J Mol Evol (1997) 45:549–563

JOURNAL OF MOLECULAR EVOLUTION

The Evolution of the Conserved ATPase Domain (CAD): Reconstructing the History of an Ancient Protein Module

Jonathan C. Swaffield, Michael D. Purugganan

Department of Genetics, North Carolina State University, P.O. Box 7614, Raleigh, NC 27695-7614, USA

Received: 28 January 1997 / Accepted: 8 May 1997

Abstract. The AAA proteins (*A*TPases *A*ssociated with a variety of cellular *A*ctivities) are found in eubacterial, archaebacterial, and eukaryotic species and participate in a large number of cellular processes, including protein degradation, vesicle fusion, cell cycle control, and cellular secretory processes. The AAA proteins are characterized by the presence of a 230 to 250-amino acid ATPase domain referred to as the *C*onserved *A*TPase *D*omain or CAD. Phylogenetic analysis of 133 CAD sequences from 38 species reveal that AAA CADs are organized into discrete groups that are related not only in structure but in cellular function. Evolutionary analyses also indicate that the CAD was present in the last common ancestor of eubacteria, archaebacteria, and eukaryotes. The eubacterial CADs are found in metalloproteases, while CAD-containing proteins in the archaebacterial and eukaryotic lineages appear to have diversified by a series of gene duplication events that lead to the establishment of different functional AAA proteins, including proteasomal regulatory, NSF/Sec, and Pas proteins. The phylogeny of the CADs provides the basis for establishing the patterns of evolutionary change that characterize the AAA proteins.

Key words: AAA proteins — ATPase — Proteasome — NSF — Sec — Pas — Metalloprotease

Correspondence to: J.C. Swaffield

Introduction

The evolution of multigene families continues to remain a major focus of molecular evolutionary research. Multigene families control related cellular and developmental functions, and the expansion of these families yields valuable insights into the molecular mechanisms of gene duplication and the genetic basis of functional diversification (Purugganan et al. 1995; Atchley and Fitch 1995; Ohno 1970). The divergence in function observed in members of a gene family also provides the molecular basis for studies on the evolution of developmental or metabolic pathways (Purugganan et al. 1995; Atchley et al. 1994).

The AAA family, named for *A*TPases *A*ssociated with a variety of cellular *A*ctivities (Kunau et al. 1993), is a large family of proteins that has recently been characterized in archaebacterial, eubacterial, and eukaryotic cellular systems. These proteins are involved in a wide range of cellular processes, including proteolytic degradation, vesicle fusion, and protein import. Members of the AAA protein family are characterized by the presence of the *C*onserved *A*TPase *D*omain or CAD (Swaffield et al. 1995), a large motif, 230–250 amino acids in length (Fig. 1). The ATPase motif contained within the CAD is referred to as a Walker type motif (Walker et al. 1982), and was first described after a comparison of a number of different ATPase sequences. Upon X-ray crystallographic determination of ATPase structures (Fry et al. 1986), the Walker motif was shown to consist of three structural components: (1) a hydrophobic strand of parallel β-pleated sheet ending in an aspartate, (2) an α-helix of two lysines separated by three residues, two of

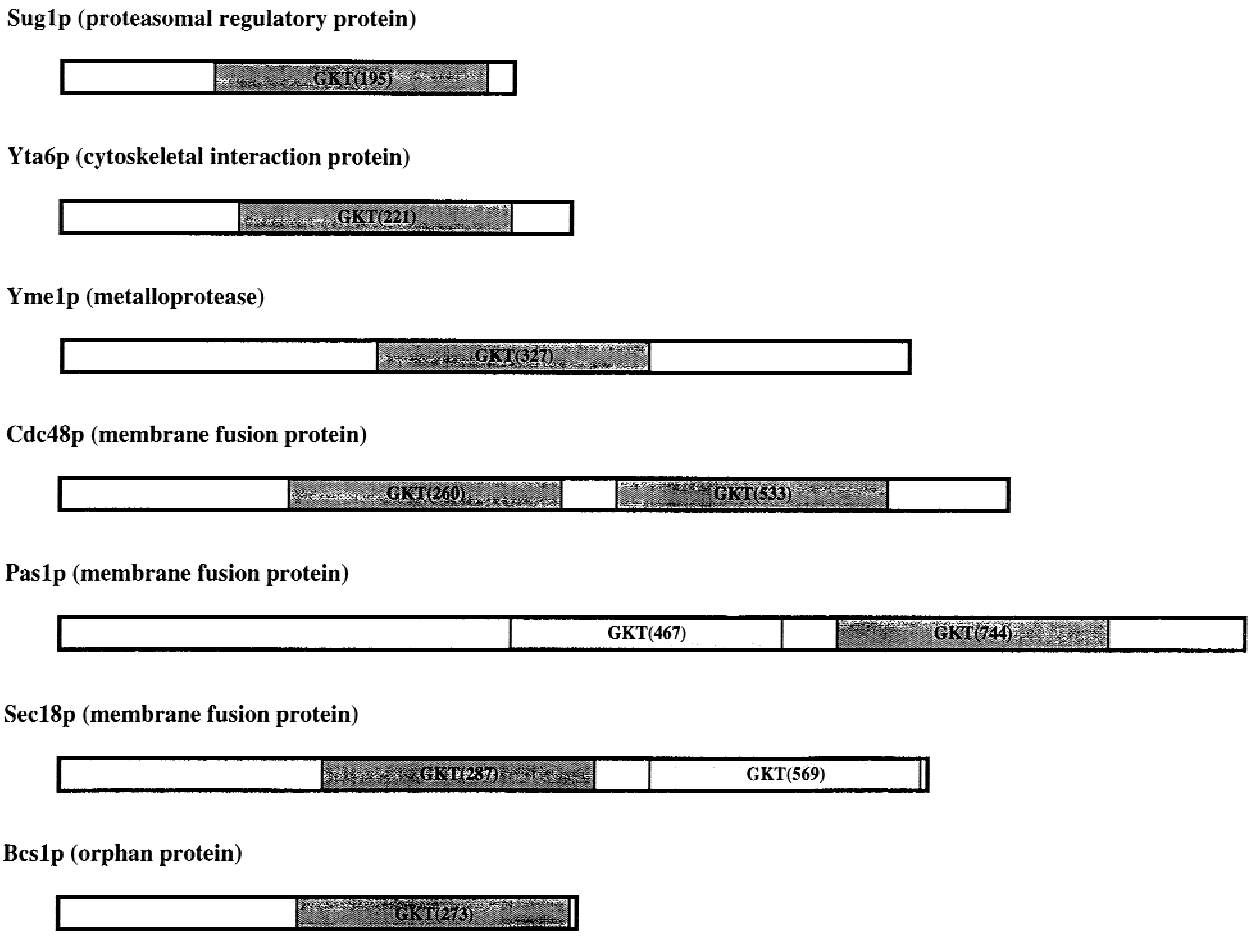

Fig. 1. Structures of typical yeast AAA proteins showing location of *C*onserved *A*TPase *D*omains (CADs). The CADs are indicated by the shaded boxes. The positions of the conserved GKT motif found in Walker type ATPases are indicated by the numbers in parentheses. In Pas1p, Sec18p, and related proteins, two CADs are present. The lightly shaded box depicts the CAD duplicate domain within these proteins that displays significant sequence divergence from other CAD sequences.

which are hydrophobic, and (3) a glycine-rich flexible ''P-loop'' containing the highly conserved ''GKT'' sequence. The lysine residue in this motif is the only amino acid within the loop that is thought to directly contact bound Mg.ATP. Mutation of this residue generally abolishes the ATPase activity and the functionality of the protein (Liu and Summers 1988; Weiss et al. 1991; Carrera et al. 1993).

AAA proteins were first isolated around 1990 (Peters et al. 1990; Frohlich et al. 1991) with the identification of several eukaryotic proteins that contained what was later termed the CAD. To date, more than 100 CAD-containing AAA proteins have been isolated, with examples from each biological urkingdom—eukaryotes, eubacteria and archaebacteria (Woese, Kandler, and Wheelis 1990). The only completely sequenced eukaryotic genome, the budding yeast *Saccharomyces cerevisiae* genome, contains genes that encode 20 clearly recognizable AAA proteins. The yeast genome also includes two less well-conserved members of this family (YB36 and BCS1), as well as 5–10 genes that have very limited similarity to other AAA proteins.

The high degree of sequence conservation seen throughout the motif between examples in each major biological group suggests that the CAD plays a critical role in cellular metabolism. The exact function(s) performed by the CAD remains a matter of debate. Precise determination of the biological role(s) for the CAD has been complicated by the extremely wide range of cellular functions AAA proteins perform. Although the domain contains a Walker Type I ATPase motif that is essential for protein function, the CAD is clearly different from other ATPase protein families such as DEAD-box helicases (Hodgman 1992), ion pumps, and ABC transporters (Fath and Kolter 1993). Members of the CAD-containing AAA family have been shown to participate in vesicle fusion (Sollner et al. 1993), peroxisomal protein import (Heyman et al. 1994), mitochondrial membrane-bound proteolysis (Guelin et al. 1994; Tomoyasu et al. 1995), cell cycle control (Frohlich et al. 1991), proteasomal regulation (DeMartino et al. 1994; Dubiel et al. 1995), mitotic spindle formation (Clark-Maquire and Mains 1994), signal transduction (Schulte et al. 1994), protein complex formation (Nobrega et al. 1992; Paul

and Tzagoloff 1995) and possibly transcription (Swaffield et al. 1995; Lee et al. 1995; Xu et al. 1995). Considering different facets of each known biological activity attributed to AAA proteins, Confalonieri and Duguet (1995) have proposed two unifying functions for CADs: (1) an ATP-dependent proteasomal or chaperone function and (2) an ability to act as an ATP-depending anchorage, or ''protein clamp.''

The wide range of roles performed by CAD-containing AAA proteins permits us to explore the evolutionary diversification of function in a large gene family across all major organismal lineages. We have undertaken a phylogenetic analysis of the evolutionary relationships within the AAA protein family, and dissected the patterns of sequence change associated with the expansion of these proteins during the evolution of eubacteria, archaebacteria, and eukaryotes. This analysis also provides us with a sound phylogenetic framework to classify the diverse array of AAA proteins found in different species.

