

27. Muller, I., Zimmermann, M., Becker, D. & Flomer, M. Calendar life span versus budding life span of *Saccharomyces cerevisiae*. *Mech. Ageing Dev.* **12**, 47–52 (1980).
28. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
29. Weindruch, R. H., Walford, R. L., Fligiel, S. & Guthrie, D. The retardation of aging in mice by dietary restriction: Longevity, cancer, immunity, and lifetime energy intake. *J. Nutr.* **116**, 641–654 (1986).
30. Roth, G. S. Caloric restriction in primates: will it work and how will we know? *J. Am. Geriatr. Soc.* **47**, 896–903 (1999).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

# Acknowledgements

We thank R. Cook, A. Park, and H. Amoroso at the MIT Biopolymers lab for the HPLC and electron-spray mass spectroscopy, and P. Matsudaira for the MALDI mass spectroscopy. We also thank S. Inamoto for bacterial strains, and H. Tissenbaum and E. Ford for comments on the paper. This work was funded by The Human Frontier Science Program Organization Long-Term Fellowship to S. L., a NIH predoctoral grant to C.A. and M.K., and grants from the NIH, Seaver Foundation, Ellison Medical Foundation, and Howard and Linda Stern Fund to L.G.

Correspondence and requests for materials should be addressed to L.G. (e-mail: [leng@mit.edu](mailto:leng@mit.edu)).

## The structures of HslU and the ATP-dependent protease HslU–HslV

Matthias Bochtler<sup>\*†</sup>, Claudia Hartmann<sup>\*†</sup>, Hyun Kyu Song<sup>\*</sup>, Gleb P. Bourenkov<sup>†</sup>, Hans D. Bartunik<sup>†</sup> & Robert Huber<sup>\*</sup>

<sup>\*</sup> Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Planegg, Germany

<sup>†</sup> Arbeitsgruppen für strukturelle Molekularbiologie der Max-Planck-Gesellschaft, Notkestr. 85, D-22603 Hamburg, Germany

<sup>†</sup> These authors contributed equally to this work

The degradation of cytoplasmic proteins is an ATP-dependent process<sup>1</sup>. Substrates are targeted to a single soluble protease, the 26S proteasome<sup>2,3</sup>, in eukaryotes and to a number of unrelated proteases in prokaryotes<sup>4,5</sup>. A surprising link emerged with the discovery of the ATP-dependent protease HslVU (heat shock locus VU)<sup>6–8</sup> in *Escherichia coli*. Its protease component HslV shares 20% sequence similarity<sup>6</sup> and a conserved fold<sup>9</sup> with 20S proteasome  $\beta$ -subunits. HslU is a member of the Hsp100 (Clp) family of ATPases. Here we report the crystal structures of free HslU and an 820,000 relative molecular mass complex of HslU and HslV—the first structure of a complete set of components of an ATP-dependent protease. HslV and HslU display sixfold symmetry, ruling out mechanisms of protease activation that require a symmetry mismatch between the two components. Instead, there is conformational flexibility and domain motion in HslU and a localized order–disorder transition in HslV. Individual subunits

of HslU contain two globular domains in relative orientations that correlate with nucleotide bound and unbound states. They are surprisingly similar to their counterparts in *N*-ethylmaleimide-sensitive fusion protein<sup>10,11</sup>, the prototype of an AAA-ATPase. A third, mostly  $\alpha$ -helical domain in HslU mediates the contact with HslV and may be the structural equivalent of the amino-terminal domains in proteasomal AAA-ATPases.

We expressed and purified histidine-tagged variants of HslV and HslU from *E. coli*. Mixing the two components was sufficient to recover ATP-dependent protease activity, as assessed by enzymatic assay with both the chromogenic peptide Z-GGL-AMC<sup>7</sup> and radio-labelled <sup>14</sup>C-methylcasein<sup>12</sup>; the latter was used as a model system for large, denatured proteins (data not shown). Three crystal forms grew (Fig. 1; and Table 1). The crystals of space group *P*<sub>6</sub><sub>3</sub><sub>2</sub><sub>2</sub> contain free HslV and HslU fully bound with AMP-PNP (5'-adenylylimidodiphosphate) on 32-point symmetry positions (Fig. 1a). In the absence of both HslV and exogenous nucleotide, HslU crystallized in space group *P*<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub>, with a full hexamer in the asymmetric unit. Four subunits are complexed with nucleotide, presumably from the cellular environment. The remaining two subunits are located on opposite sides of the ring and contain bound sulphate from the crystallization buffer (Fig. 1b). The addition of the non-hydrolysable AMP-PNP induced the growth of trigonal crystals under identical conditions with two independent hexamers located on threefold crystallographic axes with nucleotide bound to alternating subunits on the ring (Fig. 1c). Empty subunits have been modelled without sulphate and show local main-chain conformational changes of the phosphate-binding loop (P-loop).

