In eukaryotes, the covalent attachment of ubiquitin chains directs substrates to the proteasome for degradation. Recently, ubiquitin-like modifications have also been described in the archaeal domain of life. It has subsequently been hypothesized that ubiquitin-like proteasomal degradation might also operate in these microbes, since all archaeal species utilize homologues of the eukaryotic proteasome. Here we perform a structural and biochemical analysis of a ubiquitin-like modification pathway in the archaeon *Sulfolobus acidocaldarius*. We reveal that this modifier is homologous to the eukaryotic ubiquitin-related modifier Urm1, considered to be a close evolutionary relative of the progenitor of all ubiquitin-like proteins. Furthermore we demonstrate that urmylated substrates are recognized and processed by the archaeal proteasome, by virtue of a direct interaction with the modifier. Thus, the regulation of protein stability by Urm1 and the proteasome in archaea is likely representative of an ancient pathway from which eukaryotic ubiquitin-mediated proteolysis has evolved.
ubiquitin (Ub) and the related ubiquitin-like proteins (Ubls), belong to a family of modifiers that covalently attach to a diverse array of cellular targets, and orchestrate a wide variety of regulatory processes in the eukaryotic cell. The role of ubiquitylation (also referred to as ubiquitination) in proteasome-mediated proteolysis pathways has been well characterised. In addition to the roles in protein turnover, Ub and Ubl conjugation also operates as a post-translational signal to regulate diverse cellular processes and pathways including DNA replication, DNA damage repair, transcription, cell-cycle control, chromatin modification, protein trafficking, autophagy and innate immunity.

Ubls were originally thought to be confined to the eukaryotic domain of life. However, the discovery of an unanticipated structural homology between the bacterial sulphur-transfer proteins ThiS and MoaD, and the eukaryotic Ub/Ubl proteins led to speculation that primitive prokaryotic homologues are the antecedents of the eukaryotic Ub/Ubl family. These prokaryotic ubiquitin-like homologues share the characteristic Ub/Ubl-like β-grasp structural fold and are activated with similar catalytic chemistry to Ub/Ubls. However, their primary biological role is to mobilize sulphur during biosynthetic reactions.

The eukaryotic Urm1 (ubiquitin-related modifier-1) protein, which displays both structural and amino-acid sequence homology to the prokaryotic ThiS and MoaD, has been proposed as a candidate for the evolutionary ‘missing link’ between the eukaryotic Ub/Ubl family and the ancestral prokaryotic sulphur-transfer proteins. Although ubiquitylation reactions proceed via three well-characterised enzymatic steps, known as the E1-E2-E3 cascade, urmylation events are apparently less complex and only require an E1-like enzyme to generate covalently modified substrate. In addition, Urm1 also mediates sulphur mobilization reactions during post-translational tRNA modification reaction. This functional duality led to further speculation that some prokaryotic sulphur-transfer Ubl homologues might also play roles in substrate conjugation.

Bona fide ubiquitin-like modifications were eventually identified beyond the eukaryotic domain with the discovery of the SAMP (small Ubl archaeal modifier protein) moieties. Covalent attachment of the modifier, referred to as SAMPylation, occurs between the conserved C-terminal di-glycine motif of SAMP, and specific lysine residues on a variety of target proteins, in a manner reminiscent of urmylation and ubiquitylation. It has been revealed that the E1-like SAMPy catalysts adenylate the terminal glycine of the modifier before the covalent attachment to the substrate. Furthermore, the Haloferax volcanii and Methanosarcina acetivorans SAMP crystal structures confirmed the predicted Ub/Ubl-like β-grasp fold. Originally identified in the halophilic archaeon H. volcanii, bioinformatic analyses have since indicated that SAMPs are widely distributed across all the archaeal kingdoms.

By comparison, bacteria of the Actinobacteria and Nitrospirae families employ an analogous covalent modification system referred to as pypylation. These bacterial species unusually possess homologues of the eukaryotic/archaeal type proteasomal machinery, and in addition to the more commonly distributed bacterial proteolytic apparatus. Attachment of the modifier Pup (prokaryotic ubiquitin-like protein) directs substrate delivery to the proteosome, but unlike ubiquitin-mediated proteolysis the Pup tag can also be destroyed along with the target. Bacterial Pup does not belong to the β-grasp family, however, and has presumably evolved independently from the eukaryotic Ub and archaeal SAMP proteins.

Although the eukaryotic ubiquitylation and archaeal SAMPylation pathways have evolved from a common antecedent, and despite previous studies suggesting a link between SAMPylation and proteasomal degradation, direct evidence of an interaction between the modifier and the proteasome is still lacking. To investigate the possible crosstalk between SAMPylated substrates and the archaeal proteosome, we investigated these processes in the biochemically tractable thermophilic archaeon Sulfolobus acidocaldarius. Here we solve

---

**Figure 1 | Crystal structure of the S. solfataricus Urm1 protein, and comparison with β-grasp fold structural homologues.**

- **a.** S. solfataricus Urm1 structure shown as a ribbon, and coloured according to secondary structure elements; loops are coloured grey, α-helices are shown in wheat, while the β-sheet is depicted in pale blue (all four α-helices and all four β-strands are numbered). (b-g) Ribbon representation of the M. musculus Urm1, S. cerevisiae Urm1, M. acetivorans SAMP1, H. volcanii SAMP1, T. thermophilus MoaD and A. orientalis CysO β-grasp fold homologues, respectively; all homologues were identified by DALI and VAST searches. (h) Ribbon representation of H. sapiens ubiquitin. All panels are coloured as described in (a). Figures generated using PyMOL.
the crystal structure of an Urm1/SAMP homologue from the related species *S. solfataricus*, which reveals similarities to both eukaryotic Urm1 proteins, and other archaenal SAMPs. Using tandem mass-spectrometry, we demonstrate that this modifier is conjugated to a variety of targets both in *vivo* and in *vitro*. Furthermore, we demonstrated that modified substrates are recognized directly by the *S. acidocaldarius* 20S core proteasome, and by the proteasome-activating nucleotidase (PAN) ATPase.44 Taken together, our data suggest that Urm1/SAMP modification acts as a signal for substrate recognition by the archaean proteasome. We discuss our findings in relation to the evolution of Urm1/SAMPs, other Ub/Ubl modifiers, and proteasomal targeting systems.

Results

Crystal structure of an Urm1 homologue in *S. solfataricus*. We initially searched for eukaryotic-like Ub and Ubl homologues in the genomes of the thermophilic archaea *S. acidocaldarius* and *S. solfataricus*. A eukaryotic Urm1 homologue was identified in both crenarchaeal species by BLAST searches (Supplementary Fig. 1), in agreement with an earlier bioinformatics study.30 Although Urm1 also shares amino-acid sequence homology with both the *Escherichia coli* ThiS and MoaD sulphur transfer proteins,16 PHYRE2 homology searches revealed that these archaeal Urm1 proteins appear more closely related to eukaryotic counterparts, than to the bacterial ThiS/MoaD sulphur transfer components (Supplementary Table 1).

We substantiated our comparison of the archaenal Urm1/SAMP and eukaryotic Urm1 family proteins by determining the crystal structure of *S. solfataricus* Urm1 at 2.2 Å resolution (Fig. 1a and Supplementary Fig. 2; also see Table 1 for refinement and model statistics). Dali5 and VAST46 fold recognition searches of the *S. solfataricus* Urm1 structure revealed homology to a number of eukaryotic Urm1 homologues, as well as to the archaenal *M. acetivorans* and *H. volcanii* SAMP1 proteins (Supplementary Table 2; Fig. 1a–e). Additionally, homology was found to *T. thermophilum* MoaD38, and the bacterial sulphur transfer protein CysO from *Mycobacterium tuberculosis* and *Amycolatopsis orientalis* (Supplementary Table 2). Unlike other bacterial MoaD/ThiS family homologues, the *T. thermophilum*, and *A. orientalis* proteins displayed exactly the same connectivity of secondary structural elements as the eukaryotic/archaenal Urm1/SAMP modifiers (Fig. 1f,g).

