repeat-containing proteins and tetratricopeptide repeat proteins is proposed.

Recently, a new class of protein structures that contain a continuous superhelical or solenoid arrangement of repeating structural units has emerged (1, 2). The solenoid proteins with their elongated and repetitive structure contrast the structure of globular proteins, which fold in more complex ways. The repeating units of the solenoids consist of one to several segments of secondary structure, among which are α-helices, β-strands, and 3_10-helices. There are purely α-helical or β-structural solenoid proteins and others having a mixture of the secondary structures.

The solenoid proteins adopt a variety of non-globular shapes ranging from straight rods to curved shapes resembling a horseshoe. In many instances, the curvature of these proteins predetermines the repertoires of molecules that can bind to them. For example, the α/β-solenoids of leucine-rich repeat (LRR) proteins have a horseshoe shape, and their concave surfaces can embrace certain globular protein domains (3). On the other hand, the β-solenoids have a rod-like shape, and their binding counterparts themselves may require concave surfaces to bind the β-helices (2). Therefore, understanding of the interactions that control the curvature is important for correct prediction of the structure and function of the solenoid proteins. For some of them, the reasons for the curvature are well established. Indeed, the pronounced curvature of the α/β-solenoids formed by LRRs is the result of different diameters of the α-helices and β-strands when packed in a parallel arrangement (4, 5). On the other hand, the purely β-helical structures do not have any curvature because their intercoil distance is determined by the distance between hydrogen-bonded β-strands, which remains the same on all sides of the solenoid (2). The highest variety of shapes is provided by α-helical solenoids. They can be straight (6), slightly curved (7), or fold into horseshoe-like structures (8, 9). However, the interactions that control their shape are still poorly understood.

In this work, I systematically analyzed known three-dimensional structures of the α-helical solenoid proteins to identify interactions and amino acid sequence patterns that lead to their curvature. This analysis revealed two modes of interactions that are common for the curved structures as follows: (i) the helices that form the concave surface have small Ala, Ser, and Gly residues in certain positions, and these residues reduce their interhelical distances; (ii) the “concave” helices also have large hydrophilic or aromatic side chains that intercalate between the helices of the convex surface and increase their interhelical distances. Reduction of the interhelical distances on one side of the solenoid and their increase on the opposite side results in the curved superhelical structure.

To test the predictive power of the principles described above, I constructed a sequence pattern characteristic for the strongly curved α-helical solenoids and ran this pattern against the sequence data bases. It turned out that this sequence pattern matches several proteins containing so-called proteasome/cyclosome (PC) repeats (10). The members of this protein family include the two largest subunits of the 26 S proteasome and a protein of anaphase-promoting complex (APC). The 26 S proteasome is a major intracellular protease in eukaryotes (11–15). Because of its central role in the ubiquitin-mediated proteolytic pathway, this large ATP-dependent enzyme is involved in a wide variety of cellular processes, including cell cycle regulation, antigen presentation, and inflammation. The three-dimensional structures and functions of its protein subunits containing PC repeats are not known. The match between the sequence pattern of the curved α-solenoid and sequences of the PC repeat-containing proteins leads to a model of the three-dimensional structure for the Rpn1 and Rpn2 proteins of the 26 S proteasome. In accordance with this prediction, the repeat-
itive parts of these proteins fold into a structure with a novel architecture resembling an α-helical toroid. The Rpn1 and Rpn2 toroidal structures have a central pore of about 20 Å in diameter that would allow the passage of a peptide substrate. The axial symmetry and the central pore of the Rpn1 and Rpn2 structures suggest that they are aligned along the axial pore of the ATPase hexamer and the 20 S core particle and form an “antechamber” of the 26 S proteasome. The toroidal structures and their arrangement within the proteasome are consistent with the data of electron microscopy (16, 17) and protein subunit interaction maps (13, 18–21). It is proposed that the function of this antechamber is to assist the ATPases in the unfolding of protein substrates prior to their proteolysis.

