Cooperativity and Communication in Archaeal Cdc48•20S, An Ancient Proteolytic Machine

by

Jonathan Dean Grabenstatter
B.A., Miami University (2004)

Submitted to the Department of Earth, Atmospheric and Planetary Sciences in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

© Massachusetts Institute of Technology. All rights reserved

Signature of Author

Signature redacted

Department of Earth, Atmospheric and Planetary Sciences
August 27, 2015

Signature redacted

Certified by

Signature redacted

Robert T. Sauer
Salvador E. Luria Professor of Biology
Thesis Supervisor

Signature redacted

Accepted by

Signature redacted

Robert D. van der Hilst
Schlumberger Professor of Earth Sciences
Head, Department of Earth, Atmospheric and Planetary Sciences
Cooperativity and Communication in Archaeal Cdc48-20S, 
An Ancient Proteolytic Machine

by

Jonathan Dean Grabenstatter

Submitted to the Department of Earth, Atmospheric and Planetary Sciences on June 18, 2015, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract
ATP dependent proteolysis is a process essential for life and is carried out by AAA+ proteases. AAA+ unfoldases use the energy of ATP hydrolysis to power the unfolding and translocation of protein substrates into compartmentalized peptidases for regulated proteolysis. Cdc48 is a highly conserved AAA+ homohexameric unfoldase which is made up of two AAA+ rings. Each ring can, in principle, bind and hydrolyzing ATP, but it is unclear what roles are played by each ring and how they coordinate their activities. A regulatory N domain functions to control the activity of the enzyme and binding to its partner peptidase, the 20S proteasome. In this thesis I present experiments which investigate the role of inter-ring communication in ATP hydrolysis, protein unfolding, and allosteric interactions with the 20S and show how these features affect enzyme function. Experiments also show how the N domain controls D1-D2 interactions that govern ATP hydrolysis and substrate unfolding. Finally, I present experiments that take steps toward developing a system for screening protein substrates of Cdc48-20S and identify several substrates from E. coli lysates.

Thesis Supervisor: Robert T. Sauer
Title: Salvador E. Luria Professor of Biology
Acknowledgements

The last thing anyone needs is another overwrought acknowledgements section. So I will limit this to a few sentences thanking those most important and influential in my graduate experience.

The bulk of the thanks must go to my wife, Christine, for her loving support during graduate school. Without her all of this would be meaningless. Thank you to my parents for making it possible for me to follow my interests. Thank you to my advisor Bob Sauer for helping at many stages along the way and for providing a great environment in which to do science. Thanks to Roger Summons for support during my initial years at MIT and for allowing me to explore crazy ideas. Thanks to my friends in EAPS and Biology departments. I will not list you all by name, but you know who you are. It was great going through these important years with you.
# Table of Contents

Abstract 3  
Acknowledgements 4  
Table of Contents 5  

Chapter 1: Cdc48: A Functionally Diverse Protein Unfoldase 6  
  
  Protein Degradation 7  
  Ubiquitin proteasome system 10  
  Architecture of the 20S proteasome 12  
  AAA+ partners of the 20S 13  
  Cdc48/p97 function 15  
  Cdc48/p97, autophagy, membrane sorting, and chromatin 18  
  AAA+ family ATPases: A functionally diverse group of molecular motors 21  
  Single and double ring AAA+ enzymes: similarities and differences 22  
  Archaeal Cdc48-20S forms and alternative proteolytic complex 26  
References 29  

Chapter 2: Domain interactions govern function of Cdc48-20S 35  
  
  Introduction 36  
  Results 37  
  Discussion 50  
  Materials and Methods 54  
References 58  

Chapter 3: Screening Methods for Identifying Substrates of Cdc48-20S 60  
  
  Introduction 61  
  Results 62  
  Discussion 68  
  Materials and Methods 70  
References 73  

Appendix A: 75
Chapter 1:

Cdc48: A Functionally Diverse Protein Unfoldase
Protein degradation

Protein synthesis involves several anabolic steps which must occur in order to yield a properly folded and functional protein. First an mRNA is translated by the ribosome which begins the formation of a nascent polypeptide. As the nascent chain grows, specific chaperones assist in holding the incomplete protein in an unfolded state until translation of a domain or the entire protein the transcript is complete, at which point additional chaperones often help the domain or full-length protein to fold into its final and active structure (Preissler and Deuerling, 2012).

Protein degradation, the regulated process by which proteins are removed from the cell, is essentially the opposite process of protein synthesis. A family of broadly conserved proteases, with at least one component belonging to the class of AAA+ enzymes (ATPases Associated with Diverse Cellular Activities), recognizes specific proteins as substrates for degradation (Sauer and Baker, 2011). In eubacteria, recognition usually involves binding of at least one amino-acid sequence, known as a degron or degradation tag, on the target protein to the AAA+ enzyme. Next, the AAA+ enzyme unfolds the tagged protein, one domain at a time, and then translocates the unfolded peptide chain into a partner peptidase. Upon entry into the peptidase, the protein is degraded into small peptide fragments, which are further hydrolyzed to single amino acids by other cellular enzymes allowing recycling.

This process of protein degradation can consume from 30 to more than 1000 molecules of ATP per 100 amino acids (Kenniston et al., 2003; Gur and Sauer, 2009; Iosefson et
al., 2015). Thus, in some cases, more ATP is used to degrade a protein than was required for its synthesis (four ATP/GTP per amino acid). As a consequence, it is important for intracellular degradation to be accurate, degrading only the correct protein targets; efficient, consuming only as much energy as necessary; and relatively fast on cellular-time scales.

![Diagram](image)

Figure 1. AAA+ proteases across the three domains of life.

AAA+ proteases are found in all three domains of life (Figure 1), with the eubacterial enzymes being the best studied (Striebel et al., 2009; Sauer and Baker, 2011; Matyskiela and Martin, 2013). The *Escherichia coli* ClpXP system has been used as a model for understanding the principles of AAA+ protease structure and machine function (Baker and Sauer, 2012). The AAA+ ClpX enzyme assembles as a hexamer of
six identical subunits, each with the ability to bind and hydrolyze ATP. ClpP is its proteolytic partner. An axial pore in the ClpX hexamer binds the degradation tags of protein substrates (Figure 2). Repeated cycles of ATP binding and hydrolysis then pull the attached native domain against the ClpX ring, creating an unfolding force. Once the substrate unfolds, which can require 100's of pulling events, the unfolded polypeptide is actively translocated through the axial pore and into ClpP for degradation (Fig. 2).

Figure 2. Basic mechanism of a AAA+ protease. (A) A protein substrate possess a short degradation tag that targets it to the AAA+ unfoldase. (B) The degradation tag binds in the axial channel of the ATP-bound AAA+ unfoldase and is engaged by flexible pore loops. (C) ATP hydrolysis powers unfolding and translocation of the substrate into the peptidase chamber (After Sauer and Baker, 2011).

ClpP is a serine protease. The active enzyme consists of two homo-heptameric rings, stacked tail-to-tail, that enclose a barrel-shaped chamber and the peptidase catalytic
sites (Wang et al., 1997). A gated axial portal in each ClpP\textsubscript{7} ring prevents folded proteins from entering the degradation chamber in the absence of ClpX (Lee et al., 2010). One ClpX hexamer can bind to each ClpP\textsubscript{7} ring in an ATP-dependent reaction that involves docking of conserved loops on ClpX into clefts of the surface of the ClpP ring (Grimaud et al., 1998; Kim et al., 2001; Martin et al., 2007). Some of the ClpX-ClpP binding energy drives a conformational change of the gating residues of ClpP, opening the axial portal and allowing the unfoldase to thread substrates into the proteolytic chamber for peptide-bond hydrolysis.

The mode of action exemplified by ClpXP is broadly conserved among AAA+ proteases. ClpXP, ClpAP, HslUV, Mpa•20S, Lon, and FtsH are the major AAA+ proteases in eubacteria, and also function in some eukaryotic organelles (Striebel et al., 2009; Sauer and Baker, 2011). The first four enzymes consist of separate self-compartmentalized peptidases and AAA+ partner enzymes, whereas the AAA+ and protease domains are encoded on single polypeptides for Lon and FtsH. Archaea contain the AAA+ Lon, PAN•20S, and Cdc48•20S AAA+ proteases, whereas the 26S proteasome is the major ATP-dependent protease in eukaryotes (Barthelme and Sauer, 2012; Gur, 2013; Matyskiela and Martin, 2013).

**Ubiquitin proteasome system**

In the eukaryotic cytosol and nucleus, the ubiquitin-proteasome system (UPS) serves as a major means of protein degradation and plays important roles in cell-cycle control, regulation of transcription, signal transduction, inflammation, apoptosis, and general
protein-quality control (Finley, 2009; Matyskiela and Martin, 2013). The 26S proteasome recognizes and degrades protein substrates that have been targeted for degradation by addition of poly-ubiquitin chains. A set of enzymes – known as E1, E2 and E3 ligases – work in series to attach multiple copies of ubiquitin, a small protein, to a target protein destined for degradation (Figure 3). The initial linkage of one ubiquitin to the substrate is mediated by an isopeptide bond between a C-terminal diglycine of ubiquitin and one or more lysine side chains on the substrate. Subsequent ubiquitins are attached through isopeptide bonds to ubiquitins already attached to the substrate. A minimum of four ubiquitin chains is thought to be required a substrate to be recognized and degraded by the 26S proteasome.

Figure 3. Summary of the ubiquitin-proteasome system. A three-enzyme cascade catalyzes the covalent addition of the small protein ubiquitin to substrate proteins. E1 ubiquitin activating enzymes form a high-energy thiol-ester E1-Ub linkage that is transferred to the E2-ubiquitin conjugating enzyme. Finally, E3 ubiquitin protein ligases transfer the ubiquitin moiety from E2 to a protein substrate, numerous times if
necessary. Depending on the length and location of the ubiquitin linkage, substrates are targeted for proteasomal degradation, cellular signaling pathways or for structural remodeling. Deubiquitinating (DUB) enzymes can shorten and edit the Ub chains to alter targeting (After Matyskiela and Martin, 2013).

Before conjugation to targets, the C-terminus of ubiquitin is activated by an E1 enzyme, first using ATP to create an adenylated form of ubiquitin, which is then attached to E1 via a thioester linkage. This activated form of ubiquitin is handed off to an E2 enzyme in a multidomain complex with an E3 enzyme. These E2-E3 complexes are responsible for the specificity of ubiquitin addition to protein substrates. Humans contain ~35 E2 enzymes and more than 600 E3 enzymes. The end result of their activity is addition of ubiquitin chains to target protein, which can then be degraded by the 26S proteasome.

**Architecture of the 20S proteasome**

The 20S proteasome consists of four stacked heptameric rings ($\alpha_7\beta_7\beta_7\alpha_7$), which enclose a proteolytic chamber (see Fig. 4A). In bacteria and archaea, the $\alpha_7$ and $\beta_7$ rings consist of single types of $\alpha$ and $\beta$ subunits, respectively, and the $\beta$ subunits contain the threonine catalytic sites for peptide-bond cleavage. The active-site threonine is at the N-terminus of the $\beta$ subunit and is normally revealed by autocleavage of a propeptide segment. In eukaryotes, by contrast, seven distinct $\alpha$ and seven distinct $\beta$ subunits assemble to form the $\alpha_7$ and $\beta_7$ rings, and only a subset of $\beta$ subunits are catalytically active (Borissenko and Groll, 2007). N-terminal residues in the $\alpha$ subunits form a gate that restricts access to the proteolytic chamber (Groll et al., 2000). By itself,
the 20S peptidase can cleave small peptides but larger peptides or unfolded peptides are cleaved very slowly. However, deletion of residues 1-11 of the α subunits removes the gate and allows faster cleavage of larger peptides and polypeptides (Smith, et al., 2007).

**AAA+ partners of 20S**

The archaeal PAN enzyme was initially identified as a proteolytic partner of archaeal 20S (Benaroudj and Goldberg, 2000). PAN is a type-2 AAA+ enzyme, defined as containing one AAA+ module or nucleotide-binding domain (NBD) per subunit. The active oligomeric form of PAN is a hexamer, which appears to be stabilized by an N-terminal domain (Figure 4). PAN associates with the 20S peptidase in a reaction requiring ATP or ATPγS, with the latter analog being hydrolyzed extremely slowly or not at all by PAN (Zwickle et al. 1999). ADP does not support complex formation. Tripeptides at the C-terminus of PAN have a hydrophobic (Hb), tyrosine (Y), any-residue (X) sequence and dock into hydrophobic pockets on the α7 ring of the 20S peptidase. Addition of synthetic HbYX peptides to 20S stimulates degradation of peptides, a reaction called “gate” opening, and deletion of the PAN HbYX residues is sufficient to abolish gate-opening activity (Smith et al. 2005). The bacterial Mpa enzyme, a PAN homolog, docks with bacterial 20S using related HbYX motifs (Darwin et al., 2003).

The eukaryotic 26S proteasome consists of the 20S peptidase and the 19S regulatory particle (RP) (Matyskiela and Martin, 2013). The 19S complex contains a AAA+ Rpt1-6
ring, in which six distinct PAN-like proteins assemble to form a hexamer that mediates interactions with 20S (again through C-terminal tails), as well as substrate unfolding and translocation. In addition, the RP contains several scaffold proteins, ubiquitin receptors, and enzymes that can cleave ubiquitin from the substrate polypeptide chain prior to degradation. The ability of 19S to bind and stimulate peptide cleavage by the 20S enzyme is also ATP dependent (Lander et al., 2012). The complexity of the 19S RP compared to PAN/Mpa seems to be an evolutionary adaptation that allows protein degradation to depend upon marking substrates for degradation by poly-ubiquitination. In fact, ubiquitin and the E1-E3 ligase system is absent in bacteria and most archaea, but some ubiquitin ligases were found in the genome of a complex archeon bearing several other cellular features thought to be hallmarks of eukaryotes (Spang et al., 2015). Although, recognition of poly-ubiquitin chains is a major determinant of degradation by the 26S proteasome, protein substrates must also contain an unstructured polypeptide that can bind in and be engaged by the axial pore of the Rpt1-6 ring to allow unfolding (Prakash et al., 2004).

Recently, archaeal Cdc48 (known as p97 in mammals) was also found to be capable of binding the 20S peptidase, stimulating gate opening, and catalyzing protein degradation (Barthelme and Sauer, 2012). Cdc48 is a type-1 or double-ring AAA+ enzyme, defined as containing two AAA+ modules that form distinct D1 and D2 rings in the active hexamer as observed in crystal structures and cryo-EM structures. Both archaeal and eukaryotic Cdc48/p97 enzymes contain C-terminal HbYX motifs which are necessary for robust 20S-gate-opening activity. Cdc48/p97 also contains an N-terminal domain
that contacts the D1 ring in crystal and EM structures. Deletion of this N-terminal domain increases basal ATP-hydrolysis activity markedly (Gerega et al., 2005).