As predicted, the *S. solfataricus* Urm1 adopts a ubiquitin-like β-grasp fold (Fig. 1f), formed from a four-stranded β-sheet packed against a central α-helix (Fig. 1a). Similar to Ub and other Ubls, the β-strands are organized in the order β2-β1-β4-β3. The inner surface of the sheet, facing the α-helix, harbours several buried residues that form the hydrophilic core of the β-grasp fold. Four α-helices are also evident, with the first connecting strands β1 and β2, the second and third located between strands β2 and β3, and the final helix positioned on a long loop between strands β3 and β4 (Fig. 1a). This order of secondary structural features is also conserved in eukaryotic Urm1 proteins, and in the *M. acetivorans* SAMP1 structure (PDB:2L52)30 (Fig. 1a–d). The bacterial MoaD family of proteins, and the *H. volcanii* SAMP1 fold (PDB:3P00)31, also display the same order of secondary structural elements, up to the final α-helix, which is often absent (Fig. 1e)13,47. Interestingly, this final helix, linking the last two β-strands, is a feature shared by the Ub/Ubl and ThiS families.13 Inspection of the surface hydrophobicity of the *S. solfataricus* Urm1 structure revealed the presence of a solvent-exposed hydrophobic patch in close proximity to the C-terminal tail (Fig. 2a). This surface is composed of the conserved hydrophobic residues L52 and L54 on strand β3, I77 on strand β4, and I79 on the C terminus of the protein. The conserved residues G8 and I11 also contribute to an adjacent hydrophilic surface, although the K7 lysine residue, conserved across all of the *Sulfobolus* species, bisects these two solvent-exposed hydrophobic regions (Fig. 2a). Equivalent hydrophobic patches have been described previously in ubiquitin, Urm1 and MoaD, where these surfaces are essential for the interaction with partner proteins, including their cognate E1 activating enzyme.3,18,47

A second exposed hydrophobic surface is also present in *S. solfataricus* Urm1, formed by the residues I19 and V21 on strand β2, I33 at the end of helix α2, and F16, on the loop adjoining helix α1 and strand β2 (Fig. 2b). Notably the I33 residue, conserved in the *Sulfobolales* family, was somewhat sunken, creating a shallow hydrophobic depression. This second shallow hydrophobic surface is separate from the first hydrophobic patch, which is located on the opposite face of the central β-sheet (Fig. 2b). We observed in the structure that a leucine (L67) on the final loop on one Urm1 subunit protrudes into the hydrophobic depression of a neighbouring Urm1 chain to form a dimerization interface (Fig. 2c). However, this arrangement may simply be an artefact as a result of packing in the crystal lattice.

The structure also reveals a narrow hydrophobic channel, juxtaposed to the first hydrophobic patch abutting the C-terminal tail. This trough is lined with the surface exposed conserved residues L27, L47 and G50, and extends from the base of the C-terminal tail to the opposite end of the structure (Fig. 2d). The conserved residues W60 and R61, on helix α4, form one of the sidewalls of this groove, separating it from the nearby hydrophobic patch 1. Interestingly, a similar groove has been reported previously in the *M. acetivorans* SAMP1 structure (PDB:2L52)30, and this feature can also be observed in the eukaryotic Urm1

| Table 1 | Data collection and refinement statistics. |
|---|---|
| **Sso Urm1** |
| Data collection |
| Space group | P 2₁, 2₁, 2₁ |
| Cell dimensions |
| a, b, c (Å) | 31.49 65.19 109.36 |
| α, β, γ (°) | 90 90 90 |
| Wavelength (Å) | 1.5418 |
| Resolution (Å) | 28 – 2.20 (2.30 – 2.20)* |
| R<sub>merge</sub> / R<sub>merge</sub> |
| 0.0642 (0.3046) |
| R<sub>work</sub> |
| 0.2726 (0.6505) |
| Completeness (%) | 99.8 (99.3) |
| Redundancy | 12.71 (8.55) |
| Wilson B factor | 25.60 |
| Refinement |
| No. reflections | 12,034 |
| R<sub>work</sub>/R<sub>free</sub> |
| 0.1933/0.2278 |
| No. atoms |
| Protein | 1,330 |
| Ligand/ion | 5 |
| Water | 89 |
| B-factors |
| Protein | 38.80 |
| Ligand/ion | 92.70 |
| Water | 39.40 |
| R.m.s. deviations |
| Bond lengths (Å) | 0.003 |
| Bond angles (°) | 0.66 |
| Ramachandran favored (%) | 98.7 |
| Ramachandran outliers (%) | 0.0 |
| MolProbity Clashscore | 0.36 |

*Values in parentheses are for highest resolution shell.*
protein structures (PDB:2AX5)\(^{18}\), but is absent from the bacterial MoaD-like family of proteins (PDB:1JW9)\(^{47}\). A functional role for this groove has yet to be ascribed in eukaryotic and archaeal Urm1/SAMP homologues.

The Urm1 modifier is activated by an ELSA homologue. All eukaryotic Ub and Ubl proteins must be bound and activated by a cognate adenylating enzyme, often referred to as the E1 enzyme. This hydrolyses ATP and catalyses the adenylation of the terminal glycine of the modifier, before substrate conjugation\(^{24}\).

In the case of the ancestral MoaD/ThiS sulphur carriers, this step is also necessary for sulphur transfer, and is performed by the MoeB and ThiF E1-like enzymes, respectively\(^{14,15}\). Archaeal E1-like/ELSA homologues have been identified previously in both \(M.\ acetivorans\) and \(H.\ volcanii\) species\(^{29,30,48}\). We identified a single ELSA homologue (SacI0179) in the \(S.\ acidocaldarius\) genome. This displayed approximately equal sequence identity to both the bacterial ThiF enzyme and also to the eukaryotic Uba4p protein, the E1-like homologue for the eukaryotic urmylation pathway. However, a C-terminal rhodanese domain, harboured in eukaryotic E1 enzymes, was absent in the \(S.\ acidocaldarius\) genome.

**Figure 2** | **Architecture of \(S.\ solfataricus\) Urm1, identifying exposed hydrophobic surfaces and dimerization interfaces.** (a) Left: \(S.\ solfataricus\) Urm1, showing the exposed residues (red sticks) that form the C-terminal exposed hydrophobic patch. The conserved K7 lysine is also shown in dark blue. Right: surface view, identifying the exposed hydrophobic patch 1, coloured in red. (b) Left: view of the opposing edge of the \(\beta\)-sheet, on Urm1 chain B, revealing a second hydrophobic patch composed of four exposed residues (F16, I19, V21 and I33; red sticks). Right: Surface view displaying the hydrophobic patch 2, in red. (c) Left: view of Urm1 chains A and B, showing the F16, I19, V21 and I33 patch on both chains. The L67 residue of chain A protrudes into the I33 centred hydrophobic patch on chain B. Right: Surface view of the Urm1 dimer, showing the location of the surface exposed I33 residue on chain A. (d) Left: a hydrophobic groove, lined with the residues G50, L47 and L27 (red sticks) lies adjacent to the C-terminal exposed hydrophobic patch. Right: Surface view of the C-terminal groove. In all panels secondary structure elements are coloured as in Fig. 1 (loops in grey, \(\alpha\)-helices in wheat, and the \(\beta\)-sheet in pale blue). Figures generated using PyMOL\(^{68}\).
Figure 3 | Production of covalent ATP-dependent Urm1 conjugates by the ELSA/Uba4p/ThiF E1-like enzyme. (a) Auto-urmylation of the ELSA/Uba4p/ThiF enzyme. 30 μg N-terminally His-tagged Urm1 protein and 90 μg ELSA/Uba4p/ThiF enzyme were incubated together at 70 °C for 1 h in either the presence or absence of 2.5 mM ATP, and retrieved by Ni-NTA agarose pull-down. Left; Coomassie stained gel of the pull-down. Right; western blot of a duplicate gel, probed with an anti-His antibody. Lanes 1 and 4: Urm1 only controls, without or with ATP, respectively; lanes 2 and 5: untagged ELSA/Uba4p/ThiF controls, without or with ATP, respectively; lanes 3 and 6: Urm1 plus ELSA/Uba4p/ThiF incubations, without or with ATP, respectively. (b) Urmylation of the Saci0666 thermosome subunit. 30 μg N-terminally His-tagged Urm1 protein plus 15 μg ELSA/Uba4p/ThiF enzyme and 75 μg C-terminally His-tagged Saci0666 were incubated together and pulled-down as described in (a). Lane 1: reaction in the presence of 2.5 mM ATP; lane 2: no ATP control. Left; Coomassie stained gel of the pull-down. Middle and Right; western blot of duplicate gels, probed with an anti-His, or anti-Urm1 antibodies, respectively. (c) Example MS/MS spectra of the diglycine modified peptides from the in vitro urmylation assay (Saci0671 [FBP] and Saci0669 [Urm1]) and in vivo Urm1 overexpression (Saci0051 [Rad50] and Saci0656 [PAN]). m/z values of the precursor ions are shown in the top left of each panel. (+2) or (+3) indicates doubly or triply charged precursor ions, respectively. Spectra show the annotated peaks that are due to C-terminal y (coloured red) and N-terminal b (coloured blue) fragment ions. In each case, the m/z values of the precursor ions and the m/z values of the fragment ions are consistent with diglycine modified lysine residues. The amino-acid sequence of the chymotrypsin-generated peptide, including the di-glycine modified lysine, is shown within the grey box in each example. Other examples are shown in Supplementary Fig. 5.
enzyme, as has been reported previously for other archaeal and bacterial E1-like proteins.