High sequence similarity predicts that there is a conserved fold in all members of the Hsp100 (Clp) family of ATPases. In contrast, the identification of the Hsp100s as AAA proteins<sup>13</sup> has been controversial<sup>14</sup>. Comparison of the HslU structure with the only atomic-resolution structure of an AAA protein available so far<sup>10,11</sup> strongly supports this classification (see Fig. 2). HslU is folded into three distinct domains. We observe the same topology in the N-terminal domain (D2) of *N*-ethylmaleimide-sensitive fusion (NSF) protein and the N-terminal domain of HslU. In HslU, this domain comprises residues S2–K109 and I244–L332 and will be referred to as the N domain. Residues M110 to A243 are missing in other Hsp100 proteins and emerge from the globular domain. The structural equivalence of the HslU and NSF D2 backbones is greatest (r.m.s. deviation of Ca bond lengths = 0.7 Å) for a central, parallel  $\beta$ -sheet that forms the core of the protein in both structures, having strand order  $\beta$ N2,  $\beta$ N3,  $\beta$ N6,  $\beta$ N1,  $\beta$ N7 in HslU (see Fig. 2; and Table 2). Ca carbons of corresponding residues in the sequence alignment are nearly coincident, with the exception of residues PFIK of strand  $\beta$ N2 in HslU, which correspond to residues FIKI in NSF. Lysine 80 is extended towards the nucleotide in HslU. The central  $\beta$ -sheet in both HslU and NSF is sandwiched in between two layers of  $\alpha$ -helices as in other P-loop proteins. The P-loop is located between strand  $\beta$ N1 and helix  $\alpha$ N4 in HslU and is in contact with nucleotide as expected<sup>15,16</sup>. As in NSF, and unlike in other P-loop triphosphate

**Figure 1** Summary of the three crystal forms (a–c) that were used for structure determination. Subunits in the respective asymmetric units are numbered 1–6.



hydrolases<sup>17</sup>, the nucleotide fits best in *syn* conformation in all three HslU structures. Further away from the central  $\beta$ -sheet, there is substantially more structural variability, including an insertion of the two short antiparallel  $\beta$ -strands  $\beta$ N4 and  $\beta$ N5 in HslU in place of a loop in NSF.

The intermediate domain (I domain) emerges from the N domain with residue M110 and returns to it with residue A243. It is ordered and defined by electron density to varying extents in the subunits, as shown in Table 2. Where the intermediate domain is ordered, it is a loosely folded structure with coiled helices, one from residues M110–I136 ( $\alpha$ I1/ $\alpha$ I2), and the other from residue K217–

A243 ( $\alpha$ I4/ $\alpha$ I5), which is kinked at R120 and N235, respectively. The emerging helix  $\alpha$ I2 ends with the hydrophobic sequence ILDVLI (131–136) and bends sharply at two consecutive prolines (137–138). Its residues I131, L132, L135, I136 pack with M222 and with F159 and L163 of the connecting helix  $\alpha$ I3 to form a small hydrophobic nucleus most distal from the N-domain. The following strand  $\beta$ I1 forms a highly and oppositely charged ladder with  $\beta$ I2. The two strands may act as a clamp for the mobile loop of residues I175–Q209. These residues cannot be traced with confidence in any of the subunits, although broken electron density is seen in the HslU lumen enclosed by the I domains.

**Figure 2** Comparison of HslU and NSF main chains. **a**, Superposition of the ligand-bound (coloured) and free (white) HslU forms. Chains 1 and 2 of the P321 crystals (see Fig. 1c) are shown. The N domains (shown in green and red) have been superimposed (r.m.s.d.

Ca bond lengths = 0.5 Å for the central  $\beta$ -sheet, r.m.s.d. Ca bond lengths = 1.2 Å for the whole domain). For clarity, the N and I domains of the free form have been omitted. **b**, Stereo diagram of NSF D2.