Materials and Methods

All gene sequences used in these analyses were obtained from the GenBank and EMBL protein data bases (Table 1). These sequences were initially identified with BLAST searches using the *S. cerevisiae Sug1p* and *Sug2p* CADs as the query sequences. The data base searches identified 105 proteins previously identified by others as containing CAD sequences, and these proteins were retained in our data set.

Several other divergent proteins were identified in the BLAST searches, but the presumptive CAD sequences in these proteins possessed less than 20% amino acid similiarity to the reference query sequences. Previous analyses did not identify these distantly related proteins as CAD-containing AAA proteins. The limited similarity of these divergent sequences to the reference yeast CAD proteins (Fig. 1) made it difficult to confidently assign amino acid site homologies, and these proteins were excluded from the data set. There are, however, divergent CAD sequences within the data set that are identified as duplicate domains within proteins that belong to the NSF/Sec and Pas class of AAA proteins. These regions have been previously identified by other workers as CAD sequences, and site-directed mutations within these divergent duplicate CAD copies disrupted gene function (Whiteheart et al. 1994).

The amino acid sequences of the CAD regions were aligned using CLUSTAL W under standard alignment parameters (Thompson et al. 1994). The initial computer analyses revealed that several CAD protein sequences had regions of high divergence, and the alignment in these areas were characterized by the presence of numerous interspersed gaps of varying lengths. Further visual refinement of the aligned sequences was undertaken by (1) consolidating gaps across sequences and (2) aligning nonidentical amino acids based on functional group similarities.

The effect of the visual refinement on our subsequent analyses was tested in two ways. First, a gap exclusion analysis was conducted, wherein all gaps were ignored across all sequences in the phylogenetic analysis. This procedure would remove most of the very divergent positions in the data set. Second, the protein phylogeny was estimated using the raw CLUSTAL W alignment prior to the visual refinement process. Neighbor-joining trees were estimated with these test data sets using 100 bootstrap replicates of the data. The tree topologies using both these test data sets were nearly identical to the tree topology based on the visually refined alignment; any changes in branching order were primarily at the tips and did not involve alterations in major group relationships that are the basis for our conclusions. Furthermore, no major differences were observed in the bootstrap confidence support of the trees based on the analyses of the gap-excluded, raw and visually refined alignments. The changes in bootstrap support levels observed differed between 0 and 10 points for basal nodes of the major AAA protein groups (see Results and Discussion). The visually refined alignment, however, showed stronger overall support for the major groups, and we retained this data set for all subsequent analyses; this alignment is available from the authors upon request.

Sequence distance calculations were carried out using the Molecular Evolutionary Genetic Analysis package (Kumar et al. 1994). Levels of amino acid replacements were calculated as the proportion of amino acid identities between sequences (p), and a Poisson-corrected distance (d) which corrects for multiple site substitutions (Nei 1987). For the phylogenetic analyses, a 339-amino acid region encompassing the CAD was utilized. CADs are generally 230–250 amino acids in length, and the larger size of the analyzed sequences is due to the addition of nonoverlapping insertions and deletions during the alignment procedure that are necessary to align different CADs. Phylogenetic analyses using distance-based procedures employed the neighbor-joining method (Saitou and Nei 1987), with deletions ignored in pairwise comparisons. Support of nodes was estimated for 500 bootstrap replicates of the data set. Preliminary phylogenies revealed that all eubacterial sequences formed a monophyletic group (the eubacterial metalloprotease group); the final reconstruction used these eubacterial genes as the outgroup clade.

Results and Discussion

Large-scale Relationships Between CAD Sequences

The evolutionary analysis of the CAD was undertaken with sequences from 105 genes in 38 species, spanning all the major kingdoms (eubacteria, archaebacteria, plants, fungi, and animals) [see Table 1]. From the 105 genes utilized in this study, a total of 133 CAD sequences were obtained, and include loci whose products contain two separate CADs that appear to have arisen through domain duplication in the course of evolution.

The largest number of CAD sequences from a single species were the 22 AAA proteins obtained from the yeast *Saccharomyces cerevisiae*, whose genome has recently been completely analyzed. Our evolutionary analyses were also able to utilize sequence information from one archaebacterial and two eubacterial genomes that have also been completely sequenced. The archaebacterium *Methanococcus jannaschii* (Bult et al. 1996) contains three AAA family members, while the eubacterial *Haemophilus influenzae* (Fleischman et al. 1995) and *Myocoplasma genitalium* (Fraser et al. 1995) each contain only a single AAA family member. The ability to utilize complete genome data sets from organisms representing each of three biological kingdoms provides us with an unprecedented ability to characterize patterns of protein family evolution. As additional complete genome sequences become available, especially from other eukaryotic genome projects (Brown 1996), it may become possible to explore a significant component of the evolutionary diversity of AAA proteins. Multiple members

Table 1. List of genes

Eubacteria	
Bacillus subtilis (Bs)	
ftsH	D26185
Escherichia coli (Ec)	
ftsH	U18997
ORF300	U23723
Haemophilus influenzae (Hi)	
ftsH	U32812
Lactococcus lactis (Ll)	
Tma1	X69123
Mycoplasma genitalium (Mg)	
ftsH	U39732
Mycoplasma leprae (Ml)	
A2126A	U00017
Eukaryote	
Arabidopsis thaliana (At)	
CDC	U37587
Candida albicans (Ca)	
Sec18	X66467
Capsicum annum (Can)	
PFTF	X80755
ATPase	X90472
Caenorhabditis elegans (Ce)	
K04D7.2	Z69664
C24B5.2	U53149
Mei1	L25423
M03C11.5	Z49128
F23F12.6	U12965
C41C4.8	Z48334
C06A1.1	Z49886
F11A10.1	Z68297
F54B3.3	Z48583
F54C9.6	Z49967
F10B5.5	Z48334
C11H1.6	Z70205
Cricetulus griseus (Cg)	
NSF	X15652
Dictyostelium discoideum (Dd)	
TBP10	L16579
TBP2	L16578
TBPα[a]	
Drosophila melanogaster (Dm)	
19DC4Z	L36286
S4	U39303
NSF	U09373
NSF2	U30502
Glycine max (Gm)	
VCP	U20213
Homo sapiens (Hs)	
Trip1	D44467
Mss1	D11094
TBP1	M34079
S4	U39303
TBP7	P43686
p42	D78275
NSF	U03985
PXAAA1	U56602
Lycopersicom esculentum (Le)	
MA1	X74426
Mus musculus (Mm)	
SKD1	U10119
SUG1	Z54219
Mss1	S39349
S4	U39302
CIP21	L76223
SKD2	U10120
VCP	M30143
Manduca sexta (Ms)	
18–56	U43728
MS73	Z38135
Naegleria fowleri (Nf)	
TBP	U41812
Oryza sativa (Os)	
TBP1	D17788
TBP2	D17789
Archaebacteria	
Halobacterium salinarium (Hs)	
CDCH	X79560
Methanococcus jannaschii (Mj)	
S4	U67560
CDC48	U67558
S8	U67590
Methanobacterium thermoautotrophicum (Mt)	
YHEA	X83691
Sulfolobus acidocaldarius (Sa)	
SAV	L17042
Odontella sinensis (Osi)	
ftsH	Z67753
YCF46	Z67753
Plasmodium falciparum (Pf)	
TBP	X77914
SCDCATP	M96757
Pichia pastoris (Pp)	
Pas5	Z22556
Pas1	Z36987
Porphyra purpurea (Ppu)	
ftsH	U38804
Rattus norvedigus (Rn)	
PAF2	D63673
ERATPASE	U11760
Saccharomyces cerevisiae (Sc)	
MSP1	X81069
Sap1	U18796
Yta6	X81071
Vps4	U25842
Yme1	L14616
Afg3	X76643
Rca1	Z49259
Sug1	X66400
Cim5	X73571
Yta1	X73569
Yta5	X81070
Yta2	U06229
Sug2	U43720
Afg2	L14615
Sec18	M20662
Pas8	L20789
Pas1	M58676
CDC48	X56956
Yta7	X81072
Bcs1	S47190
YB36	Z36055
YLL034C	Z73139
Schistoma mansoni (Sm)	
SATPase	Z29947
Schizosaccharomyces pombe (Sp)	
SUP	L33456
Let1	U02280
Mts2	Z29366
Sus scrofa (Ss)	
TBP10	X89718
VCP	M30143
Spermophilus tridecimlineatus (St)	
CADp44	U36395
Solanum tuberosum (Stu)	
ATP1	U43398
Synechocysts sp (Syne)	
SLR0228	D64000
SLR0374	D63999
0480	D64001
Xenopus laevis (Xl)	
Sug1	X81986
Mss1	X80157
p97	X54240
Yarrowia lipolytica (Yl)	
Pay4	L23858

GENBANK accession numbers of genes used in this study are enumerated. The abbreviations of species names are enclosed in parentheses

[a] The reference for the sequence of *Dictyostelum* TBPα is Mol Cell Biol 15:1725–1736 (1995)

of the AAA family are already known in other eukaryotes, especially humans and mice. Indeed, higher eukaryotes may contain more than the 22 CAD-containing ATPases indentified in yeast as revealed by the presence of at least two forms of NSF present in *Drosophila* [NSF and NSF2] (Pallanck et al. 1995).

Based on our neighbor joining analysis, it appears that the CADs found in living organisms are organized into discrete phylogenetic groups (Fig. 2). We have identified a total of six protein classes whose members are evolutionarily related and/or include proteins with similar functions. Five of these groups correspond to monophyletic protein clades; these include the metalloproteases, proteasomal regulatory factors, cytoskeleton interaction proteins, and two membrane fusion CAD-containing AAA groups. The sixth group is referred to as the orphan protein class, and includes several highly divergent CAD sequences.