To test whether the ELSA/Uba4p/ThiF homologue was indeed the E1 activating enzyme, we purified untagged ELSA/Uba4p/ThiF, and an N-terminally His-tagged Urm1; both *S. acidocaldarius* proteins were expressed heterologously in *E. coli*. In order to promote complex formation, the two proteins were then incubated together, in either the presence or absence of ATP, before the potential interaction was examined by pull-down assay (Fig. 3a). We observed the untagged ELSA/Uba4p/ThiF associating with the tagged Urm1 protein independently of the addition of ATP. However, the inclusion of the nucleotide resulted in the formation of a conjugate between the two components. Importantly, this conjugate survived boiling in the presence of reducing agent and SDS, indicative of a covalent attachment (Fig. 3a; lane 6). The conjugate observed here is consistent with the previously reported covalent modification of SAMPs in other archaean meditated by the ELSA, following adenylation of the modifier.

Having demonstrated that the ELSA/Uba4p/ThiF homologue was competent to activate the Urm1 protein, ultimately resulting in the auto-urylation of the E1-like enzyme, we next examined the ability of the complex to modify an exogenous target protein. Given the strong genomic linkage observed between Urm1 and the β-subunit of the thermosome (Saci0666) (see Supplementary note and Supplementary Fig. 3), we purified this chaperonin component in *E. coli* and included the protein in the *in vitro* Urm1 activation reaction. Addition of this protein in the assay resulted in the formation of a clear ATP-dependent covalent conjugate, confirming that the ELSA/Urm1 complex can modify exogenously added substrates *in vitro* (Fig. 3b).

Mass spectrometry identification of urmylated substrates. It has been established that Ub, eukaryotic Urm1, and archaeal SAMP conjugates are formed between the C-terminal glycine of the modifier and the ε-amino-group of acceptor lysine residues on the target proteins. *S. acidocaldarius* Urm1 is also covalently attached to substrates by the same linkage. *S. acidocaldarius* cell-free extract was subjected to an urmylation reaction, using His-tagged Urm1 and untagged ELSA/Uba4p/ThiF enzyme, in the presence of ATP. The resultant His-tagged conjugates were retrieved by pull-down, separated by reducing SDS–polyacrylamide gel electrophoresis (SDS–PAGE), and subsequently digested in-gel with chymotrypsin, before analysis by tandem mass spectrometry (GeLC-MS/MS) (Supplementary Fig. 4A). In parallel to this *in vitro* approach, we also overexpressed His-tagged Urm1 protein from an exogenous expression vector in *S. acidocaldarius* cells. The *in vivo* generated Urm1 conjugates were collected by pull-down following cell disruption, and the resulting material was also analysed by GelC-MS/MS (Supplementary Fig. 4B).

*Chymotrypsin* treatment resulted in the hydrolysis of the *S. acidocaldarius* Urm1 peptide backbone between the H82 residue and the penultimate G83 residue of a C-terminal di-glycine motif (G83G84). Hence, if Urm1 was conjugated to a target protein, the chymotrypsin treatment resulted in a fragment harbouring the di-glycine motif, branching from a lysine residue on the linear peptide backbone of the substrate. The addition of a di-glycine motif does not increase the size of the peptides beyond the detection limits of the mass spectrometer, thereby facilitating identification of conjugates.

Analysis of both the *in vivo* and *in vitro* urmylation reactions revealed the covalent attachment of the di-glycine motif via an isopeptide bond to the ε-amino-group on a wide variety of substrate lysine residues (Fig. 3c, Supplementary Fig. 5, and Supplementary Table 3A and B). In total, 29 distinct substrate modifications were identified by the *in vitro* approach, while 25 modifications were observed following the *in vivo* Urm1 overexpression (Supplementary Table 3A and B). Interestingly, several of these components tallied with the proteins predicted to be involved in urmylation pathways by a bioinformatic analysis of the genomic context surrounding the Urm1 gene (see Supplementary Note and Supplementary Fig. 3). For example, *in vitro* modifications were identified on the proteasome β-subunit (Saci0662), and fructose 1,6-bisphosphatase (FBP), while conjugates on a proteasome assembly chaperone (PAC2; Saci0658), PAN (Saci0656), an ATPase RNAasel inhibitor (RL1), and a Hera/FtsK/TrwB superfamily protein (Saci0667) were observed in the *in vivo* analysis (Supplementary Table 3A and B). Furthermore, *in vivo* and *in vitro* modifications were also detected on several ribosomal subunits, and also on a number of tRNA synthases (Supplementary Table 3A and B). Conjugates were also detected on a number of metabolic enzymes, in addition to FBP, including phosphoglycerate kinase (PGK), phosphoenolpyruvate (PEP) synthase, aldehyde dehydrogenase, alcohol dehydrogenase, and acetyl-CoA synthetase (Supplementary Table 3A and B). Thus, it appears that urmylation may also be involved in the regulation of key metabolic pathways.

The identification of Urm1 modifications on the core 20S proteasome, the PAN proteasome regulatory ATPase and the PAC2 proteasome assembly chaperone provided further evidence to suggest a link between urmylation and the proteasome degradation pathway. Modifications were also observed on the Urm1 protein itself in both the *in vitro* and *in vivo* analyses (Supplementary Table 3A and B). These findings were suggestive that multiple lysines on Urm1 can be modified, and polymeric
chains could be formed, consistent with the SAMP linkages that have been reported previously.\textsuperscript{27–29} Since ubiquitin chains have well-established roles in proteasome-mediated degradation, it remains possible that Urm1/SAMP1 chains may also be used for targeting to the proteasome in the archaea. However, it should be noted that the lysine residues modified in this study do not appear to be conserved across archaeal species (Supplementary Table 3A and B; Supplementary Fig. 1).

\textbf{In vitro assembly of the S. acidocaldarius 20S proteasome.} In order to explore the possible physical and functional association between Urm1 conjugates and archael proteasome apparatus we biochemically reconstituted the 20S core proteasome from \textit{S. acidocaldarius}. Archaeal 20S proteasomes, like the homologous eukaryotic apparatus, are formed from four stacks of heptameric rings, creating an elongated hollow chamber. Proteasomes from the archaea \textit{Thermoplasma acidophilum}\textsuperscript{51,52}, \textit{Archaeoglobus}...
fulgidus\textsuperscript{53} and Methanosarcina thermophila\textsuperscript{54} had previously been reconstituted by recombinant expression in \textit{E. coli}. In these cases the proteasome is formed from only two 20S core proteasome proteins; an \(\alpha\)-subunit, which produces the self-assembling homo-heptameric \(\alpha\)-ring outer rings, and a catalytic \(\beta\)-subunit that constitutes the innermost rings. In contrast, the \textit{S. acidocaldarius} genome encodes two catalytic \(\beta\)-subunit homologues (Saci0662 and Saci0909), in addition to a single \(\alpha\)-subunit homologue (Saci0613).

We performed heterologous co-expression trials in \textit{E. coli} with the Saci0613 \(\alpha\)-subunit, in combination with either the Saci0662 or Saci0909 \(\beta\)-subunits. In both cases the \(\beta\)-subunit was C-terminally His-tagged, while the \(\alpha\)-subunit was native and untagged. We reasoned that if the 20S core subunits were expressed with the correct stoichiometry in \textit{E. coli} then the entire 28-subunit complex should purely over immobilized metal ion affinity chromatography (IMAC). Indeed, a seemingly stoichiometric complex of both \(\alpha\) and \(\beta\) components was purified following co-expression of Saci0662 with the \(\alpha\)-subunit. By contrast, the \(\beta\)-subunit did not co-elute with the tagged Saci0909 \(\beta\)-subunit following co-expression and purification over the Ni-NTA agarose column (Fig. 4a). Although the Saci0613 \(\alpha\)/0662 \(\beta\) proteasome initially appeared stoichiometric, the complex was not stable over size exclusion chromatography (Supplementary Fig. 6). We noted that during 20S assembly in both eukaryotic and prokaryotic proteasomes, an N-terminal pro-peptide is commonly cleaved from the \(\beta\)-subunits to expose a catalytic threonine (T1)\textsuperscript{55}. Equivalent threonine residues were identified in both of the \textit{S. acidocaldarius} \(\beta\)-subunits (Fig. 4a), and the \(\beta\)-subunit expression constructs were therefore redesigned to exclude the preceding N-terminal pro-peptides. The truncated \(\beta\)-subunits were then co-expressed with the \(\alpha\)-subunit. As observed with the full-length Saci0909 \(\beta\)-protein, the truncated version of this \(\beta\)-subunit failed to form a stable complex with the \(\alpha\)-subunit (Fig. 4a). In contrast, truncation of the 0662\(\beta\)-subunit stabilised the 20S assembly, and this material remained stoichiometric during size-exclusion chromatography (Fig. 4b), eluting at a volume consistent with a 28-subunit complex of 660 kDa. Furthermore, examination of the purified complex by transmission electron microscopy (TEM) revealed a cylindrical arrangement of four stacked heptameric rings, reminiscent of the overall shape of the previously determined \textit{T. acidophilum} 20S core proteasome crystal structure\textsuperscript{51} (Fig. 4c).

