A link between sequence conservation and domain motion within the AAA+ family

Graham R. Smith,[a,1] Bruno Contreras-Moreira,[a,1] Xiaodong Zhang,[b,*] and Paul A. Bates[a,*]

[a] *Biomolecular Modelling Laboratory, Cancer Research UK London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK*
[b] *Centre for Structural Biology, Department of Biological Sciences, Imperial College London, Flowers Building, South Kensington, London SW7 2AZ, UK*

Abstract

The AAA+ family of proteins play fundamental roles in all three kingdoms of life. It is thought that they act as molecular chaperones in aiding the assembly or disassembly of proteins or protein complexes. Recent structural studies on a number of AAA+ family proteins have revealed that they share similar structural elements. These structures provide a possible link between nucleotide binding/hydrolysis and the conformational changes which are then amplified to generate mechanical forces for their specific functions. However, from these individual studies it is far from clear whether AAA+ proteins in general share properties in terms of nucleotide induced conformational changes. In this study, we analyze sequence conservation within the AAA+ family and identify two subfamilies, each with a distinct conserved linker sequence that may transfer conformational changes upon ATP binding/release to movements between subdomains and attached domains. To investigate the relation of these linker sequences to conformational changes, molecular dynamics (MD) simulations on X-ray structures of AAA+ proteins from each subfamily have been performed. These simulations show differences in both the N-linker peptide, subdomain motion, and cooperativity between elements of quaternary structure. Extrapolation of subdomain movements from one MD simulation enables us to produce a structure in close agreement with cryo-EM experiments.
© 2003 Elsevier Inc. All rights reserved.

Keywords: AAA+ family; Structural bioinformatics; Molecular dynamics; Cooperativity; Protein evolution

1. Introduction

ATPase associated with various cellular activities (AAA) proteins were first identified as a subfamily of P-loop ATPases (Kunau et al., 1993). In the absence of structural information, Beyer performed a detailed sequence analysis and established consensus sequences for this family (Beyer, 1997). A few years later Neuwald et al. identified a broader AAA+ family of proteins, and defined a characteristic set of sequence motifs (Neuwald et al., 1999). These versatile proteins are found in all life forms and, as their name implies, are involved in

numerous cellular activities. In general, AAA+ proteins are believed to unwind or disassemble proteins or nucleic acids. They differ from other P-loop ATPases in that they contain at least one additional α-helical subdomain. Two major types of AAA+ proteins have been identified so far: those with one AAA+ cassette (type I) and those with two AAA+ cassettes (type II; we will refer to the cassettes as D1 and D2). Larger numbers of consecutive AAA+ cassettes have been reported but are rare (see http://www.sanger.ac.uk/Pfam). Substrate recognition domains have been found to be C- and N-terminal, as well as inserted within the AAA+ domain. For a recent review of the AAA+ superfamily, see Lupas and Martin (2002). Some example members of the superfamily are p97/cdc48p/VCP and NSF/Sec18p, membrane fusion ATPases (May et al., 2001), p97 being involved in Golgi, ER and nuclear envelope formation, and NSF in vesicle transport; HslU/ClpY, a component

of the bacterial chaperone–protease system; the other Clp chaperones, involved in resolubilization of aggregated proteins; Lon proteases; and the proteosome regulatory subunit (PRS) proteins.

AAA+ proteins typically exist as oligomers, usually hexamers. Nucleotide binding has often been reported to be required for oligomerization in some examples (Ogura and Wilkinson, 2001) but not in others (Hattendorf and Lindquist, 2002). By using various visualization techniques, including electron microscopy and X-ray crystallography, large conformational changes can be observed in different AAA+ proteins such as NSF, p97, HslU, and ClpX upon ATP binding/hydrolysis (Hanson et al., 1997; Kondo et al., 1997; Ortega et al., 2000; Rouiller et al., 2000; Wang et al., 2001). These conformational changes are suggested to transfer the ATPase activity to chaperone functionality directly or via adaptors.

Among the AAA+ proteins whose structures have been characterized, we selected *Escherichia coli* HslU (Bochtler et al., 2000; Trame and McKay, 2001), murine p97 (Zhang et al., 2000), and murine NSF (Lenzen et al., 1998) as representatives of the whole family. HslU is of type I whereas p97 and NSF are of type II, although the available structures contain only one AAA+ cassette, D1 in the case of p97 and D2 in the case of NSF. p97 has also a N-terminal substrate binding domain, N, and HslU a domain inserted within the α/β subdomain, I.

Using a combination of protein sequence and structure information to construct alignments, we identify precise domain boundaries and a conserved "N-linker" region at the N-terminal end of the AAA+ cassette, lying either between the AAA+ cassette and the N-terminal substrate binding domains, or between consecutive AAA+ cassettes. Previously, the AAA+ superfamily has been divided into true AAA and other AAA+ from the presence or absence of the AAA minimal consensus motif (also known as the second region of homology, or SRH), consisting of sensor I plus Arg-finger (Karata et al., 1999; Neuwald et al., 1999). We use instead the N-linker region as a distinguishing feature, define AAA+ subfamilies accordingly, and compare with the previous division made on the basis of the SRH.

The AAA+ proteins are thought to undergo conformational changes on binding/hydrolysis of nucleotide, and to transmit this to movements of the attached effector domains which act upon the substrate proteins. The conserved nature of the linker and its location between domains and (as we shall see) wrapping around the nucleotide binding pocket clearly suggest it may be involved in these.

In the case of HslU, X-ray structures exist for two differently liganded states, giving much detailed information about the conformational changes on ligand binding, and accordingly these are discussed in the first part of Section 2. For the other proteins, molecular

dynamics (MD) computer simulation provide a way of obtaining information about conformational changes, and such simulations have been carried out on p97 D1, NSF D2, and HslU hexamers, in a variety of ligand occupancy states, and also with some mutations. The analysis of these simulations produces the remainder of our results.

Because these systems are large we were only able to run each simulation for the relatively short period of 1 ns. Nucleotide hydrolysis cannot be addressed by classical MD, in which bond connectivity is fixed, only by Quantum Mechanics, and in any case occurs in AAA+ systems on a vastly longer timescale than we are investigating here ($k_{cat} \sim 1 \, s^{-1}$; Hattendorf and Lindquist, 2002). The timescales of large-scale domain motions or nucleotide binding and unbinding are also too long to be directly observed in 1 ns MD simulations. Therefore, a hierarchy of analyses of the MD has been used to bring out motions of different types. First the details of motions in regions expected to be particularly significant for function, such as changes in the volumes of the ATP binding site and changes in the N-linker backbone angles, have been examined. The root mean square fluctuations (rmsf) of the protein have also been calculated. In this, large domain movements appear as large maxima and minima extending over long regions of the peptide chain, on which smaller maxima and minima corresponding to regions of local flexibility are insterspersed. These compare to structural or sequence motifs (such as the alternative N-linkers). As well as from the rmsf plot, the size of the subdomain motions can also be identified by examining the movements of the centers of mass of the subdomains in the course of the simulation.

To better reveal the large-scale domain motions, we "factored out" the atomic-level details of the motions by applying a technique based upon the principal components analysis of atomic fluctuations, essential dynamics (ED) (Amadei et al., 1993; Garcia, 1992) to the MD trajectory. In addition, motion along a principal component can be extrapolated beyond the extremes that are sampled during the simulation, allowing an extension of the type of motion seen in the simulation into a regime inaccessible due to timescale limitations.

Moreover, previous analyses of the X-ray structures have suggested there may be differences in the way that the subunits cooperate in the functional hexameric state: in particular, that there may be a distinction between an allosteric motion of HslU (Bochtler et al., 2000) and a concerted motion of p97 (Zhang et al., 2000), upon nucleotide binding. Correlations remaining in the MD trajectory after ED analysis can be analyzed with the DynDom program (Hayward and Berendsen, 1998) to identify "dynamic domains," thus addressing this question of cooperativity.

Taken together, the above MD analyses indicate that large conformational changes between subdomains are

likely to take place upon nucleotide binding or release, particularly for the N-terminal domains of p97, and give some indication of what the changes are. They can be correlated with other structural information (different X-ray structures in the case of HslU and cryo-EM information in the case of p97) with the type of N-linker, Gly–Gly or Hydrophobic-Gly.

2. Results and discussion

2.1. Structural comparison

The overall fold of the AAA+ family (α/β plus α-helical subdomains) is conserved among its members, despite their low sequence identity (see Fig. 1, and the supplementary material for sequence alignments of the family). On average, sequences from X-ray crystal structures share less than 20% identity. However, up to 50% of Cα positions can be equivalenced under 2 Å of rmsd. When p97, HslU, and NSF are superposed [also cdc6 (Liu et al., 2000; pdb code 1fnn), RuvB (Yamada et al., 2001, pdb code 1hql), and subunits in the clamp loading complex of DNA polymerase III (Jeruzalmi et al., 2001, pdb code 1jr3), data not shown], the central β-sheet appears to be the common frame of the family, but there are some differences in the three-dimensional location of secondary structure elements as well as in the relative subdomain orientation. This structural superposition defines a precise boundary for the AAA+ cassette N-terminus, starting just before α-helix 4 (labeling according to p97; Zhang et al., 2000), in a short patch that comprises the "N-linker region." This linker is the first common element in the fold of the family, separating the AAA+ cassette from additional N-terminal domains or previous AAA+ cassettes. It has been identified as a conserved motif since the early analyses of the AAA+ family (Beyer, 1997), but its importance, since there was then no structural data, could not be recognized. The C-terminal boundary of the AAA+ cassette is the last helix in the α-helical subdomain.

The regions in the vicinity of the nucleotide binding pocket superimpose more closely than, for example, regions involved in hexamerization, such as α-helix 9 and α-helix 5 (Zhang et al., 2000). The α-helical subdomain also shows regions of good superposition near the nucleotide binding site, particularly the bridge between subdomains, the sensor II loop, and α-helix 10.

