Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 Å resolution

Steven C. Johnston, Christopher N. Larsen¹, William J. Cook², Keith D. Wilkinson¹ and Christopher P. Hill³

Biochemistry Department, University of Utah, Salt Lake City, UT 84132, ¹Department of Biochemistry, Emory University, Atlanta, GA 30322 and ²Department of Pathology, University of Alabama, Birmingham, AL 35233, USA

³Corresponding author
e-mail: chris@msscc.med.utah.edu

Ubiquitin C-terminal hydrolases catalyze the removal of adducts from the C-terminus of ubiquitin. We have determined the crystal structure of the recombinant human Ubiquitin C-terminal Hydrolase (UCH-L3) by X-ray crystallography at 1.8 Å resolution. The structure is comprised of a central antiparallel β -sheet flanked on both sides by α -helices. The β -sheet and one of the helices resemble the well-known papain-like cysteine proteases, with the greatest similarity to cathepsin B. This similarity includes the UCH-L3 active site catalytic triad of Cys95, His169 and Asp184, and the oxyanion hole residue Gln89. Papain and UCH-L3 differ, however, in strand and helix connectivity, which in the UCH-L3 structure includes a disordered 20 residue loop (residues 147–166) that is positioned over the active site and may function in the definition of substrate specificity. Based upon analogy with inhibitor complexes of the papain-like enzymes, we propose a model describing the binding of ubiquitin to UCH-L3. The UCH-L3 active site cleft appears to be masked in the unliganded structure by two different segments of the enzyme (residues 9–12 and 90–94), thus implying a conformational change upon substrate binding and suggesting a mechanism to limit non-specific hydrolysis.
Keywords: crystal structure/cysteine protease/substrate specificity/ubiquitin C-terminal hydrolase

Introduction

Ubiquitin is a small (8.6 kDa) highly conserved protein that is best known for its role in targeting proteins for degradation by the 26S protease. Recent reviews include Ciechanover and Schwartz (1994), Hershko and Ciechanover (1992), Jentsch (1992) and Wilkinson *et al.* (1995). Ubiquitin has been implicated in numerous cellular processes, including cell cycle control, oncoprotein degradation, receptor function, apoptosis, regulation of transcription, stress responses, maintenance of chromatin structure, DNA repair, signaling pathways, antigen presentation and the degradation of abnormal proteins. Monomeric ubiquitin is activated by E1 (ubiquitin-activating enzyme), which forms a thiolester bond with the ubiquitin C-terminus. Families of E2 (ubiquitin-conjugating) and

E3 (ubiquitin ligase) enzymes then catalyze ligation of the ubiquitin C-terminus to lysine side chains of acceptor proteins. Acceptor proteins can be modified with a single ubiquitin attached to one or more different lysine side chains. Alternatively, acceptor proteins can be polyubiquitinated, with a lysine side chain of the first ubiquitin conjugated to the C-terminus of the next, to form long chains attached to the target protein. Efficient targeting for degradation by the 26S protease appears to require polyubiquitination (Chau *et al.*, 1989; Gregori *et al.*, 1990). In addition to targeting proteins for degradation by the 26S protease, other roles of ubiquitination include modification of chromatin structure (Bradbury, 1992), lysosomal targeting (Hicke and Riezman, 1996) and regulation of a kinase activity (Chen *et al.*, 1996).

In addition to isopeptide linkages to the lysine side chains of acceptor proteins, the ubiquitin C-terminus is also found attached to α -amino groups in peptide bonds, since all known ubiquitin genes encode fusion proteins in which ubiquitin is followed by a C-terminal extension (Özkaynak *et al.*, 1987). Proteolytic processing at the ubiquitin C-terminus is catalyzed by DeUBiquitinating (DUB) enzymes. Such processing is likely to be required for several different functions, including liberation of monomeric ubiquitin from the polyprotein precursors, release of polyubiquitin chains from the remnants of 26S protease substrates, disassembly of polyubiquitin chains to allow recycling of monomeric ubiquitin, reversal of regulatory ubiquitination, editing of inappropriately ubiquitinated proteins and regeneration of active ubiquitin from adducts with small cellular nucleophiles (such as glutathione) that may be produced by side reactions. Additionally, several ubiquitin-like proteins that occur as fusions or conjugates have been identified, at least some of which appear to undergo a similar processing (Olvera and Wool, 1993; Haas *et al.*, 1996; Matunis *et al.*, 1996; Narasimhan *et al.*, 1996; Mahajan *et al.*, 1997).

