FEBS Letters 354 (1994) 45–49

FEBS 14695

Predicted secondary structure of the 20 S proteasome and model structure of the putative peptide channel

Andrei Lupas, Abraham J. Koster, Jochen Walz, Wolfgang Baumeister*

Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany

Received 21 September 1994

Abstract Secondary structure prediction has made great progress in recent years due to the incorporation of evolutionary information, and may be close to a point where (in combination with biochemical and low-resolution structural data) it can guide the modelling of tertiary structure in cases where no model building is possible by homology. Towards this goal it is important to gather information on the performance of prediction methods in advance of the publication of new structures. In anticipation of the soon-to-be-released structure of the 20 S proteasome from *Thermoplasma acidophilum*, we have applied several widely used secondary structure prediction methods to proteasome sequences and have attempted to model the putative channel in the outer proteasome rings (α-rings) based on the obtained predictions.

Key words: Proteasome; *Thermoplasma acidophilum*; Secondary structure prediction; Molecular modelling; Electron microscopy

1. Introduction

The 26 S proteasome is an ATP-dependent protease that is central to the ubiquitin pathway of protein degradation and has a highly conserved structure in all eukaryotes [1–3]. The core of the elongated 45 nm complex is formed by the barrel-shaped 20 S proteasome, an ATP-independent enzyme that degrades only unfolded proteins [4]. The 20 S proteasome, also known as the multicatalytic proteinase, consists of four stacked seven-membered rings that are formed by 14 related, but different, subunits. The subunits fall into two families, α-type subunits forming the outer and β-type subunits the inner rings [5]. The 20 S proteasome has not only been found in eukaryotes but also in the archaebacterium *Thermoplasma acidophilum*, where it is composed of only two proteins, α and β, that have given their names to the eukaryotic subunit families [6]. There is considerable circumstantial evidence (reviewed in [7]) that the proteolytic activity of proteasomes resides in an inner cavity formed by the β-subunit rings while the α-rings that close off the barrel control the access of substrate molecules to the inner proteolytic compartment. From electron microscopy studies with nanogold-labeled substrates (T. Wenzel and W. Baumeister, unpublished) it would appear that they achieve this by means of a central channel that operates primarily by size exclusion but may also facilitate the diffusion of peptide chains.

The 20 S proteasome of *Thermoplasma* has been crystallized [8] and the solution of its 3D structure is imminent. In an effort to obtain more information on the reliability of current secondary structure prediction methods and on the possibility of extrapolating from these to the tertiary structure, we have subjected proteasome sequences to some of the most widely used prediction methods and have generated a number of structural hypotheses for comparison with the actual structure.

2. Materials and methods

2.1. Sequence alignment

Proteasome sequences were identified in GenBank by progressive BLAST searches [9] using electronic mail (blast@ncbi.nlm.nih.gov) and were extracted using ENTREZ (National Center for Biotechnology Information, NIH, Bethesda MD, USA). The alignments were made in MACAW [10] using the BLOSUM80 matrix. Final manual adjustments were made to gap regions in MALIGNED [11].

2.2. Secondary structure prediction

All predictions are consensus predictions and were obtained separately for alignments of α- and β-type subunits. Predictions according to Chou and Fasman (CF) [12], and Garnier, Osguthorpe and Robson (GOR) [13] were obtained with the PEPTIDESTRUCTURE program of the Wisconsin package, version 7.3 (Genetics Computer Group Inc., Madison, WI, USA). Predictions according to Niermann and Kirschner (NK) [14], which represent a modified GOR method, were obtained from T. Niermann (University of Basel, Switzerland). Predictions according to Presnell and Cohen (PC) [15] were obtained with MacMatch [16] without a preset tertiary structure class. Turn predictions according to Presnell and Cohen (T) were obtained with MacMatch using an α/β pattern and were averaged over a window of 3 residues. Predictions according to Rost and Sander (RS) [17] were obtained by electronic mail from the PhD server at EMBL (PredictProtein@EMBL-Heidelberg.de) by submitting alignments of α- and β-type proteasome sequences. Predictions according to Jenny and Benner (JB) [18] were obtained by electronic mail (cbrg@inf.ethz.ch).

2.3. Molecular modelling

Models were built in Insight/Discover 2.7 (Biosym Inc., San Diego, USA). Secondary structure elements from verotoxin (Protein Data Bank: 1BOV) [19] were excised and recombined to yield the schematic architecture shown in Fig. 3. Connecting loops were extracted from PDB using the method of Jones and Thirup [20].