**Cdc48/p97 function**

Cdc48/p97 is found in all eukaryotes and archaea. It was initially purified from *Xenopus* extracts by size-exclusion chromatography and was shown to form a 97-kDa protein complex with ATP-hydrolysis activity (Peters et al., 1990). Negative-stain electron microscopy revealed ring-shaped oligomers. Cdc48/p97 is highly abundant, accounting for ~1% of total cellular protein. The *cdc48* gene is essential in yeast and mice (Moir et al., 1982; Muller et al., 2007). From archaea to humans, Cdc48/p97 orthologs display ~50% amino-acid sequence identity.

The broad conservation and distribution of Cdc48/p97 enzymes suggest a fundamentally important role in cell biology, and experiments link Cdc48/p97 to functions as diverse as membrane dynamics, ubiquitin-mediated degradation, the ER-stress response, cell-cycle regulation, and chromatin remodeling (Baek et al., 2013). These activities underlie a multitude of cellular pathways which control cellular physiology and are factors in human health and disease.

Currently, Cdc48/p97 is viewed as a molecular motor whose function is directed and modulated by association with a large set of adaptor proteins. Some of these adaptor proteins interact with the N-terminal domain of Cdc48/p97, whereas others interact with the C-terminal HbYX tails (Meyer et al., 2012). A major group of adaptors contain a UBX
domain, which adopts a ubiquitin-like fold and interacts with the Cdc48/p97 N-domain. Although the ubiquitin-like fold suggests a link to the UPS, the lack of C-terminal diglycine motif makes it unlikely that UBX domains are covalently attached to substrates.

The p47 protein was the first member of the UBX family found to interact with Cdc48/p97. The UBX fold defines a modular domain that can interact with the N-terminal domain of Cdc48/p97. Through structure-based alignments, proteins containing UBX domains (including p47, FAF1, SAKS1, and UBXD1) were found in all eukaryotic species. Mutational studies identified a conserved binding motif consisting of an R...FPR surface patch that binds to a hydrophobic surface at the junction of two subdomains of the Cdc48/p97 N-terminal domain. Several other families of adaptors have been found to function in similar fashion, including the UBX-like (UBX-L) containing Ufd1-Npl4 complex, which directs Cdc48/p97 to the cytoplasmic side of the ER, where it extracts poly-ubiquitinated substrates from the membrane for eventual delivery to the proteasome. Most studies have linked Cdc48/p97 to binding ubiquitinated clients which are then remodeled or unfolded, thus facilitating subsequent downstream steps. Cdc48/p97 can also organize the ubiquitination and ubiquitin-editing steps by interacting with a series of E3 ubiquitin ligases and cullin RING scaffolds. For example, the N-terminal domain of Cdc48/p97 N-domain recruits the E4B/Ufd2 complex, which functions to lengthen short ubiquitin chains (Jentsch and Rumpf, 2007). Alternatively, several deubiquitinating enzymes can function with Cdc48/p97 to edit the length of
ubiquitin chains, possibly to tailor different substrates for recognition by specific proteasome adaptors (Wang et al., 2004; Rumpf and Jentsch, 2006).

A network of adaptors functionalize Cdc48/p97 to control important cellular processes (Schuberth and Buchberger, 2006; Yeung et al., 2008). One of the best-studied pathways is endoplasmic-reticulum-associated-degradation (ERAD), in which a poly-ubiquitin tag mediates Cdc48/p97 binding of misfolded or aberrant proteins in the ER membrane which then catalyzes retrotranslocation into the cytosol, where the proteins undergo degradation by the 26S proteasome. In eukaryotic cells, the ER is the organelle where proteins fold and assemble as part of the secretory pathway or for membrane insertion, handling roughly one third of all cellular proteins. Thus, a robust quality-control system has evolved to maintain protein homeostasis within this organelle. The final and crucial step of this quality-control pathway requires the extraction of a misfolded polypeptide from the ER membrane. This step relies on the Ufd1-Npl4 cofactor which binds polyubiquitin chains and mediates Cdc48/p97 substrate binding in a nucleotide-independent fashion (Yi et al., 2012). Interestingly, Cdc48/p97 can interact with non-ubiquitinated ER substrates, but in this case ATP binding is required, suggesting that adaptors recruit substrates via binding motifs, but engagement by Cdc48/p97 requires nucleotide binding. In a fashion similar to ERAD, Cdc48/p97 participates in the mitochondrial protein-quality control pathway. Here, the Vms1 protein forms a heterodimer with Npl4 to direct Cdc48/p97 to ubiquitinated substrates on the cytoplasmic side of the mitochondrial outer membrane (Xu et al., 2011).
Cdc48/p97, autophagy, membrane sorting, and chromatin

After discussing the role of Cdc48/p97 in 26S-mediated degradation of ERAD substrates it is surprising to then learn of its requirement for lysosomal degradation. The lysosomal pathway is the alternate means of protein degradation in the eukaryotic cell, but shares little in common with the UPS system except for the use of ubiquitin as a signal for recognition by Cdc48/p97. Both the endocytic pathway and autophagy also make use of ubiquitin as a cargo tag and as a sorting signal. In the case of endocytosis, the protein caveolin-1 (CAV1) becomes mono-ubiquitinated which mediates recognition by the Cdc48/p97-UBXD1 complex, a requirement for late endosome formation and fusion with the lysosome. For autophagy, a double membrane first forms and binds a lipidated form of the ubiquitin-like protein LC3 (Bug and Meyer, 2012). This complex forms an autophagosome upon engulfment of a cellular target, be it mitochondria, protein aggregates, ribosomes or any number of other cellular structures. Cdc48/p97 is required for proper maturation of autophagosomes and their eventual fusion with the lysosome. The exact mechanism of Cdc48/p97 in these two cellular pathways has yet to be determined, but the evidence suggests it acts to dislocate target proteins associated with membranes.

Lysosomes are organelles which contain hydrolytic enzymes capable of breaking down many kinds of biomolecules and are the ATP-independent alternative to 26S proteasomal degradation for recycling of proteins. Genetic studies of human disorders and yeast show that functional Cdc48/p97 is required for the proper targeting of autophagosomes to the lysosome (Tresse et al., 2010). In patients with Paget's
disease, a disorder caused by missense mutations in Cdc48/p97, an accumulation of incomplete autophagy intermediates results in inclusion body myopathy and early onset dementia associated with neurodegeneration of the frontal lobe (Guinto et al., 2007; Watts et al., 2004; Kimonis et al., 2008). In yeast it has been shown that the adaptors Ufd3 and Ubp3 are required for autophagy of ribosomes. Together, these phenotypes imply that Cdc48/p97 is required for autophagosome formation and their fusion with lysosomes. In autophagy, a multi-step pathway directs cytoplasmic contents targeted for turnover to be engulfed by autophagosomes whose contents are then shuttled to the lysosome where degradation takes place. Cells lacking functional Cdc48/p97 accumulate autophagy intermediates that fail to fuse with the lysosome. Similarly, Cdc48/p97 functions in targeting substrates to the lysosome through endosomal trafficking.

In addition to its function in the cytosol, Cdc48/p97 has also been shown to act on chromatin and nucleoprotein complexes. In response to DNA damage, Cdc48/p97 was shown to control DNA replication by coordinating the degradation of replication initiation factor Cdt1 (Raman et al., 2011). Cdc48/p97 also has been shown to interact with several other DNA repair proteins. The yeast mating-type switching, a regulated DNA repair process, has been linked to Cdc48/p97•Ufd1•Npl4 activity. The Mat-α2 transcriptional repressor undergoes degradation by the 26S proteasome during switching to the α-mating type, but Mat-α2 must be first removed from promoters in a Cdc48/p97-dependent process that requires ubiquitination of Mat-α2 (Dantuma and Hoppe, 2012).
Several aspects of the eukaryotic cell cycle depend upon functional Cdc48/p97. Most significantly, during mitosis Cdc48/p97 is needed for accurate and faithful segregation of chromosomes. Depletion of Cdc48/p97•Ufd1•Npl4 results in failure of mitotic-spindle disassembly during the mitotic-exit pathway (Cao, 2003). At this point in the cell cycle, removal of Aurora B from the chromatin reduces local kinase activity, which then allows decondensation of chromosomes and reformation of the nuclear envelope. Aurora B is modified with ubiquitin chains and then becomes a substrate for the Cdc48/p97•Ufd•Npl4 complex. In the absence of Cdc48/p97, Aurora B persists on the chromatin and proper chromosome segregation cannot occur.

It is clear that Cdc48/p97 serves many biological roles in eukaryotes, and more pathways may be found to rely on Cdc48/p97. However, patterns have begun to emerge that suggest a unified picture of Cdc48/p97 biology. Whether functioning within the scope of the UPS or in proteasome-independent pathways, Cdc48/p97 functions by recognizing specific tags on target complexes, which it then remolds to facilitate downstream events. Depending on the cellular pathway, Cdc48/p97 will cooperate with a specific adaptor or groups of adaptors to recognize ubiquitin or ubiquitin-like moieties. Upon recognition and engagement, ATP-dependent movements catalyze the extraction of targets which can either be recycled, degraded, or further modified. Sometimes the target protein is the functionally important part of the pathway, as in ERAD when misfolded proteins must be removed from the ER membrane to restore protein homeostasis. Other times, the Cdc48/p97 substrates are part of large complexes that must be remodeled or dislocated from their parent complex so that organelle maturation
pathways can proceed, as in autophagy and endocytosis. At the heart of these biological functions is Cdc48/p97, which functions with a diverse repertoire of adaptors as the upstream arbiter of pathways that require the power of a molecular motor to extract or remodel client proteins.

**AAA+ family ATPases: A functionally diverse group of molecular motors**

Most AAA+ ATPases function as oligomers, with hexamers being most common, and contain at least one core ATP-binding module of roughly 200-250 amino acids (Neuwald et al., 1999). Within the AAA+ module, conserved Walker-A and Walker-B sequence motifs play essential roles in enzyme function. For example, a lysine in the Walker-A sequence contacts one or more phosphates of an ATP molecule and is required for strong nucleotide binding. The Walker-B motif consists of two conserved acidic residues following four hydrophobic residues (hhhhDE). The glutamic acid (E) is thought to activate a water molecule that is critical for ATP hydrolysis (Hanson and Whitehart, 2005). A third important sequence feature of the AAA+ module is an arginine finger, which resides in a region of homology on the C-terminal side of the Walker-B motif. The Arg finger from one subunit can make contacts with an ATP bound to a neighboring subunit, which is believed to communicate nucleotide occupancy and to coordinate of ATP hydrolysis within the AAA+ ring.

At the axis of Cdc48/p97 and other AAA+ rings that function in protein unfolding, translocation, or remodeling is a channel or pore, which is lined with conserved loops that help to engage and apply force to a polypeptide substrate (Sauer and Baker, 2011).
For example, the pore-1 loops typically contain a conserved aromatic-hydrophobic dipeptide, and mutations of these residues results in defects in substrate unfolding and translocation (Martin et al., 2008; Iosefson et al., 2015). All AAA+ enzymes couple the chemical energy of ATP binding, hydrolysis, and product release to conformational changes that can perform mechanical work. The molecular details of this chemomechanical coupling are being actively investigated but are not yet clear for any AAA+ enzyme.

**Single and double-ring AAA+ enzymes: similarities and differences**

The operating principles of AAA+ enzymes have been most intensively characterized using bacterial ClpXP as a model (Baker and Sauer, 2012). ClpX contains a single AAA+ module and assembles into a homohexamer. The axial pore of the AAA+ ring can recognize a peptide sequence in the ssrA tag, which is added as a degradation signal to proteins whose synthesis on ribosomes cannot be completed normally (Keiler et al., 1996). ClpX alone is capable of recognizing, unfolding, and translocating ssrA-tagged substrates, and ClpXP can degrade these proteins (Gottesman et al., 1998; Kim et al., 2000; Baytshtok et al., 2015).

Studies of several AAA+ enzymes show that mutation of either the Walker-A or Walker-B motifs abolishes ATPase activity, rendering the enzyme inactive (Hanson and Whiteheart, 2005). Mutations in the Walker-B motif of ClpX permit ATP binding, binding to ssrA-tagged substrates, and binding to ClpP (Hersch et al., 2005). By contrast, Walker-A mutations in ClpA unfoldase and the ClpB chaperone prevent ATP binding.
and activities that require nucleotide binding (Kim et al., 1998; Singh and Maurizi, 1994). By engineering, six ClpX subunits can be genetically linked to produce a covalent hexamer, which allows testing the effects of different numbers and combinations of ATPase-dead subunits within the hexameric ring (Martin et al., 2005). Interestingly, ClpX hexamers with only one or two ATPase-active subunits retain some unfolding, translocation, and degradation activity. This result suggests that ClpX can hydrolyze ATP and unfold substrates in a probabilistic manner in which individual subunits can function relatively independently of other subunits in the ring. Although all ClpX subunits normally have the same sequence, structural studies show that only a subset adopt conformations that allow ATP binding (Glynn et al., 2009). Moreover, ClpX subunits must be able to switch between ATP-binding and non-binding conformations to allow mechanical functions like unfolding and translocation (Stinson et al., 2013; 2015). During the unfolding of very stable proteins, the vast majority of power strokes fail. For example, ClpX hydrolyzes an average of ~600 ATPs during unfolding of an ssrA-tagged titin$^{\text{127}}$ domain (Kenniston et al., 2003). Thus, ATP hydrolysis, conformational movements, and unfolding cannot be tightly coupled. The switching of ClpX subunits between ATP-binding and non-binding conformations may help to create an enzyme that is robust to failure and thus capable of unfolding/translocating substrates with very different amino-acid sequences and a wide range of native protein stabilities (Stinson et al., 2013).

In the ClpAP protease, ClpA, a different AAA+ partner, functions with ClpP (Grimaud et al., 1998; Striebel et al., 2009). Like Cdc48/p97, ClpA is a double-ring or type-1 AAA+
enzyme, consisting of discrete D1 and D2 rings, in addition to a family specific N-terminal domain, which helps modulate substrate specificity and adaptor binding. An outstanding question in the field is what is the functional difference between single and double-ring AAA+ ATPases. Do both AAA+ modules function in ATP hydrolysis? If so, are the functions independent or linked cooperatively? In ClpA, ATP binding to the D1 ring appears to be required to form a functional hexamer, whereas ATP hydrolysis is carried out primarily by the D2 ring. For example, Weber-Ban and colleagues found that Walker-B mutations in the D1 ring had little effect on ATPase rates, whereas the same mutations in the D2 ring resulted in a 10-fold reduction of the ATP-turnover rate (Kress et al., 2009). Not only was the D2 ring of ClpA responsible for most ATP hydrolysis, its activity was also required to unfold and degrade very stable proteins, such as GFP-ssrA.