Although co-expression of the 0613 \(\alpha\)-subunit and the truncated 0662 \(\beta\)-subunit resulted in a macromolecular assembly that resembled the structures of the active \textit{T. acidophilum}, \textit{M. thermophila} and \textit{H. volcanii} proteasomes\textsuperscript{51,54,55}, this material showed no catalytic activity (Fig. 4d). Consequently, both the N-terminally truncated 0662 and 0909\(\beta\)-subunits were simultaneously co-expressed with the \(\alpha\)-subunit. This method resulted in formation of another stable complex, again consistent with a 28-subunit complex, as verified by TEM and SEC–MALS (size-exclusion chromatography–multi-angle laser light scattering) (Fig. 4c,e). However, the complex was now composed of all three subunits, with the Saci0909 \(\beta\) subunit incorporating into the \(\beta\)-subunit ring at a lower ratio than the Saci0662 subunit (Fig. 4a). Critically, the inclusion of the Saci0909 \(\beta\)-subunit into the 20S core conferred catalytic activity to the complex, as demonstrated by the direct degradation of the Urm1 modifier by the core 20S proteasome (Fig. 4d).

\textbf{Urm1 conjugates engage with the active 20S proteasome.} Our bioinformatic and mass-spectrometry analyses were suggestive of a link between Urm1 and the proteasome, consistent with previous studies\textsuperscript{27,33,44}. However, no categorical demonstration of a physical interaction between an archaeeal Ubl and the archaeeal 20S proteasome has been shown thus far. Having reconstituted the active \textit{S. acidocaldarius} 20S core proteasome in vitro, we proceeded to perform a series of biochemical assays to explore the potential association between the proteasome and the urmlyation pathway. Unexpectedly, we found that the active 20S core proteasome complex, with the incorporated catalytic Saci0909 \(\beta\)-subunit, was competent to process the Urm1 protein directly (Fig. 4d). By contrast, we found that a thermally stable (superfolder) green fluorescent protein (GFP)\textsuperscript{56} control protein was not processed under the same conditions (Fig. 5a), suggesting that the Urm1 protein is recognized specifically by the cylindrical proteasome complex. Furthermore, we also expressed the Saci0656 PAN protein, a homohexameric ATPase homologous to the Rpt hetero-hexameric ring of the eukaryotic proteasome, which interacts with the \(\alpha\)-ring of the core proteasome\textsuperscript{57}. SEC–MALS analysis confirmed the expected hexameric status of this complex (Supplementary Fig. 7). We observed that inclusion of PAN in the assay resulted in a clear stimulation of the Urm1 degradation (Fig. 4d).

After demonstrating that the Urm1 itself is processed by the archaeeal proteasome, we next investigated if fusion of Urm1 to a target protein would stimulate degradation of this substrate. N-terminal fusion of Urm1 to the superfolder GFP resulted in substrate association with the 20S proteasome and clear degradation of the Urm1 portion of the substrate; crucially the GFP protein was not degraded in the absence of the fused Urm1 tag (Fig. 5a). Furthermore, addition of two or three Urm1 chains, linked in tandem to the GFP substrate, resulted in enhanced substrate processing (Fig. 5a). Circular dicroscopy (CD) spectroscopy and thermal shift analyses verified that the GFP fusion proteins remained correctly folded, and were not destabilised upon fusion with Urm1 (Fig. 5b,c). In each reaction, an intermediate product was generated, slightly larger than the untagged GFP substrate (Fig. 5a). Mass spectrometry analysis confirmed that this product consisted of a short C-terminal Urm1 peptide attached to the full length GFP protein (Supplementary Fig. 8). This result suggested that the Urm1 tag of these fusion proteins engaged and entered the 20S core, but the processing reaction paused, or stalled, as the fused GFP substrate accessed the proteasome (Fig. 5a, Supplementary Fig. 8).

Although it was apparent that the Urm1 region of the GFP fusion proteins were efficiently processed by the active 20S core, it was less obvious if a proportion of the fused GFP also entered the 20S proteasome. We therefore performed the degradation assays in triplicate using equimolar ratios of the GFP substrates, and then quantified the Coomassie-stained intermediate products, and compared these with the equivalent band in the untagged GFP control. Unexpectedly, this quantification suggested that a proportion of the tagged GFP substrates were also entering and being processed by the active 20S core (Fig. 6a,b). This finding was also confirmed by directly measuring the GFP degradation in real-time by following the change in GFP fluorescence throughout the reaction (Fig. 6). It should be noted that in other systems the stably folded GFP proteins are resistant to processing by core proteasomes in the absence of an ATP-driven unfoldase\textsuperscript{58,59}. To our knowledge, this is the first time that the \textit{S. acidocaldarius} 20S proteasome has been examined to date, and whereas previous studies using other archaeal 20S cores were performed at 45°C (ref. 58), our assays were performed at the more physiological temperature of 70°C. A control reaction using AMP-PNP (a non-hydrolysable analogue of ATP) ensured that the observed GFP processing was not mediated by a contaminating \textit{E. coli} ATPase (Supplementary Fig. 9).

We also examined the ability of the PAN regulatory ATPase to recognize and unfold the Urm1-GFP fusion proteins, measuring...
Figure 5 | N-terminal fusion of Urm1 to a GFP substrate results in targeting to the 20S proteasome. (a) N-terminal fusion of Urm1 (or a tandem chain of two, or three, Urm1 subunits) to a thermally stable GFP56 substrate results in processing by the 20S proteasome. 10 μg of substrate was incubated with 20 μg of the proteasome complex for 1 h at 70 °C. Reactions using the Saci0613/Saci0662D/N/Saci0909D/N active complex are shown in lanes 2 (GFP), 4 (1XUrm1:GFP), 6 (2XUrm1:GFP) and 8 (3XUrm1:GFP); lanes 1, 3, 5 and 7: negative controls for each of the GFP substrates, respectively, using the inactive Saci0613/Saci0662D/N complex. The asterisk denotes an intermediate product in the processing reaction (mass spectrometry analysis of this product is provided in Supplementary Fig. 8). (b) Circular dichroism spectra of GFP and the Urm1:GFP fusion proteins, revealing that Urm1 fusion does not affect GFP folding. The CD spectrum for GFP alone shows negative and positive Cotton effects at about 216 and 198 nm, respectively. This indicates a β-strand dominated structure as expected. Addition of Urm1 domains to GFP resulted the appearance of another minima at about 208 nm which is due to the parallel component (relative to α-helix axis) of the π to π* transition in Urm1 α-helices. In addition, there is a shift in the positive cotton effect maximum from 198 nm towards shorter wavelength, which is probably due to the perpendicular component of the π to π* transition in Urm1 α-helices. The absence of a clear negative band at 222 nm due to the π to π* transition, which is a signature of α-helices, is most likely an indication of the relative small amount of helices in Urm1. (c) Thermal denaturation assay analyses of GFP and the Urm1:GFP fusion proteins, measuring the innate fluorescence of (superfolder) GFP56 (490-nm excitation with emission at 510 nm). The first derivative plot is displayed (change in relative fluorescence units (RFU) with time [dRFU/dT]). The maximum dRFU/dT occurs at 85 °C in all samples, indicating that addition of Urm1 to the N terminus of GFP does not affect the folding or stability of the GFP substrate.
Figure 6 | Quantification of the Urm1-GFP fusion substrate processing by the core 20S proteasome, and unfolding by the PAN ATPase. (a) 20S core processing of the superfolder GFP protein and the 1XUrm1-GFP, 2XUrm1:GFP and 3XUrm1:GFP N-terminal fusion substrates in triplicate and visualized by SDS-PAGE and Coomassie staining. Equimolar quantities (6 μM) of the substrates were incubated with 10 μg Saci0613/Saci0662AN/Saci0909AN active 20S complex at 69 °C. Lanes 1-4 display the reactions before heating, while lanes 5-8 represent the reactions upon termination after 75 min at 69 °C. (b) Quantification of the GFP-sized intermediates (boxed region) after resolution by SDS-PAGE. Gels were scanned using the ‘Coomassie Brilliant Blue Digitization’ mode on a Typhoon Imaging system (GE Healthcare). The histograms display the mean of three independent repeats and the error bars one s.d. The data were quantified using the ImageQuant software (GE Healthcare) and plotted in Microsoft Excel. (c) Verification of the direct degradation of the Urm1-GFP fusion substrates by the 20S core proteasome (at 69 °C), demonstrated by the reduction in the GFP fluorescence. Untagged GFP is not unfolded by PAN, even in the presence of ATP. In (c,d) the data points represent the mean of three independent repeats and the error bars one s.d. Data were plotted in Microsoft Excel. The GFP fluorescence was detected at 485 nm excitation and 520 nm emission wavelengths, respectively, using a PheraStar (BMG LABTECH) plate reader.