2.4. Electron microscopy and image processing

Complexes of *Thermoplasma acidophilum* α-subunits expressed in *Escherichia coli* were imaged by electron microscopy after negative staining (2% uranyl acetate) as well as after aurothioglucose embedding (2%). Top-view orientations of the α-rings were obtained after glow discharging the carbon film for 30 s and adding 0.5 M $CaCl_2$ to the suspension prior to staining or embedding. Images were recorded in digital format (512 × 512 pixels of 0.36 nm) using a slow scan CCD camera attached to a Philips CM 20 FEG transmission electron microscope (TEM) operating at 120 kV, at a magnification of 27,500 and an electron dose of about 1000 e/nm^2. Typically, series of 36 images were recorded containing several thousand α-ring top-views using CCD spot-scan imaging [21]. Images of negative stain preparations were taken at an underfocus of about 1 μm. Spot-scan series were taken of the aurothioglucose-embedded preparations at 0.3 μm defocus. The EM-SYSTEM software package [22] was used to extract the top-views from the images and to employ multivariate statistical analysis (MSA) followed by correlation averaging. The MSA clearly showed the sevenfold symmetry of the α-ring for both the negative stain and aurothioglu-

*Corresponding author. Fax: (49) (89) 8578 2641.

SSDI 0014-5793(94)01082-X

cose preparation. After seven-fold symmetrization, the average projection image was practically identical for both preparation techniques. The projection image of the negative stain preparation shown in Fig. 4 was obtained after averaging 3096 particles. Resolution according to the phase residual criterion is 1.6 nm.

3. Results and discussion

3.1. Secondary structure prediction

The incorporation of evolutionary information contained in the sequences of related proteins has recently led to an improvement in the performance of secondary structure prediction methods from an accuracy per residue of under 60% to over 70% [17]. Optimal performance typically requires a multiple alignment of more than 10 sequences with a full range of sequence identities from over 90% to the 'twilight zone' below 30%. In this sense, proteasome sequences are ideal prediction candidates since they form a large family of over 60 sequences with identities from > 90% to < 30% [23]. In addition, they fall