2.2. The N-linker region is part of the nucleotide binding pocket and transfers conformational changes induced upon nucleotide binding

The nucleotide is bound within a chamber which consists of (see Figs. 1 and 2): N-linker, bridge between subdomains, sensor II from the α-helical subdomain, a hydrophobic patch on α-helix 10, Walker A motif (P-loop), Walker B motif and AAA minimum consensus (from the adjacent monomer in the hexamer). The first four regions and the helical region of the Walker A motif (beginning of α-helix 6) form a hydrophobic pocket that binds the adenine ring of the nucleotide. The later two and the loop region of the Walker A motif involve predominantly polar and electrostatic interactions with the ribose and poly-phosphate tail of the nucleotide. Structural changes upon nucleotide binding and hydrolysis are probably driven by all these elements. Here we consider only conformational changes occurring around the adenine ring and the contribution the N-linker makes to those changes. The role of the other parts within the chamber has been described elsewhere (Lenzen et al., 1998; May et al., 2001; Zhang et al., 2000).

Fig. 2 shows structural details around the nucleotide binding site. The HslU "Ile–Gly" N-linker region (residues 18–19 (Bochtler et al., 2000)) interacts with the nucleotide adenine ring (Fig. 2B, right). Analyzing HslU nucleotide free and bound crystal structures, Wang et al. observed subdomain reorientations in this protein (Wang et al., 2001). Moreover, Bochtler et al. postulated that alternating bound and free nucleotide states can simultaneously exist within the HslU hexameric ring structure, as shown by one of their X-ray structures (Bochtler et al., 2000). It is also proposed that this alternating binding pattern is required in key steps of the translocation mechanism for HslU: feeding polypeptide chains through the central aperture (Bochtler et al., 2000; Wang et al., 2001). In Fig. 2C, the superposition of ligand-bound and empty HslU subunits is shown. In support of this model, we observe the repositioning of the Ile18 side chain, in the Ile–Gly N-linker, from inside to outside the adenine ring binding pocket (Figs. 2B–C), ultimately caused by reorientation of subdomains on adjacent protomer. In the nucleotide-bound state, there would be steric clashes if this Ile side chain remained in the same position as in the nucleotide free state. To compensate this side chain flip the ϕ, ψ torsion angles for the Gly in the free/bound crystal structures moves from an allowed to a disallowed region, for a non-Gly residue, of the Ramachandran plot: ($-70°$, $-130°$) to ($70°$, $-152°$). Therefore, the energetics of these coupled conformational changes seem to require a Gly in the second position of the N-linker. Additional hydrophobic residues, from both subdomains, also contribute to the hydrophobic adenine binding pocket, most notably Ile344 (α-helical subdomain), Ile17 (next to N-linker), Val61 (P-loop), and Leu66 (α-helix 6).

In the p97 structure (Zhang et al., 2000) (ADP bound; Figs. 2D–F), the N-linker region has a different "Gly–Gly" sequence (residues 207–208). It is expected to play a similar role in relating nucleotide binding to subdomain rearrangement (and N-terminal domains where present),
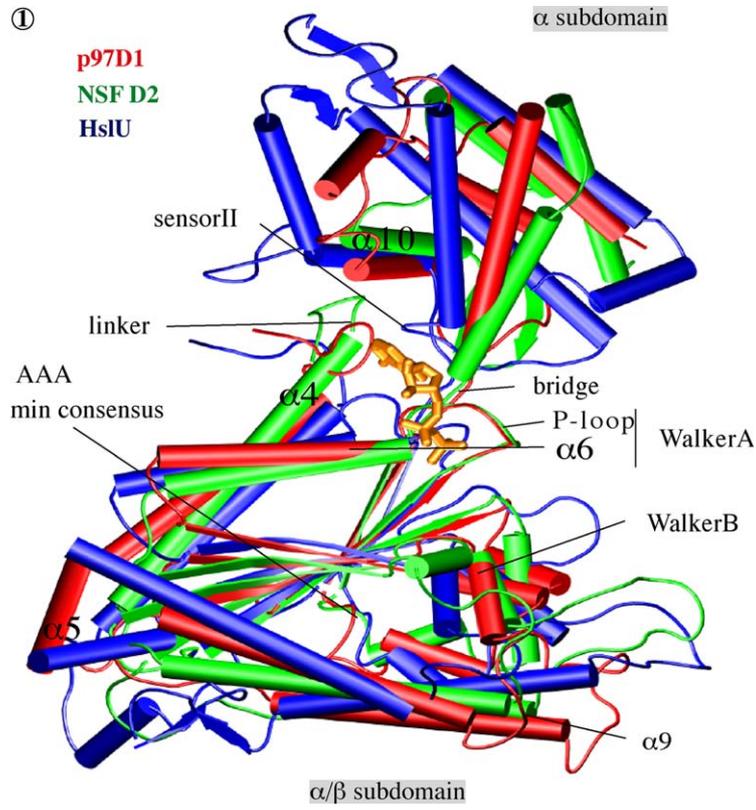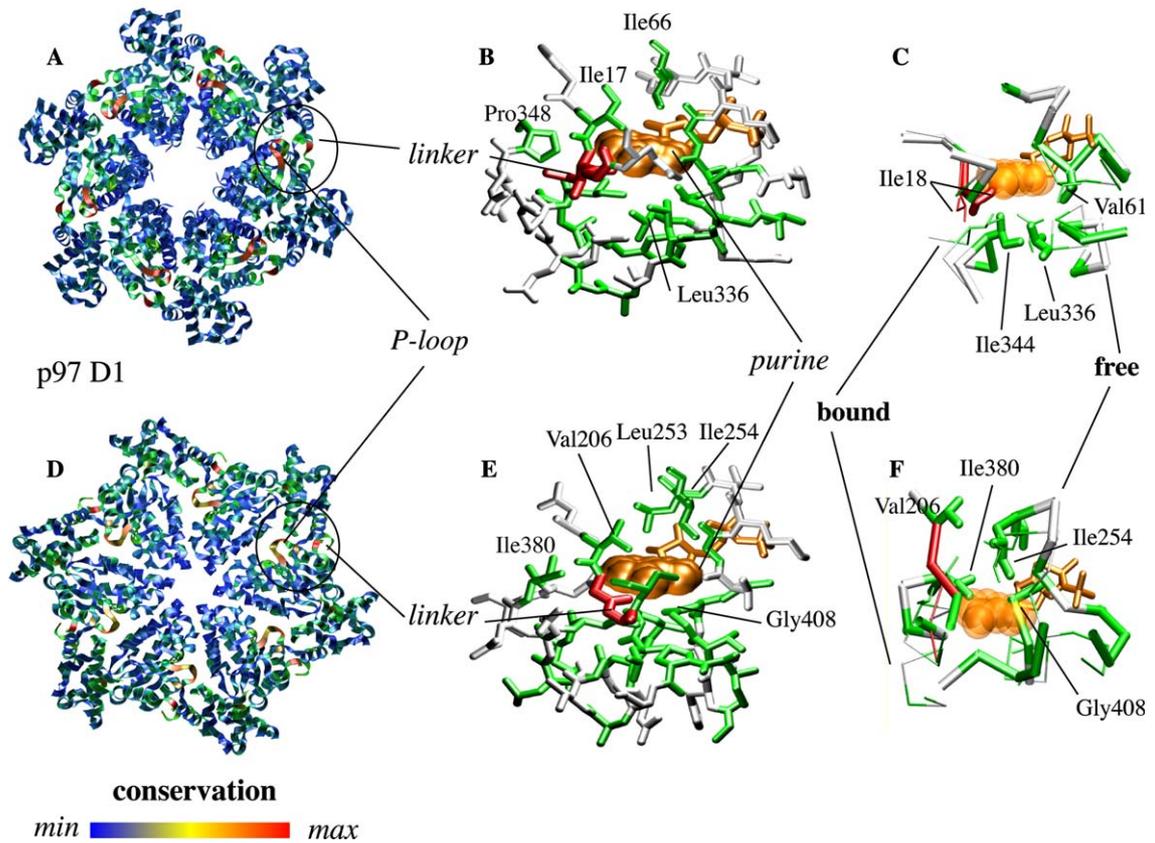
since the first Gly also makes contact to the adenine group of the nucleotide and the second Gly, as for HslU, is also in a disallowed region of the Ramachandran plot for a non-Gly residue (94°, −66°); however, the presence of a different sequence points to the likelihood of differences in the details of the mechanism. There is no nucleotide free structure to compare in this case: however, in Fig. 2F we show the superposition of the X-ray structure (thin lines) and a conformation derived from an MD simulation (thick lines). (How the MD structure is arrived at will be described later.) First, to discuss the X-ray structure alone (Fig. 2E), several hydrophobic contacts made by residues inside the chamber from both subdomains suggest that binding could change this hydrophobic network. In particular, Val206 (residue next to N-linker) forms a hydrophobic interaction with Leu253 and Ile254 (α-helix 6), while the side chain of Ile383 (α-helical subdomain) is in close vicinity. Therefore, the N-linker connects both subdomains via hydrophobic contacts around the adenine, in a similar way to HslU. In the MD conformation, the N-linker has moved upwards while maintaining these hydrophobic contacts, while residues from the C-terminal domain around Gly408 in the bridge have collapsed into the nucleotide binding pocket.

The N-linker of the D2 representative NSF has a "Trp–Gly" sequence. Based on the sequence similarity alone, we first thought the packing of Trp could be similar to the equivalent Ile in HslU, pointing away from the nucleotide in the bound state. However, inspection of the binding pocket shows that the Trp in fact makes contact with the adenine. Interestingly, it also makes hydrophobic contacts with a previously identified highly conserved Ile, Ile670 (Neuwald et al., 1999). Nevertheless the hydrophobic pocket is similar to both HslU and p97 but with the Trp side chain preventing direct contact to α-helix 10 and a small region, just before the N-linker, also making contacts with the adenine.