Bacterial ClpB and its yeast homolog, Hsp104, are double-ring AAA+ enzymes that function to solubilize and refold aggregated proteins following heat shock and other cellular stresses (Mogk et al., 2015). Unlike ClpA, ClpB/Hsp104 show allosteric communication between the D1 and D2 rings. For example, mutation in the D1 and D2 rings of Hsp104 reduced $V_{max}$ for ATP hydrolysis to ~10% and ~25% of the wild-type value, respectively (Hattendorf and Lindquist, 2002) Interestingly, these mutations had only small effects on cooperativity as measured by Hill constant, suggesting that cooperativity arises from interactions within a given AAA+ ring. In subunit-mixing experiments where wild-type monomers were mixed with ATPase mutants, it was found that ATPase activity decreased non-linearly whether single or double mutants were
titrated, revealing strong coupling between rings (Werbeck et al., 2008). Interestingly, different results were found when mixed complexes were assayed for chaperone activity.

N-ethylamine sensitive factor (NSF) is a double-ring AAA+ enzyme found in eukaryotic organisms (Nagiec et al., 1995). It functions in heterotypic membrane-fusion pathways by binding and disassembling SNARE complexes to resolve membrane fusion events. NSF is a homo-hexameric ring with an N-terminal domain responsible for SNAP-SNARE binding. The basal ATPase of this enzyme is very low but is stimulated upon addition of its substrate. Mutational studies have shown that the D1 ring of NSF is dominant in ATPase activity, resulting in approximately 80% loss of activity, whereas the same mutations in the D2 domain have minimal effects on ATP hydrolysis (Zhao et al., 2012). Additionally, nucleotide occupancy in the D2 ring helps promote hexamer formation. Although NSF and Cdc48/p97 share significant sequence and structural homology in their N-domains and D1 and D2 AAA+ modules, the D2 ring of eukaryotic Cdc48/p97 appears to be responsible for most ATP hydrolysis, whereas the D1 ring appears to be responsible for hexamerization (Wang et al., 2003). In fact the D1 ring and the linker region spanning D1-D2 is sufficient for hexamer formation, even in the absence of nucleotide. However, it was recently shown that nucleotide binding to the D2 domain increases the catalytic efficiency of D1 ATP hydrolysis markedly, both by increasing \( V_{\text{max}} \) and decreasing \( K_M \) (Chou et al., 2014).
Figure 4. Structural representations of archaeal and eukaryotic proteasomes. (A) Model of archaeal Cdc48-20S based on a low-resolution EM structure (Barthelme et al., 2014) and crystal structures of mammalian Cdc48/p97 (N domain colored in blue and D1-D2 in cyan) and archaeal 20S (α subunits in green and β subunits in orange). (B) Structural model of a docked complex based on crystal structures of archaeal PAN (cyan) and the 20S core particle (gray). (C) Cryo-EM reconstruction of the yeast 26S proteasome. The unfoldase subunits (cyan) are PAN homologs, but additional 19S regulatory subunits (orange) are also present. (D) Domain structure of the archaeal proteasomal ATPases, Cdc48 and PAN. (panels B and C adapted from Matyskiela and Martin, 2013)

Archaeal Cdc48•20S forms an alternative proteolytic complex

Many conflicting reports exist in the literature on the exact mechanism of Cdc48/p97 function, the role of structural rearrangements in the ATPase cycle, and its putative
function in concert with the 20S peptidase. Recent work from our lab has shown that archaeal Cdc48 forms a proteolytic complex with the 20S peptidase (Barthelme and Sauer, 2012; Figure 4). It uses ATP hydrolysis to unfold and translocate model substrates and has a conserved mechanism of 20S association. In Chapter 2 of this thesis, I explore the enzymatic relationship between the D1 and D2 rings of archaeal Cdc48 and how their activities and interactions govern functions such as ATP hydrolysis, protein unfolding, and 20S binding. Specifically, I use Walker-A and Walker-B mutations to selectively inhibit nucleotide binding or hydrolysis in each ring and then investigate the functional consequences. I also address the mechanism of D1-D2 communication by mutating the linker connecting these domains. My results show that ATP binding and hydrolysis in the D1 ring play an important allosteric role in Cdc48•20S complex formation. Deleting the N-domain of Cdc48 activates ATP hydrolysis and strengthens 20S binding. I explore the linkage between the N-domain, D1-D2 communication, and ATP hydrolysis, and propose a mechanism for control of 20S binding. Another question is how archaeal Cdc48 targets specific proteins for unfolding and degradation. In chapter 3, I make use of a heterologous system of substrate identification to begin to explore substrate recognition by Cdc48•20S. Together these experiments help extend and deepen our knowledge of Cdc48, a biologically essential and medically important protein. Although Cdc48 is a complicated enzyme with many layers of regulation, the experiments presented in this thesis provide a biochemical foundation for understanding the archaeal Cdc48•20S system.
References


Wang, Q., Song, C., Yang, X., Li, C.C. (2003). D1 ring is stable and nucleotide-independent, whereas D2 ring undergoes major conformational changes during the ATPase cycle of p97-VCP. J. Biol. Chem. 278, 32784-32793.


Chapter 2:

Domain Interactions Govern Function of Archaeal Cdc48•20S
Introduction

The first archaeal proteasomal ATPase discovered was PAN, which functions as single-ring AAA+ hexamer to catalyze protein unfolding and translocation into the 20S peptidase for degradation (Zwickl et al., 1999). Recently, the double-ring AAA+ Cdc48 ATPase was found to be an alternative partner for the archaeal 20S proteasome (Barthelme and Sauer, 2012; Matouschek and Finley, 2012). The $\alpha_7\beta_7\beta_7\alpha_7$ architecture of the 20S peptidase is similar in archaea and eukaryotes. However, archaeal 20S is simpler, as each ring is composed of a single type of $\alpha$ or $\beta$ subunit. In eukaryotic 20S, by contrast, seven distinct types of $\alpha$ or $\beta$ subunits comprise each ring. Archaeal and eukaryotic Cdc48 share substantial sequence and structural homology (Barthelme et al., 2014), raising the possibility that a Cdc48•20S proteasome might also function in eukaryotic organisms. However, although mammalian Cdc48 can bind and open the pore of mammalian 20S (Barthelme and Sauer, 2013), there is no compelling evidence that this complex is active in protein degradation.

All Cdc48 homologs consist of an N-terminal domain, a AAA+ module that forms the D1 ring in the hexamer, another AAA+ module that forms the D2 ring, and an unstructured C-terminal tail. In this chapter, I study the enzymology of archaeal Cdc48 to better understand this highly conserved and essential ATPase. Biochemical studies of eukaryotic Cdc48 have largely been limited to assaying changes in ATP hydrolysis, as direct assays for other activities are lacking. In contrast, assays have been developed for protein unfolding by archaeal Cdc48, binding of archaeal Cdc48 to 20S, and protein degradation by archaeal Cdc48•20S (Gerega et al., 2005; Barthelme and Sauer, 2012;
Barthelme et al., 2014). Using these assays and wild-type and mutant enzymes, I probe how the major domains of Cdc48 work together. I show that the D1 and D2 rings of archaeal Cdc48 collaborate in a cooperative ATPase cycle that is critical for robust mechanical activity and that the N domain regulates machine function, at least in part, by controlling nucleotide binding.

**Results**

**Mutations affecting ATP hydrolysis**

Archaeal Cdc48 must hydrolyze ATP to power mechanical unfolding and translocation of proteins. To test the degree to which the D1 and D2 AAA+ rings cooperate in these enzymatic activities, I introduced Glu$\rightarrow$Gln mutations\(^1\) into the Walker-B motifs of the ATP-binding pockets in either the D1 ring (E291Q), or the D2 ring (E568Q). The Walker-B glutamic acid is a conserved feature of all P-loop ATPases, with the side-chain carboxylic acid thought to activate a water molecule for nucleophilic attack on the gamma phosphate of ATP, leading to hydrolysis. In other AAA+ enzymes, Glu$\rightarrow$Gln Walker-B mutations allow ATP binding but substantially reduce the rate of hydrolysis (Hersch et al., 2005; Schaupp et al., 2007).

I first measured basal ATP hydrolysis rates in a standard buffer containing 20 mM MgCl\(_2\) (Fig. 1A). The rate for wild-type Cdc48 was only modestly higher than for the E291Q or E568Q variants. Thus, preventing robust ATP hydrolysis in either the D1 ring or the D2 ring causes a relatively small reduction in the overall hydrolysis rate under

---

\(^1\) Replacing the Glu carboxylate with the isosteric Gln amide is the most conservative mutation possible at these positions.
these conditions. The ATPase activity of wild-type Cdc48 is stimulated by high concentrations of Mg$^{++}$ (Gerega et al., 2005). In buffer containing 120 mM MgCl$_2$, I found that the hydrolysis rate for wild-type Cdc48 was ~6-fold higher than for either E291Q Cdc48 or E568Q Cdc48 (Fig. 1B). Thus, collaboration between the D1 and D2 rings is required to reach maximum ATPase rates in response to an activating stimulus. Next, I tested the effects of the Walker-B mutations in Cdc48$^{\text{AN}}$, a variant lacking the wild-type N-terminal domain, using buffer with 20 mM MgCl$_2$ (Fig. 1C). Under these conditions, the hydrolysis rate for the parental enzyme was ~9-fold higher than the E291Q or E568Q variants. Again, both the D1 ring and the D2 ring are required for rapid rates of ATP hydrolysis. It is also notable that deletion of the N-terminal domain affects the pattern of ATP hydrolysis by the parent enzymes and Walker-B mutants in a fashion similar to high Mg$^{++}$ concentration (Fig. 1B and 1C).
**Figure 1.** Effects of Walker-B mutations in the D1 and D2 rings on ATP hydrolysis and protein unfolding. (A) In a buffer with 20 mM MgCl₂, the E291Q and E568Q mutations modestly decrease the rate of hydrolysis of 10 mM ATP compared to the Cdc48 parent. Values are means (N=2) ± SEM. (B) In buffer with 120 mM MgCl₂, the E291Q and E568Q mutations reduce the Cdc48 hydrolysis rate ~6-fold. Values are means (N=2) ± SEM. (C) In 20 mM MgCl₂, the E291Q and E568Q mutations reduce the Cdc48AN hydrolysis rate ~10-fold. Values are means (N=2) ± SEM. (D) Cdc48AN unfolds Kaede-ssrA (5 µM) ~20-fold faster than E291Q Cdc48AN and ~6-fold faster than E568Q Cdc48AN. Values are means (N=3) ± SD.

To assess the importance of ATP hydrolysis in Cdc48AN machine function, I monitored unfolding of a model protein substrate, photo-cleaved Kaede with a C-terminal ssrA tag that is recognized by archaeal Cdc48 (Gerega et al., 2005). Following exposure to 340-400 nm light, Kaede is photo-cleaved at a single site, but the protein retains its native fold and shows red fluorescence (Ando et al. 2002). After unfolding, however, fluorescence is irreversibly lost. Cdc48AN unfolded photo-cleaved Kaede-ssrA at a rate ~20-fold faster than E291Q Cdc48AN and ~6-fold faster than E568Q Cdc48AN (Fig. 1D). Thus, robust ATP hydrolysis in both the D1 ring and the D2 ring is important to power high levels of unfolding activity.

**Effects of nucleotide-binding mutations**

In other AAA+ hexamers, ATP binding can be weakened substantially by Lys→Ala mutations in the Walker-A motifs (Wang et al., 2003). I constructed these Walker-A mutations in the D1 ring (K237A) or the D2 ring (K514A) of Cdc48AN. Both the K237A and K514A mutations resulted in an ~5-fold reduction in the hydrolysis rate of 10 mM ATP compared to the parent enzyme (Fig. 2A). For K237A Cdc48AN, the hydrolysis rate followed Michaelis-Menten kinetics with a $K_M$ of ~3 mM, which I assume represents the interaction of ATP with the unmutated D2 ring (Fig. 2B). For K514A Cdc48AN, by
contrast, the hydrolysis rate decreased hyperbolically with ATP concentration until reaching a plateau (Fig. 2C). This surprising behavior can be rationalized if the K514A mutation destabilizes the hexamer, a smaller oligomer has higher ATPase activity than the hexamer, and increasing ATP stabilizes the hexamer. Monomers would be inactive as the active site spans the interface between subunits, and thus it seems likely that dimers or trimers of K514A Cdc48\textsuperscript{MN} have higher ATPase activity than the hexamer. The half-maximal decrease in ATPase activity occurred at an ATP concentration of 80 nM, which is likely to represent the interaction of ATP with the unmutated D1 active sites.

**Figure 2.** Effects of ATP-binding mutations in the Walker-A motifs of the D1 and D2 rings. (A) K237A and K514A Cdc48\textsuperscript{AN} hydrolyze 10 mM ATP at a rate approximately one-fifth of the Cdc48\textsuperscript{AN} parental rate. Values are means (N=2) ± SEM. (B) ATP dependence of hydrolysis by the K237A Cdc48\textsuperscript{AN} mutant. The line is a fit to the Michaelis-Menten equation (rate = \(V_{\text{max}}/(1 + K_M/\text{[ATP]}\)) with a fitted \(K_M\) of \(3\pm0.6\) mM and a \(V_{\text{max}}\) of \(62\pm5\) min\(^{-1}\) ATP hexamer\(^{-1}\). Values are means (N=2) ± SD. (C) Increasing ATP decreases the rate of hydrolysis by K514A Cdc48\textsuperscript{MN}, suggesting that hydrolysis at low ATP concentrations results from a dimer or trimer (see text). The line is a fit to a hyperbolic function. Values are means (N=2) ± SD.

**Allosteric interactions**

Each ring of Cdc48 contains six active sites for ATP hydrolysis. In principle, therefore, ATP binding and hydrolysis in different subunits of either the D1 or the D2 rings could occur with positive cooperativity (Hill constant > 1), no cooperativity (Hill constant = 1), or negative cooperativity (Hill constant < 1). To test these possibilities, I measured initial
rates of hydrolysis by Cdc48\textsuperscript{AN} and the Walker-B variants over a range of ATP concentrations and fitted the resulting curves to the Hill form of the Michaelis-Menten equation (Fig. 3A). ATP hydrolysis by wild-type Cdc48\textsuperscript{AN} showed strong positive cooperativity (Hill constant 3.6 ± 0.6) and an apparent $K_M$ of ~80 nM. The E291Q Cdc48\textsuperscript{AN} enzyme had an apparent $K_M$ of ~60 nM and Hill constant (0.84 ± 0.16) consistent with no cooperativity or weak negative cooperativity.\textsuperscript{2} The E568Q Cdc48\textsuperscript{AN} variant showed either weak positive cooperativity or no cooperativity (Hill = 1.33 ± 0.27) and had an apparent $K_M$ of ~55 nM. Thus, robust and positively cooperative ATP hydrolysis in Cdc48\textsuperscript{AN} requires wild-type active sites in both the D1 and the D2 rings. This result suggests that ring-ring communication is important for robust Cdc48 function.