The prediction of the three-dimensional structures of Rpn1 and Rpn2 proteins and the quaternary structure of the base module of the 19 S regulatory particle and its docking onto the 20 S core particle all contribute to a better understanding of regulation in the ubiquitin-proteasome system and open possibilities for a number of experimental tests.

MATERIALS AND METHODS

Sequence Analysis—PatternFind server (www.isrec.isb-sib.ch/software/PATFND_form.html) was used to search for motifs that are characteristic for the curved α-solenoids.

Molecular Modeling—An αα-hairpin from the TPR domain of Hop protein (22) was used as an initial template to model a PC repetitive unit. The model was generated using molecular modeling modules implemented in Insight II (23). The interhelical loops were modeled by using a homology module of the Insight II program that allows search for an appropriate loop conformation in the data base of known structures. The resulting structures were energy-minimized by applying 100 steps of the steepest descent algorithm and 600 steps of the conjugate gradient algorithm. The consistent valence force field (24) and the curvature is well determined for the hairpin (25) and shown to meet all standard requirements for stereochemistry, as orientations of the internal side chains belonging to the inner helix (ii) and outer helix (oi). The decrease of the interlayer distance (d) can increase the curvature of the α-solenoid at given d_{out} and d_{in}.

RESULTS AND DISCUSSION

Interactions That Determine the Curvature of α-Helical Solenoid Proteins—To understand the origin of the curvature in α-solenoids, I analyzed systematically their three-dimensional structures. There are over 20 known three-dimensional structures of α-solenoids (1, 2). However, not all are suitable for the present analysis. The curvature is well determined for the structures formed by the uniform packing of several repeated units lying in the same plane. The majority of such structures is represented by the α-solenoids containing two-α-helix repetitive units (αα-solenoids). They are formed by αα-hairpins stacked in a parallel manner, so that their overall structures are almost “flat” and have pronounced curves (Fig. 1). These proteins were the focus of my study. Among them were proteins containing HEAT repeats (9, 27, 28), tetratricopeptide repeats (TPR) (29–31), and a recently determined MalT protein (32). A leucine-rich repeat variant protein having an αβ-solenoid structure (7) was also included in the analysis. In addition, I considered the oligomeric structure of the F_{1}F_{0}-ATP synthase (33) that is formed by several individual αα-hairpin subunits stacked in a consecutive array so that they resemble curved αα-solenoids but lack covalent connections between the hairpins.

Among the structures excluded from the analysis were α-so- lenoids, which have irregular structures without apparent curva- tures (34, 35), and twisted solenoids with three α-helices in one repeat (α/α/α-solenoids), which have little, if any, curvature (1, 2).

The leucine repeat variant (LRV) protein (7) has the smallest curvature among the analyzed structures (Fig. 2A). Its repeated structural unit is characterized by a hairpin formed by an α-helix and a 3_{10}-helix. The α- and 3_{10}-helices of each hairpin are parallel to their helix counterparts in a neighboring repeat. This creates a structure composed of a double layer of helices, with the α- and 3_{10}-helices forming the inner (concave) and outer (convex) faces, respectively. (Fig. 2A). This is a surprising arrangement because the average diameter of the α-helix is larger than the diameter of the 3_{10}-helix, and one can expect an opposite curvature. The observed curvature is partially explained by the fact that the 3_{10}-helices do not have “knob-into-hole” hydrophobic interactions with each other. As a result, the packing between the outer 3_{10}-helices is looser than between the inner α-helices. On the other hand, the analysis revealed that most of the α-helices have small Ala residues (Fig. 2A) at positions that determine their interhelical distance within the inner layer (denoted as il positions in Fig. 1). In addition, it was noticed that the α-helix has a conserved Arg residue among residues, which form the hydrophobic interior of the solenoid (il positions in Fig. 1). This Arg, intercalates between the 3_{10}-helices, and its charged guanidinium group escapes the hydrophobic environment (Fig. 2A) and forms an ionic bond with an Asp or Glu side chain. This suggests that the intercalation of the Arg residues increases the distances (d_{out}) between the 3_{10}-helices, and the small Ala residues decrease the distances (d_{in}) between the α-helices, and these are the main factors governing the curvature of LRV protein.