In light of the many different substrates, and the extensive biological consequences of ubiquitination, it is not surprising that numerous DUB enzymes have been identified. These enzymes fall into two distinct families of cysteine proteases, Ubiquitin-specific Proteases (UBPs) (Tobias and Varshavsky, 1991; Baker *et al.*, 1992) and Ubiquitin C-terminal Hydrolases (UCHs) (Pickart and Rose, 1985). Both classes of enzymes hydrolyze the peptide bond (either α - or ϵ -linked) at the C-terminus of ubiquitin. The UBP enzymes, 16 of which have been identified in yeast, were named for their ability to cleave large model fusion proteins at the C-terminus of ubiquitin. They vary in molecular weight from 50 to 300 kDa, and exhibit a broad range of substrate specificity. Roles assigned for UBPs include cleavage of ubiquitin from the remnants of degraded protein (Papa and Hochstrasser, 1993) and disassembly of polyubiquitin chains to yield

Fig. 1. Sequence alignment of UCH enzymes. Every tenth UCH-L3 residue is delineated with a space. Active site residues (Gln89, Cys95, His169 and Asp184) of UCH-L3 are indicated in red. Other invariant residues are indicated in orange. Secondary structural elements seen in the UCH-L3 crystal structure are indicated above the sequence (see Figure 3). Disordered residues that are omitted from the UCH-L3 crystal structure are indicated with broken lines. SwissProt Database entries shown are: UCH-L3 (human; SW:P15374); UCH-L1 (human; SW:P09936); UBL-DROME (*D.melanogaster*; SW:P35122); SCHPO (*Schizosaccharomyces pombe*; SW:Q10171); YUH1 (*S.cerevisiae*; SW:P35127).

functional monomers (Wilkinson *et al.*, 1995). They appear to function in cell fate determination (Huang *et al.*, 1995), transcriptional silencing (Henchoz *et al.*, 1996; Moazed and Johnson, 1996) and the response to cytokines (Zhu *et al.*, 1996).

The well-characterized UCH enzymes are generally smaller than the UBPs (25–28 kDa), although two larger sequences have been deposited in the GenBank database. Disruption or deletion of the one UCH gene identified in yeast confers no discernible phenotype, suggesting that the substrate specificity of UCH enzymes may overlap with that of the UBP enzymes (Miller *et al.*, 1989; Baker *et al.*, 1992). Biochemical studies have demonstrated that the human enzymes UCH-L1 and UCH-L3, and the UCHs from *Saccharomyces cerevisiae* and *Drosophila melanogaster*, hydrolyze ϵ -linked amide bonds at the C-terminus of ubiquitin (R.E.Cohen, personal communication) (Roff *et al.*, 1996; Wilkinson, 1997), although most studies have focused on the hydrolysis of α -linked peptide bonds and small thiolester-, ester- and amide-linked adducts (Pickart and Rose, 1986; Wilkinson *et al.*, 1986). In general, most of these small adducts are good substrates, except for peptide extensions with proline immediately following the scissile bond. UCH-L3 cleaves peptide extensions of up to 20 residues from ubiquitin with high efficiency and low sequence preference, while larger folded extensions are not cleaved (Wilkinson, 1997). Similar results have been reported for the yeast UCH (Liu *et al.*, 1989; Miller *et al.*, 1989). These data suggest that the UCH enzymes may function to regenerate active ubiquitin from adducts with small nucleophiles (Pickart and Rose, 1985). The observed tissue specificity of UCH enzymes may reflect a distinct set of substrates (Wilkinson *et al.*, 1992). UCH-L1 is identical to PGP9.5, the neuronal UCH that constitutes several percent of the total soluble

protein in mammalian brain (Wilkinson *et al.*, 1989). UCH-L2 appears to be constitutively expressed in many tissues, while UCH-L3 is expressed in hematopoietic cells.

An alignment of five UCH sequences shows that only 12% of the residues are invariant (Figure 1). Site-directed mutagenesis of invariant residues on UCH-L1 implicates Cys95 (UCH-L3 numbering) as the active site nucleophile and His169 as the general base in catalysis, with an important role also played by Asp184 (Larsen *et al.*, 1996). The UCH enzymes do not appear to share significant sequence similarity with any other protein.

In order to understand better the catalytic mechanism and substrate specificity of UCH enzymes, we have determined the crystal structure of recombinant human UCH-L3 at a resolution of 1.8 Å. This structure has some similarities with the papain family of cysteine proteases, including an active site catalytic triad and oxyanion hole. A major topological difference from papain includes a 20 residue disordered loop that spans the active site. Based upon the structure, we propose: (i) a binding orientation for ubiquitin substrates on UCH enzymes; (ii) that the UCH active site is normally closed and opens upon binding to substrate; and (iii) that the disordered loop functions to define the substrate specificity of UCH enzymes.