**Table 1** Crystallographic data statistics

|                                    | HslU–HslU                                      |          |            |            |            |            | HslU       | HslU                             |
|------------------------------------|--|----------|------------|------------|------------|------------|------------|----------------------------------|
| Data collection                    |  |          |            |            |            |            |            |                                  |
| Space group                        | P6 <sub>3</sub> 22                             |          |            |            |            |            | P321       | P2 <sub>1</sub> 2 <sub>1</sub> 2 |
| Soak                               | Ta <sub>6</sub> Br <sub>12</sub> <sup>2+</sup> |          |            |            |            |            | Thiomersal | –                                |
| <i>a</i> (Å)                       | 170.61   | 170.61   | 171.12     | 171.12     | 171.18     | 171.18     | 201.78     | 208.15                           |
| <i>b</i> (Å)                       | 170.61   | 170.61   | 171.12     | 171.12     | 171.18     | 171.18     | 201.78     | 167.70                           |
| <i>c</i> (Å)                       | 275.86   | 275.86   | 274.50     | 274.50     | 278.09     | 278.09     | 171.63     | 108.19                           |
| <i>I</i> (Å)                       | 1.0712   | 1.0712   | 1.0780     | 1.0780     | 1.0500     | 1.0500     | 1.0712     | 1.0700                           |
| Resolution (Å)                     | 20.0–2.8                                       | 20.0–2.8 | 25.0–3.0   | 25.0–3.0   | 25.0–3.2   | 25.0–3.2   | 20.0–4.0   | 35.0–3.0                         |
| Observed reflections               | 220,892  | 220,892  | 331,896    | 331,896    | 400,748    | 400,748    | 224,179    | 265,668                          |
| Sig. ind. refl.                    | 57796  | 57796    | 89272      | 89272      | 74892      | 74892      | 32750      | 74040                            |
| Data completeness (%)              | 98.4   | 98.4     | 98.3       | 98.3       | 99.7       | 99.7       | 95.8       | 96.9                             |
| <i>R</i> <sub>merge</sub> * (%)    | 14.1   | 14.1     | 8.2        | 8.2        | 8.9        | 8.9        | 10.5       | 7.9                              |
| Phasing statistics                 |  |          |            |            |            |            |            |                                  |
| Phasing power                      | –  | –        | 20.0–6.0 Å | 20.0–6.0 Å | 20.0–3.2 Å | 20.0–3.2 Å | –          | –                                |
| <i>R</i> <sub>cullis</sub>         | –  | –        | 0.71       | 0.64       | 2.02       | 2.07       | –          | –                                |
| <i>R</i> <sub>cullis,ano</sub>     | –  | –        | 0.91       | 0.92       | 0.63       | 0.62       | –          | –                                |
|                                    | –  | –        | 0.87       | 0.87       | 0.98       | 0.94       | –          | –                                |
| Refinement statistics (CNS/X-PLOR) |  |          |            |            |            |            |            |                                  |
| <i>R</i> <sub>cryst</sub> (%)      | 30.1   |          |            |            |            |            | 22.9       | 29.4                             |
| <i>R</i> <sub>free</sub> † (%)     | 34.9   |          |            |            |            |            | 28.5       | 34.2                             |
| <i>R</i> <sub>back</sub> ‡ (%)     | 16.6   |          |            |            |            |            | 17.3       | 14.3                             |
| r.m.s.d. bond lengths (Å)          | 0.009  |          |            |            |            |            | 0.007      | 0.009                            |
| r.m.s.d. bond angles (°)           | 1.4  |          |            |            |            |            | 0.9        | 1.3                              |
| r.m.s.d. NCS (Å)                   | 0.13   |          |            |            |            |            | 0.14       | 0.12                             |
| N and C domain main chain atoms    |  |          |            |            |            |            |            |                                  |

Most of the HslU crystals were macroscopically twinned and unsuitable for analysis. The specimens used for the measurement of native and derivative data were selected on the basis of their diffraction patterns but also showed reflections smeared in the *c* direction. The relatively high *R*-values in model refinement of the HslU and the orthorhombic HslU crystals are a consequence of inadequate modelling of the disordered I-domains.

Sig. ind. refl., significant independent reflections; NCS, non-crystallographic symmetry.

\* *R*<sub>merge</sub> =  $\sum |I - \langle I \rangle| / \sum I$

† The free *R*-factor was calculated from 5% of the data, removed at random before refinement was carried out.

‡ The backtransformation *R*-factor was calculated after ten cycles of twofold averaging.



The carboxy-terminal domain (C domain, residues 333–443) is a four-helix bundle (aC1/aC2, aC3, aC4, aC5/aC6) with a large bulge at residues T345–S350 and a double-stranded parallel  $\beta$ -sheet formed by  $\beta$ -strands bC1 and bC2. This forms a planar surface and is a prominent crystal contact site in all three crystal forms. The helix topology resembles that of the C domain in NSF, which is much smaller as it lacks the accessory structures mentioned above. The FIL (441–443) C terminus of HslU, which is conserved among the HslU family, is buried inside the domain with its charge balanced by R394 and by R329 of the adjacent subunit.

In our HslV–HslU crystals, HslV is in contact with HslU along and perpendicular to its sixfold axis. A complex was assigned from considerations of stoichiometry and symmetry, such that HslV and HslU hexamers form a *D*<sub>6</sub> symmetric assembly. It is a doubly capped species, as in electron microscopy images<sup>18</sup>. Our data confirm earlier data on the diameter of HslU, but disagree with the electron microscopy data on the overall size of the HslV–HslU complex. The crystal structure places the bulk of the scattering mass of HslU 5 nm away from the protease, separated by the I domains of HslU (see Fig. 3a). As a result, the overall dimensions of the complex are 25 nm in length and 12 nm in diameter. The compact shape of the molecule in electron microscopy images may be attributed to a collapsed species, but could also indicate an alternative mode of association that has the N and C domains of HslU in contact with HslV. The electron microscopy data show the presence of both hexameric and heptameric HslU species, indicating alternative modes of assembly also on this level<sup>18</sup>. The role of the I domains as spacers is maintained in both HslU crystal forms, where they span roughly the same distance between neighbouring HslU rings.

In our crystals of HslV–HslU, residues 140–149 of the I domain of HslU contact residues around 39 and 66 in HslV (see Fig. 3b). The interactions are not precisely defined, as the contacting HslU segment and the HslV segment at residue 66 are partly disordered

in the complex. Neither the sequences of the contacting segment of HslV nor those of HslU give a clue to the specificity of this interaction. We note that in both ours and other hands<sup>7</sup>, the complex in solution is rather unstable and does not survive most chromatographic procedures.