Four of these CAD-containing AAA protein groups (metalloproteases, proteasomal regulatory proteins, cytoskeleton interaction proteins, and the orphan class) include proteins that possess a single CAD (Fig. 1). The membrane fusion proteins, however, contain two CADs in tandem. These ''double CAD'' proteins are subdivided into three groups: (1) the CDC48/VCP proteins that contain two well-conserved CADs, (2) the NSF/Sec proteins that contain a well-conserved N-terminal CAD with a more divergent C-terminal domain, and (3) the Pas proteins, that contain a divergent N-terminal domain and a strongly conserved C-terminal CAD. The relationships between these CAD duplicates are complex and are discussed in greater detail below.

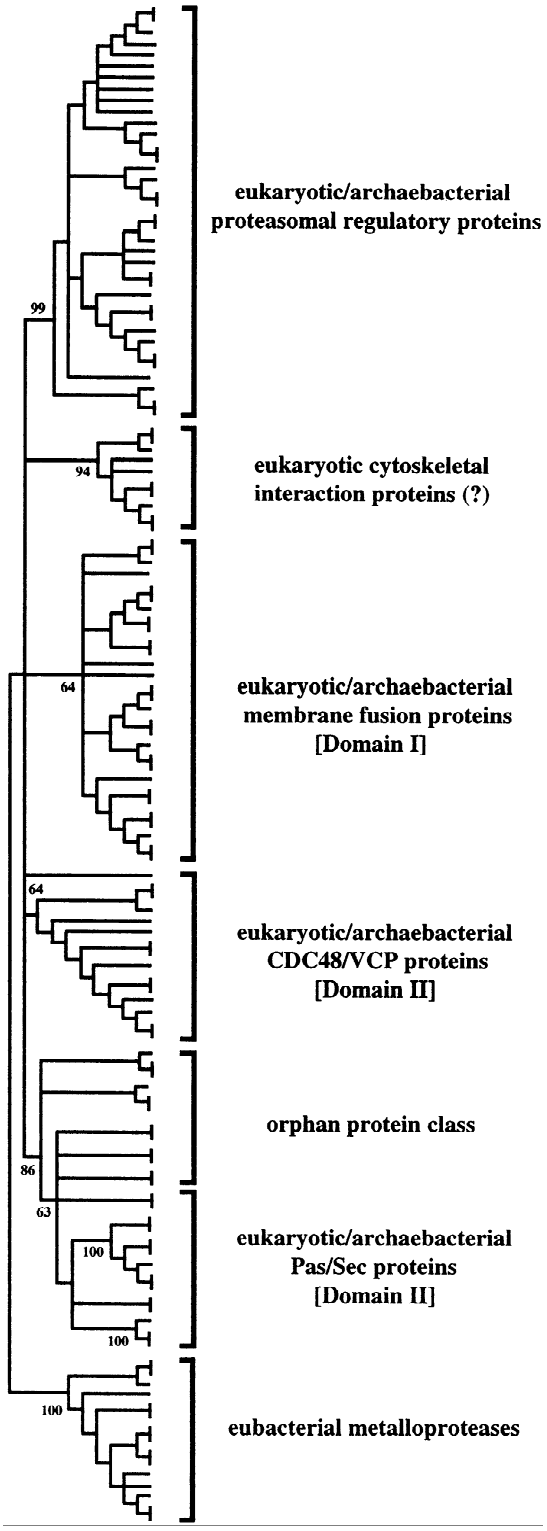

Fig. 2. Phylogenetic outline of CAD sequences. The major groups of CAD sequences are bracketed on the tree. Detailed views of different parts of the phylogeny are given in succeeding figures. The phylogeny was reconstructed using neighbor-joining analysis, based on the Poisson-corrected amino acid distance between sequences. Support for major nodes based on 500 bootstrap replicates of the data are indicated on the tree. In the phylogeny depicted in this figure, all nodes with 50% or less bootstrap support are collapsed.

The Eubacterial Metalloproteases

Nearly all eubacterial CADs belong to a monophyletic protein group that encodes membrane-bound metalloproteases, as exemplified by the *Escherichia coli* ftsH protein (Tomoyasu et al. 1995). This metalloprotease group is strongly supported as a clade in our analysis, with a bootstrap level of 100% (Fig. 3).

The eubacterial metalloproteases are characterized as containing one or two amino-terminal transmembrane domains, a single CAD, and an essential Zn-binding region. Both the ATPase activity and Zn-binding activity are required for the correct functioning of these proteins (Weber et al. 1996). Members of this protein group have been shown to degrade transcription factors such as σ32 (Tomoyasu et al. 1995; Herman et al. 1995) and phage λ cII (Arlt et al. 1996) protein and probably participate in a number of other cellular functions as well. The *E. coli* ftsH protein, for example, is also involved in bacterial septum formation (Santoss and De Almeida et al. 1975), probably via secretory processes (Kihara et al. 1995; Akiyama et al. 1996).

Members of the CAD-containing metalloproteases

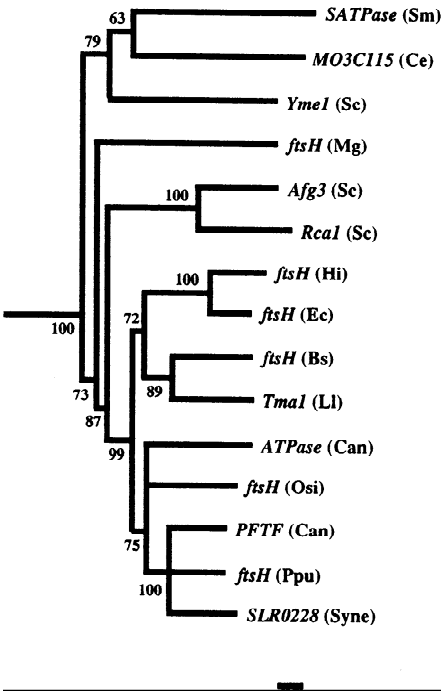

Fig. 3. The eubacterial metalloprotease protein group. The species abbreviations are shown in parentheses, with the corresponding species names enumerated in Table 1. The bar at the bottom of each phylogeny is scaled to an amino acid replacement distance of 0.05.

also contain examples isolated from eukaryotic species, but these proteins are exclusively localized in the mitochondria and/or chloroplast and are anchored in the inner organellar membrane, projecting principally into the matrix. It is interesting to note that in contrast to eubacteria, eukaryotes (as exemplified by *S. cerevisiae*) contain at least three such proteins. Mutations in the three yeast genes that encode these proteins result in multiple defects in mitochondrial function. Two of these proteins, Rcalp (= Yta12) and Afg3p (= Yta10), form an 850-kD complex anchored in the inner mitochondrial membrane that apparently does not contain any additional proteins (Arlt et al. 1996). These proteins are probably involved in protein degradation; an additional role as molecular chaperone has also been suggested for these proteins, as assembly of respiratory complexes appears to also require metalloprotease function (Arlt et al. 1996). The organellar localization and evolutionary relationship of these eukaryotic proteins with the eubacterial metalloproteases suggests that the eukaryotic metalloprotease genes arose as a consequence of ancient gene migration events that transferred these loci to the eukaryotic nucleus from the ancestral endosymbiotic organelle, followed by multiple duplication events. No archaeal members of the metalloprotease gene group have been identified, despite the availability of the complete genome sequence of *M. jannaschii.*

The phylogeny highlights the clear separation between eubacterial CADs (present only as members of the metalloprotease group) and all other CADs, indicating that a single ancestral CAD module was present in the progenitor AAA protein before the split of eubacteria from the archaebacterial/eukaryote lineage [but see discussion on *E. coli* ORF300 and *M. leprae* A2126A below]. There was subsequent diversification of CAD domain proteins in the archaebacterial/eukaryote lineage that led to the rise of paralogous gene groups with differing functions. The lack of a clear outgroup precludes the unambigous positioning of the true root of the CAD phylogenetic tree; only the identification of more distantly related CADs may allow a determination of whether the eubacterial metalloproteases are indeed the basal members of the AAA proteins used in this phylogenetic analysis.

The Proteasomal Regulatory Proteins

Eukaryotes possess a number of cellular proteolytic systems; the major route, however, for bulk cytoplasmic and nuclear proteolysis is via the ATP/ubiquitin-dependent pathway (Rock et al. 1994). Selected proteins are multiubiquitinated and targeted to the 26S proteasome for degradation. The proteasome is ubiquitous within eukaryotic cells and very abundant, comprising up to 1% of the total cellular protein (Tanaka et al. 1992). Proteasomes are not only the proteolytic component of the ubiquitin-degradation pathway, but in mammals are also responsible for peptide generation for presentation in MHC class-I expressing cells (Rock et al. 1994). Six CAD-containing AAA proteins are essential components of the major regulatory subunits of eukaryotic proteasomes.

The eukaryotic proteasome includes a catalytically latent 20S core particle complexed with activating regulatory particles to form the final 26S ''holoproteasome.'' The eukaryotic 20S core is composed of 14 subunits and forms a barrel-shaped structure of four seven-membered rings (Peters et al. 1993). A functionally and structurally similar 20S particle is present in archaebacteria. This complex has been crystallized and a detailed molecular structure has been calculated (Lowe et al. 1995). This analysis reveals that proteolytically active sites are positioned on the inner face of the central cavity of the ''barrel.'' Access to this inner cavity is restricted to two small (1.3-nm diameter) apertures at each end. The size selection produced by this restricted access requires that substrate proteins be completely unfolded in an ATP-dependent manner prior to degradation (Wenzel and Baumeister 1995). In eukaryotes, this is presumably achieved by the regulatory particles that also require ATP to associate with each end of the 20S ''barrel.''

The major regulatory particle that interacts with the 20S proteasome core is termed PA700 and contains at least 15 different proteins (DeMartino et al. 1994; Dubiel et al. 1995). While different subunits are required for substrate recognition, binding and ubiquitin recycling,

PA700 also contains six ATPases essential for activity. These proteins are AAA family members containing a single CAD and are similar to each other in size and structure, with regions of sequence similarity extending outside the CAD module. Five of the proteasomal AAA proteins contain N-terminal leucine-zipper motifs and can homo- and hetero-dimerize (Ohana et al. 1993; Russell et al. 1996). They are hypothesized to form a six-membered ring structure—similar to that observed in the alternative 20S activating particle PA28 (Gray et al. 1994)—that contacts the seven-membered ring of the terminal faces of the 20S core particle.