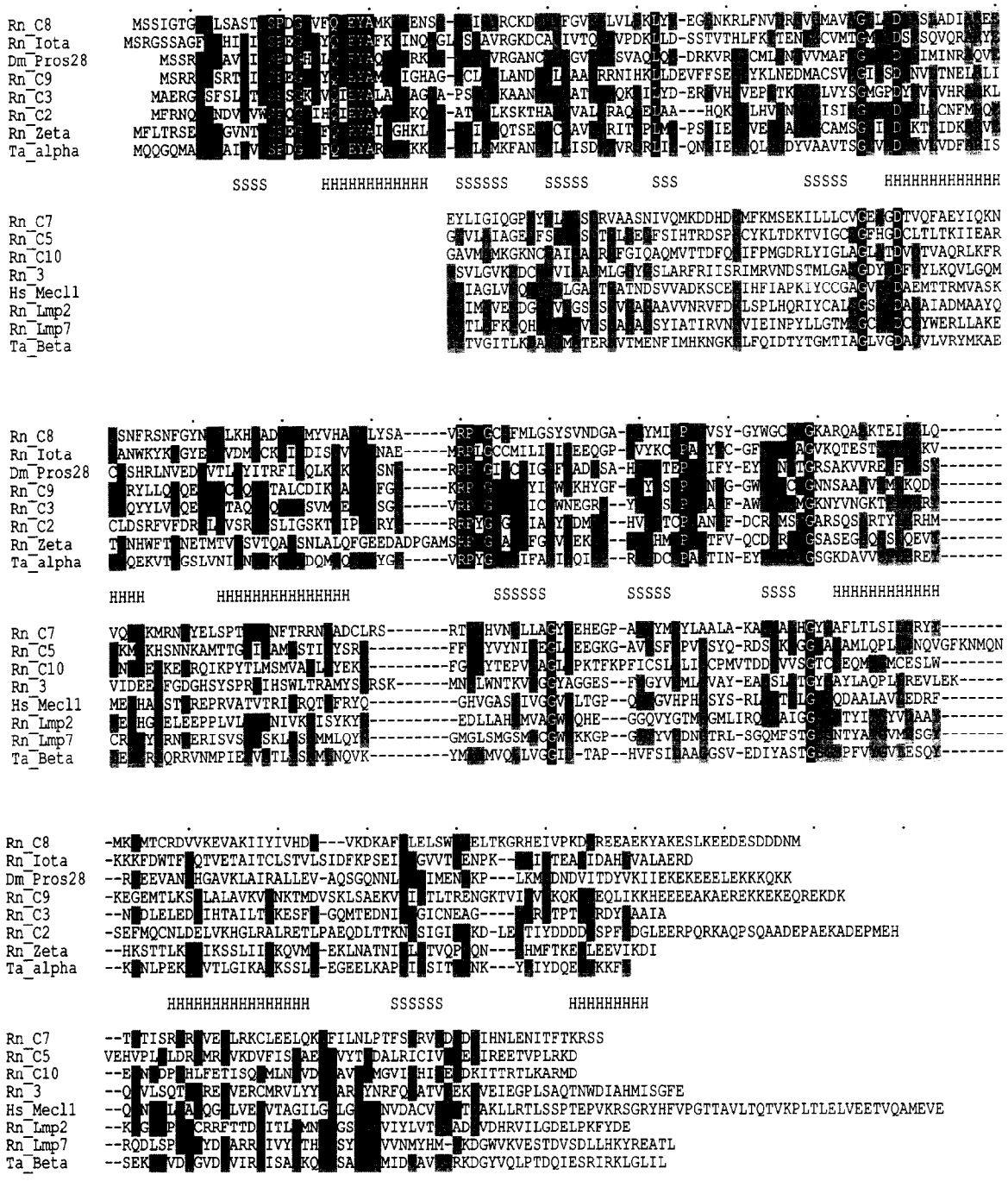

Fig. 1. Alignment of 20 S α- and β-type sequences representing the seven eukaryotic and the one archaebacterial branch within each subunit family. Where available, rat (Rn) sequences are show; in the two cases where rat subunits have not been reported yet, human (Hs) and fruit fly (Dm) sequences are shown. Residues identical within a subunit family are shown in inverted type; residues at least 50% conserved are shown shaded. Between subfamilies is shown our interpretation of the scondary structure predictions (Fig. 2). The alignment was generated in MACAW and displayed with MALIGNED.

