Available online at www.sciencedirect.com
ScienceDirect
Journal of Structural Biology 156 (2006) 2–11
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

Classification of AAA+ proteins

Moritz Ammelburg, Tancred Frickey [1], Andrei N. Lupas *

Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, D-72076 Tübingen, Germany

Received 22 December 2005; received in revised form 6 May 2006; accepted 12 May 2006
Available online 26 May 2006

Abstract

AAA+ proteins form a large superfamily of P-loop ATPases involved in the energy-dependent unfolding and disaggregation of macromolecules. In a clustering study aimed at defining the AAA proteins within this superfamily, we generated a map of AAA+ proteins based on sequence similarity, which suggested higher-order groups. A classification based primarily on morphological characteristics, which was proposed at the same time, differed from the cluster map in several aspects, such as the position of RuvB-like helicases and the inclusion of divergent clades, such as viral SF3 helicases and plant disease resistance proteins (RFL1). Here, we establish the presence of an α-helical domain C-terminal to the ATPase domain (the C-domain) as characteristic for AAA+ proteins and re-evaluate all clades proposed to belong to this superfamily, based on this characteristic. We find that RFL1 and its homologs (APAF-1, CED-4, MalT, and AfsR) are AAA+ proteins and SF3 helicases are not. We also present a new and more comprehensive cluster map, which assigns a central position to RuvB and clarifies the relationships between the clades of the AAA+ superfamily.

Keywords: AAA; AAA+; Cluster analysis; Molecular evolution

1. Introduction

AAA+ proteins form a large and diverse group of ring-shaped complexes, which are involved in a broad range of cellular processes. Indeed, the core group of this superfamily, the AAA proteins, obtained their name as an acronym for 'ATPases associated with diverse cellular activities' (Erdmann et al., 1991), which at the time reflected our basic ignorance of their function. Gradually, the group was extended to include all protease-associated ATPases (Lupas et al., 1997) and was eventually merged with a heterogeneous group of helicases, transcription activators and regulators of metabolic enzymes (Koonin, 1993) to form the AAA+ superfamily (Neuwald et al., 1999). The unifying activity of AAA+ proteins appears to involve the ATP-driven unfolding, disassociation and remodelling of macromolecules (both protein and nucleic acids) (Lupas and Martin, 2002). The mechanism by which nucleotide hydrolysis generates mechanical work is still poorly understood.

The structural core of AAA+ proteins is formed by one or more copies of an extended P-loop ATPase domain, consisting of the ATPase domain proper, followed by an α-helical subdomain (the C-domain). Most proteins have a single copy, but a sizable part of the Clp/Hsp100 and AAA groups have two, which are denoted D1 and D2 (in dynein and midasin, which have six tandem domains, these are referred to as P1–P6). The domains form ring-shaped complexes, in which the nucleotides are bound at the domain interfaces, allowing each subunit to contribute a critical residue (the arginine finger) to the active site of the next subunit in the ring (reviewed by Ogura et al., 2004). In each subunit, the presence of the γ-phosphate of ATP is sensed by a residue at the C-terminal end of strand β4 (sensor-1), which is connected to the arginine finger by a short helix, thus establishing a path of communication between adjacent active sites (Lupas and Martin, 2002). A characteristic feature, which seems unique to the extended ATPase domain of AAA+ proteins, is the C-domain,

* Corresponding author. Fax: +49 7071 601 349.
E-mail address: andrei.lupas@tuebingen.mpg.de (A.N. Lupas).
[1] Present address: Research School of Biological Sciences, The Australian National University, Canberra, ACT 2601, Australia.

1047-8477/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2006.05.002

which makes multiple contacts to the bound nucleotide and contributes an important arginine residue (sensor-2) to the active site in most AAA+ proteins (but not in Clp-D1 or the AAA group) (Ogura et al., 2004; Botos et al., 2004).

P-loop nucleotidases represent one of the largest groups of homologous proteins in nature and their sheer number precludes a classification by molecular phylogeny. Correspondingly, deep branches have been established primarily through the use of morphological traits (such as strand order of the central β-sheet, the presence of characteristic inserts, or the oligomerization state). Using such traits, AAA+ proteins have been classified together with various helicases and F-type ATPases into the unfoldase group of RecA-like nucleotidases, which themselves represent one of three basic folds within the P-loop proteins (Lupas and Martin, 2002). In a recent study, Iyer et al. (2004) proposed a classification of protein families within the AAA+ superfamily, based mainly on morphological traits. They recognized six clades and additionally grouped the two largest clades, comprising half the families, into the pre-sensor-1 hairpin superclade, named thus for an extended β-hairpin inserted at the C-terminus of helix α3, adjacent to the arginine finger.