Several α-solenoid structures with well defined curvature belong to a family of HEAT repeat-containing proteins (27). Among them are importin-β (28) and PR65/A subunit of protein phosphatase 2A (9) (Fig. 2B). Analysis of these structures showed that a highly conserved Pro residue in the middle of the outer α-helix induces formation of a π-helical turn, which makes the α-helix wider. The inner α-helix is also distorted in the middle by a couple of 3_{10} conformations that makes the inner α-helix thinner. Forces generated by packing of the helices probably induce the distortion of the inner helix. The difference in the diameters of the inner and outer helices causes the curvature. The side chains also contribute to the curvature. Similarly to the LRV protein, the HEAT proteins have conserved Arg/Lys residues within the apolar cluster of the inner helix, which intercalates between the outer helices. They also contain small Ala residues in il positions that determine the interhelical distances within the inner layer (Fig. 2B).

The repetitive units of the curved superhelical structure of TPR proteins consist of regular α-helices (Fig. 2C). The analysis shows that TPR proteins, analogous to LRV and HEAT
proteins, frequently have small Ala and Ser residues in several $il$ positions controlling $d_{in}$ distance (Fig. 2C). The $ii$ positions of the TPR inner helices have preference for bulky aromatic Trp and Tyr residues that intercalate between the outer helices similarly to the Arg/Lys of LRV and HEAT protein structures. Thus, the decrease of $d_{in}$ distance due to the location of small side chains in the $il$ positions and increase of $d_{out}$ distance due to the intercalation of Trp or Tyr protruded from the $ii$ positions are probably the reasons for the curvature. Small Ala residues in the $oi$ positions of the outer helix allow outer and inner $\alpha$-helical layers to be closer to each other (Fig. 2C), and this can increase the curvature.

Recently, an $a/a$-solenoid structure of the transcription factor MalT from Escherichia coli has been determined (32). It has a strong curved combination with a small twist (Fig. 2D). Again, small Ala often occupies two $il$ positions of the inner helix, whereas one of the three $ii$ positions always has Gln, Arg, or Tyr residues intercalating between the outer helices. Similarly to TPR proteins, MalT protein also has Ala in two $oi$ positions (Fig. 1) and, in its turn, reduced distance between the helical layers.

The three-dimensional structure of the transmembrane oligomer of the $H^+$-transporting ATP synthase from E. coli has been modeled (33) by using the solution structure of monomeric subunit “c” (36) and several intersubunit distances derived from cross-linking experiments (Fig. 2E). It is a hollow cylinder formed by 12 individual $a/a$-hairpin subunits stacked in a consecutive array so that their packing is identical to the one in the curved $a/a$-solenoids with the only difference that in the solenoids these hairpins are covalently linked to each other. The analysis shows that the inner helices are packed tighter than the outer ones due to a high content of Ala and Gly residues in almost all $il$ positions. There are also residues with small side chains in $ii$ and $oi$ positions that reduce the interlayer distance and contribute to the increase of the curvature.
Thus, despite the fact that the curvature of some \( \alpha \)-helical solenoids is partially driven by specific interactions (e.g. \( \pi \) and 3\( \gamma \)-turns of HEAT proteins), they share common structural features as follows: first, frequent occurrence of small Ala, Ser, and Gly residues in \( il \) positions that reduces the distance between the inner helices; and second, an intercalation of large hydrophilic or aromatic side chains from the inner helix to the space between the outer helices that increases the distance between outer helices. It is worth mentioning that a similar conclusion has been made previously about the origin of the curvature of the ankyrin repeat proteins (37, 38). Although these proteins are not completely \( \alpha \)-helical, the major part of their structure consists of pairs of antiparallel \( \alpha \)-helices stacked side by side. It was suggested that conserved non-polar side chains associated with the inner helices tend to be smaller in volume than those associated with the outer helices. At the same time, in accordance with our conclusion, a fragment of clathrin, a straight rod-like \( \alpha \)/\( \delta \)-solenoid (6), does not have small residues between the helices. In contrast, its hydrophobic core has an unusually high number of aromatic residues.