2.3. The N-linker signatures are well conserved through the family

The above structural comparison suggests an important role for the N-linker in AAA+ proteins. To investigate how the N-linkers of HslU on one hand, and p97 D1 and NSF D2 on the other, are representative of the AAA+ family in general, two multiple sequence alignments were constructed using ~800 members of this family in the PFAM database (Bateman et al., 2000), separating type I and II proteins. In this alignment the N-linker region emerged as a differential signature (shown in a simplified form in Fig. 3A), permitting a natural split of the AAA+ family into two groups. The class of proteins represented by p97 has a well-conserved "Gly–Gly" N-linker region, while the other, represented by HslU, has an alternative "Hydrophobic-Gly" pattern. The existence of the two types of N-linker region was confirmed using PSI-BLAST (see supplementary materials).

Since NSF D1 has high sequence similarity to p97 D1, including the Gly–Gly N-linker sequence, NSF-like proteins have been clustered together in the p97 group (total of 71 sequences). However, NSF D2 and its orthologues have a non-Gly residue in the first position of the N-linker. This is perhaps not surprising since NSF D2 shows clear differences in many of the other conserved features within the AAA+ cassette, such as the P-loop and the AAA minimum consensus for sequence alignments of the family (Neuwald et al., 1999). Indeed, as shown in supplementary materials, the D2 N-linker is generally less conserved than D1, suggesting different roles for them (Hattendorf and Lindquist, 2002). This may be reflected in the different catalytic properties and low affinity for nucleotide that have been observed for NSF D2 (Matveeva et al., 1997).

As previously reported (Neuwald et al., 1999), the nearly identical position of the Walker motifs in these proteins suggest a monophyletic origin, and our analysis agrees on this point. But we also found that the N-linker region shows a primitive dichotomy in the family: "Gly–Gly" and "Hydrophobic-Gly" groups. To test if this was a consequence of our manually corrected alignments, sequence profiles for type I AAA+ proteins (94 representative sequences), type II D1 (114), and type II D2 (114) were constructed. With these profiles, sequence similarity searches were performed against the NCBI non-redundant protein database using the program PSI-BLAST (Schaffer et al., 2001). New sequence profiles
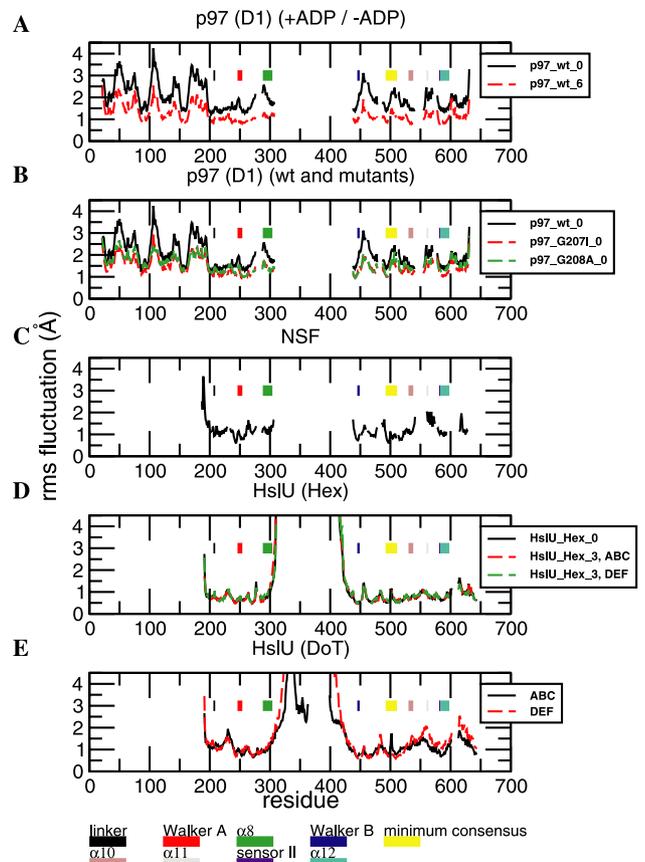
Fig. 1. Structural comparison of representative AAA+ domains. Cartoon representation of p97 D1 (red), NSF D2 (green), and HslU (blue). Nucleotide, ADP for p97, is shown in gold. Some of the AAA+ family motifs are labeled (Neuwald et al., 1999; Zhang et al., 2000).

Fig. 2. Sequence conservation and adenine binding site geometry. (A) Sequence conservation mapped to the hexameric structure of HslU. N-linker and P-loop are clearly highly conserved, see color bar. (B) The N-linker region of the hydrophobic binding pocket is colored red, all other hydrophobic residues are colored green. Key hydrophobic residues from both subdomains are labeled (see text). (C) Superposition of the "free" and "bound" ATP binding sites from the X-ray structure of HslU (Bochtler et al., 2000). Thin bonds represent the bound state, thicker bonds the free state. Hydrophobic residues are colored green, all other residue types are white. Ile18 can be clearly seen to penetrate the adenine ring (gold) of the bound state. The reduction in the volume of the binding pocket between bound and unbound states, (see Section 4 for how the volumes are calculated), is 16%. (D) Sequence conservation mapped onto the hexameric structure of p97. (E) Hydrophobic binding pocket. Coloring scheme same as for (B) above. (F) Superimposition of X-ray ADP-bound structure and projected structure derived from MD simulation. Thin bonds represent the crystallographic structure, thicker bonds the simulation. Coloring scheme as for (C) above. The figure was made with VMD (Humphrey et al., 1996).
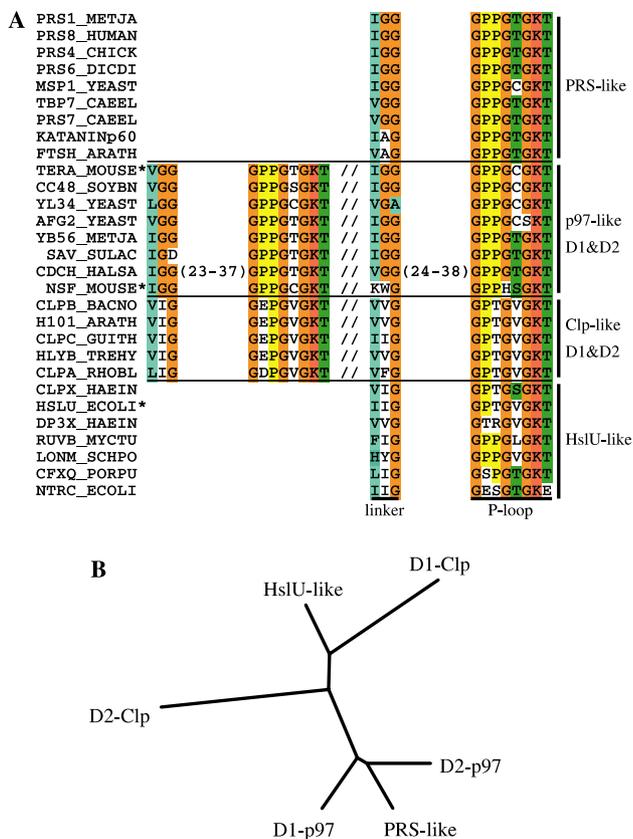
Fig. 3. N-linker sequence conservation and primitive phylogenetic tree. (A) Multiple sequence alignment of N-linker regions and P-loop from representative AAA+ proteins. Residues are colored by similarity. Four groups are labeled according to their overall sequence similarity (complete AAA+ sequences): PRS-like, p97-like, Clp-like and HslU-like. PRS-like and HslU-like have just a single AAA+ cassette (type I), while p97-like and Clp-like have two consecutive AAA+ cassettes (type II). Double forward slashes indicate the break between consecutive AAA+ cassettes. Note that PRS-like and p97-like proteins cluster together, including the N-linker motif, Gly–Gly. The same is true for Clp-like and HslU-like proteins but now the N-linker has a hydrophobic residue in place of the first Gly. The residue before the conserved N-linker pattern is a hydrophobic residue that is part of the hydrophobic cluster around the adenine ring (see text). The second residue of the N-linker "Gly" is highly conserved (90%) across the whole family. The first tends to be conserved for p97-like proteins (80%), but less (50%) for PRS-like proteins, where it is often (32%) substituted by Ala. D1 and D2 show different levels of conservation. NSF is a good example, with a Trp in place of Gly. (B) The sequence similarity based tree suggests gene duplication events (see text). Note that the N-linker distribution correlates with the overall branching. The full sequence alignments are available in the supplementary material.

Fig. 4. Local fluctuations: The root mean square fluctuations per residue during each 1 ns simulation. They are calculated for the Cα atom of each residue, from the standard deviation of its position after least-squares superposition (on the entire hexamer) of all the coordinate sets saved from the trajectory. These fluctuations are then averaged over corresponding residues in the six protomers, or each set of three symmetry-related protomers in those simulations (HslU_Hex_3 and HslU_DoT_0) which had 3-fold symmetry. (A) p97_wt_0 (empty of nucleotide) and p97_wt_6 (with 6 ADP); (B) p97_wt_0 and N-linker mutants p97_G207I_0 and p97_G208A_0; (C) NSF; (D) HslU HslU_Hex_0 (empty of nucleotide) and HslU_Hex_3 (3 ADP) (the ADP are present in chains A, B, and C); (E) HslU HslU_DoT_0 (in the crystal structure chains A, B and C contain nucleotide while D, E and F are unliganded, though the simulation contained no nucleotide). The discontinuities in the rmsf plot mostly reflect the insertions and deletions in the sequence alignment required to bring corresponding regions into registration (residue numbers on the X-axis alter accordingly): the polypeptide chain is, in each simulation, continuous. The only exception is HslU, where there were missing residues in the X-ray density of both 1g41 and 1do2: these were treated as described in Section 4. The missing residues, which are at the end of an exposed loop in the structure, are responsible for the high rmsf in the simulation just before and after. The colored bars show the locations of certain sequence features of interest, as marked on the legend.

were then automatically generated by the program with all the confident homologues found (1195, 1429, and 1367, respectively). Interestingly, only two strongly conserved sequence patterns arose from these profiles: Walker A and the two kinds of N-linker; this is shown in Figs. 2A and D which presents the information content, i.e., the relative conservation, of these profiles mapped over the three-dimensional structures of the HslU and

p97 D1. The NSF D2 map is very similar to p97 D1 (data not shown).