Changes in the GFP fluorescence as a readout for substrate unfolding. In contrast to the untagged GFP control, which remained stable even when ATP was included in the reactions, the Urm1-GFP fusion protein was unfolded by the PAN protein in an ATP-dependent manner (Fig. 6d). Taken together, our results provide the first biochemical evidence that Urm1/SAMP tagged substrates are targeted to the archaeal 20S proteasome for destruction, by either direct interaction with the 20S core or via the PAN regulatory ATPase.

Although ubiquitin may sometimes become conjugated to the N terminus of substrates in eukaryotes, in addition to the well-characterised modifications on internal lysine residues, it is not yet known if Urm1/SAMP proteins can attach to the N terminus of target proteins in vivo. We therefore performed an in vitro urmylation of the FBP protein, to generate intermediates conjugated to the lysine residues on the substrate and then presented this modified material to the 20S proteasomes. While the unmodified FBP protein was not processed by the proteasome (Fig. 7a), the urmylated FBP material associated with the 20S core, resulting in the processing of at least the Urm1 portion of the fused substrate (Fig. 7b); notably this processing was stimulated in the presence of PAN regulatory ATPase (Fig. 7b). Furthermore, an N-terminal fusion of the Urm1-tag to FBP also resulted in targeting of the substrate to the 20S core proteasome (Fig. 7c). Again, an intermediate product accumulated indicative of the reaction stalling at the FBP region of the substrate. It therefore appeared that, in the absence of a regulatory unfoldase, the Urm1-FBP fusion substrates stalled at the 20S proteasome, even in the presence of ATP, and were not unfolded, as has been shown in vivo for the FBP-E1 precursors.

Therefore, our study provides evidence that Urm1/SAMP proteins might be able to target substrates to the 20S proteasome in vivo, via direct interaction with the proteasome. It will be interesting to test whether Urm1/SAMP proteins can be specifically directed to the 20S proteasome in vivo, perhaps by a regulatory mechanism that involves ATP-dependent unfolding of the fusion substrate.
proteasome complex are displayed in lanes 2 (FBP) and 4 (Urm1:FBP). Lanes 1 and 3: negative controls for FBP and Urm1:FBP fusion, respectively, using the gel of PAN ATPase, and a further 12.5 mM ATP. Products were purified by Ni-NTA pulldown before resolution by SDS-PAGE. Visualized by Western blot probed with an anti-Urm1 antibody. Lane 1: Urmylated FBP reaction incubated with the inactive proteasome (negative control); lane 2: urmylated FBP reaction with processing by the active 20S proteasome; lane 3: stimulation of the processing reaction by the active proteasome following addition of the PAN ATPase. Arrows indicate the urmylated FBP product and also an urmylated ELSA/Ub4p/ThiF band. The band visible at 25 kDa (indicated by an asterisk) is an urmylated truncation product of the ELSA/Ub4p/ThiF protein. Panel (a) is a duplicate Coomassie-stained control gel of (b) demonstrating that the non-urmylated FBP substrate is not processed. (c) N-terminal fusion of a single Urm1 moiety to an FBP substrate results in targeting to the 20S proteasome. 10 μg of substrate was incubated with 20 μg of the proteasome complex for 1 h at 70°C. Reactions using the active proteasome complex are displayed in lanes 2 (FBP) and 4 (Urm1:FBP). Lanes 1 and 3: negative controls for FBP and Urm1:FBP fusion, respectively, using the inactive proteasome.

Discussion

In eukaryotic cells, the ubiquitin-mediated proteasomal degradation pathway plays critical roles in controlling diverse molecular processes through the regulation of protein turnover. Similarly, analogous bacterial systems, using the Pup modifier, target a variety of substrates for destruction at the archaeal/eukaryotic-analogous bacterial systems, using the Pup modifier, target a family of proteins to mark substrates for proteasomal destruction is a feature common to eukaryotes and archaea, and also in the 20S proteasome-bearing bacterial species.

Unlike the eukaryotic ubiquitin associated degradation system, where the modifier is cleaved from the target and recycled at the proteasome, the archaeal Urm1 protein, is destroyed by the S. acidocaldarius core 20S proteasome in vitro. Thus, the modifier apparently leads the process, guiding covalently conjugated substrates into the proteasomal cylinder (Fig. 8a). Our quantification of the 20S proteasomal processing assays suggested that a proportion of the substrates entered the proteasomal 20S core, even in the absence of a regulatory ATPase. This contrasts with previous studies of other archaeal proteasome complexes, and may reflect the fact that our assays are performed at the physiological temperature of 70°C, rather than the 45°C temperature used in earlier investigations. The mechanism of the Urm1 modifier leading the substrate destruction is highly reminiscent of the bacterial Pupylation pathway, in which the N terminus of the Pup modifier is first threaded into the core of the protease, via the Mpa regulatory ATPase, leading to the degradation of the modifier and subsequently the covalently attached substrate. Interestingly, deconjugation of Pup from modified targets, by the Dop protein, has been reported previously, while SAMPylated proteins in H. volcanii can be de-SAMPylated by JAB1/MPN/MOV34 metalloenzymes.

