Rethinking Proteasome Evolution: Two Novel Bacterial Proteasomes

Ruben E. Valas · Philip E. Bourne

Abstract The proteasome is a multisubunit structure that degrades proteins. Protein degradation is an essential component of regulation because proteins can become misfolded, damaged, or unnecessary. Proteasomes and their homologues vary greatly in complexity: from HslV (heat shock locus v), which is encoded by 1 gene in bacteria, to the eukaryotic 20S proteasome, which is encoded by more than 14 genes. Despite this variation in complexity, all the proteasomes are composed of homologous subunits. We searched 238 complete bacterial genomes for structures related to the proteasome and found evidence of two novel groups of bacterial proteasomes. The first, which we name Anbu, is sparsely distributed among cyanobacteria and proteobacteria. We hypothesize that Anbu must be very ancient because of its distribution within the cyanobacteria, and that it has been lost in many more recent species. We also present evidence for a fourth type of bacterial proteasome found in a few β-proteobacteria, which we call β-proteobacteria proteasome homologue (BPH). Sequence and structural analyses show that Anbu and BPH are both distinct from known bacterial proteasomes but have homologous structures. Anbu is encoded by one gene, so we postulate a duplication of Anbu created the 20S proteasome. Anbu's function appears to be related to transglutaminase activity, not the general stress response associated with HslV. We have found different combinations of Anbu, BPH, and HslV within these bacterial genomes, which raises questions about specialized protein degradation systems.

Keywords Novel proteasomes · HslV · 20S proteasome · Phylogeny · Anbu · BPH

R. E. Valas (✉)
Bioinformatics Program, University of California, San Diego, 9500 Gilman Drive, MC 0743, La Jolla, CA 92093, USA
e-mail: rvalas@ucsd.edu

P. E. Bourne
Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Introduction

The proteasome is a complex, multisubunit protein assembly which forms a barrel with multiple internal active sites that function together to recognize and degrade proteins (reviewed in Groll et al. 2005). All archaea and eukaryotes have a 20S proteasome as well as some actinobacteria, but most bacteria have a simpler homologous structure heat shock locus v (HslV). Although proteasomes are found across the tree of life, there are many bacterial species that lack them entirely. The 20S proteasome and its HslV homologues function either to degrade misfolded proteins (Goldberg 2003), as occurs under conditions of heat shock, or as a precise regulatory mechanism by degrading proteins, usually defined by a ubiquitin tag (Glickman and Ciechanover 2002).

There is an evolutionary progression in structural complexity of the proteasome. Several Protein Data Bank (PDB) structures (Berman et al. 2000) of the 20S proteasome from all three superkingdoms and HslV from bacteria all form barrel structures that have the active sites on the inside of the barrel. All of these barrels consist of inner

subunits responsible for cleavage and outer subunits responsible for protein recognition. The core of the 20S proteasome is a four-layered barrel found in archaea, eukaryotes, and actinobacteria. Each layer comprises a heptameric ring. All 28 subunits are in the same structural family in the Structural Classification of Proteins (SCOP) database (Murzin et al. 1995), implying that they all share a single common ancestor gene.

In the actinobacteria and archaea the 20S proteasome is usually encoded by two genes, the α and β subunits. The β subunits are catalytically active and form the two middle layers. The α subunits that form the outer two layers are catalytically inactive and act as scaffolding for the β subunits. The α subunits form an antechamber that restricts access of the substrate to the proteolytic chamber (Rabinovich et al. 2006). Some actinobacteria and archaea contain more than one type of α and β subunit. The eukaryotes are more complicated still. The core of the yeast 20S proteasome is coded by 14 different genes (7 α and 7 β), with only 3 of the β subunit genes having catalytic activity (Groll et al. 1997). The 20S proteasome can be knocked out from archaeal cells under normal conditions, but it is essential for surviving heat shock conditions (Ruepp et al. 1998).

Many bacteria contain a simpler proteasome homologue, HslV (Coux et al. 1996). HslV is a heat shock protein and is expressed as part of a general response to stress that causes proteins to misfold. Unlike the heptameric rings of the 20S proteasome, HslV is made up of two layers of hexameric rings, which are encoded by only one gene. According to SCOP, the subunits of HslV are also in the same structural family as the subunits of the 20S proteasome. Because of its simpler structure, HslV is a good model system for the 20S proteasome.

The 20S proteasome and HslV both associate with ATPases that use ATP to unfold proteins and translocate them into the proteasome or HslV structure, respectively (reviewed in Smith et al. 2006). There is a corresponding increase in complexity in the ATPases and other regulatory proteins associated with the proteasome in moving from the bacteria to the eukaryotes. A hexameric ring of ATPases known as heat shock locus U (HslU) binds to each side of HslV. Likewise, hexameric ATPases bind to either side of the 20S proteasome interfacing with the α subunits. Unlike the subunits of HslV and the 20S proteasome, the subunits HslU and the ATPases associated with the 20S proteasome are in different structural superfamilies (Iyer et al. 2004). HslU is related to ClpX, the ATPase of the ClpP protease. The 20S proteasome ATPases are related to the ATPase domain of the protease FtsH. In HslU and the 20S proteasome of actinobacteria and archaea, all six ATPases are encoded by a single gene (Darwin et al. 2005; Zwickl et al. 1999). This structure is more complicated in

eukaryotes, which encodes six different homologous ATPases and at least 11 other proteins in the PA700 complex (also known as the 19S regulatory cap) (Bochtler et al. 1999). Eukaryotes also have two alternative caps, PA28 (also known as the 11S cap) (Hill et al. 2002) and PA200 (Ustrell et al. 2002), that do not use ATP or recognize ubiquitin. Both the 20S proteasome core and different combinations of the 20S core and its caps are found within eukaryotic cells in significant numbers (Tanahashi et al. 2000).

There is also a progression in complexity of the targeting systems (i.e., recognizing which proteins to degrade) in the various proteasomes. In eukaryotes most proteins are targeted for degradation by ubiquitin tagging (Glickman and Ciechanover 2002; Hershko 2005), although there is a growing number of proteins found to be degraded in a ubiquitin-independent manner (Orlowski and Wilk 2003). No such tagging is known in prokaryotes, although it has recently been shown that many bacteria have some homologues to this tagging pathway (Iyer et al. 2006). Some proteins contain a tag in their N-terminus, such as the ARC protein, which targets them for degradation by HslV (Burton et al. 2005). We speculate that similar targeting may be used with other proteasomes in species that lack ubiquitin.