\textsuperscript{2} A Hill constant less than one can indicate true negative cooperativity (binding to initial sites makes binding to subsequent sites weaker) or indicate that there are two sets of non-interacting binding sites with different affinities.
Figure 3. Effects of Walker-B and linker-insertion mutations on ATP hydrolysis. (A) Walker-B mutations in either the D1 ring (E291Q) or D2 ring (E568Q) substantially reduce $V_{\text{max}}$ for ATP hydrolysis. The lines are fits to the Hill equation ($\text{rate} = \frac{V_{\text{max}}}{1 + (K_M/[\text{ATP}])^n}$). The Hill constant ($n$) was $-3.6$ for Cdc48$^{\Delta N}$, $-0.84$ for E291Q Cdc48$^{\Delta N}$, and $-1.33$ for E568Q Cdc48$^{\Delta N}$. (B) In the crystal structure of mammalian Cdc48/p97 (PDB 3CF2), a linker of ~20 amino acid connects the D1 and D2 rings. (C) Alignment of linker sequences from T. acidophilum (Ta), S. cerevisiae (Sc), M. musculus (Mm), and H. sapiens (Hs). The site of the 10-residue insertion in Cdc48$^{\text{ins10}}$ is shown above the alignment. (D) In full-length Cdc48, the insertion decreased $K_M$ from ~580 to ~16 nM and increased the Hill constant from ~0.58 to ~0.97. Values are means (N=2) ± SD. (E) Compared to Cdc48$^{\Delta N}$ in panel A, the insertion reduced $K_M$, $V_{\text{max}}$, and the Hill constant (see text). (F) A shift from 20→120 mM MgCl$_2$ stimulates ATP hydrolysis by Cdc48 but not by Cdc48$^{\text{ins10}}$, Cdc48$^{\Delta N}$, or Cdc48$^{\Delta N/\text{ins10}}$. In panels A, D, and E, values are means (N=2) ± SD.

In crystal structures of mammalian Cdc48 (DeLaBarre and Brunger et al., 2005; Davies et al., 2008), a linker of ~20-residues connects the D1 and D2 domains (Fig. 3B). The length of this linker is conserved and there is also substantial sequence homology in the Cdc48 linkers from archaea to humans (Fig. 3C), supporting an important role. To investigate linker function, I deleted residues 456-465 or inserted an extra 10 residues (ins10) after position M458 (Fig. 3C). Both mutants were constructed in full-length and Cdc48$^{\Delta N}$ backgrounds. The deletion mutants failed to express, presumably as a consequence of folding defects, but the insertion mutants expressed well and were purified.

Compared to the full-length Cdc48 parent, the insertion mutation resulted in much faster ATP hydrolysis at significantly lower ATP concentrations (Fig. 3D). For Cdc48$^{\text{ins10}}$, the apparent $K_M$ was $16 \pm 1$ nM, $V_{\text{max}}$ was $156 \pm 2$ ATP min$^{-1}$ hexamer$^{-1}$, and the Hill constant was $0.97 \pm 0.05$. For full-length Cdc48, the apparent $K_M$ was $580 \pm 140$ nM, $V_{\text{max}}$ was $67 \pm 41$ ATP min$^{-1}$ hexamer$^{-1}$, and the Hill constant was $0.58 \pm 0.14$. Thus, the insertion reduces $K_M$ ~35-fold, increases $V_{\text{max}}$ ~2.5 fold, and changes the reaction from
one showing negative cooperativity to no cooperativity. From fitting of the Cdc48^{ΔN/ins10}
data (Fig. 3E), $K_M$ apparent was 42 ± 3 nM, $V_{\text{max}}$ for hydrolysis was 148 ± 3 ATP min$^{-1}$
hexamer$^{-1}$, and the Hill constant was 1.29 ± 0.09. Compared to the Cdc48^{ΔN} parent (Fig.
3A), Cdc48^{ΔN/ins10} showed reductions in $K_M$ apparent, $V_{\text{max}}$, and the Hill constant.
Because the insertion causes opposite changes in these parameters in Cdc48^{ΔN}
compared to full-length Cdc48, both the N-domain and interactions between the D1 and
D2 ring disrupted by the insertion appear to regulate activity. For example, the N-
domain seems to inhibit ATP hydrolysis in the D1 and D2 rings, but this inhibition can be
reversed by perturbing interactions between these rings by the linker insertion.

Although Cdc48^{ins10} and Cdc48^{ΔN/ins10} hydrolyzed concentrations of ATP higher than 0.1
mM at ~60% of the rate of Cdc48^{ΔN} (Fig. 3A, 3D & 3E), neither insertion variant had
substantial Kaede-ssrA unfolding activity. The Cdc48^{ins10} unfolding rate was 4% that of
Cdc48^{ΔN}, and the Cdc48^{ΔN/ins10} unfolding rate was 8% of the Cdc48^{ΔN} parent (data not
shown). These results demonstrate that the ATP-hydrolysis cycle of Cdc48 can be
largely uncoupled from machine function by increasing the length of the D1-D2 linker.
Thus, proper communication between the D1 and D2 rings is required to coordinate
conformational changes driven by the ATPase cycle with mechanical work.

As noted above and shown in Fig. 3F, shifting from 20 to 120 mM Mg$^{++}$ stimulates the
ATPase activity of full-length Cdc48 ~7-fold. By contrast, neither the Cdc48^{ins10},
Cdc48^{ΔN}, nor Cdc48^{ΔN/ins10} variants hydrolyzed ATP appreciably faster in the high Mg$^{++}$
buffer (Fig. 3E). Again, the linker insertion seems to alter effects that the N-domain
normally has on repression of ATP hydrolysis in the D1 and D2 rings.
ATP transactions and formation of a high-affinity Cd48•20S complex

Binding of archaeal Cdc48 to the 20S proteasome requires ATP or ATPγS, although binding is relatively weak with the latter slowly-hydrolyzable analog (Barthelme et al., 2014). Two regions of Cdc48 contact 20S directly: HbYX tripeptides at the C-terminus of its flexible tail and the pore-2 loops of the D2 ring (Barthelme et al., 2013). Cdc48\textsubscript{AN} binds 20S substantially more tightly than full-length Cdc48, but the N-domain is more than 50 Å from 20S in the complex (Barthelme and Sauer, 2012; Barthelme et al., 2014). Thus, allosteric effects of the N-domain on ATP binding or hydrolysis may account for weakening of 20S binding. To address this possibility, I tested whether the Walker-A and Walker-B mutations in the D1 and D2 rings affect complex formation.

![Graph A](image1.png)

![Graph B](image2.png)

**Figure 4.** As measured by nonapeptide cleavage, ATP transactions in the D1 or D2 rings influences Cdc48\textsubscript{AN} affinity for 20S (10 nM). All experiments contained 10 μM nonapeptide and 5 μM ATP, and data were fit with a quadratic equation for near-stoichiometric binding. (A) Slowing ATP hydrolysis in the D2 ring strengthens 20S binding, whereas slowing hydrolysis in the D1 ring weakens 20S binding. (B) Perturbing ATP binding to the D2 ring strengthens 20S binding, whereas perturbing binding to the D1 ring weakens binding. Note that the K514A and K237A mutations also reduce the maximal rate of peptide degradation in comparison to the Cdc48\textsubscript{AN} control in panel A. In both panels, values are means (N=2) ± SD.
To assay binding, I titrated increasing concentrations of Cdc48\textsuperscript{\textalpha N} or variants against a fixed concentration of 20S and monitored complex formation by stimulation of 20S cleavage of a fluorogenic nonapeptide, which depends upon Cdc48 opening the axial gate into the 20S chamber.\textsuperscript{3} Fitting these data gave apparent binding constants ($K_{app}$) of $29 \pm 7$ nM for Cdc48\textsuperscript{\textalpha N}, $290 \pm 57$ nM for E291Q Cdc48\textsuperscript{\textalpha N}, and $1.4 \pm 1$ nM for E568Q Cdc48\textsuperscript{\textalpha N} (Fig. 4A). Thus, slowing ATP hydrolysis in the D1 ring weakens 20S binding, whereas slowing hydrolysis in the D2 ring strengthens 20S binding. For the Walker-A mutants, $K_{app}$ was $480 \pm 110$ nM for K237A Cdc48\textsuperscript{\textalpha N} and $3 \pm 1$ nM for K514A Cdc48\textsuperscript{\textalpha N} (Fig. 4B). These results show that ATP binding and hydrolysis in the D1 ring, which contacts the N-domain in full-length Cdc48, facilitates assembly of 20S complexes, potentially by properly positioning elements in the D2 ring for better 20S interactions. By contrast, both ATP binding and hydrolysis in the D2 ring weaken 20S affinity, presumably through direct effects on Cdc48-20S contacts.

If nucleotide binding/hydrolysis by Cdc48 is important for 20S complex formation, then 20S binding should alter the ATPase activity of Cdc48. To test for bidirectional allostery, I assayed the ATP dependence of hydrolysis by Cdc48\textsuperscript{\textalpha N} in the presence of a saturating concentration of 20S, which resulted in biphasic dependence of the hydrolysis rate on ATP (Fig. 5A). The first phase had an apparent $K_M$ of $11 \pm 1$ $\mu$M and a Hill constant of $2.7 \pm 0.6$, whereas the second phase had a $K_M$ of $140 \pm 12$ $\mu$M and was non-cooperative. For comparison, Fig. 5A shows hydrolysis in the absence of 20S (same

\textsuperscript{3} One caveat is that gate opening is an indirect measurement of binding, and only enzymatically active complexes are detected. Moreover, 20S complexes with different Cdc48 variants may have different properties. Indeed, maximal gate-opening activity was lower for the Walker-A mutants (Fig. 4B) than for the Walker-B mutants (Fig. 4A).
data as in Fig. 3A. For full-length Cdc48, addition of 400 nM 20S (reduced $K_M$ apparent from 1.3 to 0.2 mM and increased the Hill constant from $0.74 \pm 0.06$ to $1.35 \pm 0.13$ (Fig. 5B). By contrast, 600 nM 20S did not substantially alter the ATP dependence of hydrolysis by E291Q Cdc48 (Fig. 5C), even though this mutant bound 20S in a gate-opening assay (data not shown). I conclude that 20S binding largely affects ATP binding/hydrolysis in the D1 ring, a result consistent with the deleterious effects of the Walker-B E291Q D1 mutation on 20S binding (Fig. 4A). Clearly, however, this effect must be transmitted through the D2 ring.

Figure 5. Effects of the 20S peptidase on ATP hydrolysis by Cdc48 and variants. Each plot shows the ATP dependence of hydrolysis with and without 20S (400 nM). (A) Hydrolysis by Cdc48$^{AN}$ and Cdc48$^{AN}$-20S. The presence of 20S results in a biphasic curve with one phase showing a lower $K_M$, and the other phase showing a higher $K_M$ compared to Cdc48$^{AN}$ alone (see text). (B) For full-length Cdc48, addition of 20S decreased $K_M$ ~6-fold and increased the Hill constant from ~0.7 to ~1.3. (C) Addition of 20S to E291Q Cdc48 resulted in little change in the ATP dependence of hydrolysis, suggesting that hydrolysis in the D1 ring is responsible for the activating influence of 20S in panel B. In each panel, values are means (N=2) ± SD.

N domain control of nucleotide binding and dissociation

The N domain of Cdc48 decreases $V_{max}$ for ATP hydrolysis and increases the apparent $K_M$ (Figs. 5A and 5B). Moreover, 120 mM Mg$^{++}$ markedly stimulates ATP hydrolysis by Cdc48 but modestly represses hydrolysis by Cdc48$^{AN}$ (Fig. 3E). As shown in Fig. 6 for
full-length Cdc48, increasing concentrations of Mg\(^{++}\) increased \(V_{\text{max}}\) for ATP hydrolysis and decreased \(K_{M}\) apparent. The Hill constants were 0.54 ± 0.17 (20 mM Mg\(^{++}\)), 2.7 ± 0.3 (60 mM Mg\(^{++}\)), and 3.3 ± 0.7 (120 mM Mg\(^{++}\)), reflecting a switch from negative to strong positive cooperativity. Thus, increasing concentrations of Mg\(^{++}\) cause similar changes in the kinetic parameters for ATP hydrolysis as deletion of the Cdc48 N domain. Because the N domain normally packs against the D1 ring (Barthelme et al., 2014), increasing Mg\(^{++}\) may disrupt this interaction, resulting in altered enzymatic properties.

![Figure 6](image)

**Figure 6.** Effects of 20, 60, and 120 mM MgCl\(_2\) on ATP hydrolysis. Increasing MgCl\(_2\) concentrations decreased \(K_{M}\) for ATP and increased the cooperativity of the reaction, as measured by the Hill constant (see text). Values are means (N=2) ± SD.

An ADP analog with an N-methylanthraniloyl (mant) fluorophore attached to the 2\(^{'}\) or 3\(^{'}\) hydroxyl groups of the ribose sugar can be used for binding studies, as fluorescence increases in the non-polar environment of a nucleotide-binding site (Hersch et al., 2005; Sun et al., 2005). The fluorescence of mant-ADP increased upon binding to Cdc48\(^{\Delta N}\), with smaller increases observed upon binding to the Walker-A mutants (Fig. 7A). The latter result is consistent with reduced occupancy caused by weaker binding. In manual
mixing experiments, I found that addition of 4 mM ADP caused very rapid dissociation of mant-ADP from Cdc48\textsuperscript{AN} but substantially slower dissociation from full-length Cdc48 (Fig. 7B).

**Figure 7.** N-domain effects on mant-ADP binding. (A) As judged by enhanced fluorescence, mant-ADP (12 μM) bound best to Cdc48\textsuperscript{AN}, slightly more poorly to K237 Cdc48\textsuperscript{AN} or K514A Cdc48\textsuperscript{AN}, and worst to K237/K514A Cdc48\textsuperscript{AN} (all protein concentrations were 1.6 μM). (B) Dissociation kinetics of mant-ADP from Cdc48 or Cdc48\textsuperscript{AN} following manual mixing with an equal volume of buffer containing unlabeled ADP (4 mM). (C) Dissociation kinetics of mant-ADP from Cdc48 as a function of MgCl\textsubscript{2} concentration measured in a stopped-flow instrument. (D) Dependence of the half-life of mant-ADP-Cdc48 complexes on MgCl\textsubscript{2} concentration. Half lives were determined by exponential fits of the data in panel C. (E) Kinetics of dissociation of mant-ADP from Cdc48\textsuperscript{AN} in buffers containing 20 or 120 mM MgCl\textsubscript{2} measured by stopped-flow experiments. The lines are exponential fits with half-lives of ~13 msec (20 mM MgCl\textsubscript{2}) and ~ 7 msec (120 mM MgCl\textsubscript{2}).