![Figure 7](image-url) | FPB substrates are directed to the 20S proteasome as a result of either Urm1 N-terminal fusion, or by modification on internal lysines using the ELSA/Ub4p/ThiF E1 enzyme. (a,b) In vitro urmylation of FBP and subsequent proteasomal processing by the core 20S and PAN-20S complexes. 100 μg FBP, 30 μg N-terminally His-tagged Urm1 protein and 7.5 μg ELSA/Ub4p/ThiF enzyme, plus ATP at 3 mM were incubated together at 70°C for 1 h in a 200 μl reaction. 50 μl of the urmylated reaction was then incubated at 70°C for 1 h with 20 μg 20S proteasome, with or without the addition of 20 μg PAN ATPase, and a further 12.5 mM ATP. Products were purified by Ni-NTA pulldown before resolution by SDS-PAGE. (b) Urmylated products were visualized by Western blot probed with an anti-Urm1 antibody. Lane 1: Urmylated FBP reaction incubated with the inactive proteasome (negative control); lane 2: urmylated FBP reaction with processing by the active 20S proteasome; lane 3: stimulation of the processing reaction by the active proteasome following addition of the PAN ATPase. Arrows indicate the urmylated FBP product and also an urmylated ELSA/Ub4p/ThiF band. The band visible at 25 kDa (indicated by an asterisk) is an urmylated truncation product of the ELSA/Ub4p/ThiF protein. Panel (a) is a duplicate Coomassie-stained control gel of (b) demonstrating that the non-urmylated FBP substrate is not processed. (c) N-terminal fusion of a single Urm1 moiety to an FBP substrate results in targeting to the 20S proteasome. 10 μg of substrate was incubated with 20 μg of the proteasome complex for 1 h at 70°C. Reactions using the active proteasome complex are displayed in lanes 2 (FBP) and 4 (Urm1:FBP). Lanes 1 and 3: negative controls for FBP and Urm1:FBP fusion, respectively, using the inactive proteasome.
**Figure 8 | Proposed models for the archaeal Urm1-proteasome system.** (a) Covalent attachment of Urm1 to substrates results in association with either the core 20S proteasome (blue dashed arrow) or the PAN regulatory ATPase (red dashed arrow) in vitro. The Urm1 tag leads the entry of the fused GFP substrate (green) into either complex. (i) For illustrative purposes the *T. acidophilum* 20S proteasome crystal structure (PDB:1PMA) is displayed. The 20S outer α-rings are coloured purple (equivalent to Saci0613), while the inner β-rings are shown in yellow (representing Saci0662). Three of the α-subunits are coloured red to represent the Saci0909 active catalytic subunit (Inset; yeast homologues of the *S. acidocaldarius* β-subunits identified by BLAST, see Supplementary Table 4). The *Methanocaldococcus jannaschii* PAN complex (grey) is also displayed in a view perpendicular to the central channel (upper domains; the PAN N-terminal hexameric ring (PDB:3H43 (ref. 69)); below; the monomeric PAN nucleotidase domain structure (PDB:3H4M) modelled onto the coordinates of hexameric HslU (PDB:1DO0)69); (ii) cutaway view of the proteasome cylinder to demonstrate processing of the Urm1 tag of the Urm1:GFP fusion proteins by the core 20S, in a manner reminiscent of the bacterial Pup-proteasome system; (iii) this study suggests that some stably-folded Urm1-tagged substrates may enter the 20S core even in the absence of an unfoldase at physiological temperatures. (b) Genomic context associations between eukaryotic Urm1 gene (blue) and the *Nas2* proteasome assembly chaperone gene (light grey) in a wide variety of yeast species, suggestive of a putative role for Urm1 in assembly of the eukaryotic proteasome. ORFs positioned above the midline are transcribed left to right, and those below the line transcribed right to left. (c) Schematic model of the archaeal Urm1-proteasome system, and the hypothetical role in the assembly and maintenance of the archaeal ribosome. Urm1 modifications have been detected on several ribosome subunits in this study, while ubiquitin is proposed to play roles in eukaryotic ribosome maintenance70. Our study also implicates the archaeal Urm1 protein in proteasome assembly pathways, tRNA modification, metabolism, and the maintenance of genomic stability by DNA repair (see Supplementary Note, Supplementary Fig. 3 and Supplementary Tables 3 and 5 for further details).
Our demonstration of the physical and functional association between the Urm1 protein and the archaeal proteasome necessitated the reconstitution of the S. acidocaldarius 20S assembly in vitro for the first time. This revealed that the central catalytic β-rings of this complex are composed of two different subunits. In contrast, the β-rings of previously purified archaeal 20S proteasomes from H. volcanii, T. acidophilum and M. thermophillum are homo-heptameric. In S. acidocaldarius, the first of the two β-subunits is seemingly catalytically inert, and instead appears essential for interaction with the outer α-rings to form the 20S cylinder. The second β-subunit then appears to interperse within the β-ring, at a low stoichiometry relative to the other subunits, to provide catalytic activity (Figs 4a,d and 8a). Parallels can be made with the eukaryotic 20S complex in which only three of the seven β-subunits confer catalytic activity.39–52. Indeed, BLAST searches reveal that S. acidocaldarius catalytic β-subunit is most similar the S. cerevisiae β5 catalytic subunit, whereas the other structural β-subunit is more homologous to the non-catalytic β7 in yeasts (Fig. 8a and Supplementary Table 4).

The structural, bioinformatic and functional data presented here add further support to the theory that the archaeal and eukaryotic Urm1 proteins are closely related, and have evolved from a shared common ancestor.40. However, in eukaryotes, a direct association between the Urm1 modifier and the proteasome complex has not been identified to date. It is clear that the proteasome-mediated degradation pathway has been appropriated by the more sophisticated and adaptable ubiquitination system in complex eukaryotic organisms.24. However, has the ancient association of Urm1 with the proteasome that we still observe in the archaeal domain of life been entirely usurped by the ubiquitin system in the eukaryotes? In a final bioinformatic observation, we note that in a variety of yeast species the genes encoding for Urm1 and the Nas2 proteasome assembly chaperone are immediately juxtaposed (Fig. 8b). This genomic linkage is suggestive of an involvement of Urm1 in the eukaryotic proteasome assembly process. Interestingly, there is some precedent for Urm1 playing a role in the assembly of the proteasome, as we detected in vivo and in vitro Urm1 modifications on the S. acidocaldarius PAC2 proteasome assembly chaperone (Supplementary Table 3A and B). It is tempting to speculate that in the ancestral archaeal system, Urm1 may play roles in both proteasomal assembly and substrate destruction pathways. Similarly, it remains plausible that in eukaryotic organisms Urm1-conjugated proteins could still interact with proteasome, at least during assembly of the protease complex.

The Urm1/SAMP mediated proteasome degradation pathway identified by this study is representative of a streamlined and presumably ancestral version of the eukaryotic ubiquitin-proteasome degradation system. It remains to be determined if Urm1 marked substrates are recognized directly by the core 20S proteasome in vivo, in accordance with our observations using the in vitro reconstituted S. acidocaldarius system. This may represent an additional processing mechanism operating alongside the established degradation routes that utilize the regulatory ATP-dependent unfoldases, as reported in other systems.39.

Similarly, further studies should also reveal if the conjugation of a single Urm1 protein is sufficient to target substrates to the proteasome in vivo, or whether multiple Urm1 moieties are required for efficient processing. It will also be important to investigate if Urm1 is deconjugated and recycled at the proteasome before substrate delivery, or destroyed in vivo along with the target, as we observe with the reconstituted 20S proteasome. Our findings to date are also suggestive that the S. acidocaldarius Urm1 modifier may regulate a variety of diverse cellular processes and pathways (see Supplementary Note and Fig. 8c). Intriguingly, many equivalent biological functions in eukaryotic organisms are also regulated by the Urm1 and Ub modifiers. Thus, the adoption of simple and biochemically tractable archaeal models to investigate ubiquitin-like pathways should provide us with valuable future insights into these key conserved mechanisms.

Methods

**Protein purification.** Open reading frames (ORFs) were amplified by PCR from S. acidocaldarius DSM6363 genomic DNA, cloned into either pET30alpha/pET30b, pET28a, pCDF-Duet or pET-DUET vectors (Novagen), and expressed in E. coli Rosetta (DE3) pLyS cells (Novagen). Cell lysates were heat-clarified at 70°C for 20 min. His-tagged soluble proteins were subsequently purified by Ni-NTA IMAC and size exclusion chromatography. Full experimental details of the cloning strategy and purification procedure for each protein (including the untagged Urm1 and ELSA/Ubap1/Thi4 proteins) are provided in the Supplementary Methods and Supplementary Table 6.

**Cryocrystalization and X-ray structure determination.** The crystal structure of S. solfataricus Urm1 was determined by molecular replacement of a native crystal data set, leading to a complete structure refined to a resolution of 2.2 Å. Full details of the structure determination and refinement are given in the Supplementary Methods.

**In vitro urmylation assays.** Target protein (75 μg), 30 μg N-terminally His-tagged Urm1 protein and 15-μg ELSA/Ubap1/Thi4 enzyme were incubated together at 70°C for 1 h in either the presence or absence of 2.5 mM ATP in 200-μl urmylation reaction buffer (20 mM Tris acetate (pH 8.0), 100 mM NaCl, 5 mM MnCl2, 10 mM MgOAc, 50 mM KOAc, 1 mM Zn(OAc)2, 5% glycerol and 0.02% β-mercaptoethanol). Urmylated products were then retrieved by Ni-NTA pulldown or size exclusion chromatography. Full experimental details of the cloning and purification procedure for each protein (including the untagged Urm1 and ELSA/Ubap1/Thi4 proteins) are provided in the Supplementary Methods and Supplementary Table 6.

**In vitro urmylation of S. acidocaldarius cell extract for MS analyses.** S. acidocaldarius cell-free extracts were prepared by collecting 11 of cells at OD600nm at 0.6, washed in 1X TBS (10 mM Tris (pH 8.0), 150 mM NaCl), and resuspended in 5 ml 1X TBS, 0.1% Triton-X-100, 0.1% β-mercaptoethanol and 1X EDTA-free protease inhibitors (Roche). Cells were disrupted by sonication and the lysate clarified by centrifugation. 250 μl of the cell-free extract was added to 500-μl urmylation reaction buffer, supplemented with 375 μg N-terminally His-tagged Urm1 protein, 375 μg ELSA/Ubap1/Thi4 enzyme and 15 mM ATP, and incubated for 1 h at 70°C. Urmylated products were purified by Ni-NTA pulldown, separated by SDS-PAGE, and identified by western blot.