It is interesting also to compare the division on the basis of the N-linker with that on the basis of the Arg-finger in the SRH (Karata et al., 1999). It is suggested (Lupas and Martin, 2002) that the SRH is missing in those AAA+ proteins lacking the Arg-finger. However,

there are still strong similarities between the classical SRH and the sequence of the HslU-like AAA+ proteins in this region, which become particularly pronounced in structural alignments. At the position in the SRH corresponding to the first Arg, the HslU-like class has a 100% conserved Gln, although this Gln, unlike the Arg, does not directly contact the nucleotide in the adjacent protomer; rather, it forms a water-bridged hydrogen bond to an Arg in that protomer. Both contain another 100% conserved Arg three or four residues C-terminal of the Gln or first Arg, and both have a series of charged residues N-terminal to it.

There are, therefore, at least two possible ways of classifying AAA+ proteins: either by the SRH motif or by the N-linker motif defined here. The two criteria seem to correlate strongly: the SRH motif containing Arg–Arg is present in 80% of confident homologues of p97-D1 and 90% of those of p97-D2, while the Gln–Arg SRH motif is present in 100% of confident homologues of HslU. At the same time, the N-linker of type Gly–Gly is present in 90% of confident homologues of p97-D1, and 70% of those of p97-D2, whereas the type hydrophobic-Gly is present in 90% of confident homologues of HslU. This is consistent with two different although related roles for the two motifs, the SRH being more concerned with nucleotide *hydrolysis* and the N-linker with nucleotide *binding* and inter-subdomain motion.

2.4. A primitive evolutionary tree for the AAA+ family

Using the N-linker sequence pattern, a total number of six groups can be proposed from the type I and II multiple alignments mentioned above, combining the two types of N-linker with the three possible AAA cassettes, single, D1 and D2. The groups are named HslU-like, D1-Clp, D2-Clp, PRS-like, D1-p97, and D2-p97, according to the proteins they contain (see Fig. 3). To test if this family grouping has any evolutionary meaning, we constructed a phylogenetic tree based on sequence identity over the whole AAA+ cassette. The tree (Fig. 3B) shows that D1-p97 and D2-p97 cluster closely together with the PRS-like group, suggesting that they may have evolved through gene duplication of PRS-like proteins. Similarly, D1-Clp and D2-Clp cluster with HslU-like proteins. The comparatively longer distances reflect the higher sequence diversity of the HslU-like group, but they seem to be as well-duplicated genes.

Since sequence identity, one possible metric, was used to build the tree, the relative distances between groups could change if a different metric or algorithm was chosen (Schirmer et al., 1996). Nevertheless, it seems clear that the groups cluster in two well-resolved subfamilies, PRS-like and HslU-like.

2.5. MD analysis gives insights into domain movements upon nucleotide binding and release

MD studies have been performed on p97 D1, NSF D2, and HslU hexamers. In the case of HslU the simulations were performed without the I domains, which are not well resolved in the X-ray crystal structures.

The simulations that have been carried out are shown in Table 1. In most simulations, the nucleotide was initially removed, and, since the simulations were run free of any external forces, can be considered to show at least the initial stages of the spontaneous relaxation from "stressed" (protein in nucleotide bound conformation) to "relaxed" (nucleotide free) states. In the case of p97, simulations were initiated from the pdb structure 1e32 (Zhang et al., 2000). For this protein, control simulations have also been carried out

Table 1
Conditions of the various MD simulations, indicating the PDB files from which the initial protein conformation was taken, any mutations made to the protein, and the nucleotide occupancy in the protein as crystallized and in the simulations

AAA+	Simulation	PDB code	Nucleotide state (PDB)	Nucleotide state (simulation)	N-linker (PDB)	N-linker (simulation)
p97	p97_wt_0	1e32 (Zhang et al., 2000)	6 ADP	0	VGG	VGG
	p97_wt_6			6 ADP		
	p97_G207I_0			0		VIG
	p97_G208A_0			0		VGA
NSF	nsf_0	1nsf (Lenzen et al., 1998)	6 ATP	0	KWG	KWG
HslU	HslU_Hex_0	1g41 (Trame and McKay, 2001)	6 ADP	0	IIG	IIG
	HslU_Hex_3			3 ADP[a]		
	HslU_DoT_0	1do2 (Bochtler et al., 2000)	3 AMP-PNP[a]	0		

[a] These molecules have 3-fold symmetry. Three nucleotides are present, lying in alternate subunits, chains lettered AB and C (in all simulations, the chains are lettered A–F and the angular order is AFBDCE).
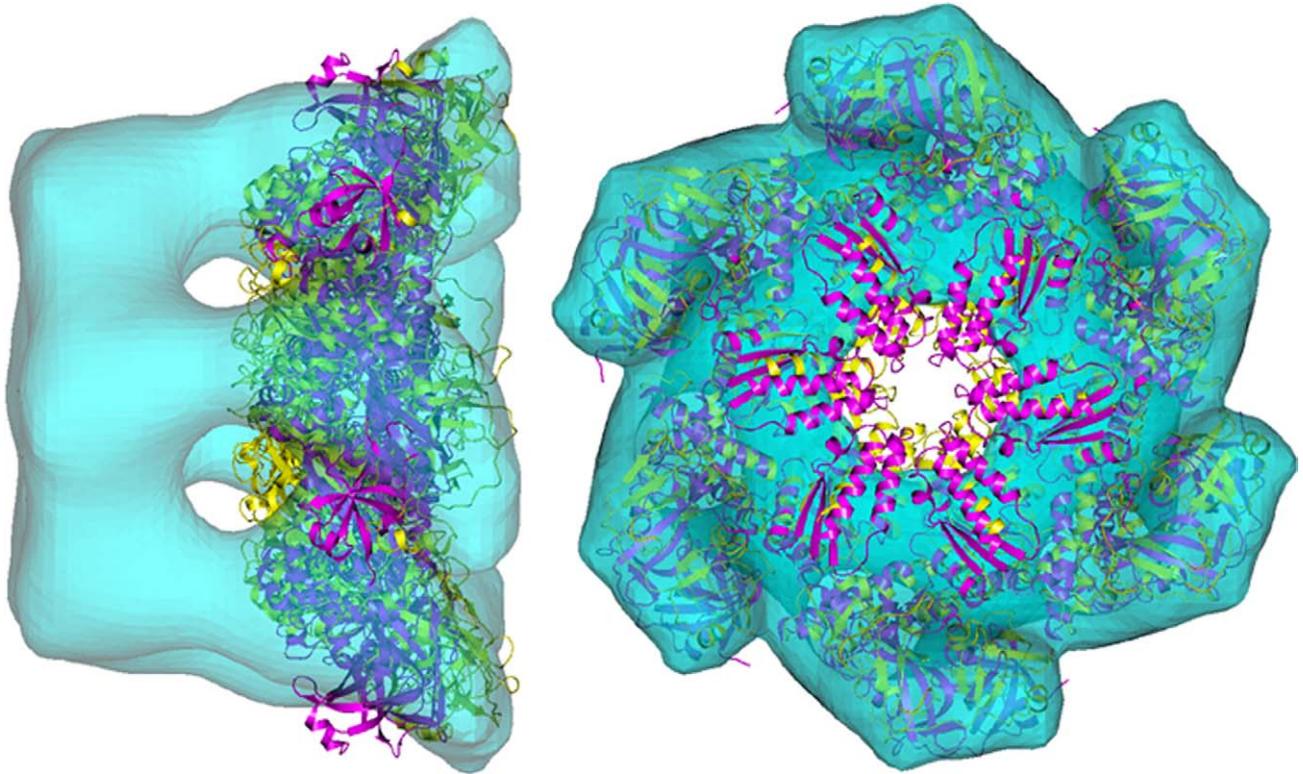
Fig. 5. X-ray crystal structure (magenta) and projected MD structure (yellow) for p97, both superimposed onto cryo-EM electron density (transparent cyan) in the region thought to be occupied by the D1 ring. Plan and side elevations are shown. The density corresponding to the D2 ring is left unoccupied.

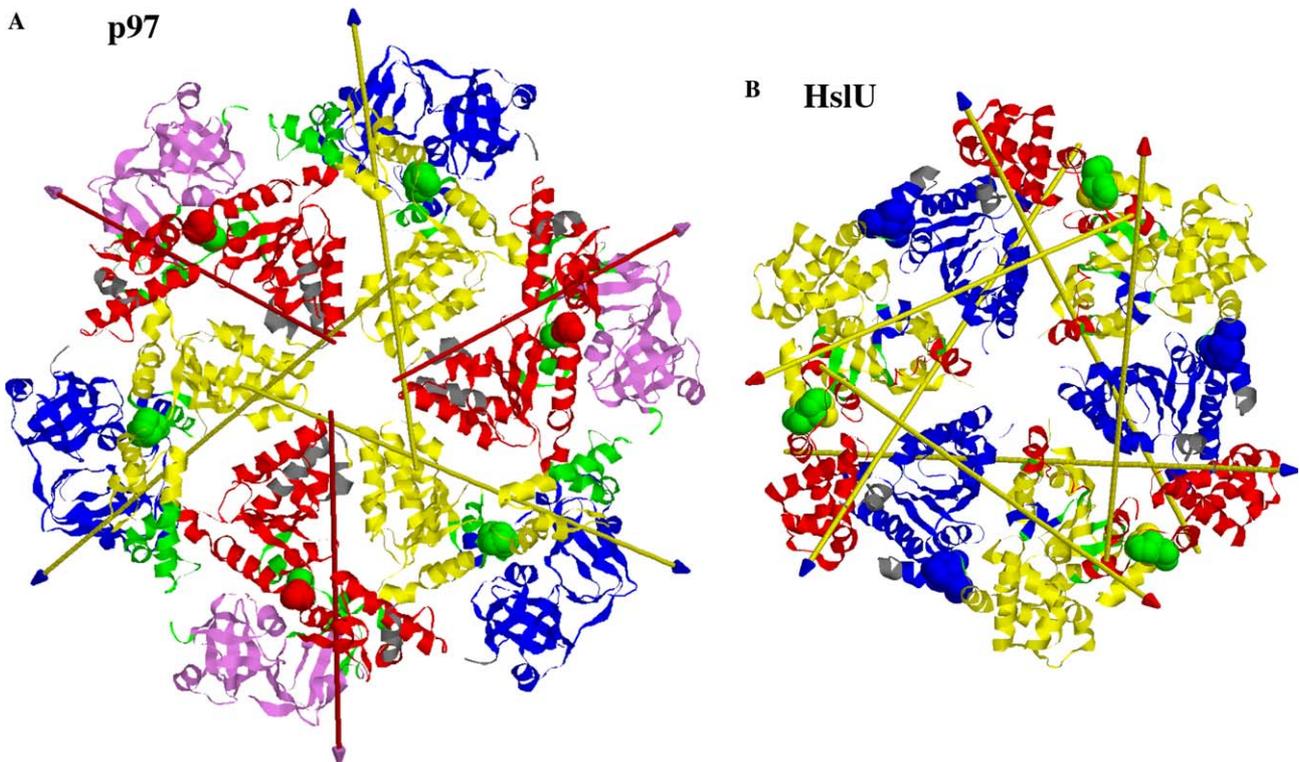Fig. 6. Dynamic domains (red, yellow, blue, and pink) and bending/hinge residues (green) identified by DynDom for p97 (A) and HslU (B). Arrows show axes of domain movements; colors of arrowhead/shaft indicate the domains to which the arrow corresponds. A single protomer is outlined in black for each protein to facilitate comparison of the protomer boundaries with the boundaries of the dynamic domains.