The discovered principles of the solenoid curvature, in turn, propose amino acid substitutions that can change the original curvature of the \( \alpha \)-solenoid proteins. These mutations can be used to understand the importance of the curvature for proper biological function of such proteins. It also allows the design of novel \( \alpha \)-solenoids with a desirable curvature.

Proteins with Proteasome/Cyclosome Repeats Contain Sequence Motifs Characteristic for the Curved \( \alpha \)-Solenoids—Understanding of the molecular basis of the \( \alpha \)-solenoid curvature opens a possibility for prediction of yet unknown curved \( \alpha \)-solenoid structures. In order to search for such structures, the three-dimensional information about the location of small residues in \( il \) positions of the curved \( \alpha \)/\( \delta \)-solenoids was translated into a sequence motif. Constructing the sequence pattern, I also took into consideration the fact that the known \( \alpha \)/\( \delta \)-solenoids have repetitive sequences with the repeat length ranging between 30 and 40 residues (1, 39). Therefore, the Ala/Gly-rich motif was repeated 3 times and looked as follows in PROSITE format: \( [AG]^{-p}(2)-[LIMVF]-[AG]^{-p}-[AG]-[LVMIF]-[AG]^{-p}(2)-[LIMVF]-[AG]^{-p}-[AG]-[LVMIF]-[AG]^{-p}(2)-[LIMVF]-[AG]^{-p}-[AG]-[LVMIF] \) (where \( p \) is any residue except proline). It turned out, that this sequence pattern matches several proteins containing PC repeats (10). The best characterized members of this protein family include the two largest subunits of the 26 S proteasome and a protein of anaphase-promoting complex. I further examined the proteins of the 26 S proteasome termed Rpn1 and Rpn2 in the Saccharomyces cerevisiae nomenclature or S2 and S1 in the mammalian one using sequence analysis and molecular modeling in order to find additional support for the hypothesis that Rpn1/S2 and Rpn2/S1 have strongly curved \( \alpha \)-solenoid structures.

Structural Models of Rpn1 and Rpn2 Subunits of the 26 S Proteasome—The 26 S proteasome is formed by the association of two subcomplexes, a 20 S proteolytic core and a regulatory 19 S particle that caps the core at both ends (11, 13, 14). Although the three-dimensional structure and enzymatic mechanisms of the 20 S particle are known (40, 41), the structure of the 19 S cap is still unknown, and its functions are less defined.

The two largest proteins of the 19 S particle called Rpn1/S2 and Rpn2/S1 have tandem arrays of similar 34–40 residue repeats within their C-terminal halves (10). The structures and functions of these repetitive domains are not known. Previously, it was suggested (10) that the three-dimensional structures of Rpn1/S2 and Rpn2/S1 are similar to the one of the LRR proteins. However, significant differences in the lengths and sequence patterns of LRR and PC repeats and an inability of the LRR-like models to explain the conservation of Gly/Ala in certain positions of PC repeats cast doubt on the validity of this model. On the other hand, a number of successful predictions of proteins with repeats below 40 residues suggest that structural prediction for proteins with repeats can be simpler and more reliable than prediction for globular proteins (39, 42–44). This, together with the identification of the sequence motif characteristic for the curved \( \alpha \)-solenoids in the PC repeat-containing proteins, prompted a more comprehensive sequence and modeling analysis of the repetitive regions of Rpn1/S2 and Rpn2/S1 proteins.