**In vitro urmylation of S. acidocaldarius cell extract for MS analyses.** S. acidocaldarius cell-free extracts were prepared by collecting 11 of cells at OD600nm at 0.6, washed in 1X TBS (10 mM Tris (pH 8.0), 150 mM NaCl), and resuspended in 5 ml 1X TBS, 0.1% Triton-X-100, 0.1% β-mercaptoethanol and 1X EDTA-free protease inhibitors (Roche). Cells were disrupted by sonication and the lysate clarified by centrifugation. 250 μl of the cell-free extract was added to 500-μl urmylation reaction buffer, supplemented with 375 μg N-terminally His-tagged Urm1 protein, 375 μg ELSA/Ubap1/Thi4 enzyme and 15 mM ATP, and incubated for 1 h at 70°C. Urmylated products were purified by Ni-NTA pulldown, separated by SDS-PAGE, and identified by western blot.

**In vivo overexpression of Urm1 for MS analyses.** N-terminally hexa-His-tagged Saci0669 (Urm1) was overexpressed in S. acidocaldarius strain MW001 with the vector pCMalLacS57, following induction with 0.4% maltose. Urm1 conjugates were purified from cell lysate by IMAC on Ni-NTA agarose. Full experimental details are described in the Supplementary Methods.

**GeLC-mass spectrometry and MS data analysis.** All GeLC-MS/MS experiments were performed using a nanoAcquity UPLC (Waters Corp., Milford, MA) system and an LTQ Orbitrap Velos hybrid ion trap mass spectrometer (Thermo Scientific, Waltham, MA). Full experimental details including sample preparation and data analysis are provided in the Supplementary Methods.

**Size exclusion chromatography-multi-angle laser light scattering.** Samples analysed by SEC-MAKS (100 μl protein complex at 2 mg/ml) were passed over a Superdex 200 10/300 Increase Gl column (GE Healthcare), in 20 mM Tris (pH 8.0), 300 mM NaCl. The column output was fed into a DAWN HELEOS II MALS detector with a laser source at 664 nm, and eight fixed angle detectors (Wyatt Technology), followed by an Optilab T-r EX differential refractometer, using a 664 nm LED light source at 25°C (Wyatt Technology).

**Reconstitution of the S. acidocaldarius 20S proteasome.** Full length untagged Saci0669 1 ORF was coexpressed with N-terminally truncated Saci0662 β, with or without the N-terminally truncated Saci0999 β catalytic subunit in E. coli Rosetta (DE3) pLyS cells (Novagen). 28-subunit (~ 660 kDa) complex was purified by heat clarification of the extract, Ni-NTA agarose IMAC, and size exclusion chromatography. Full experimental details of the cloning strategy and purification procedure are provided in the Supplementary Methods.
Electron microscopy. S. acidocaldarius 20S proteasome complexes were visualized with an FEI Philips CM100 transmission electron microscope at the Advanced Imaging Centre, University of Cambridge. 10 μl of the 20S proteasome complex at 0.035 mg ml⁻¹ was applied to glow-discharged carbon-coated electron microscopy grids, pre-coated with 10 μl 0.1% poly-L-lysine (Sigma), and negatively stained with 10 μl 2% (w/v) uranyl acetate.

N-terminal Urm1 fusion protein protease processing assays. Ten microgram of substrate (GFP, Urm1:GFP fusion, FBP, Urm1:FBP fusion or native Urm1 protein) was incubated, for 1 h at 70 °C, with 20 μg of either active (including the SacChar099 catalytic β subunit) or inactive (without SacChar099) 20S proteasome complex in 70-μl proteasome reaction buffer (20 mM Tris acetate (pH 8.0), 100 mM NaCl, 5 mM MnCl₂, 10 mM MgOAc, 50 mM KOAc, 1 mM Zn(OAc)₂, 5% glycerol and 0.02% β-mercaptoethanol). After addition of 2X laemml protein loading buffer, the products were then separated by SDS-PAGE, and visualized with 20% boiling in 2X laemmli protein loading buffer, separated by SDS-PAGE, and 3 mM ATP. 50 μl of the urmylated reaction was then incubated, at 70 °C for 1 h, with 20 μg of either catalytically active, or inactive 20S proteasome complex, with or without the addition of 20 μg of the PAN ATPase, in the presence of an additional 12.5 mM ATP, in a final volume of 80 μl. Products were then retrieved by Ni-NTA pulldown, washed in TBST (10 mM Tris (pH 8.0), 150 mM NaCl and 0.1% Tween20) supplemented with 15 mM imidazole. Proteins were eluted by boiling in 2X laemml protein loading buffer, separated by SDS-PAGE, and visualized with Coomassie stain.

In vitro urmylation and proteasome processing assay. 100 μg FBP, 30 μg N-terminally His-tagged Urm1 protein and 7.5 μg ELSA/Uba4p/ThiF enzyme were incubated together at 70 °C for 1 h in 200-μl urmylation reaction buffer (20 mM Tris acetate (pH 8.0), 100 mM NaCl, 5 mM MnCl₂, 10 mM MgOAc, 50 mM KOAc, 1 mM Zn(OAc)₂, 5% glycerol and 0.02% β-mercaptoethanol) and supplemented with 3 mM ATP. 50 μl of the urmylated reaction was then incubated, at 70 °C for 1 h, with 20 μg of either catalytically active, or inactive 20S proteasome complex, with or without the addition of 20 μg of the PAN ATPase, in the presence of an additional 12.5 mM ATP, in a final volume of 80 μl. Products were then retrieved by Ni-NTA pulldown, washed in TBST (10 mM Tris (pH 8.0), 150 mM NaCl and 0.1% Tween20) supplemented with 15 mM imidazole. Proteins were eluted by boiling in 2X laemml protein loading buffer, separated by SDS-PAGE, and visualized with Coomassie stain.

GFP fluorescence assay measuring 20S degradation and PAN unfolding. For the 20S proteasome core degradation reactions, equimolar quantities (6(μM) of the GFP, 1XUrm1-GFP, 2XUrm1-GFP and 3XUrm1-GFP were incubated with 10 μg active 20S complex at 69 °C and 0.5 nm steps. The fluorescence of the superfolder GFP (Sandia National Laboratories, USA). The fluorescence of the superfolder GFP (Sandia National Laboratories, USA). CD spectra between 250 and 185 nm were recorded on an J. Biol. Chem. 273, 16555–16560 (1998).

Leimkuhler, S., Wuebbens, M. M. & Rajagopalan, K. V. Characterization of Escherichia coli MoE and its involvement in the activation of molybdenopterin synthase for the biosynthesis of the molybdenum cofactor. J. Biol. Chem. 276, 34695–34701 (2001).

Burroughs, A. M., Balaji, S., Iyer, L. M. & Aravind, L. Small but versatile: the evolutionary roots of ubiquitin. Nature 458, 422–429 (2009).

Burroughs, A. M., Iyer, L. M. & Aravind, L. The natural history of ubiquitin and ubiquitin-related domains. Front. Biosci. (Landmark Ed) 17, 1433–1460 (2012).

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

Taylor, S. V. et al. Thiamin biosynthesis in Escherichia coli. J. Biol. Chem. 267, 47649–47648 (2002).

Pedrioli, P. G., Leidel, S. & Hofmann, K. Urm1 at the crossroad of ubiquitin-related proteins. Proc. Natl Acad. Sci. USA 103, 11625–11630 (2006).

Schmitz, J. et al. The sulfurtransferase activity of Uba4 presents a link between ubiquitin-like protein conjugation and activation of sulfur carrier proteins. Biochemistry 47, 6479–6489 (2008).

Petroski, M. D., Salvesen, G. S. & Wolf, D. A. Urm1 couples sulfur transfer to ubiquitin-like protein function in oxidative stress. Proc. Natl Acad. Sci. USA 108, 1748–1750 (2011).

Wang, F., Liu, M., Qiu, R. & Ji, C. The dual role of ubiquitin-like protein Urm1 as a protein modifier and sulfur carrier. Protein Cell 2, 612–619 (2011).

Hochstrasser, M. Biochemistry. All in the ubiquitin family. Science 289, 563–564 (2000).

Chowdhury, M. M., Dosche, C., Lohmannsroben, H. G. & Leimkuhler, S. Dual role of the molybdenum cofactor biosynthesis protein MOCS3 in tRNA thionation and molybdenum cofactor biosynthesis in humans. J. Biol. Chem. 287, 17297–17307 (2012).

Schleker, C. D., Van der Veen, A. G., Damon, J. R., Spooner, E. & Ploegh, H. L. A functional proteomics approach links the ubiquitin-related modifier Urm1 to a tRNA modification pathway. Proc. Natl Acad. Sci. USA 105, 18252–18260 (2008).

Humbard, M. A. et al. Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii. Nature 463, 54–60 (2010).

Miranda, H. V. et al. Archaeal ubiquitin-like SAMP3 is isopeptide-linked to proteins via a Ub-α-dependent mechanism. Mol. Cell Proteomics 13, 220–239 (2014).