(a) with all nucleotides present (p97_wt_6) and (b) with mutations in the N-linker (p97_G207I_0 and p97_G208A_0); these make the p97 N-linker similar to that of HslU. In the case of HslU, there are X-ray crystal structures in different levels of ligand occupancy: we have performed simulations starting from the structure of pdb 1g41 (Trame and McKay, 2001), which is a symmetric hexamer, and 1do2 (Bochtler et al., 2000), a "dimer-of-trimers." The hexameric structure of 1g41 contains 6 ADP, whereas the simulations were performed with three (HslU_Hex_3) or none (HslU_Hex_0); and the crystallographic structure of 1do2, the dimer-of-timers, contains 3 ADP, whereas the simulation (HslU_DoT_0) was performed with none. Thus the extent to which the hexameric and dimer-of-trimer conformations interconvert can be investigated, as well as the domain movements within and between protomers. Where three nucleotides are present, they are in alternate protomers.

As outlined in Section 1, we begin with a residue-level analysis of the nucleotide binding site (its volume and N-linker backbone angles) to investigate how the various N-linker sequences might be required for binding. We then proceed to the analysis of global fluctuations in the fold, which display features corresponding both to local motifs and to domain movements. Lastly we analyze domain and subdomain movements, so that these may be compared and contrasted in view of the type of N-linker present. In particular, we compare the geometric movements between the beginning and the end of the simulations, and analyze two properties based on essential dynamics, one indicating the type of cooperativity and the other possible movements beyond the extremes seen in these simulations.

2.6. Volume of binding pocket

For all three AAA+ proteins, removal of nucleotide induces collapse of conserved hydrophobic residues (Fig. 2, right and Neuwald et al., 1999) surrounding the nucleotide binding pocket, reducing the volume previously occupied by the adenine ring. We quantify the volume of the hydrophobic part of the binding pocket from the number of water molecules it contains. The ATP binding pocket has a largely enclosed volume around the adenine ring, but this is continuous with the region highly accessible to solvent in the vicinity of the phosphate tail, which eventually opens out into the cleft between subunits and from there into the bulk medium. Therefore, Voronoi polyhedra cannot be used to define the binding pocket uniquely; hence our analysis based upon the number of nearby waters.

For p97, the initial volume of the site (i.e., in the X-ray structure with nucleotide removed) is $214 \pm 34\,\text{Å}^3$ per protomer (calculated by the end of equilibration but before the 1 ns production part of the simulation, see

Section 4). After 1 ns this has decreased to $189 \pm 40\,\text{Å}^3$. In HslU the corresponding data are 239 ± 37 and $164 \pm 49\,\text{Å}^3$, while for NSF they are 104 ± 72 and $39 \pm 36\,\text{Å}^3$. So, there is a consistent trend of volume reduction upon nucleotide removal, or by symmetry, a gain upon binding.

For HslU, much of this volume change comes from the previously mentioned tendency of the N-linker to collapse into the nucleotide binding pocket, with associated readjustments of the local ϕ, ψ angles, as described above; this explains why the change is smaller for p97 (Gly–Gly linker) than HslU (Ile–Gly linker). Interestingly, NSF, the protein with the rare sequence pattern within the N-linker, shows the greatest fractional volume change (though HslU is slightly larger in absolute terms). Inspection of the coordinates after 1 ns shows much of the reduction in volume comes predominantly from a readjustment of the region just before the WG motif, that winds around the adenine ring. There is no equivalent to this region in either HslU or p97 and we believe that this region only makes contact with the adenine ring as a consequence of NSF not fitting the general sequence profile for a type II D2 structure. Moreover, this could be the reason for the low affinity for ADP (Matveeva et al., 1997), since the highly collapsed binding pocket in NSF is more favorable for ADP release than for either p97 or HslU.

2.7. N-linker backbone angles

As the N-linker forms part of this cavity, we now proceed to investigate the detailed behavior of the backbone angles in this region during the MD simulations. The overall changes in the ϕ, ψ torsion angles between the X-ray structures and the final set of coordinates after 1 ns show some interesting trends that, in the case of HslU may be directly compared with the changes in N-linker angles between the symmetric hexamer and dimer-of-trimers discussed in Section 2.1. In HslU_Hex_0, two of the six N-linker ϕ, ψ angles for Gly moved from the disallowed region for a non-Gly residue to the allowed ϕ, ψ region, similar to the changes observed in the X-ray structure of the bound and unbound nucleotide state of HslU, as discussed in the previous section. These are correlated with movements of Ile18 into the nucleotide binding site. The other four ϕ, ψ Gly regions moved towards the allowed region but remained in the disallowed region. In contrast the ϕ, ψ regions for both Gly residues in the N-linker of p97 showed little movement from their original positions, indicating a more subtle transmission of the subdomain reorientations through this linker. For NSF, the Gly residue within the linker also sometimes shows transitions between low and high energy regions of the Ramachandran plot, this time between allowed

to non-allowed regions; and also substantial ϕ, ψ changes are observed in the residues preceding the N-linker, residues that wind around the adenine ring. These motions, which tend to visit regions of the Ramachandran plot accessible only to glycine, thus establish a possible reason for conservation of this residue in the N-linkers.

2.8. Fluctuations

We now consider the local root mean square fluctuations (rmsf) produced in the MD simulations (Fig. 4). The first point to note is that, despite the low sequence similarity (sequence identity ∼20%), the local maxima and minima of the rmsf show a remarkably similar pattern during the MD simulations, and the sequence motifs, shown as colored bars in the figure, tend to behave in the same way in all the simulations. The ATP binding regions, that include the N-linker, show the least degree of local motion, further suggesting a pivoting role for this region. There is notably a large jump in rmsf for p97 in the boundary between the N-terminal domain (residues 1–204) and the α/β subdomain in the region of the Gly–Gly N-linker (Figs. 4A–B). This is a consequence of large movements of the N-terminal domains relative to the hexamer, thus supporting the hypothesis that nucleotide binding/release could induce N domain relocation that pivots around the conserved Gly residues (Zhang et al., 2000). In NSF D2 and HslU the observed fluctuations have very similar trends, despite some differences in the magnitude of the movements (Figs. 4C–E). One of the largest differences is around α-helix 8 in HslU, where the I domain is inserted; the presence of the linker to this domain produces very large local fluctuations, so it is not comparable either to p97 D1 or NSF D2. Comparing the different HslU simulations, the fluctuations in those simulations started from the symmetric hexamer 1g41 (Fig. 4D) are generally smaller than those started from the dimer-of-trimers (Fig. 4E), reflecting the tighter packing in the symmetric hexamer. In the symmetric hexamer simulations, both HslU_Hex_0 and HslU_Hex_3 have almost identical patterns of fluctuation; neither do the ABC chains of HslU_Hex_3 (where ADP is present) differ appreciably from the DEF chains (no nucleotide). On the MD timescale, therefore, no tendency is observed for the symmetric hexamer to convert to the dimer-of-trimers. Similarly, the pattern of fluctuation in the ABC chains of HslU_DoT_0 (nucleotide present in the crystal structure, though not in the simulation) and the DEF chains is very similar.

In the p97_wt_6 simulation, (Fig. 4A, red line), the same local displacements as p97_wt_0 are in general observed, as expected, identifying the same flexible and pivoting regions, but with an overall fluctuation 1.6 times smaller, seemingly a result of smaller movements

between the domains. This result supports the tense-to-relaxed transition observed, as being a natural behavior of the hexamer. The same is true for the simulations in which mutations were introduced into the N-linkers, G207I and G208A (Fig. 4B): the effect is to reduce the magnitude of the fluctuations, especially to the N-terminal domain, while the pattern remains the same. In a much longer simulation it might be hoped that clear differences with p97_wt_0 would emerge, but they were not seen in the current studies.