Previous work has analyzed genomic data to study the evolution of the proteasome (Gille et al. 2003). The authors looked at 61 complete and 60 incomplete genomes. They found that several protists contain both HslV and a 20S proteasome. Some actinobacteria also contained a 20S proteasome with distinct α and β subunits. The authors state this was probably due to horizontal transfer, but they could not identify the source. They found that several bacteria had no homologue to the proteasome or HslV, so they must use other proteases instead. They also noted that two bacteria, *Magnetospirillum magnetotacticum* and *Enterococcus faecium*, had two copies of HslV. They conclude that *Magnetospirillum magnetotacticum* had a recent duplication of HslV and *Enterococcus faecium* acquired a second copy through horizontal transfer. They found no bacteria that have both HslV and a 20S proteasome.

We are able to extend these results by analyzing many more complete genomes and using data derived from protein structure. Since structure is more conserved than sequence, this facilitates studies over long evolutionary time scales. Support for this approach comes from our recent work in constructing the tree of life based on the presence or absence of superfamilies derived from structure (Yang et al. 2005). The structure data for our work come from the Superfamily database, which holds hidden Markov model (HMM) predictions of structural domains, families, and superfamilies based on the SCOP

classification scheme (Murzin et al. 1995), for completed proteomes (Gough et al. 2001). Superfamily was used to determine which of the 238 completed bacterial genomes had multiple genes predicted to be proteasomal subunits. Most bacteria had one gene which hit the proteasome subunit family, which was usually HslV. The actinobacteria with known 20S proteasomes had two hits, as one would expect. To our surprise there was also a large number of proteobacteria with two hits, and several β-proteobacteria had three genes encoding proteasome subunits. There were also some genomes with no hits to this family as had been observed before.

Magnetospirillum magnetotacticum was one species that had two hits. We are able to analyze when these genes were duplicated by looking at how the additional hits cluster in a phylogenetic tree. The two proteins from *Magnetospirillum magnetotacticum* clustered on opposite ends on the tree, each with other sequences. This led us to the conclusion that this is not just a second copy of HslV but, rather, a representative novel proteasome homologue which we named Anbu. We also found a distinct cluster of sequences in some β-proteobacteria, which we name β-proteobacteria proteasome homologue (BPH). No species containing BPH from this group was mentioned in Gille et al. (2003), so this group is entirely novel. We found that our two novel clusters match two unannotated clusters in the NCBI Protein Clusters database: CLS882959 is Anbu, and CLS856934 is BPH.

Our trees show distinct clusters but do not show an unambiguous history of the proteasome. Since these sequences diverged billions of years ago, it is not surprising that it is difficult to get a clear phylogenetic signal. However, structural inference can be linked to sequence, so we can combine structural information with these trees to better re-create the evolutionary relationship of these families. Threaded structure predictions were created for two representative sequences from both Anbu and BPH. Anbu sequences were taken from *Rhodopseudomonas palustris* and *Hahella chejuensis*, and BPH sequences were taken from *Thiobacillus denitrificans* and *Ralstonia Metallidurans*. These are high-quality predictions because each prediction was created from several known structures of HslV and the 20S proteasome. We compared these predictions to other known structures to determine the evolutionary history of the different proteasome homologues.

HslV and the 20S proteasome are clearly evolutionarily related from their common structures. HslV is a good model system for the 20S proteasome from that fact alone. However, the question of which proteasome came first has interesting implications for evolution. If HslV is ancestral to the 20S proteasome, then the archaea must be younger than the bacteria, as all archaea have a 20S proteasome (Cavalier-Smith 2006). Since there were no other known simple proteasome homologues as potential predecessors,

this seemed reasonable. The introduction of Anbu changes this view. We show that Anbu is a more probable candidate than HslV as the ancestor of the actinobacterial 20S proteasome based on its position in the phylogenetic tree and its structural features. Further study of Anbu will shed more light on the function of the 20S proteasome rather than studying just HslV.

Methods

The Superfamily (Gough et al. 2001) database was used to identify 216 bacterial proteins (Supplemental Table 3) in the SCOP (Murzin et al. 1995) proteasome family. All hits had e-values <0.0001 at the superfamily level. We took all hits to the proteasome family regardless of e-value because we are interested in proteins that are not represented by known structures. The hit from *Deinococcus radiodurans* was not included because this sequence was a multidomain protein, while all other sequences included only a proteasome subunit domain. This protein may include the N-terminal nucleophile aminohydrolase domain, as it weakly hits that superfamily. Since it does not align with any other proteasome subunit, including those from *Thermus thermophilu*s, this sequence is probably not a proteasome subunit. *Frankia*, an actinobacteria, had a hit in addition to the 20S proteasome. This sequence did not align well with any of the five clusters, and is also probably an N-terminal nucleophile aminohydrolase, but not a proteasome subunit. Excluding this sequence increased the quality of the multiple alignment.

Sequences were aligned using MUSCLE (Edgar 2004), part of the STRAP (http://www.charite.de/bioinf/strap/) suite of programs. Multiple structural alignments were performed using Combinatorial Extension (Shindyalov and Bourne 1998), also packaged in STRAP. All trees were built using PHYML (Guindon and Gascuel 2003) with the JTT model of evolution, estimated variance and gamma, and four substitution rate categories. PHMYL was packaged as part of Geneious (Drummond et al. 2006) (http://www.geneious.com/). Each tree was bootstrapped from 100 replicates.

Representative proteasome subunits were taken from the PDB (Berman et al. 2000). These structures came from various species in all three superkingdoms. 1JJW, 1E94, and 1M4Y are HslV structures. 1Q5Q and 2FHG are actinobacterial 20S proteasomes. Two sequences from Anbu (from *Rhodopseudomonas palustris* and *Hahella chejuensis*) and two sequences from BPH (from *Thiobacillus denitrificans* and *Ralstonia metallidurans*) were threaded using the Phyre web server, which is the successor of 3D-PSSM (Kelley et al. 2000). Each predicted model was created from several known structures. All resulting structure

predictions have very high structural similarity to known proteasome subunit structures. The predicted models were aligned to create a phylogenetic tree. All structural images were created in Protein Workshop, part of the Molecular Biology Tool Kit (Moreland et al. 2005).