Because the presence of the N domain in full-length Cdc48 slows dissociation of an ADP analog, increasing Mg\textsuperscript{++} might stimulate hydrolysis by increasing the rate of ADP dissociation. To test this model, I measured rates of mant-ADP dissociation from full-length Cdc48 in buffers containing different concentrations of MgCl\textsubscript{2} using a stopped-
flow instrument (Fig. 7C). Notably, increasing MgCl$_2$ accelerated mant-ADP dissociation, with the half-life decreasing from $\sim$18 s in 20 mM MgCl$_2$ buffer to $\sim$3 s in 120 mM MgCl$_2$ buffer (Fig. 7D).

If high concentrations of Mg$^{++}$ relieve N-domain repression of ATP hydrolysis by accelerating ADP release, then dissociation of ADP from Cdc48$^{\Delta N}$ should be substantially faster and relatively independent of Mg$^{++}$ compared to dissociation from full-length Cdc48. Both predictions were confirmed. In the Cdc48$^{\Delta N}$ background, mant-ADP dissociated with half-lives of $\sim$13 and $\sim$7 msec in 20 mM and 120 mM MgCl$_2$ (Fig. 7E). At both Mg$^{++}$ concentrations, reasonable fits of the dissociation trajectories required a double-exponential function with similar amplitudes but one time constant $\sim$7-8 larger than the other. This result suggests the existence of two classes of nucleotide binding sites. In principle, these classes could represent dissociation from subunits in the D1 or the D2 rings or from subunits with different nucleotide-binding properties within single rings (Hersch et al., 2005; Stinson et al., 2015).

To ask if the D1 and D2 rings behave differently with respect to mant-ADP release, I initially measured dissociation rates using the K237A and K514A Walker-A variants in the Cdc48$^{\Delta N}$ enzyme (Fig. 8A). In these experiments, the dissociation trajectories were similar for both mutants, and dissociation was $\sim$40-fold slower (half-lives $\sim$ 1 s) than from the Cdc48$^{\Delta N}$ parent (Fig. 7E). Moreover, both curves were fit reasonably well by a single-exponential function ($R > 0.996$). Thus, when nucleotide binding to either the D1 ring or the D2 ring is severely weakened by a Walker-A mutation, the nucleotide-binding properties of the unmutated ring are altered.
**Figure 8.** Nucleotide dissociation from the D1 and D2 rings. mant-ADP (12 μM) was preincubated with K237A or K514A Cdc48\(^{\Delta N}\) (1.6 μM; panel A) or K237A or K514A Cdc48 (1.6 μM; panel B) and dissociation was monitored by changes in fluorescence following dilution into an equal volume of buffer containing 4 mM ADP.

Next, I measured dissociation for the K237A and K514A mutants in otherwise wild-type Cdc48 (Fig. 8B). For K237A Cdc48, mant-ADP dissociation fit well to a single exponential (\(R = 0.993\)) with a half-life of \(~1.2\) s. This value presumably reflects dissociation from the unmutated D2 ring and is similar to the value observed in K237A Cdc48\(^{\Delta N}\). The N domain therefore appears to have little effect on nucleotide dissociation from the D2 ring. By contrast, mant-ADP dissociation from the unmutated D1 ring of K514A Cdc48 was much slower and required a double-exponential fit. Thus, the N domain of Cdc48 principally effects nucleotide dissociation from the D1 ring, which it directly contacts.

**Discussion**

AAA+ proteases use the binding and hydrolysis of ATP to drive conformational changes which power mechanical movements used to do the work of protein unfolding and translocation. Here I have taken a biochemical approach toward expanding our
understanding of the interactions between domains of Cdc48 and how coordination of these interactions is important to function. The rigid topology of AAA+ hexamers provides a means for allosteric communication of nucleotide bound state around the ring (Hersch et al., 2005; Stinson et al., 2013). By contrast, it is unclear if and how double-ring AAA+ proteases like Cdc48 use communication between D1 and D2 rings. In Cdc48^AN I observed an interdependence of ATP hydrolysis between the rings. When hydrolysis was impaired in one ring I observed a significant drop in hydrolysis in the other ring and this was reflected similarly in the unfolding assays. Full length Cdc48 was only modestly affected by single-ring E→Q mutations, but at 120 mM Mg^{++} show a 6-fold defect in ATP hydrolysis rate. I observed a similar pattern in mutations that impaired nucleotide binding in either ring. Thus, nucleotide binding and hydrolysis is required in both rings for robust rates of ATP hydrolysis and unfolding. Similar experiments of p97 have consistently found that the D2 ring is dominant in ATP hydrolysis finding that >80% of activity remained when hydrolysis was impaired in the D1 (Song et al., 2003; Chou et al., 2014). D2 dominance was also observed in bacterial ClpA (Kress et al., 2009). In contrast, studies of Hsp104, a homolog of the bacterial ClpB chaperone, showed hydrolysis is important in both rings, finding large defects in activity with E→Q mutations in either D1 or D2 (Hattendorf and Lindquist, 2002). In contrast, another double ring AAA+ called N-ethylmaleimide sensitive factor (NSF) showed D1 dominance. When either Walker A or Walker B mutations were introduced to the D1 enzyme activity was completely abolished (Whiteheart et al., 1994). Thus, double ring AAA+ enzymes have evolved multiple modes of domain activity and even the closely related homologs Cdc48 and p97 display large differences.
Cooperativity is a feature often seen in multimeric enzymes and is a property which can coordinated the activities of individual subunits. Cdc48 must coordinate the activity of six subunits, with each subunit containing two active sites. Full length Cdc48 is noncooperative (Hill constant < 1) in ATP hydrolysis, but Cdc48\(^{AN}\) enzyme shows a marked increased in Hill coefficient (Hill constant > 3.5). Impairing ATP hydrolysis in either D1 or D2 ring abolished cooperativity which also produced defects in unfolding activity. Hsp104 has similar ATP hydrolysis characteristics to Cdc48 (above) thus it would be reasonable to guess it would have similar allosteric cooperativity. Indeed wild type Hsp104 demonstrates positive cooperativity (Hill constant = 2.3), but mutations impairing hydrolysis in either ring only modestly affect cooperativity (Hill constant = 1.9).

So Cdc48 seems to rely heavily on the cooperative ATP hydrolysis (Hattendorf and Lindquist, 2002). We reasoned that inter-domain communication must be important in this process. A conserved linker connects the D1 with D2 and I hypothesizing that the linker could play a role in communication. By lengthening the linker with a flexible 10 amino acid stretch I could assay the change in inter-ring communication. This resulted in reduced hydrolysis rate and reduced Hill constant in Cdc48\(^{AN}\), indicating that tight contacts between D1 and D2 are important for robust hydrolysis and cooperativity. The same mutation in the full length parent had the opposite effect, the V\(_{max}\) and Hill constant was increased. We can rationalize this behavior with a model in which the D1 and D2 interact during the ATPase cycle, a process which is also regulated by N-domain interactions with the D1 ring. Breaking D1-D2 communication diminishes their ability to hydrolyze ATP, but also diminishes the repressive interaction of the N domain.
Studies attempting to measure activity of p97 truncations found that N-D1 fragment had no ATPase activity, but when the linker was included (N-D1-linker) nearly 100% was restored (Chou et al., 2014). They suggest the linker positions a water molecule in the D1 active site, readying it for hydrolysis. Thus, linker could function in D1 activity, which in turn promotes activity in the D2 ring.

Forming a high affinity complex with 20S requires ATP hydrolysis in the D1 ring while hydrolysis in the D2 is detrimental. Cdc48 interacts with 20Sα subunit through C-terminal HbYX motif, but also makes contacts via axial pore-2 loops in the D2 ring (Barthelme and Saur, 2013). In ClpX both static and dynamic connections are made with ClpP. Peripheral IGF loops are static and nucleotide independent whereas the pore-2 loops are nucleotide dependent and mutations therein affect ATPase rate (Martin et al., 2007). The pore-2 loops are located near the Walker-B site of the D2 ring which would allow them to respond dynamically to changes in nucleotide state throughout the ATPase cycle. This is one explanation for how hydrolysis in D1 and D2 can control interactions with 20S.

Finally, we found that the N domain is critical in regulating nucleotide bound state of the D1 ring. Using Walker A mutations in D1 and D2 rings, mant-ADP displayed similar dissociation kinetics in the Cdc48\(^{N}\) parent, however the full length parent showed a 15-fold slower rate in the D1 ring. The dissociation of mant-ADP from the D1, which we believe behaves similarly to unlabeled ADP, was accelerated by high MgCl\(_2\) concentrations. These experiments link the function of the N domain and ATPase cycle with the nucleotide state of the D1 ring. The cryo-EM reconstruction of Cdc48\(^{N}\)20S
complexes showed that >95% of cross-linked complexes had N domains in a coplanar position, packed against the D1 ring (Barthelme et al., 2014). Crystal structures of the mammalian enzyme indicate the N-domain packs tightly into a groove near the D1 nucleotide-binding pocket (DeLaBarre and Brunger, 2005) where aspartate and asparagine residues mediate contacts between the two domains, several of which have been implicated in disease phenotypes and, when mutated, are known to increase the rate of ATP hydrolysis (Tang and Xia, 2013; Niwa et al, 2012).

Collaboration between AAA+ rings appears critical to Cdc48 function. Disrupting the elements important to this collaboration results in loss of cooperative ATP hydrolysis and an uncoupling of ATP hydrolysis and machine function. In the future it would be useful to obtain structures of Cdc48^AN and full-length Cdc48 to identify structural features governing ATP hydrolysis rate and 20S binding and how these changes are transmitted across the length of the enzyme.

**Materials and Methods**

*Cloning, expression and purification of proteins*

Proteins encoding *Thermoplasma acidophilum* Cdc48, 20Sα, and 20Sβ were cloned with His6-tag (Barthelme and Sauer, 2012) and expressed in *E. coli* BL21 DE3 RIL cells, which were grown at 18 °C to OD600 = 0.8, induced with 1 mM isopropyl thiogalactoside (IPTG), and then grown overnight. 20S genes were co-expressed and purified as a complex. Cells were harvested by centrifugation, washed, and resuspended in 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 20 mM imidazole. For the Cdc48 purification, this buffer was supplemented with 1 mM PMSF. Cells were lysed by freeze-thaw and
sonication for 3 min on ice. Lysates were centrifuged at 30,000 x g for 30 min, and the supernatant was placed in a water bath at 60 °C. After 20 min, the lysate was centrifuged again to clear precipitated proteins, and the supernatant was incubated with Ni²⁺-NTA beads for 30 min at room temperature. Bound proteins were eluted from the Ni²⁺-NTA beads with 20 mM Tris-HCl (pH 8.0), 100 mM 300 mM imidazole. The eluted material was diluted by addition of an equal volume of 20 mM Tris-HCl (pH 8.0) and was applied to MonoQ column for anion-exchange chromatography. Purification was performed by running a linear gradient from 0-600 mM NaCl in 20 mM Tris-HCl (pH 8.0). Fractions containing the desired protein were identified by SDS-PAGE, concentrated to 2 mL using a Millipore Centricon 25 kDa cutoff concentrators, and applied to a Superdex S200 gel-filtration column equilibrated in HBS (50 mM HEPES-KOH (pH 7.5), 20 mM MgCl₂). Appropriate fractions were identified by SDS-PAGE, concentrated, quantified by absorbance at 280 nm, and frozen in liquid nitrogen.

His₆-Kaede-ssrA was expressed in E. coli strain BL21 DE3 RIL as described above, but the culture was grown for only 4 h after addition of IPTG, and the incubation step at 60 °C following cell lysis was omitted. Following Ni²⁺-NTA and MonoQ purification steps, the purified protein was exposed to direct sunlight by placement on a south facing window for 2-3 days. Photo cleavage was monitored by following appearance of an absorbance band at 580 nm after excitation at 572 nm.

Measurements of ATP hydrolysis

Rates of ATP hydrolysis were measured using a system in which pyruvate-kinase catalyzed conversion of ADP and phosphoenolpyruvate to ATP and pyruvate was
coupled to lactate-dehydrogenase catalyzed conversion of pyruvate to lactate with concomitant oxidation of NADH to NAD (Norby, 1988; Barthelme and Sauer, 2012). During assays, decreases in absorbance at 340 nm resulting from the oxidation of NADH were measured using a plate reader. Unless noted, reactions were performed at 45˚C in HBS buffer using 0.1 μM Cdc48 hexamer.

Measurements of protein unfolding
Kaede-ssrA unfolding assays were performed in a fluorescence plate reader (excitation 508 nm; emission 580 nm) using Corning half area black plates. Reactions were performed in a 50 μL reaction volume in HBS buffer with Cdc48 or variants (0.3 μM hexamer), 10 mM ATP, and different concentrations of Kaede-ssrA. Reactions were started by adding ATP to a mixture of the other assay components that were preincubated at 45˚C for 10 min.

Cdc48 stimulation of 20S peptide cleavage
The rate of 20S cleavage of a fluorogenic 9-residue peptide (Mca-AKVYPYPME-Dpa(Dnp)-amide) was measured by changes in fluorescence (excitation 340 nm; emission 405 nm) in buffer containing 50 mM HEPES-KOH (pH 7.5) and 10 mM MgCl₂ as described (Barthelme and Sauer, 2012). Different concentrations of Cdc48 or variants and 20S (10 nM) were preincubated for 10 min at 45˚C, and reactions were started by addition of ATP (5 mM) and peptide (10 μM).
Nucleotide binding and dissociation

Experiments with mant-ADP were performed in HBS buffer at 22 °C, typically using 12 μM of mant-ADP (Jena Bioscience) and 1.6 μM of Cdc48 or variants, and were monitored by changes in fluorescence (excitation 320 nm; emission 405 nm). In some cases, reactions were supplemented with additional MgCl₂. For steady-state binding, measurements were taken in a fluorescence plate reader. Manual-mixing dissociation experiments in a 70 μL cuvette were performed in a fluorimeter. Stopped-flow dissociation experiments were performed in a KinTek SF-300X instrument following 1:1 mixing of Cdc48 or variants prebound to mant-ADP with a solution containing 4 mM ADP.
References


Wang, Q., Song, C., Yang, X., Li, C.C. (2003). D1 ring is stable and nucleotide-independent, whereas D2 ring undergoes major conformational changes during the ATPase cycle of p97-VCP. J. Biol. Chem. 278, 32784-32793.

Chapter 3:

Screening Methods for Identifying Substrates of Cdc48•20S
Introduction

Cdc48 is a AAA+ unfoldase with two ATPase rings. In archaea, Cdc48 works with the 20S peptidase to degrade protein substrates (Pamnani et al., 1997; Barthelme and Sauer, 2012). Many archaea have another AAA+ unfoldase, known as PAN, that also functions with 20S to degrade proteins (Zwickl et al., 1999). PAN, its eukaryotic homologs (the proteasomal RPT1-6 subunits), and most Cdc48 proteins have a C-terminal hydrophobic-tyrosine-anything (HbYX) motif that assists in 20S docking. In fact, all archaea with sequenced genomes have genes encoding 20S and at least one Cdc48 enzyme with a HbYX tail. PAN, however, is present in only a subset of archaea species. Thus, archaeal Cdc48•20S appears to serve an important biological role.