Our evolutionary analysis reveals that the proteasomal regulatory CADs in eukaryotes are closely related to each other and form a monophyletic gene group with 99% bootstrap support (Fig. 4). This proteasomal regulatory protein group is further divided into six paralogous subgroups, each of which are named for the yeast proteins that are members of these clades. These include the Sug1p, Sug2p, Cim5p, Yta1p, Yta5p, and Yta2p subgroups—members of each subgroup are apparently required for the correct functioning of proteasomes (with the possible exception of Yta5p). From the phylogeny, it appears that the Sug1p and Cim5p subgroups are more closely related to each other than they are to the other proteasomal regulatory protein subgroups; together, these two subgroups form a larger clade with 90% bootstrap support. There is also moderate support for joining the Yta2p and Yta5p subgroups together (84% bootstrap support).

The diversification of the proteasomal regulatory proteins into subgroups apparently occurred after the divergence of the archaebacteria from the eukaryotes 1.7 billion years ago, but prior to the split of plants, animals, and fungi (Doolittle et al. 1996). The divergence of eukaryotic proteasomal regulatory proteins into distinct subgroups may have accompanied the diversification of the core 20S subunits of the proteasome. While the eukaryotic 20S proteasomal particle contains 14 different subunits, the archaebacterial 20S particle contains only two distinct subunits (α and β). Previous analysis (Lupas et al. 1994) suggests that the eukaryotic subunit sequences can be divided into two groups of seven members each; each subunit group appears to be related to the archaebacterial sequences and may be classified as either ''α-like'' or ''β-like.'' Structural analyses indicates that members of these two eukaryotic subunit groups occupy similar positions in the eukaryotic 20S proteasomal core as their counterparts in the archaebacterial 20S particle (Hegerl et al. 1991; Lowe et al. 1995). The diversification of the 20S subunits and regulatory CAD-containing AAA proteins may have occurred in parallel, and this possibility is supported by current models of proteasomal structure which call for the regulatory AAA proteins to directly contact the ends of the 20S barrel composed of a ring of α-like subunits.

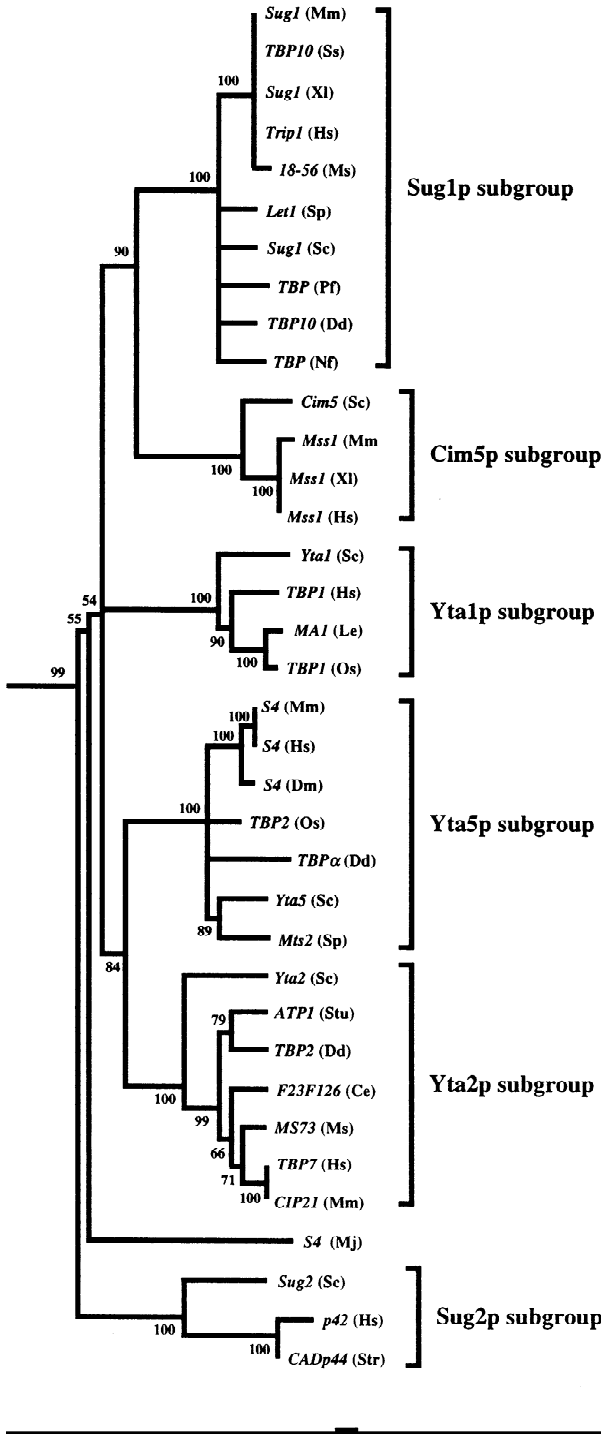

Fig. 4. Phylogeny of the eukaryotic/archaebacterial proteasomal regulatory protein group. In eukaryotes, members of this group are integral components of the proteasome.

The proteasomal regulatory protein group also includes one protein found in the archaebacteria *Methanoccocus jannaschii* (Bult et al. 1996). It does not appear to belong to any of the subgroups and may, in fact, be the basal member of the proteasomal regulatory protein group. This suggests that proteasomal regulatory proteins evolved prior to the separation of the archaebacteria from

the eukaryotes but presumably after the eubacterial lineage split away. A novel ATPase has recently been identified in *E. coli.* It consists of a ''hemi-20S particle'' composed only of β subunits complexed, not with a CAD protein, but with a Clp-like protein (Rohrwild et al. 1996). Clp proteins constitute a eubacterial proteolytic system with structural and functional similarities to eukaryotic CAD protein-containing proteasomes (Squires and Squires 1992; Schirmer et al. 1996).

The Membrane Fusion Proteins

The membrane fusion proteins are a class of CAD-containing proteins found in both archaebacteria and eukaryotic cells. The apparent diversity of cellular roles of this group of AAA proteins initially made it difficult to define a common functional feature that characterizes this group. The CDC48 proteins were first identified in yeast as cell cycle control mutants (Frohlich et al. 1991), while the Sec proteins were deficient in cellular secretion processes (Novick et al. 1980). Pas proteins are implicated in the proliferation of peroxisomes during growth (Kunau et al. 1993).

Although marked by its functional diversity, this group of proteins appears to be primarily involved in membrane-dependent processes. Yeast Cdc48p is required for homotypic membrane fusion during mitosis (Latterich et al. 1995). NSF/Sec18p is necessary for the fusion of transport vesicles from the endoplasmic reticulum with the Golgi apparatus (Sollner et al. 1993). NSF/Sec18p is also involved in the reformation of Golgi stacks after mitotic fragmentation. This process further requires the activity of another member of this group of AAA proteins, VCP/p97 (the homologue of yeast Cdc48), at a later stage of the process (Acharya et al. 1995; Rabouille et al. 1995). VCP/p97 also interacts with the vesicle-coating protein clathrin (Pleasure et al. 1993) and in human T-cells is tyrosine-phosphorylated in response to proliferation signals (Schulte et al. 1994). The Pas genes affect the import of protein into peroxisomes, and appear to be involved in membrane addition to the peroxisome. *pas1* and *pas5* mutants in *P. pastoris* result in small peroxisomes, and double mutants accumulate small vesicles.

The evolution of this group of proteins differs in certain respects from other AAA proteins. Unlike the other CAD-containing AAA proteins, the membrane fusion protein group of proteins possess two CADs. These two domains are referred to as N-terminal (N) and C-terminal (C) CADs. The CAD-N and CAD-C domains display varying degrees of sequence conservation. Both CAD-N and CAD-C found in CDC and VCP proteins show strong conservation with each other. In the Pas and Sec proteins, however, one CAD shows significant sequence divergence and consists mainly of Walker-type ATPase motifs. Interestingly, it is the N-terminal CAD domain which is divergent in the Pas proteins, but the C-terminal domain that shows increased sequence variation in the NSF/Sec 18 proteins.

Several CAD domains from the membrane fusion protein group form a monophyletic group, with a bootstrap support level of 64% (Fig. 5). This clade is referred to as membrane fusion protein domain I group, and comprises several distinct subgroups. A CDC48/VCP subgroup is strongly supported (100% bootstrap support) and includes the yeast CDC48 CAD-C domain. Based on relatedness in CAD-C domains, yeast CDC48 appears to be orthologous to the mammalian VCP and *Xenopus* p97 proteins. There is also moderate support for grouping the peroxisome proliferation proteins (Pas proteins) together (bootstrap support 64%). The positions of the yeast Pas8 and Pas1 CAD-C domains suggest these two fungal genes separated prior to the divergence of animals and fungi. A CAD from animal NSF and yeast Sec 18 proteins also forms another group with strong support in the phylogeny (100% bootstrap support); it appears, however, that the CDC48/VCP and Pas CAD-C domains are related to the N-terminal CAD of NSF/Sec (CAD-N). Finally, the archaebacterial members of this gene group appear closely related to each other (93% support), and include the SAV proteins (Confalonieri et al. 1994) of *Methanococcus janaschii* and *Sulfolobus acidocaldarius,* and the CDCH protein of the halophilic bacteria *Halobacterium salinarium* (Bibikov and Osterhelt 1994).

The formation of these distinct subgroups appears to have occurred fairly early in evolution. The formation of the NSF/Sec, Pas and CDC48/VCP subgroups probably occurred after the divergence of archaebacteria from eukaryotes, but before the split of the animal–fungal lineage from plants. The conclusions regarding the relative timing of these divergences is based on the separation of the archaebacterial lineage prior to the formation of the eukaryotic subgroups, and the inclusion of several plant proteins in the CDC48/VCP subgroup.

The CAD in membrane fusion proteins appears to have duplicated into CAD-N and CAD-C prior to the separation of the archaebacterial and eukaryotic lineages. Unlike the membrane fusion protein domain I group, however, there is no support for monophyly of the duplicate copy of the CAD domains in the membrane fusion protein group of proteins. Our phylogenetic analysis indicates that the CAD-N domains of the archaebacterial proteins and the eukaryotic CDC48/VCP subgroup are monophyletic, forming CDC48/VCP domain group II [bootstrap support level 61%] (Fig. 6). The CAD-C domains of the NSF/Sec subgroup, and CAD-N domains of the Pas subgroup, however, display a greater degree of sequence divergence, and form a very weak group referred to as Pas/Sec domain group II (Fig. 7); our phylogenetic analysis fails to place these subgroups together with the CAD-N CDC48/VCP proteins (Fig. 2).