BLAST (Altschul et al. 1990) searches were performed using HslU from *Ralstonia solanacearum* (GI:17427050) and proteasome-associated ATPase from *Mycobacterium tuberculosis* (GI:113700393) against cyanobacteria and β-proteobacteria to find potential ATPases for Anbu and BPH. Table 2 was created using Superfamily's predictions for the transglutaminase catalytic domain. The *p*-values in Table 2 were calculated using a one-tailed *t*-test without the assumption that the variances of the groups were equal.

Results

Phylogenetic Analysis

We constructed a maximum likelihood tree from a multiple sequence alignment of sequences predicted to be in the proteasome subunit family by the Superfamily database (Fig. 1). This tree shows five distinct clusters. Three of
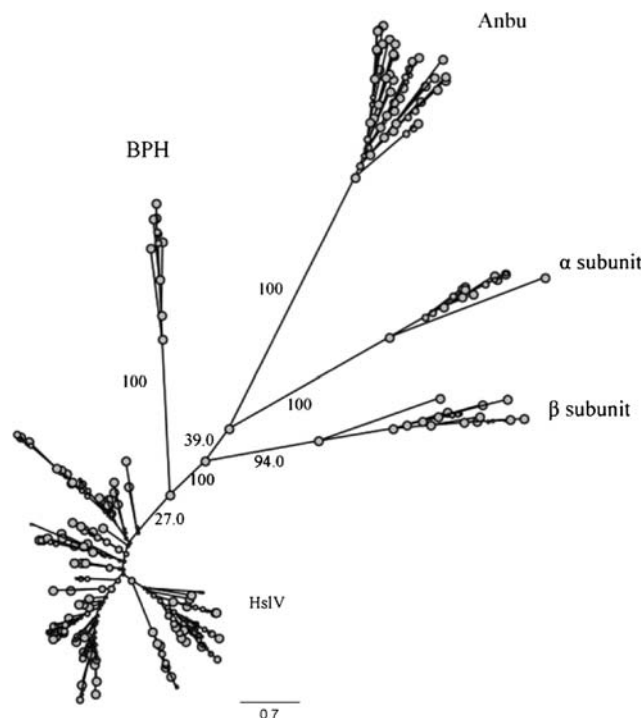
Fig. 1 Maximum likelihood tree from a multiple alignment of proteins predicted to be proteasome subunits in the Superfamily database. One hundred replicates were run to obtain bootstrap values. The Anbu and BPH clusters represent two novel proteasome homologs. Anbu's position near both subunits of the 20S proteasome implies that it is ancestral to the 20S proteasome

these clusters are known proteasome subunits; HslV, the 20S proteasome α subunit, and the 20S proteasome β subunit. There are some low bootstrap values, but most of the critical edges have high values. The two novel clusters, Anbu and BPH, are both supported as true novel groups with bootstrap values of 100%. A 100% bootstrap value also separates this tree into two groups; BPH with HslV and Anbu with the 20S proteasome. This tree strongly supports Anbu being ancestral to the 20S proteasome, not HslV, which is the current view.

Thr1, Lys33, and Gly47 are all catalytic residues in *Thermoplasma acidophilum's* 20S proteasome (Lowe et al. 1995; Seemuller et al. 1996; Seemuller et al. 1995). A deprotonated Thr1 performs a nucleophilic attack on the substrate, which is stabilized by Gly 47. Lys33 promotes the deprotonation of Thr1. The corresponding sites are universally conserved throughout Anbu and BPH with only one exception (Supplemental Tables 1 and 2). This is evidence that these novel groups function like the known bacterial proteasomes. The distribution of Anbu and BPH on the tree of life has several interesting features (Supplemental Fig. 1). Anbu is found in α-proteobacteria, β-proteobacteria, γ-proteobacteria, and cyanobacteria according to the Superfamily database. This is noteworthy since no cyanobacteria has HslV. Anbu is present in *Gloeobacter violaceus*, which is an early-branching cyanobacteria (Honda et al. 1999). It appears that Anbu was present in the cyanobacterial ancestor so it must be very ancient. A BLAST search revealed that Anbu was also present in *Leptospirillum ferrooxidans* as well as *Solibacter usitatus. Cytophaga hutchinsonii*, a sphingobacteria, was found to have two copies of Anbu. A species with a duplication of Anbu could be the precursor to the 20S proteasome. Anbu's distribution is sparse but broad, which infers it is an ancient protein that has been lost many times. This repeated loss is not unrealistic given that photosynthesis was also lost many times in the proteobacteria (Woese 1987). The BPH group only includes β-proteobacteria. This extremely narrow distribution implies that BPH is a relatively young proteasome. This, combined with BPH's position in the phylogenetic tree, implies that BPH evolved from HslV.

The current view is that bacteria either have HslV, a 20S proteasome, or no proteasome. There are no known cases of a bacterium having both HslV and a 20S proteasome. With the discovery of Anbu and BPH, it is now clear that proteasome homologues occur in many combinations in bacterial genomes (Supplemental Fig. 1, Table 1). Anbu, HslV, and BPH are present together in several genomes in different combinations, but none of them are ever found in the same genome as a 20S proteasome. However, both HslV and the 20S proteasome were found in a recent metagenomic study of *Leptospirillum* group II bacteria.

Table 1 Combinations of proteasomes in bacteria

Combination	Occurrence in genomes	Example species
No proteasome homologue	64	*Synechococcus* sp. WH 8102
HslV	132	*Escherichia coli*
20S proteasome	10	*Mycobacterium tuberculosis*
Anbu	6	*Synechocystis* sp. PCC 6803
Anbu + HslV	19	*Rhodobacter sphaeroides*
BPH	0	No occurrence
BPH + Anbu	1	*Polaromonas* sp. JS666
BPH + HslV	2	*Thiobacillus denitrificans*
BPH + Anbu + HslV	3	*Ralstonia solanacearum*

Note. Different combinations of proteasome homologues found in 238 bacterial genomes. Anbu, BPH, and HslV are found in several combinations, but never with the 20S proteasome