Miranda, H. V. et al. E1- and ubiquitin-like proteins provide a direct link between protein conjugation and sulfur transfer in archaea. Proc. Natl Acad. Sci. USA 108, 4417–4422 (2011).

Sanjan, N., Damberger, F. F., Sutter, M., Allain, F. H. & Weber-Ban, E. Solution structure and activation mechanism of ubiquitin-like small archaeal modifier proteins. J. Mol. Biol. 405, 1040–1055 (2011).
31. Jeong, Y. J., Jeong, B. C. & Song, H. K. Crystal structure of ubiquitin-like small archaeal modifier protein 1 (SAMP1) from Haloferax volcanii. Biochem. Biophys. Res. Commun. 405, 112–117 (2011).
32. Li, Y. et al. Crystal structure of the ubiquitin-like small archaeal modifier protein 2 from Haloferax volcanii. Protein Sci. 22, 1206–1213 (2013).
33. Makarova, K. S. & Koonin, E. V. Archaeal ubiquitin-like proteins: functional versatility and putative ancestral involvement in TRNA modification revealed by comparative genomic analysis. Archaea 2010, 710303, doi:10.1155/2010/710303 (2010).
34. Barandun, J., Delley, C. L. & Weber-Ban, E. The pupylation pathway and its role in mycobacteria. BMC Biol. 10, 95 (2012).
35. Pearce, M. J., Münsterer, J., Ferreyra, J., Gyg, S. P. & Darwin, K. H. Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis. Science 322, 1104–1107 (2008).
36. Mayr, I., Seemuller, E., Muller, M. A. & Baumeister, W. Late events in the assembly of 20S proteasomes. J. Struct. Biol. 124, 179–188 (1998).
37. Zuhl, F., Seemuller, E., Golbik, R. & Baumeister, W. Dissecting the assembly pathway of the 20S proteasome. FEBS Lett. 418, 189–194 (1997).
38. Raju, R. M., Goldberg, A. L. & Rubin, E. J. Bacterial proteolytic complexes as therapeutic targets. J. Mol. Biol. Drug Discovery 7, 777–789 (2012).
39. Forster, E., Unverdorben, P., Sledz, P. & Baumeister, W. Unveiling the long-held secrets of the 26S proteasome. Structure 21, 1551–1562 (2013).
40. Katayama, Y. et al. The two-component, ATP-dependent Clp protease of Escherichia coli. Purification, cloning, and mutational analysis of the ATP-binding component. J. Biol. Chem. 263, 15226–15236 (1988).
41. Kelly, F. C., Gross, M. L., Makeyev, S., Koonin, E. V. & Aravind, L. Comparative structural organization in the gradient between E. coli ClpAP protease and the eukaryotic 26S proteasome. J. Mol. Biol. 250, 587–594 (1995).
42. Striefel, F., Hunkeler, M., Summer, H. & Weber-Ban, E. The mycobacterial Mpa-proteasome unfolds and degrades pupylated substrates by engaging Pup’s N-terminus. EMBO J. 29, 1262–1271 (2010).
43. Barandun, J., Delley, C. L., Ban, N. & Weber-Ban, E. Crystal structure of the complex between prokaryotic ubiquitin-like protein and its ligase PyAa. J. Am. Chem. Soc. 135, 6794–6797 (2013).
44. Maupin-Furlow, J. Proteasomes and protein conjugation across domains of life. Nature 446, 533–539 (2007).
45. Nature Publishing Group
46. NATURE COMMUNICATIONS | DOI: 10.1038/ncomms9163
47. How to cite this article: Anjum, R. S. et al. Involvement of a eukaryotic-like ubiquitin-related modifier in the proteasome pathway of the archaeon Sulfolobus acidocaldarius. Nat. Commun. 6:8163 doi:10.1038/ncomms9163 (2015).
48. This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.
49. Rzechorzek, N. J. Proteasomal AAA ATPases: structure and function. Biochim. Biophys. Acta. 1823, 67–82 (2012).
50. Benaroudj, N., Zwicp, P., Seemuller, E., Baumeister, W. & Goldberg, A. L. ATP hydrolysis by the proteasome regulatory complex PAN serves multiple purposes in protein degradation. Mol. Cell. 14, 69–78 (2003).
51. Sauer, R. T. & Baker, T. A. AAA + proteases: ATP-fueled machines of protein destruction. Annu. Rev. Biochem. 80, 587–612 (2011).
52. Ciechanover, A. & Ben-Sadaon, R. N-terminal ubiquitination: more protein substrates join in. Trends Cell Biol. 14, 103–106 (2004).
53. Benazéq, T. & Baumeister, W. Conformational constraints in protein degradation by the 20S proteasome. Nat. Struct. Biol. 2, 199–204 (1995).
54. Reyes-Turcu, F. E., Ventii, K. H. & Wilkinson, K. D. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. Annu. Rev. Biochem. 78, 363–397 (2009).
55. Burns, K. E. et al. ‘Depupylation’ of prokaryotic ubiquitin-like protein from mycobacterial proteasome substrates. Mol. Cell 39, 821–827 (2010).
56. Imkamp, F. et al. Dop functions as a depeptase in the prokaryotic ubiquitin-like modification pathway. EMBO Rep. 11, 791–797 (2010).
57. Heslop, N. et al. Archaeal JAB1/MPN/MOV34 metalloenzyme (HvJAM1) cleaves ubiquitin-like small archaeal modifier proteins (SAMPs) from protein-conjugates. Mol. Microbiol. 68, 971–987 (2012).
58. Satoz, T. et al. Structural basis for proteasome formation controlled by an assembly chaperone nas2. Structure 22, 731–743 (2014).
59. Berkner, S., Wlodkowski, A., Albers, S. V. & Lipp, G. Inducible and constitutive promoters for genetic systems in Sulfolobus acidocaldarius. Extremophiles 14, 59–67 (2010).
60. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. www.pymol.org/.
61. Zhang, F. et al. Structural insights into the regulatory particle of the proteasome from Methanocaldococcus jannaschii. Mol. Cell. 34, 473–484 (2009).
62. Scherbik, N. & Postov, D. G. Ubiquitin and ubiquitin-like proteins in the nucleolus: multitasking tools for a ribosome factory. Genes Cancer 1, 681–689 (2010).

Acknowledgements
N.P.R. would like to thank Helen Wise for initial discussions regarding ubiquitin-like modifications in the archaea. S.M.B., R.S.A., J.K.B. and N.P.R. are grateful to Jeremy Steeper and colleagues at the Cambridge Advanced Imaging Centre for assistance and advice with the TEM. N.P.R. and R.S.A would like to thank Joseph Maman and Katherine Stott for their assistance with the CD and thermal shift analyses of the Urm1-GFP fusion proteins, and Tommaso Moschetti for advice with the plate-reader assays. We would also like to thank Len Packman (Protein & Nucleic Acid Chemistry (PNAC) Facility, Department of Biochemistry, Cambridge University) for the mass spectrometry analysis of the Urm1-GFP proteosomal processing assay. N.P.R would also like to thank Steve Jackson for his critical reading of the manuscript, and Neil Rzechorecz for assistance and advice with the preparation of the PyMOL illustrations, and superimposition of the hexameric PAN model. N.P.R. is grateful to the Department of Biochemistry and the Isaac Newton Trust (Trinity College, Cambridge) for supporting this study. The N.P.R. laboratory is funded by the Medical Research Council [Career Development Award G0701443], and also by an Isaac Newton Trust Research Grant. S.M.B. is supported by a BBSRC Doctoral Training Grant [RG53842]. I.P. and M.L.K. were funded by a Wellcome Trust Senior Fellowship Award in Basic Biomedical Sciences [grant no. 08279/Z/07/Z]. S.V.A. was supported by intramural funds of the Max Planck Society. X-ray diffraction data were collected at the Crystallographic X-ray facility at the Department of Biochemistry, University of Cambridge; we are grateful to the Facility Manager, Dr Dimitri Chirgadze, for his assistance in using these facilities.

Author contributions
N.P.R. conceived the study, collected and analysed the data and wrote the manuscript. S.M.B., R.S.A., J.K.B., M.L.K., M.A.C., B.M.F., S.L., J.A.H., M.J.D. and N.P.R. collected and analysed the data. M.A.C. and M.L.K. crystallised and solved the structure of S. solfataricus Urm1. J.A.H. and M.J.D. collected the mass-spectrometry datasets. S.V.A. provided overexpression vectors and gave assistance with the Urm1 in vivo overexpression.

Additional information
Accession codes: The coordinates and structure factors of the S. solfataricus Urm1 crystal structure have been deposited in the Protein Data Bank under accession code 4WWM.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.