Sequences of Rpn1 and Rpn2 proteins from yeast (993 and 945 residues) share significant similarities (–20% identity). Both contain nine PC repeats of 35–40 residues (10), which are preceded by two more-divergent repeats (Fig. 3A). The arrays of the repeats are flanked by aperiodic N- and C-terminal sequences of unknown but, most probably, globular structures. The repetitive region of Rpn1 is interrupted in the middle by a hydrophilic negatively charged region of 100 residues.

Alignment of the repeats showed that they consist of two highly conserved parts separated by a variable region: the first part is characterized by an alternating pattern of large aliphatic residues and Gly or Ala residues, and the second one by several conserved positions of large aliphatic residues (Fig. 3B). As described above, the first region is enriched in Gly and Ala residues and matches the sequence motif characteristic for the curved \( \alpha \)-solenoid. This suggests that this region corresponds to the inner \( \alpha \)-helix, and the small Gly and Ala residues are clustered on their interfaces. Current state-of-the-art methods of the secondary structure prediction, for example, PSIPRED (45), also predict an \( \alpha \)-helix for this region of the repeat. It is worth mentioning, that this conclusion differs from the previous result of the secondary structure predictions (10, 46), which suggested a \( \beta \)-structural conformation for this region. The second part of the repeat contains a pattern of hydrophilic and
aliphatic residues, which, as it was noticed before (10), has a periodicity of amphiphilic α-helices. It can be assigned to the outer helix of the solenoid. Two regions of variable length located in the middle and end of the repeat can be assigned to the loops connecting the inner and outer α-helices (Fig. 3B).

Comparison of the PC repeats with those of the known α/ω-solenoid structures revealed similarities between PC and tetra-tricopeptide repeats (Fig. 3B). TPR and PC repeats have similar length and consensus sequences. Furthermore, both types of proteins are involved in the same protein complexes (47) or biological pathways (30). This suggests that PC repeat-containing proteins and TPR proteins may be evolutionarily related to each other. An α/ω-hairpin of TPR was used as a template to model a PC repetitive unit. In the modeled unit, the conserved aliphatic side chains are directed inside and form the hydrophobic core (Fig. 3C), whereas variable and mostly hydrophilic residues are exposed to the solvent.

The next step was to estimate the curvature of Rpn1 and Rpn2 solenoids. Their curvature should be larger than that in the TPR structures because PC repeats have smaller residues in positions that control the distance between the inner helices. In addition, all known PC repeat proteins have a limited and almost identical number of repeats. Such constancy in the repeat number is unusual for “open” repetitive structures but can be explained by structures where the first and the last repeat interact with each other to “close” the structure. This limited and constant number of repeats suggests that the array of PC repeats folds into a toroid rather than solenoid structure. It is worth mentioning that although Rpn1, Rpn2, and the other PC-containing proteins have 8–9 well recognized repeats, they also have two additional, more divergent repeats (Fig. 3A). The current theoretical analysis cannot unambiguously predict whether the toroid contains 9 or 11 repeat units. Clarification of this point will need subsequent experimental tests.

In order to choose optimal interhelical distances, several structures of three α/ω-hairpins stacked in a consecutive array have been generated. In these structures, the initial distance between the inner helices was identical or slightly greater than the corresponding 8-Å distance of the ATP synthase oligomer (33). After a session of energy minimization (100 steps of the steepest descent algorithm and 500 steps of the conjugate gradient algorithm implemented in Insight II (23)), these structures converged to one with a distance of about 9.5 Å. This distance was chosen to model the toroids of Rpn1 and Rpn2 proteins. After the modeling of interhelical loops the structures were energy-minimized and were shown to meet all standard requirements for protein structures.