A direct comparison of the bound and free HslV species in the HslV–HslU crystals suffers from ill-defined electron density for the free species. A comparison of the HslV component of the complex with previously analysed isolated HslV shows small shifts of subunits within the dodecamer probably as a consequence of different crystal packing constraints and, prominently, a disordering of segments 65–69 and 84–93 (see Fig. 3b). The latter segment is an  $\alpha$ -helix lining the entrance channel into the HslV protease chamber.

The nucleotide in HslU subunits is in contact with residues from both the N and the C domains (see Fig. 4). Residues V61, I17 and I18 from the N domain and residues L335, I343 and A392 from the C domain form a hydrophobic pocket that accommodates the purin ring. The main-chain residues H16–I18 and G60–V61 supply hydrogen acceptors for the adenine base. The sugar is sandwiched in between E56 and H396. A number of residues of the N domain contact the phosphate moiety of ATP. The peptide amido groups of G60–E65 are close enough to the  $\alpha$ - and  $\beta$ -phosphates to form hydrogen bonds. In addition, the phosphates are coordinated by a set of polar side chains. These include K63 of the Walker A motif, K80, which lies within hydrogen-bonding distance of the  $\gamma$  phosphate, and D256. Threonine 59 and T64 face the  $\gamma$ -phosphate from opposing sides. The C domain interacts with the  $\beta$ - and  $\gamma$ -phosphates through R393, a residue that is highly conserved among AAA-ATPases<sup>13</sup> and that is part of the signature sequence V of Clp ATPases<sup>19</sup>. Two acidic residues of the adjacent subunit, E286 and E321, are also in contact with nucleotide. Although acidification of the phosphate-binding pocket as a result of oligomer assembly could be the trigger for ATP-hydrolysis in HslU, this

**Figure 3** HslV–HslU crystal complex. **a**, Ribbon plot (MOLSCRIPT) representation of the HslV–HslU crystal complex. Data from the HslU crystals has been included to model intermediate domains that were disordered in the HslV–HslU crystals. **b**, Close up view showing details of the contact region between HslV and HslU, with the HslV subunit taken

from the lower right corner of the HslV particle. Residues of HslV and HslU that are disordered in the HslV–HslU complex are shown in blue. An additional site of contact in HslV that remains ordered in the HslV–HslU complex is shown in yellow.



conclusion has to be treated with caution. Neither of the acidic residues is conserved in other AAA proteins, and NSF D2 has a lysine (K639) precisely in the position of E321. Arginine 325, a highly conserved residue<sup>13</sup> and the homologue of a putative arginine finger<sup>20</sup> in FtsH, is substantially further away ( $\approx 8$  Å) and may act indirectly.

Negative cooperativity of ATP binding in adjacent subunits is suggested by the orthorhombic (see Fig. 1b) and trigonal (see Fig. 1c) HslU crystals. It might explain the widely different apparent Michaelis constant ( $K_m$ ) for peptide and protein hydrolysis<sup>12</sup>. A decrease in ATPase activity with decreasing ATP saturation could be deduced from a superposition of the nucleotide-bound subunits. We find a retraction of the mentioned acidic groups from the  $\gamma$ -phosphate when nucleotide is absent in an adjacent subunit (data not shown). Unfortunately, electron density for the  $\gamma$ -phosphate itself is ill-defined in our structures even for the non-hydrolysable AMP-PNP analogue, owing to the absence of  $Mg^{2+}$  which is caused by an excess of EDTA in the crystallization conditions for the HslV–HslU (Fig. 1a) and orthorhombic HslU (Fig. 1b) crystals.

The absence of nucleotide leads to a relaxation of the structure, with the N and C domains rotating away from each other as rigid entities (see Fig. 2a). Looking down the sixfold axis of HslU towards the HslV complex, the C domain of one subunit is in contact with the N domain of the neighbouring subunit in counterclockwise orientation. Unexpectedly, this contact is so tight that the two domains, which are not covalently linked, behave as one rigid unit even in the context of the HslU oligomer. It appears that the

movement of the N domain is entirely driven by the change in orientation relative to the adjacent C domain (see Fig. 5). The acidic residues E286 and E321 of the N domain, which were directly in contact with nucleotide before its hydrolysis and dissociation, are so close to the centre of the overall rotation that they hardly move. In contrast, strands bN4 and bN5 are displaced by nearly 10 Å. The domain movements shift the I domain and may serve to dissociate sticky substrates and deliver them to HslV. In addition, they lead to a loss of proper sixfold symmetry in partially nucleotide-bound HslU particles. The resulting mismatch between the distorted sixfold symmetry of HslU and the proper sixfold symmetry of HslV should reduce the affinity of the two particles. Consistent with this hypothesis, we observe HslU with AMP-PNP bound to all subunits in the HslV–HslU co-crystals (see Fig. 1a) and partially AMP-PNP-bound HslU molecules in both HslU crystal forms (see Fig. 1b, c).