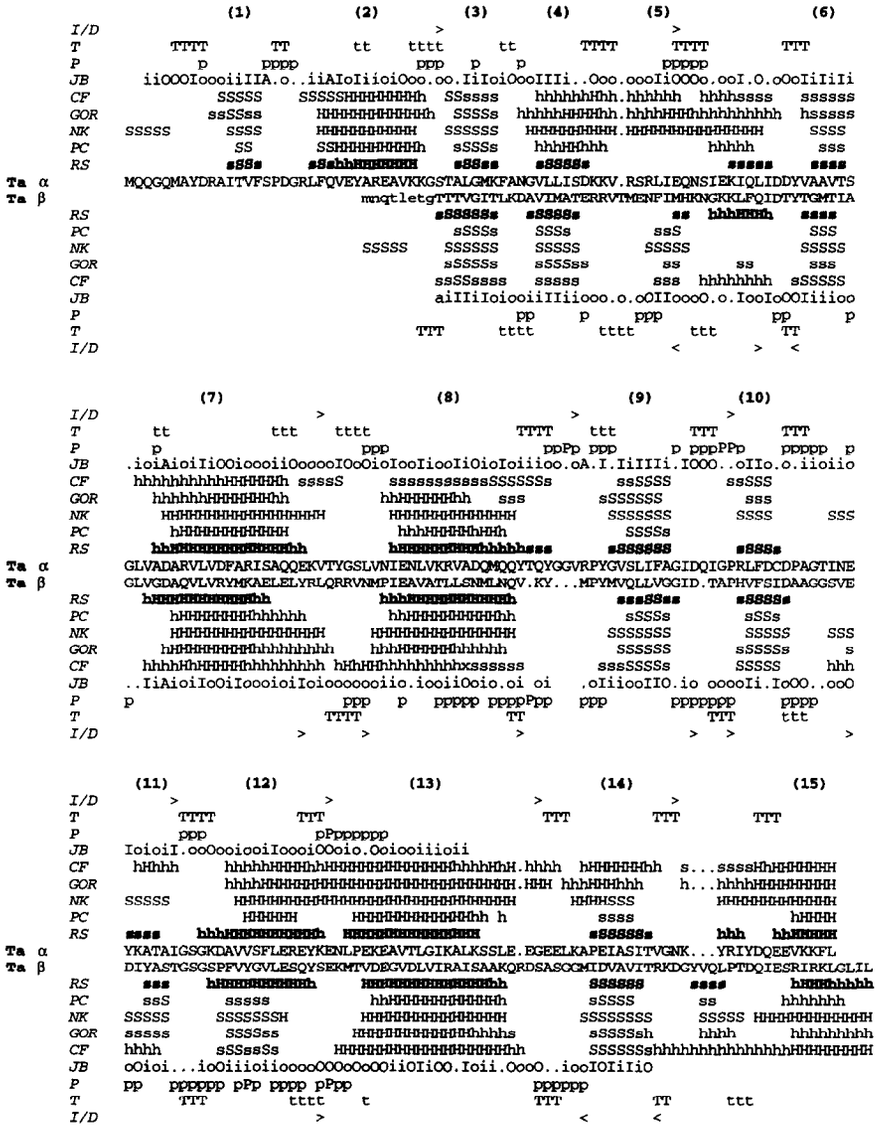

Fig. 2. Predicted secondary structure of 20 S proteasome subunits: (h, H) = α-helix; (s, S) = β-strand; (i, I) = buried residue; (o, O) = surface-exposed residue; (p,P) = parsing sites; (t, T) = turn or loop; (>, <) = sites of insertion (>) and deletion (<) occurring in members of the same subunit family. Although only the sequences of the *Thermoplasma* α and β subunits are shown, secondary structure predictions are consensus predictions obtained separately for alignments of α- and β-type subunits. The methods employed are those of Rost and Sander (RS), Presnell and Cohen (PC, T), Niermann and Kirschner (NK), Garnier, Osguthorpe and Robson (GOR), Chou and Fasman (CF), and Jenny and Benner (JB, P). JB, which generates its own alignment, could not align all sequences at their C-terminal end and only predictions for the completely aligned part are shown. Capital letters denote positions where more than 50% of aligned sequences were predicted to show the same type of secondary structure. For RS, capital letters denote the subset of positions with an expected prediction accuracy >82%. RS is shown bold as the most accurate of the current methods. Turns (T) were scored with a capital letter if they were predicted in more than 2/3 of sequences in the alignment within a window of 3 residues. Elements of secondary structure are numbered sequentially.

into two subfamilies with more than 25 members each that span the entire range of pairwise identities, while the pairwise identity between sequences from different subfamilies is typically below 30%. Thus, predictions can be generated separately for each subfamily and compared, as the two groups are sufficiently divergent to prevent extensive redundancy in the predictions.

To evaluate the performance of secondary structure prediction approaches, we have gathered predictions for the family of 20 S proteasome subunits by several widely used methods in advance of the release of the crystal structure. As a first step, we made an alignment of proteasome sequences, proceeding from alignments of the sequences within individual branches of one subfamily, over alignments of the α- and β-type subunit families to a complete alignment of all sequences. This alignment, for which representative sequences are shown in Fig. 1, was used in all further steps except in the prediction by the method of Jenny and Benner (JB) which generates its own alignment.

The methods, which are based on statistical information (CF, GOR), sequence patterns (T), neural networks (PC, RS) and evolutionary conservation (NK, JB, P) agree surprisingly well (Fig. 2). For JB, where the output provides information on inside and outside location of residues and on the parsing of