Because of its relative simplicity compared to eukaryotic homologs, the archaeal 20S proteasome was initially used as a model system for understanding the structure and function of proteasomes in general. For example, the principles of subunit architecture, catalysis, and docking with PAN were established in large part by studies of the archaeal enzymes (Zwickl et al., 1994; Semuller et al., 1995; Lowe et al., 1995; Seemuller et al., 1996; Baumeister et al., 1998; Smith et al., 2005). Many of these findings have since been extended to eukaryotic proteasomes and provided the foundation for experiments on the much more complicated ubiquitin-proteasome system (Kish-Trier and Hill, 2013; Matyskiela and Martin, 2013). In contrast to the substantial body of biochemical literature, relatively little is known about the biological role of archaeal proteasomes and their AAA+ partners. In Haloferax volcanii, which possesses genes for the 20S beta-subunit (β), two genes for the 20S alpha-subunit (α1 and α2),
and two genes for PAN (panA and panB), deletion of β, α1, or panA resulted in cells that were sensitive to high osmolarity (Zhou et al., 2008). Interestingly, deletion of panA also resulted in increased thermotolerance. In a somewhat contradictory finding, Ruepp and colleagues treated Thermoplasma acidophilum, which does not contain PAN, with a 20S inhibitor and found minimal effects under normal growth conditions but severe growth defects following heat shock (Ruepp et al., 1998). In these experiments, they were able to show 75-80% covalent modification of the 20S beta-subunits by their inhibitor. Thus, wild-type levels of functional 20S do not appear to be important for normal growth. Another study looked at the proteomes of wild-type and panA-deleted cells (Kirkland et al., 2008). Deletion of panA resulted in a two-fold increase in phosphorylated proteins and increased expression of Cdc48. To my knowledge there are no published reports of Cdc48 deletion in archaea.

To date, only model ssrA-tagged protein substrates have been shown to be unfolded by Cdc48 or degraded by Cdc48•20S (Gerega et al., 2005; Barthelme and Sauer, 2012). However, the eubacterial ssrA-tagging system is absent in archaea (Moore and Sauer, 2007). As a prelude toward the identification of archaeal substrates of Cdc48•20S, I have evaluated a method to trap proteins within the chamber of inactive 20S and have also screened an E. coli lysate for proteins degraded by Cdc48•20S.

Results

Substrate trapping

Substrates of ClpXP can be trapped within an inactive ClpP barrel and then identified by proteomics for substrate discovery (Kim et al., 2000; Flynn et al., 2003; Neher et al.,
2006). Thus, I sought to determine if a similar approach, diagrammed in Fig. 1A, might work for Cdc48•20S. For these studies, I used a His$_6$-tagged βT1A 20S variant in which the catalytic threonine of the beta subunit was mutated to alanine (Kisselev et al., 2000). To test this approach, I used ssrA-tagged GFP, which is recognized and degraded by Cdc48$^{AN}$•20S. Prior to the experiment, Cdc48$^{AN}$ and GFP-ssrA were treated with TEV protease to remove their His$_6$ tags, so that only 20S$^{βT1A}$ had a His$_6$ affinity tag. GFP-ssrA was mixed with Cdc48$^{AN}$•20S$^{βT1A}$ and ATP, and reactions were incubated at 45°C for trapping. Reactions were then passed over small Ni$^{++}$-NTA columns to bind 20S and any associated proteins and the bound fraction was eluted and analyzed by SDS-PAGE. Fig. 1B shows the results of a representative experiment in which GFP-ssrA copurified with His$_6$-20S$^{βT1A}$. Higher input concentrations of GFP-ssrA resulted in improved trapping (Fig. 1B, left lanes), and only trace amounts of GFP copurified with His$_6$-20S$^{βT1A}$ when Cdc48$^{AN}$ was omitted from the reaction (Fig. 1B, middle lanes). The right lanes of Fig. 1B show the reaction components prior to Ni$^{++}$-NTA purification.
Figure 1. Substrate trapping by inactivated 20S. (A) Cartoon depiction of trapping and identification strategy. (B) GFP-ssrA trapped with Cdc48^AN•20S^BT1A after Ni^{2+}-NTA purification (left three lanes). No Cdc48^AN reactions (middle lanes). Reaction components prior to pull-down (right lanes). (C) GFP-ssrA (10 μM) fluorescence (excitation 467 nm; emission 511) normalized to time zero of reactions containing Cdc48^AN•20S^BT1A plus 10 mM ATP (red), Cdc48^AN alone plus 10 mM ATP (purple), Cdc48^AN•20S^BT1A plus 10 μM ATPyS (orange), Cdc48^AN•20S^BT1A with no nucleotide (teal), or 20S^BT1A alone plus 10 μM ATP (green).

Upon incubation of GFP-ssrA with 20S^BT1A at 45 °C, loss of GFP fluorescence was greatest when Cdc48^AN and ATP were both present, presumably due to unfolding and trapping within the 20S^BT1A chamber (Fig. 1C). With no added ATP or with ATPyS, the rate of fluorescence loss was the same as background. Because ATPyS supports gate opening (Barthelme and Sauer, 2012), this result indicates that Cdc48^AN must actively translocate GFP-ssrA into 20S^BT1A for trapping. Omission of 20S^BT1A resulted in slightly
faster loss of fluorescence than the no-ATP control (Fig. 1C), likely as a consequence of GFP-ssrA unfolding by Cdc48^AN and failure of some fraction to refold.

I conclude that the trapping approach should work for substrate identification. However, an initial trapping experiment using a clarified cell lysate from T. acidophilum was unsuccessful, I suspect in part because only a small quantity of cell lysate was available.

**Heterologous substrate identification**

As an alternative approach for identification of additional substrates, I prepared lysates grown to mid-log phase from a *E. coli* strain lacking the ClpXP and ClpAP proteases. I then incubated this lysate with different amounts of Cdc48^AN•20S for 1 h at 45 °C. Compared to control reactions containing 20S alone, I observed several prominent bands disappear in the samples containing Cdc48^AN•20S (Fig. 2A). In total, five proteins appeared to be degraded; two with molecular weights near 150 kDa, one near 90 kDa, and two near 75 kDa (Fig. 2A). Degradation was dependent on Cdc48^AN and wild-type 20S and was not observed without 20S or when β^T1A•20S was used (Fig. 2B).

I used mass-spectrometry to identify the *E. coli* proteins degraded by Cdc48^AN•20S. Bands from the no Cdc48^AN control lane in Fig. 2C (JG1, JG2, JG3, JG4, and JG5) were excised, digested with trypsin, and subjected to mass-spectrometry for sequence identification. The two largest bands were the β subunit of RNA polymerase (RNAP; 43 peptides indentified) and the β’ subunit of RNAP (25 peptides indentified), consistent with their large molecular weights (155 and 151 kDa) and relative abundance. The three
remaining proteins were chaperones. JG3 was the AAA+ remodeling chaperone ClpB (22 peptides identified), JG4 was the Hsp70 DnaK chaperone (13 peptides identified), and JG5 was the Hsp90 HtpG chaperone (13 peptides identified).

Figure 2. Degradation of bacterial proteins. (A) SDS-PAGE shows that Cdc48AN•20S but not 20S alone catalyzed degradation of specific proteins in a clarified *E. coli* lysate supplemented with ATP. The time points were 0, 20, 40, and 60 min. The gel lanes on the right show an enlarged view, with degraded bands marked JG1-5. In these experiments, the Cdc48AN concentration was 0.3 μM and the 20S concentration was 0.9 μM. (B) Degradation of high molecular weight bands in the lysate (marked by arrow) was observed with Cdc48AN•20S but not with Cdc48AN alone or with Cdc48AN•20S PT1A. Same enzyme concentrations as in panel A.

I validated Cdc48AN•20S degradation of ClpB and subunits of RNAP using purified proteins. I purified recombinant His6-SUMO-ClpB (obtained from the Baker Lab, MIT)
and removed His-SUMO fusion by cleavage with Ulp1 protease. I found that ClpB was degraded in the presence of Cdc48\(^{\text{AN}}\)•20S and ATP, but not when any of these reaction components were omitted (Fig. 3A; data not shown). All of the ClpB protein (22 \(\mu\)M) was degraded by 3 h. Accumulation of degradation intermediates was observed between 20-60 min, but these fragments were mostly degraded by 3 h. Thus, ClpB can be degraded by Cdc48\(^{\text{AN}}\)•20S both in a lysate and in a purified reaction.

RNA polymerase holoenzyme contains five core subunits (\(\beta\beta'\alpha\omega\)) and the \(\sigma^70\) specificity subunit (Vassilyev et al., 2002). I obtained an expression construct containing the gene for the \(\beta\) subunit (rpoB) with a His\(_6\)-tag on the C-terminus (Chakroborty et al., 2012), introduced it into an expression strain, but did not induce overexpression so leaky expression yielded stoichiometric RNAP complexes (\(\beta^{\text{His}}\beta'\alpha\omega\sigma^70\)). After Ni\(^{++}\)-NTA and anion-exchange chromatography, \(\beta\) and \(\beta'\) copurified with \(\alpha\) and \(\sigma^70\) (Fig. 3B, left gel). These fractions were used for degradation by Cdc48\(^{\text{AN}}\)•20S. I observed Cdc48\(^{\text{AN}}\) and ATP-dependent degradation of \(\beta\), \(\beta'\), and \(\alpha\) (top three bands, Fig. 3B, right gel). I conclude that Cdc48\(^{\text{AN}}\)•20S can target subunits of \(E.\ coli\) RNA polymerase for degradation in whole lysates and as a purified complex.
Figure 3. Degradation of purified proteins by Cdc48<sup>ΔN</sup>·20S. Reactions contained Cdc48<sup>ΔN</sup> (0.3 μM or omitted), 20S (0.9 μM), ATP (10 mM), and an ATP-regeneration system. (A) SDS-PAGE showing degradation of purified E. coli ClpB (34 μM) when Cdc48<sup>ΔN</sup> and 20S were present (right lanes) but not when Cdc48<sup>ΔN</sup> was omitted. (B) the gel on the left shows purified RNA polymerase. The gel on the right shows degradation of purified RNA polymerase by Cdc48<sup>ΔN</sup>·20S.

Discussion

Understanding the biological function of a protease requires knowledge of its substrates. The Cdc48·20S proteasome is ubiquitously distributed throughout archaea, and understanding its function may provide insight into the role of ATP-dependent protein degradation in a precursor of eukaryotes. Cdc48·20S biochemistry can be investigated using recombinantly expressed enzymes, but we currently lack tools to address its biology and substrate specificity in archaea. Here, I have taken two initial approaches in an attempt to discover substrates of Cdc48·20S other than model ssrA-tagged GFP. The first approach makes use of proteolytically inactive 20S to trap substrates for identification by mass spectrometry. Control experiments showed that Cdc48<sup>ΔN</sup>·20S<sup>BT1A</sup> can trap purified GFP-ssrA. I did not get substrate trapping to work
with archaeal lysates, but there is no obvious reason why this approach should not work if a good source of archaeal cell lysate was available.

In a second approach, I incubated an *E. coli* lysate with Cdc48\[^{\Delta N \cdot 20S}\], used SDS-PAGE to find missing bands, indentified these proteins by mass spectrometry, and validated several of these proteins as substrates in degradation assays *in vitro*. Successful application of this approach requires a separation technique with sufficient resolution, and even then only degradation of the most abundant proteins is likely to be evident. Indeed, the substrates I identified — RNA polymerase and the ClpB, DnaK, and HtpG chaperones — are all highly abundant proteins in *E. coli*.

Substrates contain sequence elements, known as degrons or degradation tags, which are recognized and engaged by the AAA+ protease. These degrons are typically short stretches of unstructured amino acids at the N or C-terminus of a protein (Sauer and Baker, 2011). Degrons can be encoded within the gene of a given protein, or they can be added during or after translation. ClpXP favors C-terminal sequences containing an Ala-Ala dipeptide, whereas ClpAPS efficiently degrades proteins with N-terminal Tyr, Phe, Trp, and Leu residues (Flynn et al., 2003; Román-Hernández et al., 2011). The wild-type C-termini of the β (Glu-Leu-Glu-Asp-Glu) and β′ (Asp-Asn-Glu) subunits of RNAP contain an intriguing number of negatively charged amino acid residues. This observation raises the possibility that Cdc48 prefers to substrates with negatively charged C-termini. In a previous study, yeast Cdc48 was shown to mediate the degradation of Rpb1, the largest subunit of eukaryotic RNA polymerase II after UV-induced damage (Verma et al., 2011), raising the possibility that degradation of RNA
polymerase is a conserved Cdc48 function. In contrast, the C-terminal Ala-Val dipeptide of ClpB is non polar. Although, ClpB in combination with DnaK helps to disaggregate and refold misfolded proteins in the cytoplasm, it seems unlikely that bound client proteins mediate Cdc48 recognition (Mogk et al., 2015), as purified ClpB was also a good substrate for Cdc48ΔN•20S degradation.

Using bacterial lysates to find substrates of archaean proteases cannot address the biological role of Cdc48, but as demonstrated here it has the potential to point us in the right direction. Future work should focus on acquiring sufficient archaean biomass for substrate-trapping experiments and on finding ways to efficiently inhibit degradation of substrates by endogenous Cdc48•20S.

Materials and Methods

Protein expression and purification

Enzymes and GFP-ssrA were purified as previously described (Barthelme and Sauer, 2012). 20SβT1A was generated by round-the-horn PCR reaction in which the pro-peptide sequence (residues 1-8) was removed. The His6-TEV tag was removed from Cdc48 and GFP-ssrA by digestion with TEV protease prior to the final gel-filtration column.

Substrate trapping and pull-down

GFP-ssrA (10-30 μM) was incubated with Cdc48ΔN•20SβT1A (1 μM) for 30 min at 45°C in PD buffer (50 mM HEPES-KOH (pH 7.5), 100 mM KCl, 20 mM MgCl2). Reactions components were pre-incubated without ATP, and reactions were started by adding 10
mM ATP. An ATP-regeneration system (16 mM creatine phosphate, 0.32 mg/mL creatine kinase) was also present. Changes in GFP-ssrA fluorescence (excitation 467 nm; emission 511 nm) in the presence of Cdc48AN, 20S, and different nucleotides were monitored in 50 μL reactions in Corning black half area plates using a plate reader.

Pull-down assays were performed using Ni^{2+}-NTA agarose beads (Qiagen) pre-equilibrated in 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 20 mM imidazole. Samples were passed over a small column, which was washed with several volumes of buffer and then eluted in buffer supplemented with 300 mM imidazole. Eluted fractions were mixed with 3X SDS loading dye, boiled for 3 min, and separated by electrophoresis on 10% SDS polyacrylamide gels using 1X MES running buffer. Gels were stained with 0.1% Coomassie-Blue (R-250) in 40% methanol and 10% acetic acid.