The crystal structures constrain models for substrate recognition and translocation. It is incompatible with the suggested presence of PDZ-domains<sup>21</sup> in the C-terminal part of the molecule. The previously characterized “sensor and substrate discrimination” (SSD) domain<sup>22</sup> coincides with the C domain of the crystal structure. Unlike in NSF D2, this domain is substantially involved in inter-subunit contacts and does not protrude from the ring. Its molecular surfaces face either adjacent subunits or solvent. There is no area of contact with the HslU cavity. We imagine a major role for the I domains in substrate recognition. The I domains, which are mostly disordered in three different crystal forms except when they make contacts with crystallographic neighbours (see Tab. 2b), emerge

**Figure 4** Representation of the nucleotide-binding site in HslU. The nucleotide is bound at the interface between two subunits. The N and C domains of the subunit that contributes most interactions are shown in yellow and red, respectively. Although the P-loop is part of

the N-domain of this subunit, it has been highlighted in green. Residues of the adjacent subunit are shown in purple.

**Figure 5** View of HslU along the sixfold axis (indicated by a circle). Three subunits on the ring and all I domains have been omitted for clarity. Fully ligand-bound HslU (from the P6<sub>3</sub>22 crystals) shown in grey is superimposed on partially ligand-bound HslU (from the

P2<sub>1</sub>2<sub>1</sub>2 crystals) shown in blue, red and green to distinguish individual subunits. Green nucleotide is present in both species, grey nucleotide only in the fully ligand-bound form. Note the movement of the N domain on the left side of the drawing.



**Table 2 HslU secondary structure assignment and list of disordered residues in HslU in the three crystal forms**

|  |
|--|
| HslU secondary structure assignment*   |
| aN1(P6-K15), aN2(D21-Q39), aN3(E42-E47), bN1(I53-I56) [V547-E550], aN4(K63-A74), bN2(F78-V81) [I573-C576], aN5(A83-T87), aN6(V96-K109), aI1(M110-K118), aI2(Y121-I136), aI3(T146-G166), bI1(L168-E174), bI2(K210-L216), aI4(K217-L233), aI5(P236-A243), aN7(I244-H250), bN3(I252-I255) [C607-V610], aN8(S268-E286), bN4(T289-T292), bN5(G295-K298), bN6(L303-G308) [L649-T654], aN9(P320-G324), bN7(I328-E331) [T671-H674], aC1(T337-E346), aC2(I351-E362), bC1(N365-F368), aC3(D370-S386), aC4(A392-L413), bC2(N417-N420), aC5(A422-A434), aC6(E436-I442) |
| Disordered residues in HslU in the three crystal forms†  |
| HslVU (P <sub>6</sub> ,22)<br>140–152, 167–215 (4); 114–157, 164–233 (5)   |
| HslU (P <sub>2</sub> ,2;2)<br>128–233 (1); 139–149, 177–212 (2); 118–230 (3); 129–229 (4); 142–150, 177–212 (5); 118–221 (6)   |
| HslU (P321)<br>175–209 (1); 135–222 (2); 175–209 (3); 135–221 (4)  |
| * Structurally equivalent residues in NSF are indicated in square brackets for the central, parallel $\beta$ -sheet.   |
| † Chain identifiers are given in parentheses (see Fig. 1).   |

teeth-like from a base plate made up of the N and C domains and may assist entrapment of substrate molecules of variable size and shape, similar to the action of a grab dredge. Their entirely disordered segment I175–Q209 may act as a bait and/or an adaptor to substrate molecules and promote binding. Substrates inside HslU may be destabilized and partially unfolded, as has been observed in ClpA<sup>23</sup>. Association of HslU with HslV leads to the observed local disordering of the HslV entrance channel (see Fig. 3b). We note that in our crystallization conditions HslVU is expected to be less active than in the presence of ATP and low salt<sup>12</sup>. Therefore, the role of the described structures in the reaction cycle requires further characterization.

Atomic-resolution structures have been reported for 20S proteasomes from archaeae<sup>24</sup> and yeast<sup>25</sup> and for the mammalian, ATP-independent proteasome activator PA28 (ref. 26). In contrast, only biochemical and electron microscopy data are available for the 19S regulatory cap. These data show that the particle can be dissociated into subcomplexes called lid and base. Along with other subunits, the base contains the six proteasomal AAA-ATPases<sup>27</sup>. They can be aligned with the N and C domains of HslU, as expected from our structure-based classification of HslU as an AAA protein. Unexpectedly, there may even be a counterpart for the I domain of HslU: it shares 14% identity and 37% homology with residues 11–141 of the N terminus of MSS1. We suggest that the proteasomal ATPases are structurally similar to HslU but have the I domain fused to the N terminus, which is nearby. We expect a ring-like, oligomeric structure, as in HslU, with the non-ATPase subunits present to modulate its activity. Our structure does not make any predictions for the lid complex which must have evolved with the ubiquitin system.