2.9. Subdomain movements

We now quantify global changes occurring during the 1 ns duration of the simulation, using distances and angles measured between the centers of mass of the subdomains and relative to the least-squares plane fitted through the hexamer. The subdomains, α/β, α-helical and (for p97) N, are shown in Fig. 1 and defined there and in Section 1. The results are shown in detail in Table 2. For p97_wt_0, all six N-terminal domains move on average 9 Å away from the plane of the ring. The six α-helical subdomains, on the opposite side of the plane, also move in a concerted fashion but towards the plane, approx. 5 Å, maintaining hydrophobic contacts with the N-domains. They also move radially outwards by about 2.3 Å. The α/β subdomains show extremely small displacements as also does the aperture about the 6-fold axis. As remarked in the previous section, the N-linker itself shows small fluctuations, but the displacement between the start and the end of the 1 ns simulation period increases from the nucleotide binding site (residue 206) to the insertion point in the N-terminal domain (around residue 195). This study identifies another pivoting point, the bridge between the α/β and the α-helical subdomain, also identified in Beuron et al. (2003), agreeing with the argument that nucleotide binding alone can induce substantial subdomain arrangement. As remarked when discussing the fluctuations, the subdomain movements of p97_wt_6 and the two N-linker mutants are quite similar to p97_wt_0, except of smaller magnitude. The only difference is in $D\phi$ for the N-linker mutants, but the magnitude of the difference is in any case smaller than the standard deviation between protomers. Similar subdomain movements were observed for NSF (approx. 3 Å movement of the α-helical domains towards the plane), supporting the observation that concerted movement is not solely due to the additional N terminal domain in p97. For HslU only very small subdomain movements are observed; few of the movements are larger than the standard deviation between protomers and there is no tendency to observe larger movements in those protomers that do not have the same nucleotide occupancy in the simulations as in the crystal structures (HslU_Hex_0, HslU_Hex_0

Table 2
Movements of the centers of mass of the domains in the MD simulations relative to the corresponding crystal structure

AAA	Domain		Dr	Dx	$D\phi$	$D\theta$
p97_wt_6	N		−0.6(0.5)	−4.6(1.6)	−0.1(2.0)	4.2(1.4)
p97_wt_0	N		−2.5(0.9)	−8.7(2.3)	−0.1(3.3)	8.1(2.0)
p97_G207I_0	N		−1.8(1.7)	−5.0(2.3)	0.7(2.1)	4.8(2.2)
p97_G208A_0	N		−0.9(0.7)	−4.4(1.4)	0.8(3.2)	4.0(1.2)
Hslu_Hex_0	α		−0.0(0.2)	−0.0(0.7)	0.0(0.6)	0.1(1.0)
HslU_Hex_3	α	ABC	−0.3(0.5)	0.7(1.1)	0.2(0.3)	−0.9(1.4)
		DEF	−0.4(0.3)	0.1(1.0)	−0.7(0.3)	−0.2(1.2)
HslU_DoT_0	α	ABC	−0.0(0.5)	0.1(1.4)	1.2(2.1)	−0.1(1.6)
		DEF	0.5(0.5)	−2.4(1.3)	−0.7(2.0)	3.0(1.5)
p97_wt_6	α		1.0(0.6)	−2.4(1.5)	−0.4(1.2)	2.9(1.7)
p97_wt_0	α		2.3(1.7)	−4.6(2.0)	−0.2(2.4)	5.3(2.2)
p97_G207I_0	α		0.9(0.7)	−3.0(1.2)	0.3(1.9)	3.4(1.3)
p97_G208A_0	α		1.1(0.9)	−2.4(2.3)	0.1(2.3)	2.8(2.4)
nsf_0	α		1.1(0.7)	−3.2(0.5)	−0.5(1.5)	4.0(0.7)
HslU_Hex_0	α/β		0.3(0.3)	−0.0(0.4)	−0.1(0.4)	0.0(0.8)
HslU_Hex_3	α/β	ABC	0.1(0.2)	0.1(0.8)	0.5(1.0)	−0.2(1.5)
		DEF	0.4(0.2)	−0.1(0.2)	−0.5(0.4)	0.2(0.4)
HslU_DoT_0	α/β	ABC	0.1(0.7)	−0.4(1.1)	−0.1(1.1)	0.8(2.2)
		DEF	0.0(1.0)	0.4(0.3)	0.8(0.9)	−0.8(0.4)
p97_wt_6	α/β		0.4(0.3)	−0.0(1.0)	0.1(0.7)	0.0(1.8)
p97_wt_0	α/β		0.7(0.5)	−0.0(1.0)	−0.5(0.9)	−0.0(1.9)
p97_G207I_0	α/β		0.1(0.2)	−0.0(0.7)	−0.2(0.5)	−0.0(1.3)
p97_G208A_0	α/β		0.5(0.5)	−0.0(1.4)	−0.5(1.3)	0.0(2.6)
nsf_0	α/β		−0.2(0.5)	−0.0(0.6)	−0.4(1.1)	−0.0(1.1)

This is calculated for the N-terminal domain, α-helical subdomain, and α/β subdomain of p97 and for the α-helical and αβ subdomains of HslU and NSF. The X-ray hexamer structure and the MD structure from the end of the trajectory ($t = 1$ ns) are first least-squares fitted on all αβ subdomains. All displacements are calculated relative to the αβ subdomain of the same protomer in the same structure, before taking the difference between the X-ray structure and the MD structure. All displacements are then averaged over the six protomers, or each set of three symmetry-related protomers in those simulations (HslU_Hex_3 and HslU_DoT_0) which had 3-fold symmetry (in which case ABC or DEF is indicated, as appropriate; otherwise all six chains were averaged); the mean and standard deviation are shown. The movements calculated are:

Dr, displacement of the domain in the radial direction (in the plane of the hexamer, outwards from the center).

Dx, displacement of the domain perpendicular to the plane of the hexamer.

$D\phi$, angular displacement of the domain tangentially within the hexamer (i.e., about an axis perpendicular to the plane of the hexamer through its center).

$D\theta$, angular displacement of the domain out of the plane of the hexamer (i.e., about an axis in the plane of the hexamer through its center and roughly perpendicular to the vector from the center of the hexamer to the center of the domain).

(DEF chains), and HslU_DoT_0 (ABC chains)) than those that do. It seems, then, that there is little tendency for directed conformational changes to occur on a short timescale in the HslU simulations, though as we shall see information may still come from the fluctuations.

In view of the similarity of the various p97 simulations, and the various HslU simulations, we will concentrate in the rest of this section on the simulations p97_wt_0, (which reflects a possible empty state of the D1 ring), and HslU_Hex_3 (since the others are strictly not biologically relevant, given that HslU requires nucleotide to form hexamers).

2.10. Cooperativity of domain motions between subunits

We have applied essential dynamics (ED) (Amadei et al., 1993; Garcia, 1992) to the MD trajectory (see Sections 1 and 4) with the aim of removing the details of atomic fluctuations and reveals large-scale motions.

Here, we have additionally chosen to average the principal components in a way reflecting the symmetry of the molecule (i.e., 6-fold for p97_wt_0, 3-fold for HslU_Hex_3), to remove noise remaining from the simulation and ensure that any resulting structures also have that symmetry. The program DynDom (Hayward and Berendsen, 1998; Hayward and Lee, 2002) was then used on conformations along the first principal component of the ED analysis to define "dynamic domains" within the protein: regions which move as quasi-rigid units. (By "first" principal component is meant the component associated with the largest eigenvalue.) These may, but need not, correspond to domains defined on the basis of structure or sequence. The results are shown in Fig. 6: the dynamic domains identified are colored red, yellow, blue, and pink. For p97, they are confined within protomers, indicating the independence of the protomers. The basic movement is a bending upwards of the N-terminal domains relative to the αβ and α subdomains along an almost radial axis.

The residues (green), which are identified as hinges for the bending, are found to include the N-linker (space-filled). In HslU, on the other hand, the axes are more tangential and some of the dynamic domains extend *between* protomers (the yellow domains), indicating that a greater degree of cooperativity may be expected, consistent with the proposed sequential mechanism of nucleotide binding, hydrolysis, and release. Once again, the N-linker is identified as part of the hinge (spacefilled green), underlining its importance in domain movements, this time between domains in different protomers.

2.11. Essential dynamics and comparison with cryo-EM

Although no X-ray structure is available for a differently liganded form of p97 (in contrast to HslU), cryo-EM studies on p97 containing both D1 and D2 rings have been performed, in which the top (D1) ring is thought to contain either no nucleotide or ADP (Beuron et al., 2003; Rouiller et al., 2000). However, the existing p97 X-ray crystal structure does not fit perfectly into the electron density: the N-terminal domains appear to be raised out of the plane of the α/β-subdomains (Beuron et al., 2003). To explain this, an elastic network model (Ming et al., 2002) was applied to the crystal structure and successfully modeled certain aspects of the N-terminal movement. The apparent movement of the N-terminal domains is also similar to that seen in the simulations of p97 conducted here, especially p97_wt_0. However, they have moved rather further than they do in the simulations (Table 2). Therefore, we have extrapolated the conformation along the direction described by a principal component beyond the extremes that are sampled during the simulation. The result of projecting p97_wt_0 along the first symmetrized principal component (in the same direction as the domains move in the simulation) is shown in Fig. 5, fitted into the cryo-EM density of the entire D1–D2 p97. It is apparent that there is now appreciably better agreement with this density than is obtained by using the X-ray crystal structure, suggesting that the symmetrized ED analysis of the MD has successfully identified a biologically relevant motion, which in addition involves the pivoting of the N-domains about the N-linkers. Because of the extrapolation and symmetrization of the observed conformational changes, the physical accuracy of the resulting conformation is reliable only in so far as the hinge axis between domains is well defined by the simulation and does not change. Moreover, the conformations produced by extrapolation along a principal component are not as physically meaningful as those produced in the course of the simulation itself: some distortion of the protein structure occurs. Nevertheless the good agreement with the cryo-EM density suggests that the approximations work quite well in this case.

3. Conclusion

In this study a distinctive sub-classification of AAA+ proteins was achieved by reconsidering a short sequence motif that forms part of the ATP binding site, the N-linker. Two broad groups are thus defined: one represented by PRS and p97-like proteins and a "Gly–Gly" N-linker, and a second represented by HslU and Clp-like proteins, characterized by a "Hydrophobic-Gly" N-linker.