Lysate preparation and degradation assay

_E. coli_ strain BL21 ΔclpA ΔclpX ΔclpP was grown at 37 °C to mid-log phase (OD_{600} = 0.6), and cells were harvested, washed in 20 mM Tris-HCl (pH 7.5), and lysed by sonication. Lysed cells were centrifuged at 30,000 x g and then at 100,000 x g in a table top ultracentrifuge for 1 h. Clarified lysates were frozen and stored at −80 °C. Each lysate sample was tested for optimum loading by SDS-PAGE to maximize band clarity and intensity. Degradation assays were carried out at 45 °C with 10 mM ATP for 1 h. Samples were taken every 20 min and quenched by mixing with 3X SDS loading buffer. Separation and staining was performed as described above. For band identification, small slices of the gel were excised, destained, reduced with dithiothreitol, alkylated with iodoacetamide, and digested with trypsin following standard protocols. LC-MS analyses
of the tryptic peptides was performed with the LTQ mass spectrometer at the Koch Institute Proteomics Core facility. Proteins in the sample were identified by using the Mascot database search software.

A plasmid encoding His-SUMO-ClpB was obtained from the Baker Lab (MIT). His-SUMO-ClpB was expressed in *E. coli* strain BL21 and purified by Ni^{++}-NTA affinity. The His-SUMO moiety was cleaved using Ulp1 protease, and ClpB was purified by passage through Ni^{++}-NTA and by gel filtration on a Superdex S200 column. A plasmid encoding the β-His_{6} of *E. coli* RNAP under control of an IPTG-inducible promoter was obtained from Richard Ebright (Rutgers) and transformed into *E. coli* strain BL21 DE3. Expression was performed without IPTG induction, and His-tagged RNAP was purified by Ni^{++}-NTA affinity and MonoQ column as described (Chakroborty et al., 2012). Degradation assays at 45 °C were performed in PD buffer with Cdc48^{AN-20S} (0.3 μM, 0.9 μM) and assayed by SDS-PAGE.
References


Chakraborty, A., Wang, D., Ebright, Y.W., Kornlann, Y., Kortkhonjia, E., Kim, T.,


APPENDIX A

Identification of 24-n-propylidenecholesterol in a member of the Foraminiferida

Abstract

We report a study of the sterols produced by the thecate allogromid foraminiferan, *Allogromia laticollaris*, grown in culture. Gas chromatographic retention time, together with mass-spectrometric fragmentation patterns of trimethylsilyl and acetate derivatives enabled us to identify a suite of C$_{27}$-C$_{30}$ sterols. Two C$_{30}$ sterols were identified as (24E)-n-propyldenecholesterol and its (24Z)-isomer by direct comparisons with an authentic standard. The C$_{30}$ sterols were undetectable in the two algae, *Isochrysis galbana* and *Dunaliella tertiolecta*, used as food sources for the *Allogromia laticollaris* consistent with previous analyses of these organisms. The identity of (24E)-n-propyldenecholesterol was confirmed by 600 MHz $^1$H-NMR.

Introduction

Sterols are a class of lipids that is essential to the functioning of eukaryotic cell membranes and hallmarks of the difference between eukaryotic and prokaryotic life (Benveniste, 2004; Bloch, 1979). The diversification of sterol biosynthetic networks and resultant structural features of sterols are differentiated along phylogenetic boundaries making them interesting targets for studying biochemical changes among Eukarya and as taxonomic proxies in modern and ancient environments (Brocks et al., 1999; Desmond and Gribaldo, 2009; Waldbauer et al., 2009). Membrane sterols range in carbon number from 26 to 31, with the most common being C$_{27}$, C$_{28}$ and C$_{29}$.

Sterols from the contemporary oceans are useful chemotaxonomic markers for marine invertebrates (Kerr and Baker, 1991; Schmitz, 1978). Steranes, fossilized sterols, can
be used to constrain evolutionary appearances, reconstruct relative abundances of algae in the past (Brocks et al., 1999; Grantham and Wakefield, 1988; Knoll et al., 2007) and are important proxies for paleoreconstruction.

All sterols are synthesized from the 30-carbon isoprenoid precursor 2,3-oxidosqualene by action of oxidosqualene cyclases (OSC). The epoxidation of squalene and the appearance of the two sterologenic cyclases, cycloartenol synthase and lanosterol synthase, must have post-dated the advent of oxygenic photosynthesis on Earth (Bloch, 1987; Waldbauer et al., 2011). Variations in sterol structures that are associated with phylogenic placement include methylation at C-4 and alkylation in the side-chain which is primarily observed at the C-24 position but also at C-22, C-23, C-26, and C-27. Demethylations are observed at C-21, C-24 or C-27. The number and location of unsaturations also correlate somewhat with phylogeny. C_{27} sterols are found primarily in the metazoa but may occur in varying amounts in the protists either as biosynthetic intermediates or end products. C_{28} sterols are present in all major clades with exception of eumetazoa. C_{29} sterols are typically found in photosynthetic organisms including diatoms, green algae and numerous other archeoplastids as well as porifera (Volkman, 2003). Sterols containing 30 carbon atoms are exceedingly rare and have so far been primarily reported in marine sponges, pelagophyte algae and dinoflagellates, the latter clade producing the eponymous dinosterol (Giner and Wikfors, 2011; Piretti et al., 1997; Volkman, 2003; Withers et al., 1982). Because of their rarity, C_{30} sterols are useful biomarkers of organisms from both present and past environments (Summons et al., 2006). C_{30} steranes with a n-propyl substituent at C24, for example, derive from marine
pelagophyte algae (Giner et al., 2009) and are commonly used to discern organic matter of marine origin and to distinguish marine sediments and oils from non-marine counterparts (Moldowan, 1984). C_{30} steranes with an isopropyl substituent at C24, in contrast, are found in some demosponges (Bergquist et al., 1991), although they also occur in some orchids (Kikuchi 1982 Chem Charm Bull page 370), and appear in anomalous abundances in sediments and oils of late Neoproterozoic to Cambrian age (Love et al., 2009; McCaffrey et al., 1994) consistent with sponge radiation at that time (Sperling et al., 2010). It was in this context that we have been undertaking studies of the sterol complements of other classes of marine invertebrates and some protists since molecular clock data and timing of their appearances in the record of physical fossils points to a potential for finding the corresponding chemical fossils (Kodner et al., 2008b; Peterson and Butterfield, 2005; Peterson et al., 2008; Sperling et al., 2010). One group of protists that has received little attention in this regard is the foraminifera. They are of particular interest to us because they have a long fossil record which was recently elaborated to include sediments deposited during the Cyrogenian Period (Bosak et al., 2011; Bosak et al., 2012). In addition, foraminifera are an important member of the eukaryotic communities present in extant microbialites (Bernhard et al., 2013; Edgcomb et al., 2013). We wondered if thecate foraminifera, which are basal forms (Pawlowski et al., 2003), possess unique or unusual sterols.

In this study, gas chromatographic retention time was paired with mass-spectrometric fragmentation patterns of trimethylsilyl and acetate derivatives of an authentic standard to establish the structure of a C_{30} sterol isolated from a cultured foraminiferan,
*Allogromia laticollaris.* Here we report that structure as (24E)-n-propylidenecholesterol and trace amounts of the (24Z)-isomer.

**Methods**

**Cultures**

After unsuccessful experiments with a range of bacterial food sources, *Allogromia laticollaris* cultures, which were originally obtained from J. L. Travis (University of Albany), were ultimately raised on a mixed microalgal culture of *Isochrysis galbana* and *Dunaliella tertiolecta* which, in turn, were grown on Alga-Gro\textsuperscript{®} medium (Carolina Biological) maintained at \(-23^\circ\text{C}\) and in a 12 h light-dark cycle. Foraminifera were grown in 16X150 mm test tubes half filled with algal culture and diluted with 30 ppt artificial seawater (Instant Ocean\textsuperscript{®}) with strong lighting. On two occasions, foraminifera were cultured several times over six months. In the first experiment, *A. laticollaris* cells were picked out from their pre-filtered (125 µm and 63 µm filters) culture media using pre-combusted Pasteur pipettes. The collection beakers were left to rest, allowing most of the remaining algae to be removed using a pipette. About 50 culture tubes were harvested. From this we obtained samples of *Allogromia laticollaris* and culture supernatant including the algae, each of which were extracted as described below.

In a second experiment, the culture was scaled up to obtain approximately 100 tubes containing 20-40 foraminifera per tube. After approximately six months, the *A. laticollaris* were removed from the algae cultures and transferred to new culture tubes containing the cyanobacterium *Spirulina* and artificial seawater. We removed as much algal biomass as possible before transferring, but trace carry over was unavoidable or
present in *A. laticollaris* cells as ingested food. The *A. laticollaris* were grown for approximately three weeks with cyanobacteria with the aim of reducing the quantity of algal background in the lipid extracts. After three weeks, the *A. laticollaris* were picked with pasture pipettes and centrifuged to remove most of the remaining water. This cleaned sample was frozen at -20°C until freeze-drying and extraction. Samples of *Isochrysis galbana* and *Dunaliella tertiolecta* cultures were also reserved for lipid comparison with the *A. laticollaris* cultures.

**DNA Extraction, PCR, Sequencing**

Total nucleic acids were extracted from laboratory maintained algal cultures of *Isochrysis galbana* and *Dunaliella tertiolecta*, as well as from a pool of approximately 30 hand picked and rinsed *Allogromia laticollaris* foraminifera specimens using the DNeasy Plant Mini Kit (Qiagen). *Dunaliella* SSU rRNA genes were amplified using 2 primer combinations, 360f (5’-CGGAGARGGMGCMTGAGA-3’) with either 1492r (5’-GGTTACCTTGTTACGACTT-3’) or U1391r (5’-GGGCGGTGTGTACAARGR-3’), while only 360f/1492r were used for *Isochrysis*. Algal PCR amplification parameters were: 95°C for 5 minutes followed by 35-40 cycles of 95°C for 60 seconds, 50-55°C for 60 seconds, and 72°C for 90 seconds with a final extension of 72°C for 7 minutes. Because general eukaryotic SSU rRNA PCR primers miss most foraminifera, SSU rRNA genes were amplified from foraminifera extracts with both foraminifera-targeting primers (s14F3a, s14f1, and sB; (Pawlowski, 2000) as well as 360f/1492r. A Nested PCR strategy was employed to amplify foraminiferal (*A. laticollaris*) SSU rDNA genes. Primary PCR conditions using s14F3a/sB were: 5 minutes at 95°C followed by 30
cycles of 95°C for 60 seconds, 45°C for 60 seconds, and 72°C for 90 seconds with a final extension of 72°C for 7 minutes. One microliter of the primary PCR product was then subjected to a second amplification using the nested primer s14f1 with sB as described above. General eukaryotic primers (360f/1492r) were also used to test for the presence of other eukaryotes that might be present in the *A. laticollaris* cultures. The parameters were: 95°C for 5 minutes followed by 35 cycles of 95°C for 60 seconds, 55°C for 1 minute, and 72°C for 90 seconds with a final extension of 72°C for 7 minutes. PCR products were visualized on a 1% SYBR Safe (Invitrogen) stained agarose gel and purified from the gel with the Zymoclean Gel DNA Recovery Kit (Zymo Research). Gel purified PCR products were cloned into pCR4 vector with the TOPO-TA Sequencing Kit (Invitrogen). Plasmids were sequenced using the universal M13F primer at the W. M. Keck Ecological and Evolutionary Genetics Facility (MBL). A total of 75 clones (19 from *Dunaliella*, 16 from *Isochrysis*, and approximately 40 from the *A. laticollaris* cultures) were sequenced in a single direction using the M13F primer.

*Lipid extractions*

Harvested cell material was freeze-dried for 48 hours. The dried biomass was extracted by a modified Bligh-Dyer procedure (Iverson et al., 2001). Briefly, biomass was sonicated in c. 10 ml MeOH:DCM:water (10:5:4, v/v/v) solution followed by 10 ml MeOH:DCM:trichloroacetic acid (10:5:4, v/v/v) solution. The mixture was phase separated three times by addition of 20 mL 1:1 (v/v) DCM: water. The organic phase was collected and dried under nitrogen yielding a total lipid extract (TLE). No further preparation was performed and TLE was used for GC-MS analysis.
Sterol analysis

All sterols were derivatized to form either trimethylsilyl (TMS) ethers or acetate esters. The TLE was dissolved in an appropriate volume of DCM and transferred to a GC vial containing a pointed insert. This sample was dried under nitrogen and then reacted with equal volumes of pyridine and BSTFA (Sigma Aldrich) in the case of TMS derivatives or acetic anhydride (Sigma Aldrich) to form the acetates. This was usually 25 µL of each reagent and solvent but, at times, this volume was varied to compensate for the concentration of the TLE. Samples were then capped and incubated at 70°C for 1 hour. All GC-MS analyses were carried out on an Agilent GC-MSD 5975C equipped with a programmable temperature vaporizing injector PTV. Initial experiments were carried out on an Agilent J&W 60-meter DB-5 column, inner diameter 0.25 mm and 250 µm film thickness. The initial temperature was 60°C isothermal for 2 min followed by an oven ramp at 10°C/min to 150°C, zero hold, followed by another ramp of 3.5°C/min to 325°C with a 20 min hold and a total run time of 81 min. Retention time and co-elution experiments were repeated with an Agilent 30-meter DB-XLB column with a 0.25-mm inner diameter and a 250-µm film thickness.

¹H-NMR spectra were obtained using a Bruker Avance III 600 MHz instrument equipped with a 5 mm triple resonance (TXI) probe. The spectra were reference to residual solvent (CDCl₃, 7.26 ppm).

Results

Two C₃₀ sterols in *A. laticollaris* extracts

GC-MS data for *A. laticollaris* TLEs showed one C₂₇ and several C₂₈ and C₂₉ sterols including cholesterol, brassicasterol and stigmasterol (Figure 1a), identified on the basis
of their mass spectra and comparisons with literature data. As previously reported
(Volkman, 2003), these compounds were also present in extracts from the algae used
for the co-culture with A. laticollaris foraminifera ultimately making it difficult to
confidently discern the sources of the compounds. However, two isomeric sterol
derivatives with a molecular ion at 408 Da, corresponding to doubly unsaturated C$_{30}$
erterol acetates, were present in the A. laticollaris TLEs and not in the algal extracts
(Figure 1). The more abundant of these two sterols (denoted 1A) eluted approximately
0.5 min after sitosterol. The second C$_{30}$ sterol (2A) had a retention time approximately
0.2 min longer and the spectra of the two compounds were essentially indistinguishable.
When derivatized as TMS-ethers, these sterols showed a molecular ion of m/z 498
(Table 1) with major fragment ions at 129, 296, 386 and 483 Da diagnostic for $\Delta^5$-TMS
sterol, $\Delta^5$ – [sc + TMSO] and $\Delta^5$ – [sc] respectively (Brooks et al., 1968). The 296 and
386 fragments are consistent with a sterol side-chain bearing an ethyl or propyl group at
the C-24 position and a double bond between C-24 and C-28 (Brooks et al., 1972; Idler
et al., 1976). Sterols containing a double bond at position-24 have characteristic mass
spectral fragments corresponding to loss of the side-chain and H after cleavage of the
C22-23 bond with (296 Da) and without (386 Da) loss of TMSOH. Thus, the mass
spectral data support a structure of 24-n-propylidenecholesterol. Published mass
spectra of this compound from Pelagophyte algae and a scallop, Placopecten
magellanicus, are consistent with those reported here (Idler et al., 1971).