M

## Methods

### Protein purification

HslV was purified as previously described. HslU from *E. coli* was cloned into pET12b with an N-terminal MHHHHHH tag and expressed under the transcriptional control of the T7 promoter in BL21(DE3). Cells were collected after a 4-h induction with IPTG, resuspended in 10 mM Tris-HCl pH 7.5 and lysed by sonification in the presence of 1 mM phenylmethylsulphonyl fluoride (PMSF) and of 20 mg lysozyme and 1 mg DNaseA per litre of culture. The lysate was clarified by ultracentrifugation and applied to a chelating Sepharose column ( $\phi$  = 2.5 cm, h = 5 cm, Pharmacia) loaded with 2 mg ml<sup>-1</sup> of Zn<sup>2+</sup>. HslU eluted between 20 and 50 mM imidazole in buffer A (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.002% NaN<sub>3</sub>). HslU-containing fractions were pooled, supplemented with 5 mM ethylenediaminetetraacetic acid (EDTA) and applied to a Q-sepharose FF column ( $\phi$  = 3.5 cm, h = 7 cm, Pharmacia) equilibrated with 200 mM NaCl in buffer B (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM NaN<sub>3</sub>). HslU eluted at 300 mM NaCl. Fractions were pooled, concentrated and subjected to a final gel-filtration step in buffer B on a Sephacryl S300 column.

### Crystallization

Crystals of HslU grew at room temperature within a few days by sitting-drop vapour diffusion against a reservoir containing 100 mM sodium cacodylate pH 6.5, 15% glycerol, 10.5% polyethylene glycol (PEG) 8K and 500 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. Drops initially contained 2 ml reservoir solution, 2 ml protein at 16 mg ml<sup>-1</sup> in buffer B. The addition of 0.5 ml of 1.0 M guanidinium chloride and of 1 ml detergent octaethylene glycol monododecyl ether (C<sub>12</sub>E<sub>8</sub>) improved the quality of the crystals. Without the addition of nucleotide, we could grow orthorhombic crystals that diffracted to 3 Å resolution. The addition of 1 mM AMP-PNP and of 5 mM MgCl<sub>2</sub> induced the growth of trigonal crystals under otherwise identical conditions. Both crystal forms could be frozen in reservoir buffer mixed with 87% glycerol in the ratio 17:3. For crystallization of the protein complex, a 20 mg ml<sup>-1</sup> solution of HslU in buffer B with 1 mM AMP-PNP was prepared. It was mixed in a 2:1 ratio with a 16 mg ml<sup>-1</sup> solution of histidine-tagged HslV in 300 mM NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM NaN<sub>3</sub>. Crystals were grown at room temperature in sitting drops, equilibrating 2 ml of protein solution and reservoir buffer each against the reservoir buffer that contained 100 mM MES pH 6.3 and 2.0 M sodium acetate. A 17:3 mixture of reservoir solution and D(-)-2,3-butanediol served as cryoprotectant.

### Structure determination

All diffraction data were recorded at BW6, DESY, Hamburg (see Table 1). The HslV–HslU structure was solved by a combination of the molecular replacement and MAD phasing methods. Anomalous data were collected with crystals soaked with either tantalum bromide (TaBr<sub>5</sub>·2H<sub>2</sub>O) or thiomersal (C<sub>6</sub>H<sub>5</sub>HgNaO<sub>2</sub>S). Both derivatives were too weak to be localized by conventional Patterson search methods. The clue to structure determination was our assumption that one HslV particle occupied a position of 32-point symmetry in the HslV–HslU crystals. This assumption was supported by the self-rotation function and the strong pseudo-origin peaks at (2/3, 1/3, 1/2) and (1/3, 2/3, 1/2). It could be confirmed by molecular replacement methods (AMOR<sup>28</sup>) with the HslV dodecamer as the search model. Although the two independent subunits of HslV represented less than a quarter of the total scattering mass in the asymmetric unit, their placement allowed the calculation of partial phases that were good enough to identify candidate heavy atom positions in both anomalous and dispersive difference Fourier maps. All sites that could be confirmed by cross-vector verification (RSPS) were refined and used for initial phasing (MLPHARE<sup>29</sup>). The figure of merit was 0.53 to 2.8 Å resolution and could be improved to 0.55 by phase combination with partial model phases. The corresponding electron density map allowed the tracing of both HslU subunits with the exception of the intermediate domain, and showed the presence of a free HslV dodecamer in the crystals. Phases were improved by solvent flattening and cyclic averaging of the HslV–HslU electron density. The HslU model from the HslV–HslU crystals without the intermediate domains could be positioned in both HslU crystal forms by molecular replacement (AMORE). The originally weak signal improved markedly after allowing for mobility in the hinge region around residue 333. Cyclic averaging was used to improve phases in all three crystal forms. Models were built with FRODO<sup>29</sup> and refined with X-PLOR (P321) or CNS<sup>30</sup> (P<sub>6</sub>,22 and P<sub>2</sub>,2;2) with Engh & Huber parameters.

Received 21 September; accepted 23 December 1999.