Examination of the relative branch lengths on the phy-

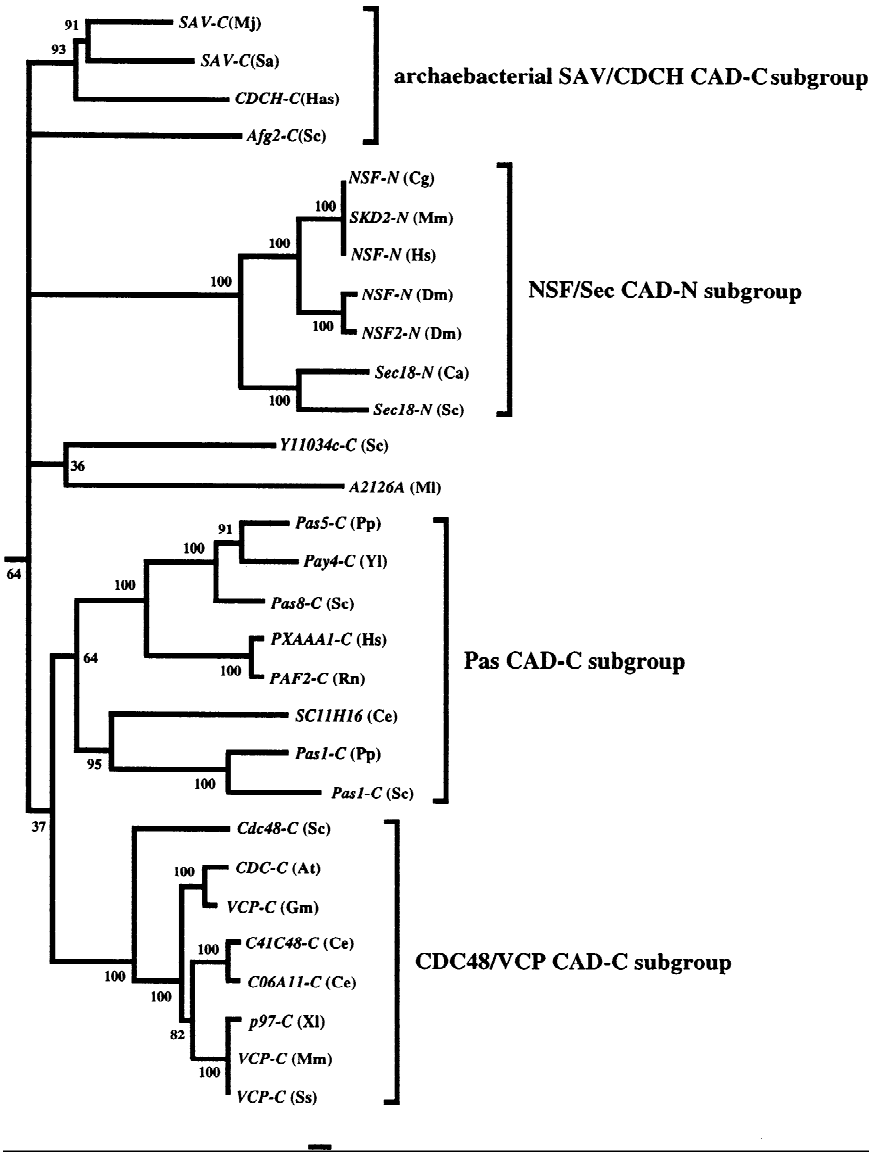

Fig. 5. The eukaryotic/archaebacterial membrane fusion protein group (domain I). This clade includes the C-terminal CAD of archaebacterial SAV and CDCH proteins, and eukaryotic peroxisome proliferation proteins (Pas), CDC and VCP sequences. This monophyletic group also includes the N-terminal CAD of the eukaryotic NSF and Sec proteins.

logenetic tree, as well as direct comparison of amino acid distances between CAD domains, indicate an acceleration in the rates of protein sequence evolution for duplicate CADs in the NSF/Sec and Pas subgroups. Figure 8 depicts a plot of Poisson-corrected sequence distance versus divergence time for CAD-N and CAD-C sequences within the CDC48/VCP and NSF/Sec subgroups. The rates of molecular evolution for both CAD-N and CAD-C are comparable within the CDC48/VCP protein subgroup, but the NSF/Sec proteins display a greater than twofold increase in the rate of CAD-C evolution. The reason for this accelerated sequence change is not apparent, but may suggest either strong selection for structural divergence between CAD-N and CAD-C or relaxation of functional constraint on the duplicate CAD domain for the NSF/Sec and Pas proteins of the AAA family. Functional analysis of *Sec18* demonstrates that mutation of the conserved ATPase motifs in the divergent CAD-C markedly decreases Sec function by inhibiting trimerization (Whiteheart et al. 1994). In contrast, replacement of K467 (within the divergent CAD-N) had no obvious effect on the biological function of yeast Pas1p (Krause et al. 1994)

The phylogeny of the membrane fusion protein CADs suggests the probable events that mark the evolution of these genes. The CAD-N and CAD-C domains of these proteins duplicated prior to the divergence of the archaebacteria and eukaryotes, and the CDC/VCP, NSF/Sec, and Pas subgroups were established after the separation of these two urkingdoms. The monophyly of the CDC48/VCP and Pas CAD-C with NSF/Sec CAD-N domains indicates that a domain reversal event probably occurred in the NSF/Sec lineage that switched the relative positions of the two CAD duplicates in these proteins. Finally, an acceleration in sequence evolution resulted in a larger degree of sequence divergence in the CAD-C and CAD-N domains of the NSF/Sec and Pas subgroup proteins, respectively.

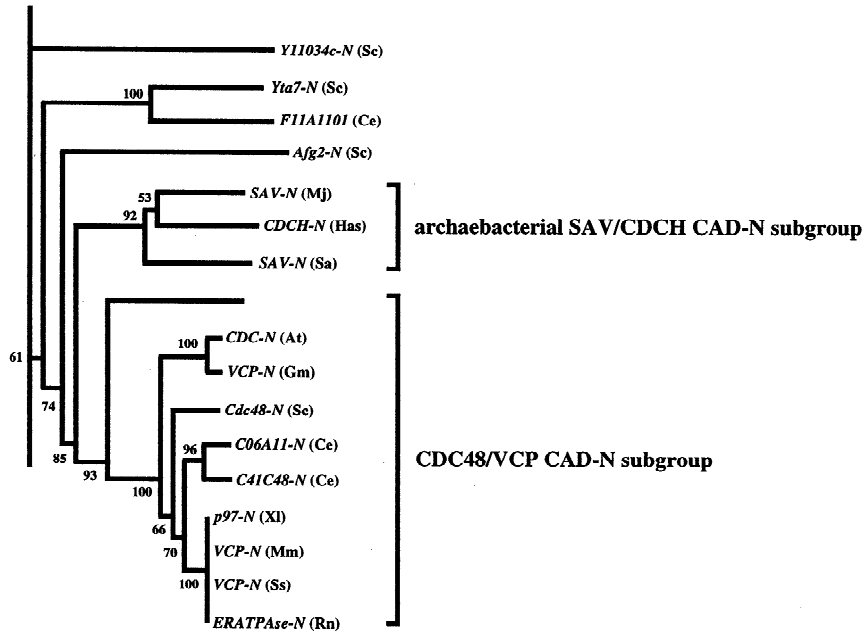

Fig. 6. The phylogeny of the eukaryotic/archaebacterial CDC48/VCP CADs (domain II). This phylogeny depicts the evolutionary relationships between the N-terminal copy of CAD found in CDC and VCP proteins. The phylogeny of the other copy (CAD-C) for these proteins are shown in Figure 5. The CAD-N of the *S. cerevisiae* Y11034c protein is also shown, although bootstrap analysis does not support a relationship between this yeast protein and other CAD-N sequences.

There are a number of other archaebacterial and eukaryotic genes, which we have termed ''orphan proteins,'' that group with the Pas/Sec domain II sequences (Fig. 7). Detailed resolution of relationships between these orphan proteins and the NSF/Sec and Pas proteins is problematic, since bootstrap support for nodes within this domain II clade are low. Many of these orphan proteins, however, which include the archaebacterial sequence *M. jannaschi* S8 and *M. thermoautotrophicum* YHEA, do not possess the duplicate domain characteristic of other members of this clade. Moreover, the protein sequences in this clade are very divergent from all other AAA CADs, and the phylogeny of this group is characterized by long branch lengths. Previous phylogenetic analyses that do not specify the eubacterial CADs as an outgroup set reveal that the placement of members of the orphan gene class, as well as Pas/Sec domain II, is unstable (data not shown). The long branch lengths in this portion of the tree does suggest that this grouping may be an artifact of the phylogeny estimation procedure (i.e., long-branch attraction).

Although we think it unlikely that the orphan protein domain class and the Pas/Sec domain II sequences form a natural group, we propose an alternate evolutionary scenario that takes this possibility into account. The species distribution of the duplicate CADs suggest, in this alternate hypothesis, that the membrane fusion protein group is a fairly complex group that evolved through multiple gene duplications and at least one domain loss. After the duplication of the CAD-N and CAD-C domains, the CDC48/VCP subgroup separated from a large subgroup that included the ancestral members of the NSF/Sec and Pas subgroups, as well as proteins in the orphan class. This divergence occurred, however, prior to the separation of the archaebacterial and eukaryotic lineage; thus, both the CDC48/VCP subgroup and the large subgroup that contains the NSF/Sec and Pas domain II CADs contain basal archaebacterial sequences. A loss of the duplicate domain in several members of the large NSF/Sec and Pas domain II subgroup gave rise to the orphan class proteins. It is unclear whether the loss of the duplicate CAD in some orphan proteins occurred once or multiple times, since rapid sequence divergence within this group makes it difficult to resolve relationships with any degree of confidence.