This N-linker is part of a group of hydrophobic residues that contact the adenine ring of the nucleotide. MD simulations suggest that there is a common hydrophobic expansion mechanism of this pocket to bind ATP and collapse to exclude ADP, and that this facilitates subdomain and, when present, N-terminal domain reorientation, without the need to consider hydrolysis explicitly. We found some differences between p97 and HslU in both subdomain reorientation and local conformational changes, particularly within the N-linker region (the readjustments of the local ϕ, ψ angles, particularly for the conserved Gly residues, and the tendency of Ile18 in HslU to point in towards the adenine ring binding cavity). It is possible that the two different N-linker sequence patterns may therefore be signatures for different modes of action in nucleotide binding/release. The DynDom analysis of the MD trajectory has provided evidence for the differential types of cooperativity that had previously been suggested on the basis of structural studies alone: an allosteric mode of binding and release of nucleotide in HslU (Bochtler et al., 2000) with close communication between adjacent protomers of the hexamer, partially initiated through the N-linker (which is identified as a hinge by DynDom); contrasting with a concerted motion in p97, where the motion (confined within each protomer) is consistent with the six ATP molecules binding and releasing simultaneously within each hexameric ring (Zhang et al., 2000). The ED analysis of the MD of p97 indicated that in this case the N-linker is important in allowing rearrangements of the N-domain, producing a conformation which appears to be close to that seen in cryo-EM. Moreover, analysis of backbone angles, domain movements and binding site volume in HslU indicates that the "relaxed" (collapsed) position of Ile18 in the N-linker upon nucleotide release could induce subdomain rearrangements to promote the "tense" position of Ile in the next protomer, allowing nucleotide binding.

Since the work described in this paper was completed, we learnt of the publication of an X-ray crystal structure of p97 consisting of a hexamer of the full length protein, containing both the D1 and D2 rings (DeLaBarre and Brunger, 2003; pdb code 1oz4). The authors of this paper believe that the D1 ring is permanently occupied with nucleotide whose binding is essential to the maintenance of the hexameric structure, and that only the D2

ring hydrolyzes ATP. This may conflict with the details of the mechanism proposed here. However, we are not considering ATP hydrolysis; rather subdomain motion on nucleotide binding/release. In addition, it is interesting to note that in our simulations with ADP present, the same pattern of fluctuations, although damped, are still present (Fig. 4A), pointing to this region as the pivotal location for subdomain motion. It may be that conformational changes, propagated in a very similar way to that we have described for the N-linker, occur in the D1–D2 linker. This linker, also identified and discussed in DeLaBarre and Brunger (2003), has the sequence IGG in p97 (residues 479–481), putting it also in the Gly–Gly class; and movements pivoting around it could move the entire D1 and N ring relative to the D2 ring. Such motions were also described by DeLaBarre and Brunger as likely to be crucial to the protein's function.

Analysis of the full p97 hexamer along the lines carried out here would also be desirable in future work; although, since the inclusion of the D2 ring would more than double the size of the system to be simulated when solvent is taken into account, it would be a demanding project indeed.

The N-linker pattern described here is a definitive signature within the AAA+ family and, indeed, is almost as conserved within the various AAA+ cassettes as the Walker A motif. However, as the Walker A motif shows little variation across the family, the N-linker is useful to make an evolutionary split in the still growing AAA+ family. It correlates with, but seems slightly different to, the presence or lack of a ''classical'' SRH with an Arg-finger: though as we have noted, the Arg-finger itself is perhaps more highly conserved in the AAA+ superfamily than has been believed hitherto.

Just why these linker patterns are so highly conserved through evolution, and whether they are indeed essential for the proteins to carry out their function, will require more X-ray structures and probably further developments in MD algorithms as well as an increase in computational power to establish. We would also welcome experimental mutagenesis studies of residues in the N-linkers, along the lines of the in silico mutagenesis that was attempted above. The analysis presented here suggests that they are involved in transmitting the fine details of ATP binding to distinctive subdomain movements and are, therefore, an indication of protein function.

4. Materials and methods

4.1. Multiple sequence alignments

Protein identifiers for the AAA+ family were extracted from the PFAM database (Bateman et al., 2000)

in May, 2001. Since the N-linker region is not included in their domain definition, we took the full sequence for each member of the family from the SWISS-PROT database (Bairoch and Apweiler, 2000) and constructed multiple sequence alignments with the ClustalX program (Thompson et al., 1997) using the default settings. Due to the difference in sequence length and identity across the AAA family, sequences had to be cut in some cases to get the common motifs aligned. Particularly difficult blocks of sequences were aligned using the profile-profile tool in ClustalX.

To check the quality of the multiple sequence alignment in the N-linker region, we followed the multiple structural superimposition of representative AAA+ proteins within the Protein Databank (Berman et al., 2000), with codes 1e32, 1do2, 1fnn, 1hqc, 1jr3 and 1nsf.

4.2. Conservation maps

Three multiple alignments of representative sequences for type I AAA+ proteins, for type II D1, and for type II D2 were constructed with ClustalX (Thompson et al., 1997) and manually corrected to put the N-linkers in register (94, 114, and 114 sequences, respectively). As reference crystal structures we put in type I HslU, in type II D1 p97, and in type II D2 NSF. Using these profiles as input (in CLUSTAL format), sequence similarity searches were performed against the NCBI non-redundant protein database (ftp://ftp.ncbi.nih.gov/blast/db, as from 19-1-2002, with 854 809 sequences) using the program PSI-BLAST (Schaffer et al., 2001) (version 2.2.1). Non-standard parameters: $-j2$ (two iterations), $-Q$ (to keep the position specific scoring matrix generated after the first iteration). This last file contains the information per residue in the aligned set of significant homologues found (1195, 1429, and 1367, respectively). Following Sander and Schneider (1991), we use this raw information content (in bits) as a measure of conservation. Re-scaling this numbers to the range [0, 99], we use them as temperature factors in the PDB coordinate files of the reference structure files to display them in space as in Figs. 2A and D. The full sequence alignments are available as supplementary material.

4.3. Phylogenetic tree

Based on the multiple sequence alignments, six groups were defined inside the AAA family (PRS-like, HslU-like, Clp-D1, Clp-D2, p97-D1, and p97-D2), each of them including 20 random representatives. All groups were compared one-to-one performing the resulting 20^2 pairwise global alignments and computing the average sequence identity. Sequence identity was calculated using the formula id = identities/((lengthA + lengthB)/2). With these data we filled a 6×6 diagonal matrix, converting identities to distances (distance = 1 − identity)

and re-scaling the averages to obtain zero distances in the diagonal. Finally a tree was calculated to represent this matrix with the FITCH program in the PHYLIP 3.5c package, without assuming an evolutionary clock. PHYLIP is a phylogeny inference package created by J. Felsenstein (Department of Genetics, University of Washington, Seattle) and is freely available at ⟨http://evolution.genetics.washington.edu/phylip.html⟩.

4.4. Molecular dynamics simulations

Molecular dynamics (MD) simulations of 1 ns duration were carried out for each of three AAA+ proteins in their hexameric state [p97 in ADP, 1e32 (Zhang et al., 2000); NSF in ATP analogue, 1nsf (Lenzen et al., 1998); HslU in ADP, 1g41 (Trame and McKay, 2001), and 1do2 (Bochtler et al., 2000), all hexamers regenerated from the crystal structures by appropriate symmetry operations]. The angular order of chains in the hexamer is AFBDCE: where three nucleotides are present, they are in alternating chains, i.e., chains A B and C. The simulations performed are described in Table 1. They were performed using the Molecular Dynamics (MD) package GROMACS versions 2.0 and 3.0 (Berendsen et al., 1995; Lindahl et al., 2001).

To set up each simulation the following procedure was followed. Crystal water molecules were retained. Nucleotides were generally removed; for those simulations where they were kept an Mg^{2+} ion was also associated with each nucleotide. Other non-protein groups (e.g., sulfate ions) were removed. Polar and aromatic hydrogens were added to the protein, and it was fully solvated with water (simple point charge) in a rectangular box. Na^+ and Cl^- ions were then added to mimic more closely physiological screening conditions. First, enough ions were added to give a salt concentration of 0.1 M, then extra positive or negative ions were added to make the overall simulation system electrically neutral. The force field used was Gromos 96 (Scott et al., 1999). Short-range electrostatic and Van der Waals interactions were cut off at 14 Å; long-range electrostatics were treated by a reaction field (Tironi et al., 1995). The simulation boxes were large enough that each molecule was separated from its periodic images by at least the short-range cut-off distance. The total number of atoms in the p97_wt_0 system was 198419, of which 26 196 were protein atoms, and the charge on the protein hexamer was −78 e. For NSF the corresponding data are: 116 250 atoms, 15 018 protein atoms, and protein charge +18 e; and for HslU: HslU_Hex_3 16 5771 atoms, 19 608 protein atoms, and protein charge −36 e. Missing residues in the X-ray density at the N-terminus were not modeled and otherwise require no special treatment; however, there were missing residues in the X-ray density in the middle of the polypeptide chain for both 1g41 (positions 87–92

and 120–226) and HslU_DoT_0 (positions 174–210). In the simulations, these were handled by using harmonic distance restraints to keep the ends of these "breaks" separated by (on average) the same distance as in the crystal.

Each simulation was run in parallel on four 866 MHz Intel Pentium III processors of a Linux farm. After building, each system was minimized, then heated to 300 K, and equilibrated for 80 ps with all heavy protein atoms restrained to their crystallographic coordinates by harmonic restraints, before the 1 ns production run. In the initial equilibration the MD integration timestep was 2 fs; but by using the dummy/heavy hydrogen features of GROMACS, it was possible to extend the timestep to 5 fs in the second equilibration and production stages of simulations without nucleotide. A 1 ns production run took approximately 1 month. Coordinates of all atoms in the system were saved every 10 ps from the production run, so 100 frames were obtained from each simulation.

4.5. Volume calculations of adenine ring binding site

The volume of the binding site was estimated by identifying residues that bind the adenine moiety of the nucleotide in the X-ray structure, then counting the numbers of water molecules within 10 Å of all these residues throughout the simulation. By multiplying the number by 29.9 Å^3, the volume of a water molecule, an estimate is obtained of the volume of the hydrophobic part of the binding pocket.