- Goldberg, A. L. & St. John, A. C. Intracellular protein degradation in mammalian and bacterial cells: part 2. *Annu. Rev. Biochem.* **45**, 747–803 (1976).
- Hershko, A. & Ciechanover, A. The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* **61**, 761–807 (1992).
- Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Maurizi, M. R. Proteases and protein degradation in *Escherichia coli*. *Experientia* **48**, 178–201 (1992).
- Gottesman, S., Wickner, S. & Maurizi, M. R. Protein quality control: triage by chaperones and proteases. *Genes Dev.* **11**, 815–823 (1997).
- Chuang, S. E., Burland, V., Plunkett, G., Daniels, D. L. & Blattner, F. R. Sequence analysis of four new heat shock genes constituting the hslu and hslv operons in *Escherichia coli*. *Gene* **134**, 1–6 (1993).
- Rohrwild, M. et al. HslV-HslU: a novel ATP-dependent protease complex in *Escherichia coli* related to the eukaryotic proteasome. *Proc. Natl Acad. Sci. USA* **93**, 5808–5813 (1996).
- Missiakis, D., Schwager, F., Betton, J. -M., Georgopoulos, C. & Raina, S. Identification and characterization of HslV HslU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. *EMBO J.* **15**, 6899–6909 (1996).
- Bochtler, M., Ditzel, L., Groll, M. & Huber, R. Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **94**, 6070–6074 (1997).
- Lenzen, C. U., Steinmann, D., Whiteheart, S. W. & Weiss, W. I. Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. *Cell* **94**, 525–536 (1998).
- Yu, R. C., Hanson, P. I., Jahn, R. & Brünger, A. T. Structure of the ATP-dependent oligomerization domain of the N-ethylmaleimide sensitive factor complexed with ATP. *Nature Struct. Biol.* **5**, 803–811 (1998).
- Huang, H. -C. & Goldberg, A. L. Proteolytic activity of the ATP-dependent protease HslVU can be uncoupled from ATP-hydrolysis. *J. Biol. Chem.* **272**, 21364–21372 (1997).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: a class of chaperone-like ATPases associated with the assembly, operation and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Feng, H. P. & Gierasch, L. M. Molecular chaperones: clamps for the Clps? *Curr. Biol.* **8**, 464–467 (1998).
- Traut, T. W. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide binding sites. *Eur. J. Biochem.* **222**, 9–19 (1994).
- Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop—a common motif in ATP—and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434 (1990).
- Smith, C. A. & Raymont, I. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys. J.* **70**, 1590–1602 (1996).
- Rohrwild, M. et al. The ATP-dependent HslVU protease from *Escherichia coli* is a four-ring structure resembling the proteasome. *Nature Struct. Biol.* **4**, 133–139 (1997).



19. Schirmer, E. C., Glover, J. R., Singer, M. A. & Lindquist, S. Hsp100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem. Sci.* **21**, 289–296 (1996).
20. Karata, K., Inagawa, T., Wilkinson, A. J., Tatsuta, T. & Ogura, T. Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. *J. Biol. Chem.* **274**, 26225–26232 (1999).
21. Levchenko, I., Smith, C. K., Walsh, N. P., Sauer, R. T. & Baker, T. A. PDZ-like domains mediate binding specificity in the Clp/Hsp100 family of chaperone. *Cell* **91**, 939–947 (1997).
22. Smith, C. K., Baker, T. A. & Sauer, R. T. Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. *Proc. Natl Acad. Sci.* **96**, 6678–6682 (1999).
23. Weber-Ban, E. U., Reid, B. G., Miranker, A. D. & Horwich, A. L. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**, 90–93 (1999).
24. Löwe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533–539 (1995).
25. Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386**, 463–471 (1997).
26. Knowlton, J. R. *et al.* Structure of the proteasome activator REG $\alpha$  (PA28 $\alpha$ ). *Nature* **390**, 639–643 (1997).
27. Glickman, M. H. *et al.* A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94**, 615–623 (1998).
28. Project, C. C. C. The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763 (1994).
29. Jones, A. T. A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272 (1978).
30. Brünger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Cryst. D* **54**, 905–921 (1998).

# Acknowledgements

We thank P. Zwickl for <sup>14</sup>C-labelled casein, for his help with the protein degradation assay and for essential discussions; S. Grazulis for sharing PDB-scripts M. Boicu for DNA sequencing; and R. Engh for critically reading the manuscript. The financial support of the Deutsche Forschungsgemeinschaft and of the Human Frontier Science Program is acknowledged.

Correspondence and requests for materials should be addressed to M. B. (bochtler@biochem.mpg.de). Coordinates have been deposited in Protein Data Bank under accession numbers 1DOO (P6<sub>3</sub>22), 1DO0 (P2<sub>1</sub>2<sub>1</sub>2) and 1DO2 (P321).