Finally, our analysis places two eubacterial sequences, the *Mycoplasma leprae* ORF A2126A and *E. coli* ORF 300, at separate positions within membrane fusion protein groups. The *E. coli* sequence (one of the more distantly related CAD sequences included in our analysis) appears to be associated with a retron insertion site within a defective prophage (Lim 1995), and may be the result of a horizontal transfer event into this bacterial species. An orthologue to this gene in the completely sequenced *H. influenzae* genome has not been identified. The *Mycoplasma* sequence is reported only as a sequence present within a cosmid, and an orthologue appears to be absent in the complete genome sequences of *M. genitalium* and *H. influenzae*. The position of these genes may reflect the result of a horizontal transfer event or could be an artifact. Detailed characterization of these eubacterial loci may resolve the anomalous placement of these sequences in the CAD phylogeny.

The Cytoskeleton Interaction Proteins

A final class of AAA proteins represent a group we have tentatively designated the cytoskeleton interaction proteins. These include proteins encoded by the *Caenorhab-*

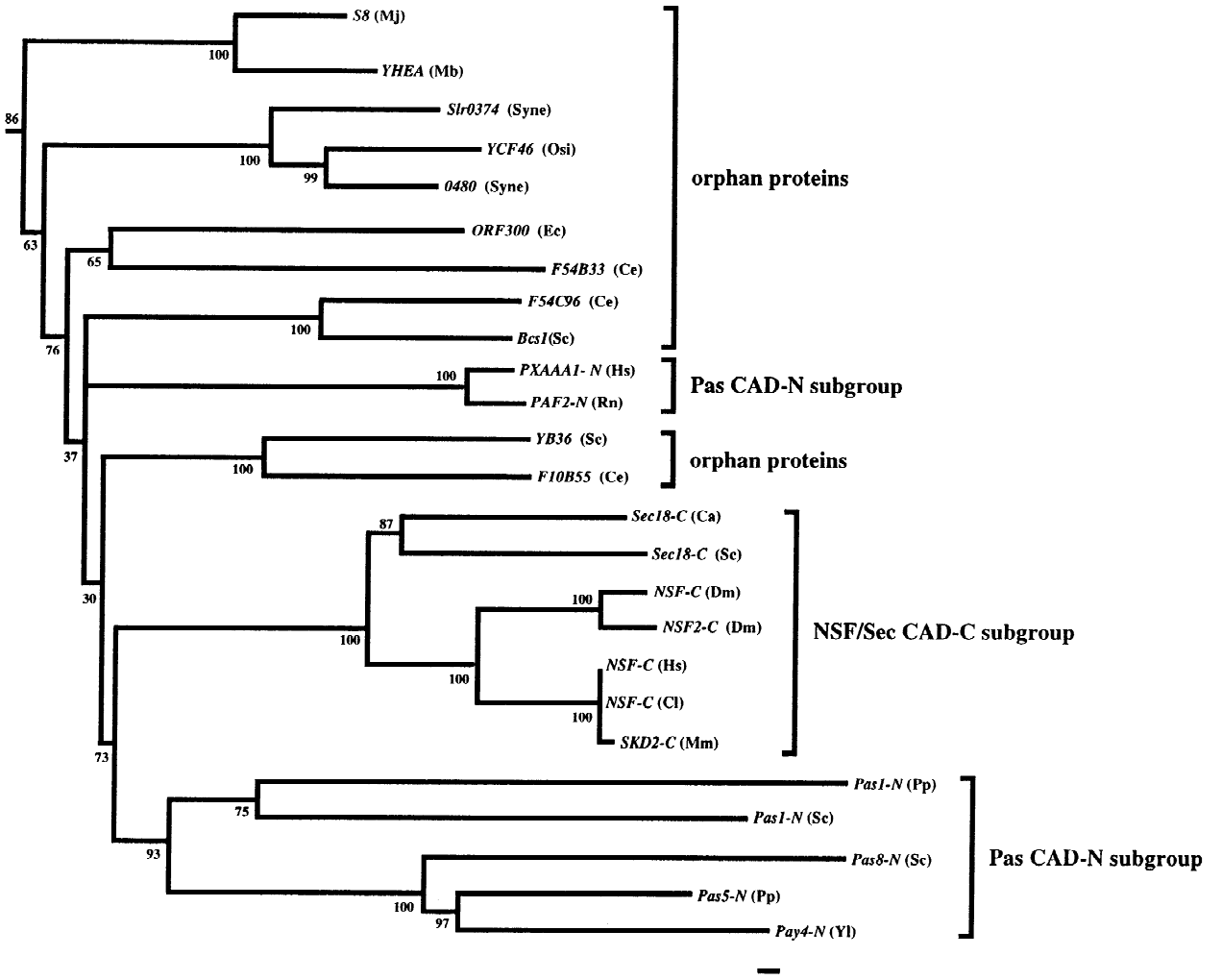

Fig. 7. Phylogeny of eukaryotic NSF/Sec and Pas CADs (domain II) as well as orphan proteins. Note that the branch lengths of the CAD-C of NSF/Sec and CAD-N of Pas proteins are longer than their duplicate copies depicted in Figure 5.

ditis elegans Mei1 gene, and the yeast *Msp1* and *Yta6* loci. This group has strong bootstrap support in our analysis (94% support level) [Fig. 9], and appears to contain only proteins found exclusively in eukaryotic organisms.

The *Mei1* gene from *C. elegans* was first identified through mutants that disrupted the transition from meiosis to mitosis in nematodes (Clark-Maquire and Mains 1994). *Mei1* appears to be required for meiotic spindle formation in the maternal germline, and the Mei1 protein is localized to the spindle during meiosis I and II, but is absent during mitosis. Indeed, cellular mitosis requires the removal of the Mei1 protein before division can proceed.

The other members of this group are not known to directly interact with the cytoskeleton; their functions, however, also suggest involvement with cytoskeletal-driven processes. Vps4 protein from *S. cerevisiae* (also known as End13p) is required for the delivery of endocytosed vesicles to the vacuole (Gammie et al. 1995; Munn and Riezman 1994), while *SUP* from *S. pombe* can suppress *bem1bud5* double mutants in *S. cerevisiae*. Bem1 and Bud5 proteins form a complex with CDC24p, CDC42p, and Bud1p (White et al. 1994). This complex is involved in bud site selection in yeast, a process that involves cytoskeletal interactions. A member of this complex, Bud1p, is a homologue of the Rap1 protein in humans; the latter is known to associate with the cytoskeleton in platelet cells.

Other proteins within this group include Sap1p from *S. cerevisiae*, which associates with the transcriptional activator Sin1p and is required for mitotic chromosome segregation (Liberzon et al. 1996). Msp1p, also from yeast, is associated with the outer mitochondrial membrane facing into the cytoplasm, and is possibly involved in mitochondrial protein sorting (Nakai et al. 1993).

The Evolutionary Divergence of the CAD

The phylogenetic relationships of the different CAD domains permit us to reconstruct the evolutionary pathways

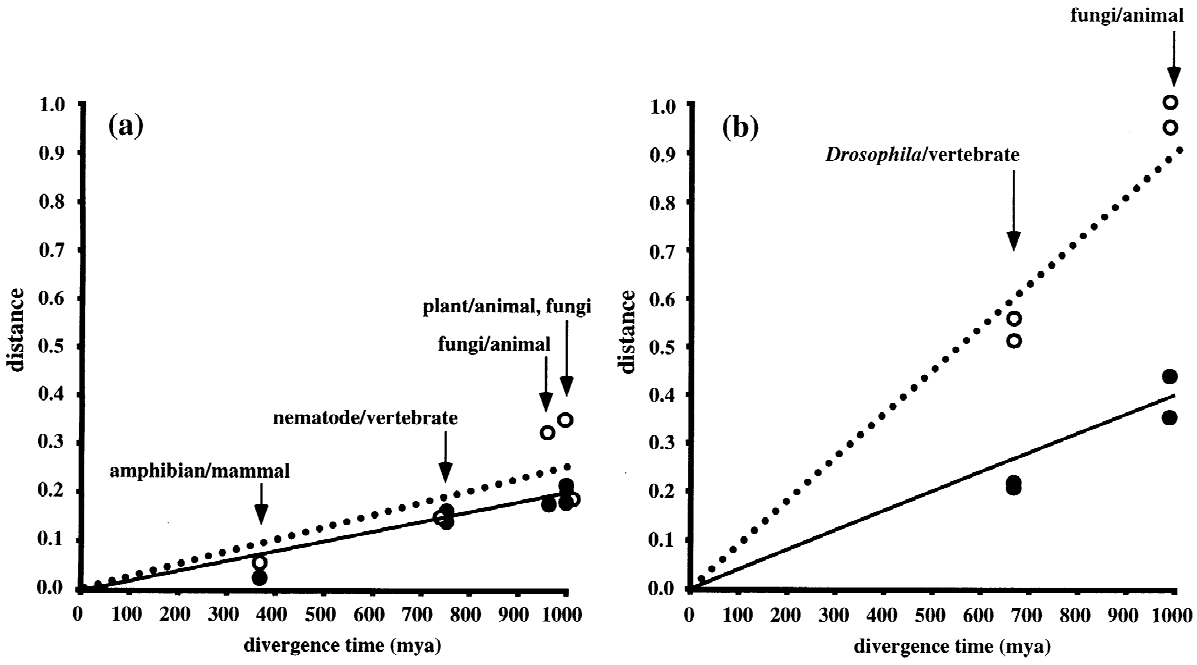

Fig. 8. The relationship between amino acid distance (Poisson-corrected) and divergence times of duplicate CADs in **(a)** CDC48/VCP and **(b)** NSF/Sec orthologues. The filled circles are the conserved CAD-N in **(a)** and CAD-C in **(b).** The open circles are for CAD-C in **(a)** and CAD-N in **(b).** The solid and dotted lines plot the relation between sequence distance and divergence times for the filled and open circled data points, respectively. The divergence times are taken from Doolittle et al. (1996).