The authors state that in this case the 20S proteasome was probably horizontally transferred from the actinobacteria (De Mot 2007). A BLAST search revealed that this metagenome also contains Anbu. Although this metagenomic sample is dominated by *Leptospirillum* group II bacteria (Lo et al. 2007), these data are not from a single species. Therefore this is not evidence that a single genome contains HslV, Anbu, and the 20S proteasome. However, it is evidence that all three of these proteasomes can be useful in the same environment. The three *Ralstonia* species in our sample have Anbu, BPH, and HslV. We believe that these three proteasomes are functionally distinct (discussed below). This raises an important question of how bacteria target a protein to a specific proteasome to be degraded without using ubiquitin. BPH is never found as the sole proteasome homologue. It can be inferred that BPH degrades proteins that cannot be degraded by one of the other mechanisms, but it does not degrade a wide enough variety of substrates on its own to replace HslV or Anbu. It would be interesting to create knockouts in these species to see how BPH functions and hence to compare the functions of BPH and HslV in these species. This would allow us to determine whether BPH's function is redundant or whether it degrades additional substrates. The 20S proteasome may be able to act on a wider variety of substrates than other homologues, so it can replace the function of different proteasome families. The idea that bacteria can only have HslV or the 20S proteasome exclusively is too simple. Instead we need to determine the specific functions of each family and how they interact in all of these different combinations.

These new proteasome families are good candidates for structure prediction using fold recognition (threading), because the PDB has several structures for the 20S proteasome from archaea, bacteria, and eukaryotes as well as HslV. We created two models from sequences of both Anbu and BPH using the Phyre web server, which is the successor to 3D-PSSM (Kelley et al. 2000). Anbu was modeled from archaeal and eukaryotic 20S proteasome structures. Anbu from *Rhodopseudomonas palustris* has

18% sequence identity to the structure of the archaeal 20S proteasome. BPH was modeled from structures of the eukaryotic proteasome and HslV. BPH from *Thiobacillus denitrificans* has 22% sequence identity to the structure of HslV. We built a multiple sequence alignment from a multiple structural alignment for each cluster using Combinatorial Extension (Shindyalov and Bourne 1998) and built a tree using maximum likelihood (Fig. 2). This was done to increase the quality of the alignment using structural information. Anbu again falls right between the α and the β subunits, and BPH clusters with HslV. This tree is in agreement with the one constructed from sequence alone. It supports Anbu being ancestral to the 20S proteasome and HslV being ancestral to BPH.

Structural Analysis

The predicted structures of both Anbu and BPH align very well with known proteasome subunits, but each has unique structural features. The areas around the active sites align particularly well (Fig. 3). This conserved catalytic triad is strong evidence that Anbu and BPH both function as proteasomes.

After a crystal structure of HslV from *E. coli* was determined, it was compared to the β subunit from the archaea *Thermoplasma acidophilum* (Bochtler et al. 1997). The authors proposed several differences that could account for HslV forming a hexamer while the 20S proteasomes forms a heptamer. The first is that the β subunits may be forced by the α subunits to form a heptamer. Helix 1, which is in contact between the α and the β subunits, is extended by five residues in the β subunit relative to HslV (highlighted in red in Fig. 4A). The β subunit also has an extra C-terminal helix (highlighted in green in Fig. 4A), which could affect the way the subunits pack together into rings. We compared our models of Anbu with known structures of HslV and the 20S proteasome. Helix 1 is extended in Anbu compared to HslV (highlighted in red in Fig. 4E). Anbu's C-terminal tail is also extended relative to HslV (highlighted in green in 4E). The threaded models of
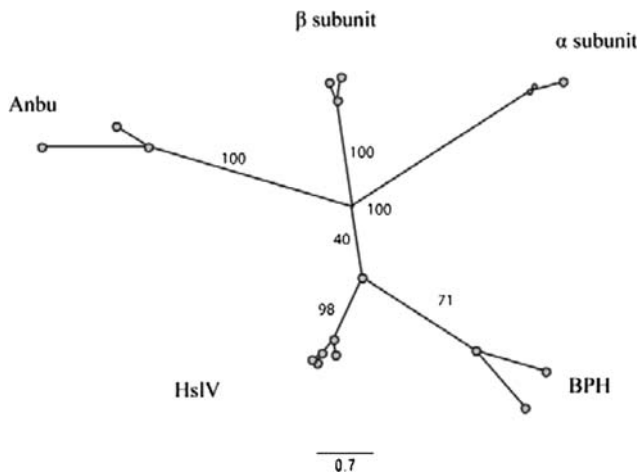
Fig. 2 Maximum likelihood tree from a structural alignment of seven proteasome subunits from the PDB and four structural predictions (two from Anbu and two from BPH). The placement of BPH and Anbu in this tree is in agreement with the tree in Fig. 1
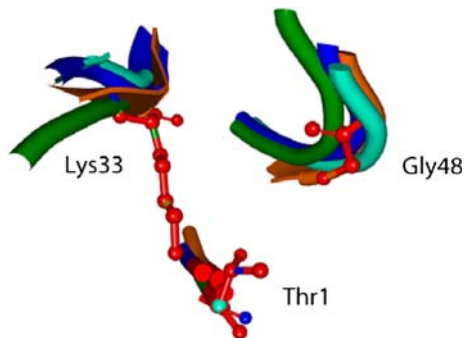
Fig. 3 Comparison of catalytic triads in different proteasomes. HslV (1E94) is green, β subunit (1Q5Q_H) is cyan, Anbu (predicted structure from *Rhodopseudomonas palustris*) is blue, and BPH (predicted structure from *Thiobacillus denitrificans*) is orange. The side chains of HslV are colored red. The corresponding backbone and neighboring residues are visible from each structure. All three sites are highly conserved in sequence as well as structure. The labels refer to the positions of these residues in 1E94

Anbu cuts out about 30 C-terminal residues that do not hit known structures. The secondary structure of this region is predicted to be a sheet followed by a helix with possible loops between them. There are several highly conserved positions in the missing section of the tail. It is possible that this region has a functional role that is not present in HslV or the 20S proteasome. Anbu has other features that are not shared by any of the other proteasome families. Both turn 3 and turn 6 have significant extensions in Anbu that could affect packing in the biological unit (highlighted in yellow in Figs. 4B and C. (These turns are colored orange in the biological unit of the 20S proteasome in Supplemental Fig. 2.) The extended loop 3 could act as a gate into the proteasome if Anbu forms two layers of rings. We cannot definitively conclude Anbu's biological unit from these

features, but they do give a strong indication that the 20S proteasome evolved from Anbu. Both the helix extension and the C-terminal tail discussed above are present in both the α and the β subunits of the 20S proteasome (Figs. 4B and D). Both structural features were probably present in the ancestor of both subunits. A duplication of Anbu would be more likely to result in a 20S proteasome-like structure than a duplication of HslV because Anbu already has both of these structural features. That, taken with Anbu's position in our trees, indicates that the 20S proteasome evolved from Anbu, not HslV.