Fig. 4A displays a model of the PC repeat-containing region of Rpn2 from yeast. Nine α/ω-hairpins corresponding to the nine PC repeats form a super-helix where the first and the last unit interacts to close the asymmetrical structure. The model explains the residue conservations that are observed in the sequence alignments of the PC repeats (Fig. 3B). Indeed, the conserved aliphatic residues of both inner and outer helices form the hydrophobic core of the toroid. Small Ala and Gly residues in the il positions allow a tighter packing of the inner helices compared with the outer helices. This leads to the strong curvature of the toroid. Similarly to the known TPR protein structures, the inner α-helix of the model is ended by a glycine-specific αi-conformation that provides a short crossover to the outer helix. The occurrence of this αi-conformation can explain why PC repeats frequently have a Gly residue in this position (Fig. 3B). The model has a central pore of about 20 Å in diameter. The wall of the pore is hydrophobic with a ring of positively charged residues bordering the pore (Fig. 4B).

It is worth mentioning that although Rpn1, Rpn2, and the other PC-containing proteins have 8–9 well recognized repeats, they also have two additional, more divergent repeats (Fig. 3A). The current theoretical analysis cannot unambiguously predict whether the toroid contains 9 or 11 repeat units. Clarification of this point will need subsequent experimental tests.

**Modeling of the Quaternary Structure of the 26 S Proteasome**—In the 26 S proteasome, the 19 S regulatory particles associate with either one or both ends of the 20 S proteolytic
core complex (11, 16). The crystal structure of the 20 S core and its main enzymatic mechanisms have been determined (40, 41). However, the overall organization and function of the 19 S regulatory particle are less defined. The 19 S particle is composed of two structurally distinct modules called the base and the lid (49). The base contains eight subunits (six ATPases of the AAA family and two large subunits Rpn1 and Rpn2) and is connected with the lid by Rpn10 protein (49). Like other members of the AAA family, the ATPases of the base should assemble into a hexameric ring that forms the interface of the 19 S complex with the 20 S core (50). A reliable structural model of the ATPase hexamer can be obtained by using the known crystal structures of the other ATPase rings (51). Together with the prediction of the repetitive structures of Rpn1 and Rpn2, this model opens a possibility to build the three-dimensional structure of the base and to dock it to the 20 S core particle.

The main part of the ATPase hexamer, which is involved in ATP binding and formation of the hexameric core, has been modeled using the crystal structure of the C-terminal part (residues 209–458) of the AAA ATPase p97 (51). These parts have ~40% of sequence identity that gives confidence to the modeling. On the other hand, the first ~100 residues of the 19 S ATPases that form the hexamer periphery do not have good templates for the modeling. However, even a crude model of this part would be instrumental to fit the ATPase/Rpn1/Rpn2 structure into the known electron microscopy density. An α-helical coiled-coil hairpin structure of seryl-tRNA synthase from Thermus thermophilus (52) was used to model these regions because most of them have two adjacent α-helical coiled-coil motifs (53). The next ~100 residues have ~20% of sequence identity to the second N-terminal domain (residues 112–190) of the ATPase p97 (51), and this four-stranded β-barrel was used for the modeling. Thus, a final model of the 19 S ATPase represents the N-terminal α-helical coiled-coil hairpin followed by the small β-barrel domain and the core domain of the hexamer. It is important to mention that the ATPases of the proteasome are related but nonetheless distinct from each other. Despite this, I modeled only one of them (Rpt1) and used it to construct a 6-fold symmetrical hexamer. Because the goal of the ATPase modeling in this work was to fit the structural models into the electron microscopy density, such approximation was appropriate.

In accordance with the results of the modeling, the ATPase hexamer and Rpn1 and Rpn2 toroids have the axial symmetry and the central pores. If this is true, then the ATPase hexamer and the toroids are probably all aligned along the axial pore of the 20 S core particle. Thus, the proposed structural model of the 19 S base placed on the top of the known 20 S core particle (40) has a common central pore of about 20 Å in diameter that would allow passage of a peptide substrate into the proteolytic chamber of the 20 S core (Fig. 5). The model of the complex nicely fits into the electron microscopy density of the 26 S proteasome (16, 17). It can also explain an observed shape of an incomplete form of the proteasome with only the base of the 19 S particle on the top of the 20 S core (49). This form of the proteasome, similar to the suggested model, has a well recognized cylinder of a smaller diameter, nested on the cylindrical holder of a larger diameter. However, the thin cylinder of the model is longer than the one obtained by averaging the complexes observed on the electron micrographs (49) (Fig. 5). This can be explained by a limited resolution of the micrographs and by a possibility that the average image includes a mixture of complexes with one and with two Rpn1 and Rpn2 molecules. Indeed, the expected volume of the base particle can be calculated from the total mass of the subunits (Rpn1, Rpn2, and Rpt 1–6), and this calculation gives a larger volume than the one obtained by the averaging of the images.