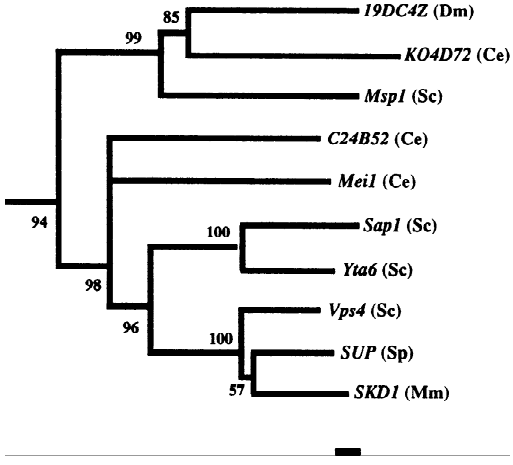

Fig. 9. Phylogeny of the eukaryotic cytoskeleton interaction protein group.

that have led to the current diversity of the CAD-containing AAA proteins. Since the derived phylogenetic tree from our analyses is unrooted, and the precise branching patterns remain unresolved in several places, it will be difficult to differentiate between several alternate evolutionary scenarios. The distribution of proteins from various kingdoms in different subgroups, however, and our ability to refer to complete genome sequence data from two eubacteria, an archaebacteria, and a eukaryote species, allows us to advance a tentative hypothesis for the patterns of diversification of the CAD. This proposed evolutionary scenario is illustrated in Figure 10.

The restriction of eubacterial CADs to the metalloprotease group suggests that these genes are basal to all other AAA proteins, which almost exclusively contain only archaebacterial and eukaryotic proteins. The major diversification events within the AAA family appear to have occurred after the separation of the eubacterial and the archaebacterial/eukaryotic lineages. The divergence of these two lineages was subsequently accompanied by the establishment of the proteasomal regulatory protein, the cytoskeleton interaction protein and the membrane fusion protein subgroups.

These major protein groups were established before the split of the archaebacterial and eukaryotic lineages. This is supported by the presence of archaebacterial proteins in the proteasomal regulatory protein and membrane fusion protein groups. It is also likely that the duplication of the CAD in the membrane fusion proteins occurred at around this time, since the archaebacterial members of this protein group possess both CAD-N and CAD-C. The cytoskeleton interaction protein group, however, does not appear to have archaebacterial representatives. This may reflect inadequate sampling from archaebacterial species, but the available complete genome sequence of *M. jannaschii* makes this argument unlikely (unless there was only limited loss of this gene in the *Methanococcus* lineage). It is possible that the archaebacteria lost the genes that encode the orthologues to the eukaryotic cytoskeleton interaction proteins. It is also possible that the cytoskeleton interaction proteins are actually related to either of the other two major ar-

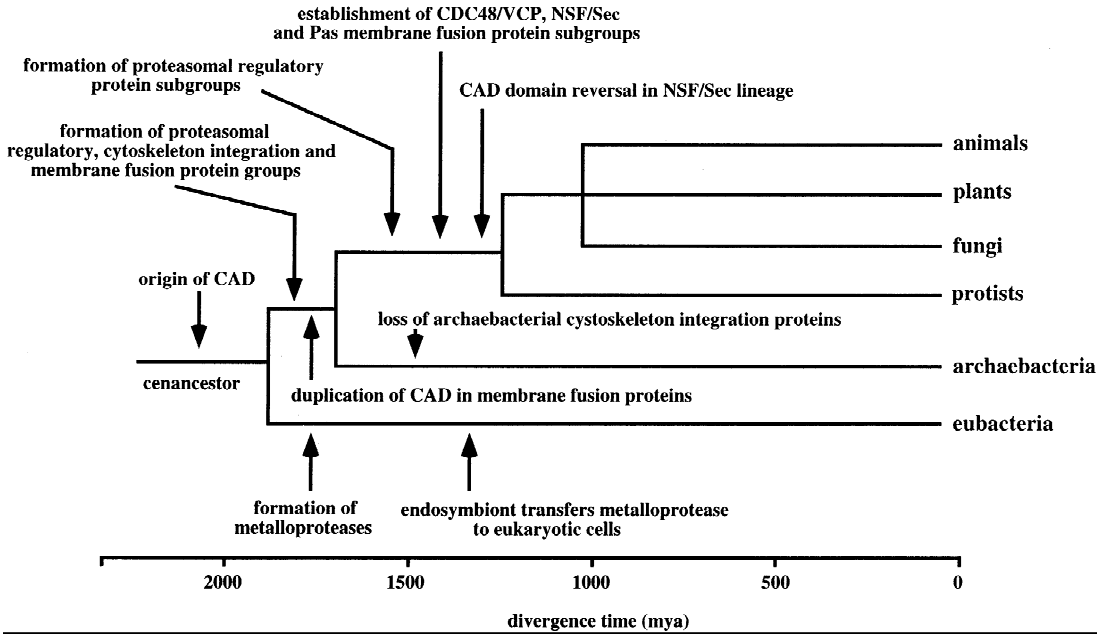

Fig. 10. Summary of the evolution of the CADs. The placement of the *arrows* only depict the relative occurrence of events in time, and are not meant to indicate absolute timing of evolutionary changes. The relative timing of these events are inferred from the patterns of divergence of species and/or protein groups as indicated by the phylogenetic analysis. The divergence times between nodes in this tree of life are taken from Doolittle et al. (1996).

chaebacterial/eukaryotic groups. Alternative topologies that place these proteins internally within either the proteasomal regulatory and membrane fusion protein groups, however, is unsupported by bootstrap analysis and is difficult to defend on either functional or structural grounds.

Subsequent to the establishment of these major protein lineages, several gene duplication events resulted in an increase in diversity in each protein group. In the proteasomal regulatory protein group, six distinct subgroups were formed after the separation of the archaebacteria and the eukaryote lineage. In yeast, these represent the proteins that comprise the PA700 regulatory particle of the proteasome. In the membrane fusion group, there was diversification into the CDC48/VCP, and the NSF/Sec and Pas protein subgroups, and this divergence occurred prior to the separation of plant, animal, and fungal lineages. Determining whether this diversification occurred before the divergence of protists from the rest of the eukaryotes will have to await the availability of additional sequence data from the membrane fusion protein group. It is also possible, however, that the diversification within the membrane fusion group proteins occurred prior to the separation of the archaebacterial and eukaryotic lines, followed by subsequent loss of a duplicate domain in several genes (see discussion above).

The presence of CADs in all organismal lineages indicates that this protein module had its origins in the cenancestor, the last common ancestor of prokaryotes and eukaryotes which was extant approximately 2 billion years ago (Fitch and Upper 1987; Doolittle et al. 1996). Since the basal genes in the AAA family are metalloproteases, and the proteasomal regulatory proteins form one of the ingroup protein clades, it may be that the ancestral CAD-containing proteins provided protease functions to the cenancestor. It is also suggestive that a recent analysis of the minimal eubacterial gene set, arrived at by a comparison of all genes held in common by the *Mycoplasma genitalium* and *Methanococcus jannaschii* genomes include AAA family members (Mushegian and Koonin 1996). The rise of cellular complexity that accompanied the origin of eukaryotes also appears to be correlated with an increase in diversity among CAD-containing proteins. The origins of this ATPase protein module early in the evolutionary history of life suggests that CAD-containing proteins are important and central players in cellular metabolism.

Acknowledgments. We are grateful to J.L. Thorne, W.R. Atchley, and E. Zuckerkandl for constructive comments on this paper. This work was supported in part by the North Carolina Agricultural Research Service.

References

Acharya U, Jacobs R, Peters JM, Watson N, Farquhar MG, Malhotra V (1995) The formation of Golgi stacks from vesiculated Golgi membranes requires two distinct fusion events. Cell 82:895–904

Akiyama Y, Kihara A, Tokuda H, Ito K (1996) FtsH is an ATP-dependent protease selectively acting on SecY and some other membrane proteins. J Biol Chem 271:31196–31201

Atchley WR, Fitch WM, Bonnerfraser M (1994) Molecular evolution of the MyoD family of transcriptional activators. Proc Natl Acad Sci USA 91:11522–11526

Atchley WR, Fitch WM (1995) Myc and Max- Molecular evolution of a family of proto-oncogene products and their dimerization partners. Proc Natl Acad Sci USA 92:10217–10221

Arlt H, Taauer R, Feldmann H, Neupert W, Langer T (1996) The YTA10-12 Complex, an AAA protease with chaperone-like activity in the inner membrane of mitochondria. Cell 85:875–885

Bibikov S, Osterhelt D (1994) Light-dependent tyrosine phosphylation of the VCP analog in the archaebacterium *Halobacterium salinarium.* 6th International Conference on Retinal Proteins, Leiden, The Netherlands, p. 67

Brown JR (1996) Preparing for the flood: evolutionary biology in the age of genomics. Trends Ecol Evol 11:510–513

Bult CJ et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii* Science 273:1058–1073

Carrera AC, Alexandrov K, Roberts TM (1993) The conserved lysine of the catalytic domain of protein kinases is actively involved in the phosphotransfer reaction and not required for anchoring ATP. Proc Natl Acad Sci USA 90:442–446

Clark-Maquire S, Mains PE (1994) Mei-1, a gene required for meiotic spindle formation in *Caenorhabditis elegans,* is a member of a family of ATPases. Genetics 1336:533–546

Confalonieri F, Marsault J, Duguet M (1994) SAV, an archaebacterial gene with extensive homology to a family of highly conserved eukaryotic ATPases. J Mol Biol 235:396–401

Confalonieri F, Duguet M (1995) A 200-amino acid ATPase module in search of a basic function. BioEssays 17:639–650

DeMartino GN, Moomaw CR, Zagnitko OP, Proske RJ, Chu-Ping M, Afendis SJ, Swaffield JC, Slaughter CA (1994) PA700, an ATP-dependent activator of the 20S Proteasome, is an ATPase containing multiple members of a nucleotide-binding protein family. J Biol Chem 269:20878–20884

Doolittle RF et al. (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271: 470–477

Dubiel W, Ferrell K, Rechsteiner M (1995) Subunits of the regulatory complex of the 26S protease. Mol Biol Rev 21:27–34

Fath MJ, Kolter R (1993) ABC transporters: bacterial exporters. Microb Reviews 57:995–1017