We also compared the predicted structure of BPH to that of HslV and Anbu. It is highly unlikely that the 20S proteasome evolved from BPH or vice versa based on their distributions in the bacteria. BPH has an extended loop 2 relative to both HslV and Anbu (highlighted in green in Fig. 5). BPH's helix 1 is also extended relative to HslV, but it does not have a C-terminal extension. Structurally BPH shares similarities with both Anbu and HslV, but it is probably not an intermediate structure because of its narrow distribution within the β-proteobacteria.

The 20S proteasome and HslV both degrade proteins in an ATP-dependent manner. The ATPase binding surfaces in these complexes are very different; the 20S proteasome is four layers and HslV is two layers (Fig. 4 in Cavalier-Smith 2006). This means that the ATPases are binding to opposite faces of the proteasome subunit in two- and four-layered proteasomes. We could postulate as to whether Anbu forms a two- or four-layered biological unit by determining whether its ATPase is more like HslU or the ATPases associated with the 20S proteasome. A BLAST (Altschul et al. 1990) search was run against cyanobacteria to find potential ATPases for Anbu. The distribution of ATPase homologues in cyanobacteria is informative since they do not have HslV and only some have Anbu. We were unable to locate a known proteasomal ATPase that matched the distribution of Anbu or BPH. This could mean that an ATPase is moonlighting or that Anbu or BPH is acting in an ATP-independent manner and only breaking peptides down. It is possible that one of the genes of unknown function associated with Anbu (discussed below) could be its ATPase.

It has been argued that HslV could not evolve from the 20S proteasome because the decrease in pore size from a heptamer to a hexamer would not be favorable (Cavalier-Smith 2006). Also, the loss of the inactive α subunits would be a major transition that would result in a proteasome with a large pore and no regulatory ATPase, which would not be favorable. By this same logic it is highly unlikely that Anbu or BPH could evolve from the 20S proteasome, as they appear to have only active subunits.

Our structural predictions infer that Anbu is the ancestor of the 20S proteasome. Larger structural features such as
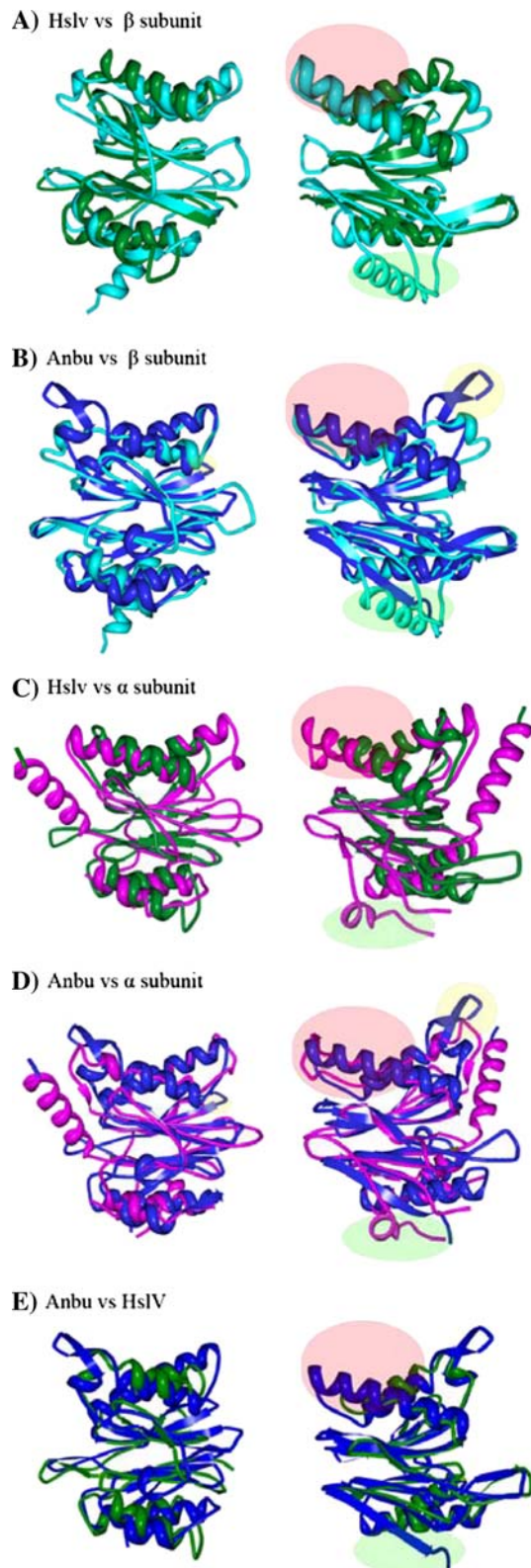
◄ **Fig. 4** Comparison of Anbu to crystal structures of known proteasomes. The image on the right is an ∼180-deg rotation of the image on the left. HslV (1E94) is green, α subunit (1Q5Q_A) is magenta, β subunit (1Q5Q_H) is cyan, and Anbu (predicted structure from *Rhodopseudomonas palustris*) is blue. Red ovals highlight an extended helix shared between Anbu and both subunits of the 20S proteasome but absent in HslV. Green ovals highlight an extended C-terminal shared between Anbu and both subunits of the 20S proteasome but absent in HslV. The yellow ovals highlight an extended turn that is unique to Anbu

to understand the biological units of Anbu and BPH. If we are correct that Anbu is the ancestor of the 20S proteasome, a structure of the complex would provide an excellent opportunity for an improved understanding of the 20S proteasome.