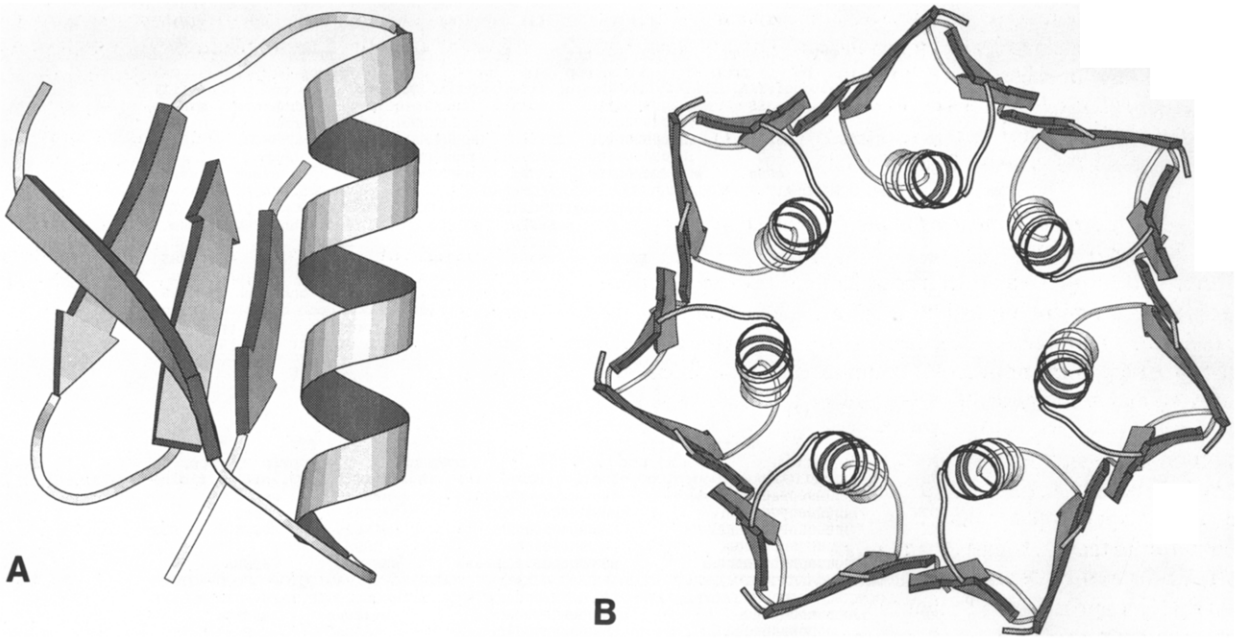

Fig. 3. Schematic representations of the α subunit N-terminal domain (residues 6–49, 55–61) and of the heptameric ring structure proposed for the complex. The drawings were generated by MOLSCRIPT [31].

the sequence, we have interpreted the patterns following the published procedure [18] but report the output as obtained from the server. Our synopsis of the prediction results (which we have based mainly on RS and JB) is shown with the alignment in Fig. 1. For the N-terminal extension present in α subunits, all methods agree on (1) a β-strand and (2) an α-helix, and several methods predict a second short β-strand immediately preceeding the helix that we incline to discount. For the part of the sequence present in α and β subunits, the methods identify the following elements (Fig. 2). (3) A β-strand. (4) A potential β-strand that is seen as a helix for α subunits by all methods except RS and JB. (5) A region of inconsistent predictions in which may be present a further short β-strand (predicted only in β subunits). (6) A β-strand. (7) An α-helix. (8) An α-helix, seen at least in part as a strand by CF. (9) A β-strand. (10) A β-strand. (11) A potential β-strand, seen as a helix by CF. (12) A potential α-helix, seen as a strand in β subunits by all methods except RS and JB. (13) An α-helix. (14) A potential β-strand, seen as a helix in α subunits by CF and GOR, followed by an area of inconsistent predictions. (15) An α-helix.