4.6. Essential dynamics

Essential dynamics (ED) was carried out using the g_covar and g_anaeig programs of GROMACS and the authors' own programs. ED is described in detail elsewhere (Amadei et al., 1993; Garcia, 1992). It is based on Principal Components Analysis of the atomic fluctuations in the trajectory and allows the extraction of large-scale motions, removing the instantaneous "flickerings" of bond angles, dihedrals and short chain segments that will be present in any particular instantaneous conformation. First all the coordinate frames of the trajectory are least-squares fitted onto the initial structure, using all Cα atoms of the protein hexamer. The covariance matrix is then constructed and diagonalized:

$$C_{ij} = \langle (s_i - \langle s_i \rangle)(s_j - \langle s_j \rangle) \rangle,$$

where $\{s_i\}$ are the x, y, and z coordinates of some subset of the atoms and the average is over all coordinate frames saved from the trajectory. The Cα atoms were used here; if more atoms are included the matrix become intractably large.

The eigenvectors (principal components) with largest eigenvalues tend to describe large-scale domain motions

of the molecule. Previous work on ED suggests that at least the "essential subspace" of the first few eigenvectors often converges quickly (Amadei et al., 1999). In the work reported here, the eigenvectors were symmetrized by 6-fold or 3-fold averaging, as appropriate to the molecule under consideration. Cα-only structures can be extrapolated along the eigenvectors beyond the extremes of fluctuation observed in the simulation, and full atomic models can be rebuilt from these by local fitting to the crystal structure in a 3-residue window (hence the original backbone ψ, ϕ and side chain χ angles tend to be retained). We remark that it is essential to have all heavy backbone atoms present before a DynDom analysis can be carried out (see below).

When projecting beyond the fluctuations observed in the simulation, for example to produce the model of p97 to fit in the cryo-EM data, it was found that projecting in a single step produced bond lengths and angles so far from normal that the structure could not be energy minimized. Therefore the projection proceeded in cycles, each of a small projection followed by a minimization.

4.7. Cooperativity in domain movement

Dynamic Domains were identified using DynDom (Hayward and Berendsen, 1998; Hayward and Lee, 2002) version 1.5, operating on the X-ray structure and a conformation projected along the lowest symmetrized principal component. There have been several applications of DynDom to analyze domain movements in proteins, using both multiple X-ray crystal structures and structures derived from ED on MD trajectories, for example (de Groot et al., 1998; Roccatano et al., 2001). Because DynDom builds up its domains progressively, a symmetrical arrangement of domains was not generated even with symmetrical input. Therefore the analysis was done for an adjacent pair of protomers and symmetrically replicated. The ratio of inter- to intra-domain fluctuation was 0.5 for HslU HslU_Hex_3, and 1.0 for p97 p97_wt_0; aside from this, default parameters were used.

Supplementary material

This consists of multiple sequence alignments, divided into the p97-like and HslU-like subfamilies, and is available at http://www.bmm.icnet.uk/supplementary/AAA/.

References

Amadei, A., Ceruso, M.A., Di Nola, A., 1999. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. Proteins 36, 419–424.

Amadei, A., Linssen, A.B.M., Berendsen, H.J.C., 1993. Essential dynamics of proteins. Proteins: Struct. Funct. Genet. 17, 412–425.

Bairoch, A., Apweiler, R., 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28, 45–48.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., Sonnhammer, E.L.L., 2000. The Pfam protein families database. Nucleic Acids Res. 28, 263–266.

Berendsen, H.J.C., van der Spoel, D., van Drunen, R., 1995. GROMACS: a message-passing parallel molecular dynamics implementation. Comp. Phys. Commun. 91, 43–56.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.

Beuron, F., Flynn, T.C., Ma, J., Kondo, H., Zhang, X., Freemont, P.S., 2003. Motions and negative cooperativity between p97 domains revealed by cryo-electron microscopy and quantised elastic deformational model. J. Mol. Biol. 327, 619–629.

Beyer, A., 1997. Sequence analysis of the AAA protein family. Protein Sci. 6, 2043–2058.

Bochtler, M., Hartmann, C., Song, H.K., Bourenkov, G.P., Bartunik, H.D., Huber, R., 2000. The structures of HsIU and the ATP-dependent protease HsIU-HsIV. Nature 403, 800–805.

de Groot, B.L., Hayward, S., van Aalten, D.M., Amadei, A., Berendsen, H.J., 1998. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. Proteins 31, 116–127.

DeLaBarre, B., Brunger, A.T., 2003. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. Nat. Struct. Biol. 10, 856–863.

Garcia, A.E., 1992. Large-amplitude nonlinear motions in proteins. Phys. Rev. Lett. 68, 2696–2699.

Hanson, P.I., Roth, R., Morisaki, H., Jahn, R., Heuser, J.E., 1997. Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy. Cell 90, 523–535.

Hattendorf, D.A., Lindquist, S.L., 2002. Cooperative kinetics of both Hsp104 ATPase domains and interdomain communication revealed by AAA sensor-1 mutants. EMBO J. 21, 12–21.

Hayward, S., Berendsen, H.J., 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins 30, 144–154.

Hayward, S., Lee, R.A., 2002. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. J. Mol. Graph. Model. 21, 181–183.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38, 27–38.

Jeruzalmi, D., O'Donnell, M., Kuriyan, J., 2001. Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. Cell 106, 429–441.

Karata, K., Inagawa, T., Wilkinson, A.J., Tatsuta, T., Ogura, T., 1999. Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. Site-directed mutagenesis of the ATP-dependent protease FtsH. J. Biol. Chem. 274, 26225–26232.

Kondo, H., Rabouille, C., Newman, R., Levine, T.P., Pappin, D., Freemont, P., Warren, G., 1997. p47 is a cofactor for p97-mediated membrane fusion. Nature 388, 75–78.

Kunau, W.H., Beyer, A., Franken, T., Gotte, K., Marzioch, M., Saidowsky, J., Skaletz-Rorowski, A., Wiebel, F.F., 1993. Two

complementary approaches to study peroxisome biogenesis in *Saccharomyces cerevisiae*: forward and reversed genetics. Biochimie 75, 209–224.

Lenzen, C.U., Steinmann, D., Whiteheart, S.W., Weis, W.I., 1998. Crystal structure of the hexamerization domain of *N*-ethylmaleimide-sensitive fusion protein. Cell 94, 525–536.

Lindahl, E., Hess, B., van der Spoel, D., 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J. Mol. Model. 7, 306–317.

Liu, J., Smith, C.L., DeRyckere, D., DeAngelis, K., Martin, G.S., Berger, J.M., 2000. Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. Mol. Cell 6, 637–648.

Lupas, A.N., Martin, J., 2002. AAA proteins. Curr. Opin. Struct. Biol. 12, 746–753.

Matveeva, E.A., He, P., Whiteheart, S.W., 1997. *N*-Ethylmaleimide-sensitive fusion protein contains high and low affinity ATP-binding sites that are functionally distinct. J. Biol. Chem. 272, 26413–26418.

May, A.P., Whiteheart, S.W., Weis, W.I., 2001. Unraveling the mechanism of the vesicle transport ATPase NSF, the *N*-ethylmaleimide-sensitive factor. J. Biol. Chem. 276, 21991–21994.

Ming, D., Kong, Y., Lambert, M.A., Huang, Z., Ma, J., 2002. How to describe protein motion without amino acid sequence and atomic coordinates. Proc. Natl. Acad. Sci. USA 99, 8620–8625.

Neuwald, A.F., Aravind, L., Spouge, J.L., Koonin, E.V., 1999. AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res. 9, 27–43.

Ogura, T., Wilkinson, A.J., 2001. AAA+ superfamily ATPases: common structure–diverse function. Genes Cells 6, 575–597.

Ortega, J., Singh, S.K., Ishikawa, T., Maurizi, M.R., Steven, A.C., 2000. Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. Mol. Cell 6, 1515–1521.

Roccatano, D., Mark, A.E., Hayward, S., 2001. Investigation of the mechanism of domain closure in citrate synthase by molecular dynamics simulation. J. Mol. Biol. 310, 1039–1053.

Rouiller, I., Butel, V.M., Latterich, M., Milligan, R.A., Wilson-Kubalek, E.M., 2000. A major conformational change in p97 AAA ATPase upon ATP binding. Mol. Cell 6, 1485–1490.

Sander, C., Schneider, R., 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 9, 56–68.

Schaffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., Altschul, S.F., 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. 29, 2994–3005.

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

Scott, W.R.P., Huenenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krueger, P., van Gunsteren, W.F., 1999. The GROMOS biomolecular simulation program package. J. Phys. Chem. A 102, 3596–3607.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Tironi, I.G., Sperb, R., Smith, P.E., van Gunsteren, W.F., 1995. A generalized reaction-field method for molecular dynamics simulations. J. Chem. Phys. 102, 5451–5459.

Trame, C.B., McKay, D.B., 2001. Structure of *Haemophilus influenzae* HslU protein in crystals with one-dimensional disorder twinning. Acta Crystallogr. D Biol. Crystallogr. 57, 1079–1090.

Wang, J., Song, J.J., Seong, I.S., Franklin, M.C., Kamtekar, S., Eom, S.H., Chung, C.H., 2001. Nucleotide-dependent conformational changes in a protease-associated ATPase HsIU. Structure (Camb.) 9, 1107–1116.

Yamada, K., Kunishima, N., Mayanagi, K., Ohnishi, T., Nishino, T., Iwasaki, H., Shinagawa, H., Morikawa, K., 2001. Crystal structure of the Holliday junction migration motor protein RuvB from *Thermus thermophilus* HB8. Proc. Natl. Acad. Sci. USA 98, 1442–1447.

Zhang, X., Shaw, A., Bates, P.A., Newman, R.H., Gowen, B., Orlova, E., Gorman, M.A., Kondo, H., Dokurno, P., Lally, J., Leonard, G., Meyer, H., van Heel, M., Freemont, P.S., 2000. Structure of the AAA ATPase p97. Mol. Cell 6, 1473–1484.