![Fig. 5. Superposition of the observed EM density of the 26 S proteasome (16) and the three-dimensional structure suggested for a complex between the 20 S core particle, ATPase hexamer, and Rpn1 and Rpn2 toroids. The crystal structure of the 20 S core particle (40) is shown in magenta; structural model of the ATPase hexamer is shown in green and two stacked toroids of Rpn1 and Rpn2 are shown in blue. In this figure, the upper toroid consists of 9 repeats, whereas the buried toroid has 11 repeats to visualize the relationship between their sizes. However, in the proteasome, both toroids assumed to have the same number of repetitive units. A red contour line, corresponds to the EM density of the base of the 19 S regulatory particle (49).](image-url)
20 S particle, was chosen based on the arrangement observed in the structure of the ATP-dependent HslVU protease (57). The peripheral β-structural and coiled-coil domains of ATPases interact with each other forming a nest for Rpn1/Rpn2 toroids (Fig. 5). As mentioned above, there are two additional “covert” repeats in the sequences of Rpn1 and Rpn2 proteins, and there is a probability for these proteins to fold into wider 11-repeat toroids (~60 Å in diameter versus ~55 Å of the 9-repeat toroids). A model of the 11-repeat toroid was also built, and its docking demonstrated that it also fit into the ATPase “nest” (see a buried toroid in Fig. 5).

The largest subunit of another important protein complex of the cell cycle, called cyclosome or anaphase promoting complex (47, 58), also has PC repeats and can be modeled similarly to Rpn1 and Rpn2 subunits of the proteasome. Given the fact that amino acid sequences of several other APC subunits are similar to the sequences of the TPR proteins with known three-dimensional structural elements (30), the three-dimensional structures of almost all cyclosome subunits can be modeled. This may open a possibility for reconstruction of the cyclosome quaternary structure and understanding of its function.

Functional Implication of Rpn1 and Rpn2 Subunits—The structural models of Rpn1 and Rpn2 and their location within the 26 S proteasome suggested that they, together with the ATPase ring, unfold proteins assigned for proteolysis. Indeed, the walls of the pores covered by hydrophobic residues provide an energetically favorable environment for unfolded protein substrates. The pores do not have enough room for partially unfolded protein structures, and this suggests that Rpn1/Rpn2 dimers unfold proteins starting from their N or C terminus. An intriguing detail of Rpn2 is a ring of positively charged residues bordering the pore. These positively charged residues are conserved among Rpn2 proteins from different organisms and most probably are important for the function. For example, the positively charged ring can favor protein unfolding starting from the negatively charged C terminus as it was suggested for the similar AAA ATPase complex (59). Further experiments can test the hypothesis about the functional importance of the charged ring of Rpn2.

In connection with the function suggested for these proteins, it is interesting to compare the 26 S proteasome with its bacterial homologue, the ATP-dependent protease HsIVU (60). The HsIVU does not have proteins similar to Rpn1 and Rpn2; however, it is fully competent in degradation and thus, presumably, in unfolding of protein substrates. The comparison shows that, instead, the HsIVU ATPase ring has six characteristic distal domains I, which protrude from it and form an open-work antechamber of the protease complex (57, 61). It was proposed previously (57, 62) that these domains I unfold the substrate polypeptide. The structures formed by bacterial domains I and by eukaryotic subunits Rpn1 and Rpn2 share similar locations and shapes, and this suggests that their functions are analogous. In both cases, once the polypeptide is clamped in the ATPase pore, concerted movements, promoted by ATP binding and hydrolysis, may facilitate further its unfolding and translocation.