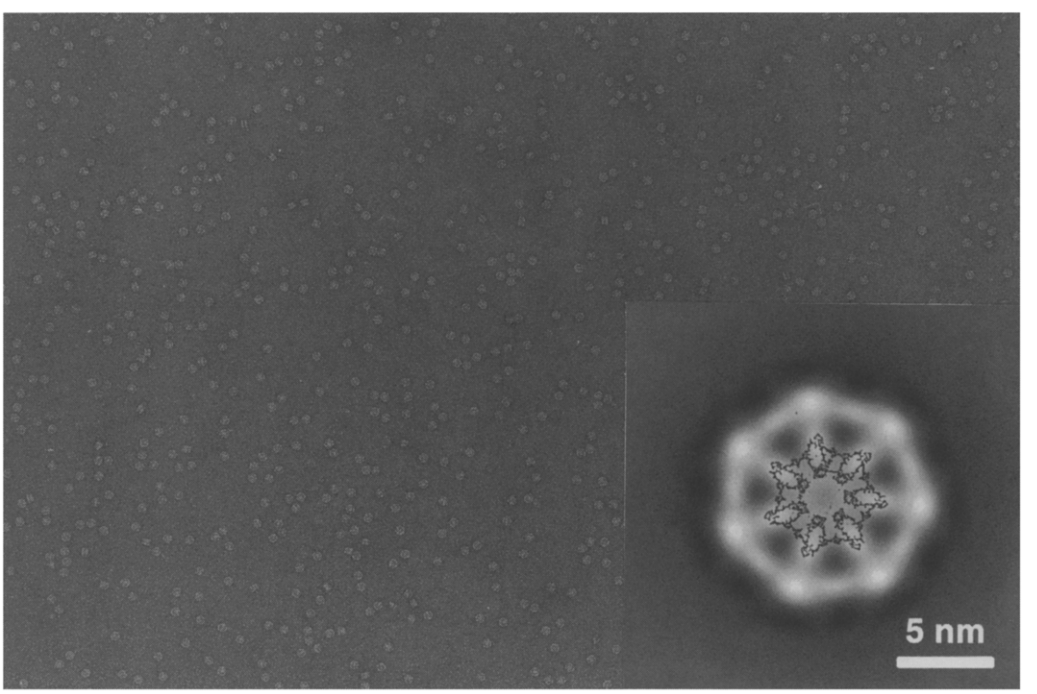

Fig. 4. Electron micrograph of *Thermoplasma* α-rings negatively stained with 2% uranyl acetate. The inset shows an average projection image after seven-fold symmetrization with a backbone trace of the model for the α N-terminal domain superposed. The resolution is 1.6 nm.

The distribution of secondary structure elements indicates that proteasome subunits belong to the $\alpha+\beta$ tertiary class, but prediction by the method of Chou and Zhang, which assigns the tertiary class on the basis of residue composition [24], yields a preference for the α/β class (averaged correlation coefficients: α 92.6%; β 89.7%; $\alpha+\alpha$ 94.2%; α/β 95.7%).

3.2. Tertiary structure models for the N-terminal domain of α subunits

Averaged projection structures of 20 S proteasomes derived from electron micrographs and a 3D reconstruction obtained by electron tomography [25] have shown the existence of a central cavity in the proteasome complex, and electron microscopy studies with nanogold-labeled substrates have raised the possibility of a channel in the outer α-rings. Since α subunits contain an N-terminal extension and α-rings have additional electron density towards their center that is not present in β-rings, it appears likely that the N-terminal domain of α subunits forms the inner part of the rings. From the inability of β-subunits and of N-terminal mutants of α-subunits to form rings [26] it appears further that the N-terminal domain mediates the main ring-forming interactions. The secondary structure analyses indicate to us that this N-terminal domain extends beyond the 35 residues not present in β subunits and probably includes the first 60 residues with a predicted secondary structure of $\beta\alpha\beta\beta\beta$. By analogy to several recently solved channel structures, all formed by $\alpha\beta$-rolls [19,27–30], we have considered the possibility of a fold related to $\alpha\beta$-rolls for the α N-terminal domain (Fig. 3). We have based the model on the structure of a pentameric AB_5 toxin [19] because its channel contains an extended peptide and thus has the dimensions we anticipate for the channel in 20 S proteasomes. Since the sequence of secondary structure elements is different in proteasomes and AB_5 toxin B subunits, we had to assume a different connectivity although the number of β-strands (5×6 in AB_5 toxins; 7×4 in the model structure) allowed us to maintain a similar size for our model. The main goal of the model is to illustrate how the peptide channel could be formed by the helices of the seven α subunits. The arrangement of the supporting β framework is considerably more hypothetical.

Our predictions with regard to the topology of proteasome subunits can be summarized as follows:

(i) α subunits contain two domains (residues 1–60 and 70–233 in *Thermoplasma* α) of which the N-terminal one extends beyond the 35 residues that are not present in β subunits. The corrolary of this is that β subunits contain a small subdomain of two (possibly three) antiparallel β-strands facing the central cavity that are only loosely attached to the main body of β- rings.

(ii) Both domains of α subunits belong to the $\alpha+\beta$ class. The N-terminal domain forms an $\alpha\beta$-roll and is oriented with its secondary structure elements parallel to the central axis of symmetry. The loops leading into (1), connecting (2) with (3), and leading away from (4) are solvent-exposed at the top of the α-rings, while the conserved loop connecting (1) with (2) is buried at the interface between α- and β-rings. The central channel is formed by 7 helices (one from each subunit) in an arrangement similar to the one seen in AB_5 toxins.