## Structure of the intact transactivation domain of the human papillomavirus E2 protein

Alfred A. Antson\*<sup>†</sup>, Julie E. Burns<sup>†</sup>, Olga V. Moroz\*, David J. Scott\*<sup>§</sup>, Cyril M. Sanders<sup>‡</sup>, Igor B. Bronstein\*, G. Guy Dodson\*<sup>¶</sup>, Keith S. Wilson\* & Norman J. Maitland<sup>‡</sup>

\* York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK

<sup>†</sup> YCR Cancer Research Unit, Department of Biology, University of York, York YO10 5DD, UK

<sup>§</sup> Department of Biology, University of York, York YO10 5DD, UK

<sup>¶</sup> National Institute for Medical Research, London NW7 1AA, UK

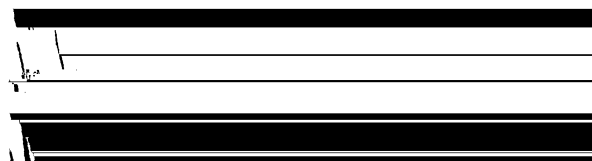
<sup>‡</sup> These authors contributed equally to this work

**Papillomaviruses cause warts and proliferative lesions in skin and other epithelia. In a minority of papillomavirus types ('high risk', including human papillomaviruses 16, 18, 31, 33, 45 and 56), further transformation of the wart lesions can produce tumours<sup>1</sup>. The papillomavirus E2 protein controls primary transcription and replication of the viral genome<sup>2</sup>. Both activities are governed by a , 200 amino-acid amino-terminal module (E2NT) which is connected to a DNA-binding carboxy-terminal module by a flexible linker. Here we describe the crystal structure of the complete E2NT module from human papillomavirus 16. The**

**E2NT module forms a dimer both in the crystal and in solution. Amino acids that are necessary for transactivation are located at the dimer interface, indicating that the dimer structure may be important in the interactions of E2NT with viral and cellular transcription factors. We propose that dimer formation may contribute to the stabilization of DNA loops<sup>3</sup> which may serve to relocate distal DNA-binding transcription factors to the site of human papillomavirus transcription initiation.**

Human papillomaviruses (HPVs) have evolved a sophisticated system of control, mediated by protein–DNA and protein–protein interactions, that involves both cellular and viral proteins. The relative molecular mass 45,000 (*M<sub>r</sub>* 45K) nuclear phosphoprotein, E2, has two key roles in this control. It functions as the principal virally encoded transcription factor and, in association with the viral E1 protein, it creates the molecular complex at the origin of viral DNA replication<sup>4</sup>. The E2 protein has three distinct modules (Fig. 1). The sequences of the N-terminal module (E2NT) of , 200 amino acids are highly conserved among the 86 known papillomavirus types (E2 Sequence Database, <http://hpv-web.lanl.gov>; see Supplementary Material), and this module is responsible for interactions with viral and host-cell transcription factors, such as Sp1, TFIIB and AMF-1 (refs 2, 5, 6). It is followed by a flexible, proline-rich, linker module and a C-terminal module (E2CT), each of , 90 amino acids. E2CT binds tightly as a homodimer to DNA sites with a consensus sequence of ACCGN<sub>4</sub>CGGT (ref. 2). The crystal structure of E2CT determined in complex with cognate DNA<sup>7</sup> shows that the protein dimer induces substantial bending (42–44°) of the DNA from its B-form double helix. The structure of a proteolytic fragment of HPV18 E2NT has been determined at 2.1 Å resolution<sup>8</sup>.

Here we report the crystal structure of the complete HPV16 E2NT module refined with data extending to 1.9 Å spacing (Table 1). There are two domains, N1 and N2, arranged to give the module an overall L-shaped appearance. The N1 domain (residues 1–92) is composed of three long  $\alpha$ -helices (Fig. 2), packed antiparallel to one another in the form of a twisted plane, with angles of about 20° and 25° between consecutive helices. There is a tight loop between helices  $\alpha$ 1 and  $\alpha$ 2 and a more extended one between helices  $\alpha$ 2 and  $\alpha$ 3. There is no structurally homologous domain in the PDB<sup>9</sup>. Domain N2 comprises residues 110–201 and is almost entirely antiparallel  $\beta$ -sheet structure. Our results and those reported for the HPV18 fragment<sup>8</sup> show that the N2 domain also has a new fold. Between the N1 and N2 domains, residues 93–109 make up two consecutive single turns of helical structure forming a fulcrum which packs tightly against both domains and could be considered to be part of either N1 or N2. Thirty-three of the total 201 residues in our E2NT construct are totally or highly conserved. Twenty-five of these appear to have a purely structural role, with 12 clustering at the N1–N2 interface reflecting the functional importance of the fulcrum in defining the relative orientation of the two domains (Fig. 2). Eight conserved amino acids are exposed to the solvent, indicating that they may have a role in intermolecular interactions. Mutational substitutions (reviewed in refs 2, 10) indicate that two of these residues are involved in transactivation (Arg 37 and Ile 73) and one in replication (Glu 39).



**Figure 1** Functional and structural assignments of papillomavirus E2. View of the N-terminal, linker and C-terminal modules indicating the known functions of each. Amino-acid numbers that delimit the modules correspond to those of HPV16 E2.

<sup>k</sup> Present address: Cold Spring Harbor Laboratory, PO Box 100, Bungtown Road, Cold Spring Harbor, New York 11724, USA.