Evolutionary Relationships of PC Repeat-containing Proteins—The sequence analysis and modeling of PC repeat-containing proteins suggest that they may be evolutionarily related to TPR-containing proteins. Indeed, TPR and PC repeats have the same length and similar consensus sequences. Some of these proteins are subunits of common protein complexes or are involved in the same cellular processes (47). Furthermore, the molecular modeling suggests that the overall structures of these proteins are similar in some differences in the curvatures of their α-helical solenoids. Previously, it was also noticed that TPR-containing proteins and 14-3-3 proteins share common structural and functional properties, despite their lack of obvious sequence similarities (63). Thus, these three protein classes may have a common ancestor diverged into different families. Similarly to the proposed evolutionary scenario of proteins containing LRRs (64), β-trefoil repeats (65), and ARM/HEAT repeats (66), the ancestor of 14-3-3, TPR, and PC repeat proteins could be a homo-multimer composed of one repeat.

Acknowledgments—I thank Dr. M. Dorée for initiation of this work; Drs. O. Coux, U. Blum-Trouvanziam, and P. J. Steinbach for helpful discussions and a critical reading of the manuscript; Dr. S. Ei leer for valuable suggestions; and Dr. D. Finley for providing data prior to publication.

REFERENCES

1. Groves, M. R., and Barford, D. (1999) Curr. Opin. Struct. Biol. 9, 383–389
2. Koe, B., and Kajava, A. V. (2000) Trends Biochem. Sci. 25, 509–515
3. Kobe, B., and Deisenhofer, J. (1993) Nature 366, 751–756
4. Kobe, B., and Deisenhofer, J. (1995) Nature 374, 183–186
5. Kobe, B., and Kajava, A. V. (2001) Curr. Opin. Struct. Biol. 11, 725–732
6. Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 282–291
7. Luo, J. M., Cejka, Z., Hearst, J. R., Kleinschmidt, J. A., and Baumeister, W. (1995) J. Mol. Biol. 234, 932–937
8. Mack, R., and Finley, D. (1991) EMBO Rep. 2, 821–828
9. Maqbool, S., and Staehlin, L. (1994) J. Biol. Chem. 269, 13689–13697
10. Masters, S. B., and Brunger, T. A. (1995) Science 268, 143–150
11. Matthews, B. W. (1968) J. Mol. Biol. 38, 455–470
12. Mikkelsen, T. J., and Finley, D. (1995) Nature Struct. Biol. 2, 319–328
13. Nakamura, Y., and Iwata, N. (1995) J. Mol. Biol. 249, 573–588
14. Nakamura, Y., and Iwata, N. (1996) J. Mol. Biol. 258, 105–115
15. Nakamura, Y., and Iwata, N. (1997) J. Mol. Biol. 265, 503–513
16. Nakamura, Y., and Iwata, N. (1998) J. Mol. Biol. 281, 611–621
17. Nakamura, Y., and Iwata, N. (1999) J. Mol. Biol. 289, 229–239
18. Nakamura, Y., and Iwata, N. (2000) J. Mol. Biol. 296, 115–130
19. Nakamura, Y., and Iwata, N. (2001) J. Mol. Biol. 307, 1077–1093
20. Nakamura, Y., and Iwata, N. (2002) J. Mol. Biol. 317, 967–977
21. Nakamura, Y., and Iwata, N. (2003) J. Mol. Biol. 335, 369–380
22. Nakamura, Y., and Iwata, N. (2004) J. Mol. Biol. 339, 229–240
23. Nakamura, Y., and Iwata, N. (2005) J. Mol. Biol. 347, 39–50
24. Nakamura, Y., and Iwata, N. (2006) J. Mol. Biol. 359, 622–635