In parallel to that effort we took a different approach, using sequence similarity to obtain a cluster map of AAA+ proteins (Frickey and Lupas, 2004b). Although this approach uses the same data as molecular phylogeny, it is not limited by the number of sequences (indeed, it becomes more accurate with increasing numbers of sequences as the larger number of pairwise relationships average out spurious matches resulting from the limitations of the search programs). Our goal was to define the AAA proteins within the AAA+ superfamily, but our cluster map turned out to contain all known AAA+ families and suggested a number of higher-order groups. It agreed with the morphological classification of Iyer et al. in many points, but differed in some important respects, including the membership of divergent proteins (such as SF3 helicases and plant disease resistance proteins) and the position of several families, most importantly the helicase RuvB.

Here, we describe a study based on both morphological traits and molecular sequences, undertaken to re-evaluate these points. We find that plant disease resistance proteins and their homologs, such as APAF-1, CED-4, and MalT/AfsR-like transcription regulators of bacteria, previously grouped into a separate superfamily (STAND; Leipe et al., 2004), are AAA+ proteins and that SF3 helicases are not. Our new cluster map places RuvB at the base of the clamp loader clade, close to the center of the map. This revised classification provides the most comprehensive integration of sequence and structure data of AAA+ proteins yet and resolves conflicts between the previous classification and the crystal structures of STAND proteins and SFIII helicases.

2. Methods

2.1. Analysis of C-domains

We extracted all C-domains from family c.37.1.20 of the SCOP database, annotated as the 'extended AAA-ATPase domain' (Andreeva et al., 2004). We then searched the Protein Data Bank for further occurrences of this domain: At the sequence level we used Hidden Markov Model comparisons implemented in HHpred, with default settings and global alignment mode (Soding, 2005; Soding et al., 2005). At the structural level we used DALI (Holm and Sander, 1993). These searches revealed the absence of the C-domain in the SF3 helicases of family c.37.1.20 (1N25, 1U0J, and 1TUE) and its presence in Orc2 (1W5S), VPS4 (1XWI), PspF (2BJV), ZraR (1OJL), APAF-1 (1Z6T), and CED-4 (2A5Y), which have not yet been classified in SCOP.

The C-domains were superimposed interactively in Swiss-PDB Viewer (Guex and Peitsch, 1997), after automated attempts using POSA (Ye and Godzik, 2005) and STAMP (Russell and Barton, 1992) failed to yield a usable output, due to the substantial structural divergence of the proteins in the dataset. RuvB from *Thermotoga maritima* (1IN4, chain A) was used as a reference structure for the superpositions, as this protein made the highest number of connections both in structure and in sequence searches. Quantitative information for the superpositions is given in Table 1. The alignment in Fig. 3 reflects the structural superposition; residues not considered equivalent appear in the alignment in lower case.

2.2. Identification of AAA+ proteins in the non-redundant sequence database

We extracted the sequences of the extended ATPase domain of AAA+ proteins, from helix α0 of the ATPase domain proper to the C-terminus of the C-domain. Initially this was done for all proteins in SCOP family c.37.1.20 minus the three viral SF3 helicases. Subsequently, after repeated sequence searches with this set consistently recovered plant disease resistance proteins and APAF-1 and the sequence and structure searches with the C-domains confirmed the presence of such a domain in these proteins, the sequence of APAF-1 was included in the search set. The sequences in the search set were aligned manually, guided by their structures, and a Hidden Markov Model (HMM) was built and calibrated using HMMer (hmmer.wustl.edu; Eddy, 1998).

To search for AAA+ proteins, we used nr70, which represents the non-redundant database reduced to a maximum pairwise identity of 70%. We obtained a first filtering by running PSI-Blast (Altschul et al., 1997) at an inclusion cutoff of 10^{-6} for 20 iterations (or to convergence) with each sequence in our search set and gathering all sequences with an *E*-value better than 10^{-5} in the last iteration. We then searched the obtained sequences with the HMM of the extended ATPase domain and gathered all sequences