Identification of 24-n-propylidenecholesterol by GC retention time analysis and co-
injection
Mass spectrometry alone is not sufficient to exclude the various alternative structures for the unknown sterols from the foraminiferan in this study. This is demonstrated by compounds A1 and A2, which have nearly identical spectra but different retention times. To exclude alternative structures we obtained an authentic standard containing (24E) and (24Z)-isomers of 24-propylidenecholesterol and 24-methylenecholesterol (3:1:1) for comparison with our unknowns (Giner et al., 2009). This standard, which had a single $^{13}$C substitution at C-22, and A. laticollaris TLE were analyzed in a series of GC-MS injections to compare relative retention times (e.g., Figure 3). Approximately 10 ng of standard injected on-column produced two well-resolved C$_{30}$ sterols with retention times of 66.20 and 66.92 minutes. Under our standard chromatographic conditions (24E) eluted prior to (24Z) so we concluded that (24E), with a retention time of 66.20 minutes, was the most abundant unknown C$_{30}$ sterol in (Figures 1, 1A). The molecular and M-15 ions of the standard and the major fragment ions of 297 and 387 Da were shifted by one dalton from the A. laticollaris sterols, consistent with the presence of a $^{13}$C atom at the C-22 position.

We next used a number of co-injection experiments to ask if the A. laticollaris sterols could be chromatographically resolved from the 24-n-propylidenecholesterol standards. First, a five-times diluted standard sample was analyzed, followed by an aliquot of A. laticollaris TLE mixed with the standard (Figures 4a, 4d). We discerned no difference in the chromatograms of standard alone and co-injection samples. The peak corresponding to the (24E) sterol remained a single peak with no broadening or shape change after addition of A. laticollaris TLE.
Comparison of select ion chromatograms could resolve a doublet if there is a shift in the distribution of diagnostic ions. Spectra of the TMS derivatives of *A. laticollaris* C₃₀ sterols have major ions at 296 and 386 Da (Table 1) which can be used as tracers to distinguish the natural product from the $^{13}$C-labeled standard which has major ions at 297 and 387 Da. This is useful in interpreting the co-injection chromatogram. In both cases the 296/297 and 386/387 peaks overlapped identically (Figure 3a-d) and confirm that the standards and unknowns share diagnostic mass spectral features and have identical behavior in gas chromatography.

The above retention time experiment was repeated with the more polar DB-XLB column and the results were consistent: the *A. laticollaris* C₃₀ sterols co-eluted with (24E) and (24Z)-n-propyldenecholesterol (data not shown) as they did on the non-polar phase. Subsequently, both chromatographic comparisons were repeated with acetate derivatives and, again, we found co-elution of the two *A. laticollaris* sterols with the two standards.

*Confirmation of 24-n-propyldenecholesterol by $^1$H-NMR*

The identity of the sterol was confirmed by 600 MHz $^1$H-NMR through detection of the H-28 signal of the 24-n-propyldene group in the olefinic region of the spectrum of the mixed sterol fraction from the foraminiferan. Thus, a triplet at 5.067 ppm ($J = 7.0$ Hz) was detected that exactly matched the corresponding signal for an authentic sample of (24E)-24-n-propyldenecholesterol (1A).
DNA sequencing of *A. laticollaris* and algal cultures

The 75 sequences from the general eukarya- and the foraminifera-targeted PCR amplifications were compared to SSU rRNA genes in GenBank by BLASTn. As expected, all of the foraminifera-targeted clone sequences showed significant BLAST alignments to allogromiid foraminifera with *Allogromia laticollaris* partial 18S rRNA gene sequence (GenBank accession # AJ311218) receiving the top score. No other sequences were retrieved as top BLAST scores. PCR amplifications using general eukaryotic primers (which miss most foraminifera) resulted in the majority of clones (35) showing significant BLAST scores to *Nannochloropsis salina* or *Nannochloropsis gaditana* (accession numbers M87328.1 and EF473733.1, respectively). Other GenBank sequences with significant similarity to our eukarya clone library sequences include two fungi: *Cladosporium* sp. (3 clones, EU167574) and *Davidiellacea* sp. (2 clones, GU250935) as well as the green alga *Dunaliella* sp. (1 clone, FJ164062).

Sixteen of the *Isochrysis* culture clones were sequenced as above and submitted to BLASTn. All of the recovered clone sequences (100%) showed significant identity to *Pseudoisochrysis paradoxa* (AM490999.2) and *Isochrysis galbana* (HM149541.1) partial SSU sequences in Genbank. No other SSU sequences were amplified from the *Isochrysis* culture with the 360f/U1392r primer pair.

Ten clones from each of the *Dunaliella* culture clone libraries were sequenced as described above. All (100%) of the sequences are most similar to a *Dunaliella sp.* partial SSU sequence found in GenBank (JX839831.1) regardless of the primer pair used.
(360f/1492r or 360f/U1391r) to amplify it. No other SSU sequences were amplified from the *Dunaliella* culture.

**Discussion**

24-Propylidenecholesterols were first identified in the scallop *Placopecten magellanicus* and subsequently identified in all Pelagophyte algae, some of which are responsible for harmful brown tides (Giner et al., 2009; Idler et al., 1971). The presence of either the 24E or 24Z sterol has been used as a biomarker to distinguish between two major clades within this class of algae. The Pelagomonodales are reported to contain exclusively the 24Z isoform while the Sarcinochysidales are found to contain only (24E)-propylidenecholesterol (Giner et al., 2009). Besides the pelagophytes, no other organisms have been shown to produce 24-propylidenecholesterol. Here we show that laboratory cultures of the foraminifera *Allogromia laticollaris* contain a mixture of (24E) and (24Z)-n-propylidenecholesterol. This result was reproduced with a second independent experiment to verify the original findings. Sterols were quantified by GC-MS and those from the first co-culture experiment and had an *E*/Z ratio approximately 9:1 while in the second experiment we detected almost exclusively 24E which may be a result of differing culture conditions, extraction procedure or an unknown variable. Previously, it had been shown that the *E*/Z ratio changes depending on the method of analysis (Giner et al. 2009), and it is hypothesized that the two isomers differ in stability to air and heat.

The sterol composition of any organism may originate from one of four sources: *de novo* biosynthesis via squalene epoxide, dietary origin without modification, dietary origin with
modification or from symbionts (Djerassi, 1981). It is now believed that the source of 24-\(n\)-propylidenecholesterol in marine invertebrates such as the scallop is a result of dietary input because mollusks are incapable of \textit{de novo} sterol biosynthesis (Kanazawa, 2001; Napolitano et al., 1993). To rule out the possibility of sterols arising from a contaminant instead of the \textit{A. laticollaris} and their eukaryotic co-culture, we generated clone libraries of SSU rRNA genes, sequenced them, and compared those sequences against GenBank reference sequences using BLASTn. We found no contaminating Pelagophyte algae in the allogromiid, the \textit{Isochrysis} or the \textit{Dunaliella} clone libraries. Using foraminifera-directed primers we generated forty clones and BLASTn all were identified as having come from \textit{A. laticollaris}. When general eukaryotic primers were used for PCR amplification of rRNA SSU genes derived from the allogromiid, and the resulting sequences were submitted for BLASTn analysis, thirty-five clones had identity to \textit{Nannochloropsis salina} or \textit{Nannochloropsis gaditana}. These heterokont algae do not contain \(C_{30}\) sterols and therefore we conclude they cannot be a contaminating source of \(C_{30}\) sterols in our analysis (Mohammady, 2004). The same analysis of \textit{Isochrysis} and \textit{Dunaliella} cultures returned BLASTn hits only from their respective algal constituents. Despite these tests, the possibility remains that a cryptic pelagophyte is present in the foraminiferan culture.

The fact that 24-\(n\)-propylidenecholesterol is present within a foraminiferan poses questions about the utility of these steroids as biomarkers for pelagophyte algae both in modern environments and in the fossil record. It would be of interest to determine if this is the result of lateral gene transfer from a pelagophyte. Foraminifera are a deeply
branching group of successful eukaryotic amoeboid protists that occupy most ecological niches in the marine realm (Goldstein, 1999). We do not presently know if A. laticollaris has the genes necessary for de novo synthesis. Unfortunately, full genomes of foraminifera are not available to our knowledge, and given their diversity and ecological variability, uniformity in putative sterol biosynthesis or modification genes among foraminiferal species is not necessarily expected. Identification of a sterol within one member of this group cannot support a conclusion that 24-n-propyldenecholesterol is characteristic of all foraminifera. Nor is it clear that the biomass of foraminiferans in any sample would be sufficient to leave a biomarker signal. In fact, as part of our extended project, we have analyzed a small selection of microbialites (stromatolites and thrombolites) known to contain a diverse foraminiferal assemblages, including thecate forms (Bernhard et al., 2013; Edgcomb et al., 2013) but have failed, so far, to detect 24-n-propyldenecholesterol or its derivatives. Nevertheless, n-propycholesterols and their fossil derivatives may be potentially useful for examining other microbialites since allogromiid foraminifera are thought to be significant bioturbators and there may be a connection between disruption of microbialite fabrics, the Neoproterozoic decline in stromatolites and the presence of these taxa in the fossil record.

The chemistry of sterol side chain biosynthesis is relatively well understood but we know little of the molecular biological and physiological drivers that lead to diverse sterol structures (Djerassi, 1981; Giner, 1993; Giner and Wikfors, 2011). It has been proposed that unusual algal sterols including 24-propyldenecholesterol interfere with the sterol requirements of the grazers that consume these algae (Giner et al. Phytochemistry
Sterol side chains are modified by successive methylation by S-adenosylmethionine methyltransferases (SMTs) to form $C_{28}$, $C_{29}$ and $C_{30}$ sterols. However, the identity of these SMTs and any accessory proteins presently remains unclear, with the exception of a few organisms related to human disease such as yeast and the parasitic protist *Trypanosoma brucei* (Ganapathy et al., 2011; Liu et al., 2011), and agriculture (corn, cotton, soy, wheat, grapes, sorghum, barley, oats, tobacco, poplar, etc.), several marine algae, fungi, sponges, and protists. Because much of the structural diversity in sterols is introduced by methylation reactions there must be an equivalent genetic diversity. A clear next step in chemotaxonomic studies of marine sterols would be to identify and characterize the genes encoding the enzymes involved in those sterol side-chain methylation reactions that yield uncommon sterols. Early work on this topic has begun but a more thorough treatment awaits (Kodner et al., 2008a). A genomic approach, in combination with over sixty years of studying the structure and distribution of sterols, will open a new era of investigation in the field.
References


Edgcomb, V.P., Bernhard, J.M., Summons, R.E., Orsi, W., Beaudoin, D., Visscher, P.T., 2013. Active eukaryotes in microbialites from Highborne Cay, Bahamas and Hamelin Pool (Shark Bay), Australia. ISME J, Submitted.


Legends to Figures and Tables

Figure 1. Identification of a C₃₀ sterol in *A. laticollaris* extracts. Gas chromatograms of acetylated from a) *A. laticollaris* extracts and b) extracts of the culture medium containing *I. galbana* and *D. tertiolecta*. Dotted lines show the peaks from the C₃₀ sterols 1A and 2A and the corresponding position in the algal chromatogram where they are absent. The inset contains the structures of candidate structures: (24E) and (24Z)-n-propylidenecholesterol.

Figure 2. Comparison of *A. laticollaris* sterols and 24-n-propylidenecholesterol. Superimposed chromatograms of sterols from trimethylsilyl derivatives of 24-propylidenecholesterol (solid line) standard and *A. laticollaris* extracts (dotted line).

Figure 3. Retention time and co-injection analysis of *A. laticollaris* sterols with authentic 24-n-propylidenecholesterol. Standard mixture containing (24E) and (24Z)-n-propylidenecholesterol derivatized as TMS ethers TIC (a) and extracted ion chromatograms for 296 and 297 Da (b). Co-injection of *A. laticollaris* TLE and ¹³C-labelled standard TIC (c) and extracted ion chromatograms for 296 and 297 Da (d). Note the enrichment of the unlabelled products in the 296 Da trace of (d) compared to the 296 trace of (c).

Table 1. Mass spectrometric fragmentation patterns of C₃₀ sterols from *A. laticollaris* and standard E and Z 24-n-propylidene sterols analyzed as acetates (A) or TMS ethers (T).
Figure 1.
Figure 2.
Figure 3.
<table>
<thead>
<tr>
<th>Compound</th>
<th>Molecular Ion</th>
<th>Base peak</th>
<th>Principal ions in mass spectrometry (% base peak)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1T</td>
<td>498</td>
<td>69</td>
<td>69 (100), 69 (53), 73 (80), 75 (37), 81 (37), 388 (78), 95 (33), 119 (33), 296 (31), 105 (28), 107 (27), 91 (23), 257 (15), 488 (2)</td>
</tr>
<tr>
<td>2T</td>
<td>498</td>
<td>69</td>
<td>69 (100), 73 (85), 129 (77), 75 (70), 388 (46), 296 (36), 81 (38), 95 (35), 119 (35), 105 (28), 107 (27), 91 (25), 257 (18), 488 (1)</td>
</tr>
<tr>
<td>1A</td>
<td>408</td>
<td>296</td>
<td>296 (100), 69 (44), 408 (26), 51 (28), 281 (27), 81 (23), 297 (22), 145 (21)</td>
</tr>
<tr>
<td>2A</td>
<td>408</td>
<td>296</td>
<td>296 (100), 297 (27), 69 (24), 55 (20), 145 (19), 105 (15), 81 (14), 408 (11)</td>
</tr>
<tr>
<td>(2,4,7)-</td>
<td>499</td>
<td>129</td>
<td>129 (100), 387 (62), 73 (89), 69 (85), 297 (64), 55 (68), 75 (49), 119 (38), 81 (34), 282 (33), 95 (32), 105 (32), 107 (31), 388 (28), 93 (27), 258 (27), 70 (26), 81 (26), 145 (26), 159 (25), 499 (6)</td>
</tr>
<tr>
<td>p-cholest eno ITMS</td>
<td>499</td>
<td>73</td>
<td>73 (100), 129 (33), 387 (69), 69 (87), 297 (71), 55 (81), 75 (50), 282 (43), 119 (39), 81 (34), 95 (33), 105 (33), 258 (32), 388 (31), 107 (30), 70 (27), 91 (27), 93 (27), 145 (26), 159 (25), 499 (6)</td>
</tr>
</tbody>
</table>

Table 1.