Function of Anbu and BPH

Anbu is found in a very diverse set of bacteria, including both oxygenic and anoxygenic phototrophs. It is also present in many species that have unique phenotypes such as *Ralstonia metallidurans*, which can withstand high metal concentrations and plays a role in the formation of gold (Reith et al. 2006), *Rhodoferax ferrireducens*, which can reduce Fe(III) (Finneran et al. 2003), and *Burkholderia xenovorans*, which is capable of degrading polychlorinated biphenyl (Goris et al. 2004). HslV expression is increased under heat shock and other stresses that cause proteins to misfold, so we searched the literature on microarray experiments to see if any of the stresses these bacteria face in these varied environments induced expression of Anbu. Anbu was not induced in *Synechocystis* sp. PCC 6803 in response to heat shock (Singh et al. 2006), UV-B light (Huang et al. 2002), salt stress, and hyperosmotic stress (Kanesaki et al. 2002). Anbu was also not differentially expressed under oxidative stress conditions (addition of H_2O_2) in *Synechocystis* (Li et al. 2004) and *Rhodobacter sphaeroides* (Zeller et al. 2005). *Pseudomonas putida* KT2440 did not induce Anbu expression in the presence of any of several different aromatic compounds, although
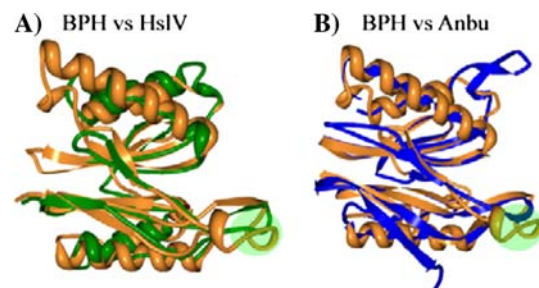
Fig. 5 Comparison of BPH (predicted structure from *Thiobacillus denitrificans*), in orange, against HslV (1E94), in green, and Anbu (predicted structure from *Rhodopseudomonas palustris*), in blue. The green oval highlights an extension unique to BPH

whether Anbu's rings are heptameric or hexameric will make for stronger evolutionary arguments. It will be necessary to get a crystal or cryoelectron microscopy structure

some triggered increased expression of HslV (Dominguez-Cuevas et al. 2006). Although these experiments do not reveal Anbu's function, they show that Anbu is not differentially expressed in several situations that HslV would be. This is functional evidence that Anbu is distinct from HslV. Future microarray experiments in these species may reveal when Anbu is induced. Unfortunately we could not find any microarray experiments with these kind of stresses for the few species that have BPH.

We compared the operons of HslV, Anbu, and BPH using the MicrobesOnline (Alm et al. 2005) operon browser (Supplemental Figs. 3–6). HslV almost always falls in the same predicted operon as HslU, and they are always predicted to be in the same regulon. We noticed that Anbu is often expressed in an operon with genes labeled as COG2307, COG2308, and COG1305 (Fig. 6a). When Anbu is not in the same operon as these three genes, they are almost always predicted to be in the same regulon. COG2307 and COG2308 are uncharacterized conserved proteins. COG2308 is predicted by Superfamily to have a glutathione synthetase ATP-binding domain. The hits to this superfamily are near the threshold of what is considered a significant hit in Superfamily. Understanding how COG2308 uses ATP may be key to understanding Anbu's function. It is possible that this uncharacterized protein interacts with Anbu, but it would have to have some other function as well since it appears in genomes that lack Anbu. COG1305 is a transglutaminase-like protein. Some bacterial transglutaminases act as proteases (Pfister et al. 1998), while others selectively cross-link proteins (Seitz et al. 2001). Either function could have interesting interactions with a proteasome. If this transglutaminase acts as a protease, it could break down the peptides that come out of Anbu into even smaller pieces. If it acts at as a cross-linker, Anbu may degrade it to regulate the levels of cross-linking in the cell. Either of these functions could also act to regulate Anbu. We compared the average number of predicted transglutaminase catalytic domains using Superfamily in genomes that have Anbu, the 20S proteasome, or neither (Table 2). Genomes that had either Anbu or the 20S proteasome both had a statistically significant

higher average occurrence of transglutaminases than genomes that had neither of these proteasomes. The species that have Anbu have over five times more transglutaminases on average than the species that lack both Anbu and the 20S proteasome. We observed the same result when we repeated this measure in genomes from just the α-proteobacteria, β-proteobacteria, and γ-proteobacteria. There was no genome that had Anbu and completely lacked transglutaminase. It should be noted that a major exception to this trend is *Rhodopirellula baltica*. It has 11 transglutaminase catalytic domains, the most of any genome in this study, but has no proteasome homologues. These proteins are predicted to have a domain with similar structure to the transglutaminase associated with Anbu, but their functions could be very different. Transglutaminases do not strictly require Anbu, but there is a definite association between them. Understanding Anbu's function will require better characterization of the different functions of bacterial transglutaminases as well as COG2307 and COG2308.

The few samples of BPH showed two operon-based patterns. In *Thiobacillus denitrificans* and *Chromobacterium violaceum*, BPH is in the same operon or regulon as ornithine carbamoyltransferase (argI or argF) and argininosuccinate synthase (argG) (Fig. 6b). Both of these proteins are involved in arginine biosynthesis, which is induced as part of the heat shock response in several species including *Bacillus subtilis* (Helmann et al. 2001) and *Desulfovibrio vulgaris* (Zhang et al. 2006). HslV and HslU are in the same operon as argF, which is next to argG in *Desulfuromonas* spp. BPH is in the same operon as heat shock protein 33 (HslO), a chaperone that is activated under oxidative stress (Akhtar et al. 2004), in *Chromobacterium violaceum*. In these species BPH appears to be acting as part of a heat shock response. This could be the result of functional conservation if we are correct that BPH evolved from HslV. Both *Thiobacillus denitrificans* and *Chromobacterium violaceum* have BPH and HslV but lack Anbu. Identifying the difference in conditions that induce expression of BPH and HslV will help explain BPH's function. However, BPH seems to play a different role in the other species that have Anbu as well. In the *Ralstonias*

Table 2 Comparison of genomic occurrence of transglutaminase-like catalytic domains

	Neither	Anbu	20S proteasome	αβγ-neither	αβγ-Anbu
Mean	0.752525253	3.17241379	2.454545455	0.722222222	3.24
Variance	1.831846383	3.14778325	3.072727273	1.596129838	3.606667
P-value		2.27E-08	0.004472988		3.42E-07