(iii) The C-terminal domain of α subunits is formed by an antiparallel β-sheet that is flanked on one side by helices (7) and (8) and on the other by helices (12), (13) and (15).

3.3. Correlation with the low-resolution structure of α-rings

Recently, we have been able to determine the low-resolution structure of α-rings, purified in recombinant form from *Escherichia coli*, to a resolution of 1.6 nm (Fig. 4). The α-rings have a diameter of 11 nm and are shaped like a spoked wheel with areas of low density in the center and between the spokes. The resolution of the projection image prevents us from unambiguously deciding whether the α-subunits form rings with a central channel. However, an overlay of the model onto the projection image shows that the data are consistent with a the existence of a channel.

References

[1] Goldberg, A.L. (1992) Eur. J. Biochem. 203, 9–23.
[2] Rechsteiner, M. (1991) Cell 66, 615–618.
[3] Hershko, A. and Ciechanover, A. (1992) Annu. Rev. Biochem. 61, 761–807.
[4] Peters, J.-M., Cejka, Z., Harris, J.R., Kleinschmidt, J.A. and Baumeister, W. (1993) J. Mol. Biol. 234, 932–937.
[5] Grziwa, A., Baumeister, W., Dahlmann, B. and Kopp, F. (1991) FEBS Lett. 290, 186–190.
[6] Dahlmann, B., Kopp, F., Kuehn, L., Niedel, B., Pfeifer, G., Hegerl, R. and Baumeister, W. (1989) FEBS Lett. 251, 125–131.
[7] Lupas, A., Koster, A.J. and Baumeister, W. (1994) Enzyme and Protein, in press.
[8] Jap, B., Pühler, G., Lücke, H., Typke, D., Löwe, J., Stock, D., Huber, R. and Baumeister, W. (1993) J. Mol. Biol. 234, 881–884.
[9] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[10] Schuler, G.D., Altschul, S.F. and Lipman, D.J. (1991) Proteins 9, 180–190.
[11] Clark, S.P. (1992) CABIOS 8, 535–538.
[12] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251:276.
[13] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[14] Niermann, T. and Kirschner, K. (1991) Methods Enzymol. 202, 45–59.
[15] Cohen, B.I., Presnell, S.R. and Cohen, F.E. (1991) Methods Enzymol. 202, 252–268.
[16] Presnell, S.R., Cohen, B.I. and Cohen, F.E. (1993) CABIOS 9, 373–374.
[17] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[18] Benner, S.A., Badcoe, I., Cohen, M.A. and Gerloff, D.L. (1994) J. Mol. Biol. 235, 926–958.
[19] Stein, P.E., Boodhoo, A., Tyrrell, G.J., Brunton, J.L. and Read, R.J. (1992) Nature 355, 748–750.
[20] Jones, T.A. and Thirup, S. (1986) EMBO J. 5, 819–822.
[21] Kleinz, J. (1994) PhD Dissertation, Technical University, Munich, Germany.
[22] Hegerl, R. and Altbauer, A. (1982) Ultramicroscopy 9, 109–116.
[23] Pühler, G., Pitzer, F., Zwickl, P. and Baumeister, W. (1994) System. Appl. Microbiol. 16, 734–741.
[24] Chou, K.-C. and Zhang, C.-T. (1992) Eur. J. Biochem. 207, 429–433.
[25] Hegerl, R., Pfeifer, G., Pühler, G., Dahlman, B. and Baumeister, W. (1991) FEBS Lett. 283, 117–121.
[26] Zwickl, P., Kleinz, J. and Baumeister, W. (1994) Nature Struct. Biol., in press.
[27] Sixma, T.K., Pronk, S.E., Kalk, K.H., Wartna, E.S., van Zanten, B.A.M., Witholt, B. and Hol, W.G.J. (1991) Nature 351, 371–377.
[28] Stein, P.E., Boodhoo, A., Armstrong, G.D., Cockle, S.A., Klein, M.H. and Read, R.J. (1994) Structure 2, 45–57.
[29] Kong, X.-P., Onrust, R., O'Donnell, M. and Kuryan, J. (1992) Cell 69, 425–437.
[30] Unwin, N. (1993) J. Mol. Biol. 229, 1101–1124.
[31] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.