Fitch WM, Upper K (1987) The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. Cold Spring Harbor Symp Quant Biol 52:759–767

Fleischman RD et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–511

Fraser CM et al. (1995) The minimal gene complement of *Mycoplasma genitalium.* Science 270:397–403

Frohlich K-U, Fries H-W, Rudiger M, Erdmann R, Botstein D, Mecke D (1991) Yeast cell cycle protein CDC48p shows full-length homology to the mammalian protein VCP and is a member of a protein family involved in secretion, peroxisome formation, and gene expression. J Cell Biol 114:443–453

Fry DC, Kuby SA, Mildvan AS (1986) ATP-binding site of adenylate kinase: Mechanistic implications of its homology with ras-encoded p21, F_1-ATPase and other nucleotide-binding proteins. Proc Natl Acad Sci USA 83:907–911

Gammie AE, Kurihara LJ, Vallee R, Rose MD (1995) *DNM1,* a dynamin-related gene, participates in endosomal trafficking in yeast. J Cell Biol 130:553–566

Gray CW, Slaughter CA, DeMartino GN (1994) PA28 activator protein forms regulatory caps on proteasomal stacked rings. J Mol Biol 236:7–15

Guelin E, Rep M, Grivell LA (1994) Sequence of the *AFG3* gene encoding a new member of the FtsH/Yme1/Tma subfamily of the AAA-protein family. Yeast 10:1389–1394

Hegerl R, Pfeifer G, Puhler G, Dahlmann G, Baumeister W (1991) The three-dimensional structure of proteasomes from *Thermoplasma acidophilum* as determined by electron microscopy using random conical tilting. FEBS Lett 283:117–121

Herman C, Thevenet D, D'Ari R, Bouloc P (1995) Degradation of σ32, the heat shock regulator in *Escherichia coli,* is governed by *HflB.* Proc Natl Acad Sci USA 92:3516–3520

Heyman JA, Monosov E, Subramani S (1994) Role of the *PAS1* gene of *Pichia pastoris* in peroxisome biogenisis. J Cell Biol 127:1259–1273

Hodgman C (1992) Patterns and specificity—nucleotide binding proteins and helicases. In: Taylor WR (ed) Patterns in proteins sequence and structure, Vol 7. Springer-Verlag, London, pp 69–84

Kihara A, Akiyama Y, Ito K (1995) *FtsH* is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. Proc Natl Acad Sci USA 92:4532–4536

Krause T, Kunau WH, Erdmann R (1994) Effect of site directed mutagenisis of conserved lysine residues upon Pas1 protein function in peroxissome biogenisis. Yeast 10:1613–1620

Kumar S, Tamura K, Nei M (1994) Molecular Evolutionary Genetic Analysis Package 1.1. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, State College

Kunau WH, Beyer A, Franken T, Gotte K, Marzioch M, Saidowsky J, Skalettz-Rorowski A, Wiebel FF (1993) Two complementary approaches to study peroxisome biogenisis in *Saccharomyces cerevisiae:* forward and reverse genetics. Biochemie 75:209–224

Latterich M, Frohlich K-U, Schekman R (1995) Membrane fusion and the cell cycle: Cdc48p participates in the fusion of ER membranes. Cell 82:885–893

Lee JW, Ryan F, Swaffield JC, Johnston SA, Moore DD (1995) Interaction of thyroid-hormone receptor with a conserved transcriptional mediator. Nature 374:91–94

Liberzon A, Shpungin S, Bangio H, Yona E, Katcoff DJ (1996) Association of yeast SAP1, a novel member of the ''AAA'' ATPase family of proteins, with the chromatin protein SIN1. FEBS Lett 388:5–10

Lim D (1995) Analysis of a retron EC86 and EC87 insertion site in *Escherichia coli.* Plasmid 34:58–61

Liu Q, Summers WC (1988) Site-directed mutagenesis of a nucleotide binding domain in HSV-1 thymidine kinase: effects on catalytic activity. Virology 163:638–642

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

Lupas A, Zwickl P, Baumeister W (1994) Proteasomal sequences in eubacteria. TIBS 19:533–534

Munn AL, Riezman H (1994) Endocytosis is required for the growth of vacuolar H(+)-ATPase-defective yeast: Identification of six new *END* genes. J Cell Biol 127:373–386

Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparisons of complete bacterial genomes. Proc Natl Acad Sci USA 93:10268–10273

Nakai M, Enddo T, Hase T, Matsubara H (1993) Intramitochondrial protein sorting. Isolation and characterization of the yeast *MSP11* gene which belongs to a novel family of putative ATPases. J Biol Chem 268:24262–24269

Nei M (1987) Molecular evolutionary genetics. Columbia, NY

Nobrega FG, Nobrega MP, Tzagoloff A (1992) *BCS1,* a novel gene required for the expression of functional Rieske iron-sulfur protein in *Saccharomyces cerevisiae.* EMBO J 11:3821–3829

Novick P, Field C, Schekman R (1980) Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. Cell 21:205–215

Ohana B, Moore PA, Ruben SM, Southgate CD, Green MR, Rosen CA (1993) The type 1 human immunodeficiency virus Tat binding protein is a transcriptional activator belonging to an additional family of evolutionary conserved genes. Proc Natl Acad Sci USA 90: 138–142

Ohno S (1970) Evolution by gene duplication. Springer-Verlag, Berlin

Pallanck L, Ordway RW, Ramaswami M, Chi WY, Krishnan KS,

Ganetzky B (1995) Distinct roles for N-Ethymaleimide-sensitive Fusion Protein (NSF) suggested by the identification of a second *Drosophila* NSF homolog. J Biol Chem 270:18742–18744

Paul MF, Tzagoloff A (1995) Mutations in *RCA1* and *AFG3* inhibit F1-ATPase assembly in *Saccharomyces cerevisiae.* FEBS Lett 373: 66–70

Peters JM, Walsh MJ, Franke WW (1990) An abundant and ubiquitous homo-oligomeric ring-shaped ATPase particle related to the putative vesicle fusion proteins Sec18p and NSF. EMBO J 9:1757–1767

Peters J-M, Cejka Z, Harris JR, Kleinschmidt JA, Baumeister W (1993) Structural features of the 26S proteasome complex. J Mol Biol 234:932–937

Pleasure I, Black MM, Keen JH (1993) Valosin-containing protein, VCP, is a ubiquitous clathrin-binding protein. Nature 365:459–462

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF (1995) Molecular evolution of flower development—diversification of the plant MADS-box regulatory gene family. Genetics 140:345–356

Rabouille C, Levine TP, Peters JM, Warren G (1995) An NSF-like ATPase, p97, and NSF mediate cisternal regrowth from mitotic Golgi fragments. Cell 82:905–914

Rock KL, Gramm C, Rothstein L, Clark K, Stein R, Dick L, Hwang D, Goldberg AL (1994) Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. Cell 78:761–771

Rohrwild M, Coux O, Huang H, Moerschell RP, Yoo SJ, Seol JH, Chung CH, Goldberg AL (1996) HslV-HslU: a novel ATP-dependent protease complex in *E. coli* related to the eukaryotic proteasome. Proc Natl Acad Sci USA 93:5808–5813

Russell SJ, Sathyanarayana V, Johnston SA (1996) Isolation and characterization of SUG2, a novel ATPase family component of the yeast 26S proteasome. J Biol Chem 271:32810–32817

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Santos D, De Almeida DF (1975) Isolation and characterisation of a new temperature-sensitive cell division mutant of *Escherichia coli* K-12. J Bacteriol 124:1502–1507

Schirmer E, Glover J, Singer MA, Lindquist S (1996) HSP100/Clp proteins: a common mechanism explains diverse functions. Trends Biochem Sci 21:289–296

Schulte RJ, Campbell M-A, Fischer WH, Sefton BM (1994) Tyrosine Phosphylation of VCP, the mammalian homologue of the *Saccharomyces cerevisiae* CDC48 protein, is unusually sensitive to stimulation by sodium vanadate and hydrogen peroxide. J Immunology 153:5465–5472

Sollner T, Bennett MK, Whiteheart SW, Scheller RH, Rothman JE (1993) A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation and fusion. Cell 75:409–418

Squires C, Squires CL (1992) The Clp proteins: proteolysis regulators or molecular chaperones? J Bact 174:1081–1085

Swaffield JC, Melcher K, Johnston SA (1995) A highly conserved ATPase protein as a mediator between acidic activation domains and the TATA-binding protein. Nature 374:88–91

Tanaka K, Tamura T, Yoshimura T, Ichihara A (1992) Proteasomes—proteins and gene structures. The New Biologist 4:173–187

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tomoyasu T, Gamer J, Bukau B, Kanemore M, Rutman AJ, Oppenheim AB, Yura T, Yamamaka K, Niki H (1995) *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor σ32. EMBO J 14:2551–2560

Walker JE, Saraste MJ, Runswick JJ, Gray NJ (1982) Distantly related sequences in the α- and β-subunits of ATPase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide-binding fold. EMBO J 1:945–951

Weber ER, Hanekamp T, Thorsness PE (1996) Biochemical and functional analysis of the *YME1* gene product, an ATP and zinc-dependent mitochondrial protease from *S. cerevisiae.* Mol Biol Cell 7:307–317

Weiss DS, Batut J, Klose KE, Keener J, Kustu S (1991) The phosphorylated form of the enhancer-binding protein NTRC has an ATPase activity that is essential for activation of transcription. Cell 67:155–167

Wenzel T, Baumeister W (1995) Conformational constraints in protein degradation by the 20S proteasome. Structural Biology 2:199–204

White GC, Crawford N, Fischer TH (1993) Cytoskeletal interactions of Rap1p in platelets. Adv Exp Med Biol 344:187–194

Whiteheart SW, Rossnagel K, Buhrow SA, Brunner M, Jaenicke R, Rothman JE (1994) N-Ethylmaleimide-sensitive fusion protein: a trimeric ATPase whose hydrolysis of ATP is required for membrane fusion. J Cell Biol 126:945–954

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Xu Qunli, Singer R, Johnston GC (1995) Sug1 modulates yeast transcription activation by Cdc68. Mol Cell Biol 15:6025–6035