Note. The average number of transglutaminases per genome was computed for genomes with different proteasome contents. Neither means genomes that have neither Anbu nor the 20S proteasome. The *p*-values are the probability of the mean being that much higher compared to the genomes that have neither proteasome. This calculation was repeated in just the genomes of α-proteobacteria, β-proteobacteria, and γ-proteobacteria. Genomes that have Anbu or the 20S proteasome have significantly more transglutaminase domains than genomes that have neither Anbu nor the 20S proteasome

and *Poloramonas* BPH was in an operon with the three
genes encoding the pyruvate dehydrogenase complex. In
Escherichia coli these genes are in the same operon as the
autoregulator pdhR. pdhR represses transcription of that
operon in the absence of pyruvate (Quail and Guest 1995).
BPH may play a similar regulatory role, degrading the
pyruvate dehydrogenase complex in the absence of pyru-
vate. It would be interesting if transcriptional regulation
was replaced by regulation at the level of degradation. It is
possible that BPH has been adapted to both regulatory and
heat shock roles, but it is hard to draw a conclusion on how
conserved these operons are from a sample of only six
species.

Discussion

Anbu's position in the trees and its hypothetical structure
make a compelling case for its being ancestral to the 20S
proteasome found in the actinobacteria. Sequence and
functional data indicate that BPH evolved from HslV.
Determining whether HslV or Anbu is older is a much
more challenging problem. Cavalier-Smith (2006) argues
that the oldest groups of bacteria are the Cyanobacteria,
Hadobacteria, and Chlorobacteria (from youngest to old-
est). Neither HslV nor Anbu has been found in any
chlorobacterial genome. Anbu is present in several

Cyanobacteria. *Thermus thermophilus*, a Hadobacteria, has
HslV. Its sequence is related to HslV of other hyper-
thermophiles, which may reflect a horizontal transfer. This
makes it hard to say which proteasome is older based on
their distribution in these bacteria. The pattern of repeated
loss of Anbu in genomes that have HslV infers that HslV
replaced Anbu. In this scenario Anbu would be the oldest
proteasome. Solving the biological units of Anbu and
probing its interactions may also help sort this out by
showing which transitions between proteasomes are the
most favorable.

Its has been argued that the actinobacteria are ancestral
to both the eukaryotes and the archaea because they are the
only group of bacteria with a 20S proteasome, while the
20S proteasome is found in all eukaryotes and Archaea
(Cavalier-Smith 2006). Although we have shown that
Anbu is more likely to be the ancestor of the 20S protea-
some than HslV, our data still support the actinobacteria
having the original 20S proteasome. A horizontal transfer
of the 20S proteasome to the actinobacteria as proposed by
Gille et al. (2003) is unnecessary with the discovery of
Anbu. Our work also shows that bacterial evolution has
tinkered with the proteasome much more than previously
thought. We have found bacteria that have many different
combinations of the 20S proteasome, HslV, and Anbu. It is
important to note that there is evidence that any protea-
some can and has been lost under the right circumstances.

Many of these conclusions can only be drawn because of the large number of genomes we looked at in this study, but this number will be considered small in a few years. There may be many groups of species-specific proteasomes like BPH in parts of the tree of life that have not been sampled. Finding a group of bacteria outside the actinobacteria with a true 20S proteasome would have major implications for the evolution of the eukaryotes and archaea, but until then the actinobacteria proteosomes seem the most plausible ancestor of eukaryotic and archaeal proteasomes.

Note Added in Proof

A recent microarray experiment revealed that Anbu and the operon we have defined are the most upregulated genes in *Pseudomonas putida* under nitrogen-limiting conditions (Hervas et al. 2008). The authors propose that Anbu and its operon may play a role in protein turnover in response to changing nitrogen availability. This confirms that Anbu is functionally distinct from HslV which was not upregulated under these conditions.

References

Akhtar MW, Srinivas V, Raman B, Ramakrishna T, Inobe T, Maki K, Arai M, Kuwajima K, Rao Ch M (2004) Oligomeric Hsp33 with enhanced chaperone activity: gel filtration, cross-linking, and small angle x-ray scattering (SAXS) analysis. J Biol Chem 279:55760–55769

Alm EJ, Huang KH, Price MN, Koche RP, Keller K, Dubchak IL, Arkin AP (2005) The MicrobesOnline Web site for comparative genomics. Genome Res 15:1015–1022

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242

Bochtler M, Ditzel L, Groll M, Huber R (1997) Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. Proc Natl Acad Sci USA 94:6070–6074

Bochtler M, Ditzel L, Groll M, Hartmann C, Huber R (1999) The proteasome. Annu Rev Biophys Biomol Struct 28:295–317

Burton RE, Baker TA, Sauer RT (2005) Nucleotide-dependent substrate recognition by the AAA+ HslUV protease. Nat Struct Mol Biol 12:245–251

Cavalier-Smith T (2006) Rooting the tree of life by transition analyses. Biol Direct 1:19

Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. Annu Rev Biochem 65:801–847

Darwin KH, Lin G, Chen Z, Li H, Nathan CF (2005) Characterization of a *Mycobacterium tuberculosis* proteasomal ATPase homologue. Mol Microbiol 55:561–571

De Mot R (2007) Actinomycete-like proteasomes in a Gram-negative bacterium. Trends Microbiol 15:335–338

Dominguez-Cuevas P, Gonzalez-Pastor JE, Marques S, Ramos JL, de Lorenzo V (2006) Transcriptional tradeoff between metabolic and stress-response programs in *Pseudomonas putida* KT2440 cells exposed to toluene. J Biol Chem 281:11981–11991

Drummond A, Kearse M, Heled J, Moir R, Thierer T, Ashton B, Wilson A, Stones-Havas S (2006) Geneious v2.5

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Finneran KT, Johnsen CV, Lovley DR (2003) *Rhodoferax ferrireducens* sp. nov., a psychrotolerant, facultatively anaerobic bacterium that oxidizes acetate with the reduction of Fe(III). Int J Syst Evol Microbiol 53:669–673

Gille C, Goede A, Schloetelburg C, Preissner R, Kloetzel PM, Gobel UB, Frommel C (2003) A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome. J Mol Biol 326:1437–1448

Glickman MH, Ciechanover A (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol Rev 82:373–428

Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. Nature 426:895–899

Goris J, De Vos P, Caballero-Mellado J, Park J, Falsen E, Quensen JF, 3rd, Tiedje JM, Vandamme P (2004) Classification of the biphenyl- and polychlorinated biphenyl-degrading strain LB400T and relatives as *Burkholderia xenovorans* sp. nov. Int J Syst Evol Microbiol 54:1677–1681

Gough J, Karplus K, Hughey R, Chothia C (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol 313:903–319

Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik HD, Huber R (1997) Structure of 20S proteasome from yeast at 2.4 A resolution. Nature 386:463–471

Groll M, Bochtler M, Brandstetter H, Clausen T, Huber R (2005) Molecular machines for protein degradation. Chembiochem 6:222–256

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Helmann JD, Wu MF, Kobel PA, Gamo FJ, Wilson M, Morshedi MM, Navre M, Paddon C (2001) Global transcriptional response of *Bacillus subtilis* to heat shock. J Bacteriol 183:7318–7328

Hershko A (2005) The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. Cell Death Differ 12:1191–1197

Hervas AB, Canosa I, Santero E (2008) Transcriptome analysis of *Pseudomonas putida* in response to nitrogen availability. J Bacteriol 190:416–420

Hill CP, Masters EI, Whitby FG (2002) The 11S regulators of 20S proteasome activity. Curr Top Microbiol Immunol 268:73–89

Honda D, Yokota A, Sugiyama J (1999) Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. J Mol Evol 48:723–739

Huang L, McCluskey MP, Ni H, LaRossa RA (2002) Global gene expression profiles of the cyanobacterium *Synechocystis* sp. strain PCC 6803 in response to irradiation with UV-B and white light. J Bacteriol 184:6845–6858

Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol 146:11–31

Iyer LM, Burroughs AM, Aravind L (2006) The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. Genome Biol 7:R60

Kanesaki Y, Suzuki I, Allakhverdiev SI, Mikami K, Murata N (2002) Salt stress and hyperosmotic stress regulate the expression of different sets of genes in *Synechocystis* sp. PCC 6803. Biochem Biophys Res Commun 290:339–348

Kelley LA, MacCallum RM, Sternberg MJ (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 299:499–520

Li H, Singh AK, McIntyre LM, Sherman LA (2004) Differential gene expression in response to hydrogen peroxide and the putative PerR regulon of *Synechocystis* sp. strain PCC 6803. J Bacteriol 186:3331–3345

Lo I, Denef VJ, Verberkmoes NC, Shah MB, Goltsman D, DiBartolo G, Tyson GW, Allen EE, Ram RJ, Detter JC, Richardson P, Thelen MP, Hettich RL, Banfield JF (2007) Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature 446:537–541

Lowe J, Stock D, Jap B, Zwickl P, Baumeister W, Huber R (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 A resolution. Science 268:533–539

Moreland JL, Gramada A, Buzko OV, Zhang Q, Bourne PE (2005) The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. BMC Bioinformatics 6:21

Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540

Orlowski M, Wilk S (2003) Ubiquitin-independent proteolytic functions of the proteasome. Arch Biochem Biophys 415:1–5

Pfister P, Wasserfallen A, Stettler R, Leisinger T (1998) Molecular analysis of Methanobacterium phage psiM2. Mol Microbiol 30:233–244

Quail MA, Guest JR (1995) Purification, characterization and mode of action of PdhR, the transcriptional repressor of the pdhR-aceEF-lpd operon of *Escherichia coli*. Mol Microbiol 15:519–529

Rabinovich E, Bajorek M, Glickman M, Bar-Nun S (2006) Proteasome channel opening as a rate-limiting step in the ubiquitin-proteasome system. Israel J Chem 46:219–224

Reith F, Rogers SL, McPhail DC, Webb D (2006) Biomineralization of gold: biofilms on bacterioform gold. Science 313:233–236

Ruepp A, Eckerskorn C, Bogyo M, Baumeister W (1998) Proteasome function is dispensable under normal but not under heat shock conditions in *Thermoplasma acidophilum*. FEBS Lett 425:87–90

Seemuller E, Lupas A, Stock D, Lowe J, Huber R, Baumeister W (1995) Proteasome from *Thermoplasma acidophilum*: a threonine protease. Science 268:579–582

Seemuller E, Lupas A, Baumeister W (1996) Autocatalytic processing of the 20S proteasome. Nature 382:468–471

Seitz A, Schneider F, Pasternack R, Fuchsbauer HL, Hampp N (2001) Enzymatic cross-linking of purple membranes catalyzed by bacterial transglutaminase. Biomacromolecules 2:233–238

Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 11:739–747

Singh AK, Summerfield TC, Li H, Sherman LA (2006) The heat shock response in the cyanobacterium *Synechocystis* sp. Strain PCC 6803 and regulation of gene expression by HrcA and SigB. Arch Microbiol 186:273–286

Smith DM, Benaroudj N, Goldberg A (2006) Proteasomes and their associated ATPases: a destructive combination. J Struct Biol 156:72–83

Tanahashi N, Murakami Y, Minami Y, Shimbara N, Hendil KB, Tanaka K (2000) Hybrid proteasomes. Induction by interferon-gamma and contribution to ATP-dependent proteolysis. J Biol Chem 275:14336–14345

Ustrell V, Hoffman L, Pratt G, Rechsteiner M (2002) PA200, a nuclear proteasome activator involved in DNA repair. EMBO J 21:3516–3525

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–71

Yang S, Doolittle RF, Bourne PE (2005) Phylogeny determined by protein domain content. Proc Natl Acad Sci USA 102:373–378

Zeller T, Moskvin OV, Li K, Klug G, Gomelsky M (2005) Transcriptome and physiological responses to hydrogen peroxide of the facultatively phototrophic bacterium *Rhodobacter sphaeroides*. J Bacteriol 187:7232–7242

Zhang W, Culley DE, Hogan M, Vitiritti L, Brockman FJ (2006) Oxidative stress and heat-shock responses in *Desulfovibrio vulgaris* by genome-wide transcriptomic analysis. Antonie Van Leeuwenhoek 90:41–55

Zwickl P, Ng D, Woo KM, Klenk HP, Goldberg AL (1999) An archaebacterial ATPase, homologous to ATPases in the eukaryotic 26 S proteasome, activates protein breakdown by 20 S proteasomes. J Biol Chem 274:26008–26014