Results and discussion

Structure determination

Crystals of native and selenomethione-substituted UCH-L3 were grown in space group $P2_12_12_1$ ($a = 48.6$ Å, $b = 60.8$ Å, $c = 81.4$ Å). There is one molecule in the asymmetric unit and the solvent content is 48%. The structure of selenomethione-substituted UCH-L3 was determined at 2.35 Å resolution by the method of multi-wavelength anomalous dispersion (MAD) (Figure 2). The

Fig. 2. Experimental electron density map. The electron density map (blue) is shown contoured at 1.0 r.m.s. deviation with the refined coordinates. Map calculation used $\lambda 1$ (0.9796 Å) structure factor amplitudes in the resolution range 10–2.35 Å. The MAD phases were refined by solvent flattening and histogram shifting. The position of the selenium atom of Met87 is apparent from the pseudo isomorphous ($\lambda 3$ – $\lambda 1$) difference map (red), which is contoured at 30 r.m.s. deviation. This figure was made with the program O (Jones *et al.*, 1991).

native structure subsequently was refined against 1.8 Å data to an *R*-value of 22.9% (free *R*-value = 29.1%) with good stereochemistry [root mean square (r.m.s.) deviation of bonds = 0.009 Å]. The current refined UCH-L3 model contains 206 of the 230 residues. Two regions of UCH-L3 lack defined electron density and have been omitted from the model (residues 1–4 and 147–166). The side chains of Arg136, Arg145 and Glu199 also lack defined density and have been included in the model with an occupancy of zero.

Structure of UCH-L3

UCH-L3 has overall dimensions of 43 Å × 32 Å × 37 Å. The structure is organized around a central six stranded antiparallel β -sheet, with α -helices packing on either side of the sheet to form a bilobal structure (see Figure 3). The left lobe, as viewed in Figure 3, contains the β -sheet and two long α -helices. His169 and Asp184, which have both been implicated in catalysis, are located at the amino- and carboxyl-terminal ends of strand 3 and strand 4 respectively. The right lobe includes a long buried α -helix (helix 4) which contains the active site nucleophile Cys95, and a cluster of smaller helices. Helix 4 makes predominantly hydrophobic interactions with the β -sheet, several helices and an extended segment. The active site of UCH-L3 is located between the molecule's two lobes, within a long cleft that appears to be closed in this unliganded structure. As discussed below, the catalytic nucleophile Cys95, the general base His169 and Asp184 form a catalytic triad that, along with other structural features,

Fig. 3. Ribbon diagram of UCH-L3. Side chains of the active site residues, Gln89, Cys95, His169 and Asp184, are shown in red and labeled Q, C, H and D. Amino- and carboxy-termini are denoted with N and C. Residues 146 and 167, which mark the ends of the large disordered loop, are indicated. Secondary structures were defined by the program PROMOTIF (Hutchinson and Thornton, 1996). Strands are colored green and the helices blue. Helix 4, which contains the active site nucleophile, Cys95, is colored cyan. Strand 1 (29–34), strand 2 (49–57), strand 3 (168–176), strand 4 (179–183), strand 5 (191–195), strand 6 (223–229). Helix 1 (residues 13–22), helix 2 (39–42), helix 3 (60–76), helix 4 (92–110), helix 5 (118–125), helix 6 (131–140), helix 7 (201–215). Helix 4 has two kinks at residues 95 and 105 that separate the large central α -helical segment from the two short 3_{10} segments at the ends of this helix. All other helices are α . Figures 3 and 5A were made with the programs MOLSCRIPT (Kraulis, 1991) and RASTER 3D (Bacon and Anderson, 1988).

resembles the well-known family of papain-like cysteine proteases (see Figure 4).

A predicted secondary structure assignment was proposed recently for UCH-L3 and other UCH isozymes (Larsen *et al.*, 1996) using the neural network program of the PredictProtein server (Rost and Sander, 1993). This analysis predicted 34% α -helical content and 17% β -sheet for UCH-L3, which is similar to our observation of 37% α -helix and 20% β -sheet in the crystal structure. The PredictProtein server correctly predicted five out of seven helices, and three out of six strands. However, a number of important secondary structural elements in the crystal structure are misidentified by the prediction, including helix 4, which contains the active site nucleophile, Cys95, and strand 4, which terminates one residue before Asp184, the third member of the catalytic triad.

Comparison with other structures

Although several well-characterized classes of enzymes are known to have active site triads that apparently function to orient and activate either cysteine or serine nucleophiles, we restrict comparisons to the papain family of cysteine proteases (Rawlings and Barrett, 1994), with which UCH-L3 shows the greatest similarity. We have compared 21 papain-like structures that have been deposited in the Brookhaven database with UCH-L3 (see Figures 4 and 5). Of the papain-like structures, three are free enzyme, four have the active Cys bound either to oxygen atoms, 2-mercaptoethanol or metal ion, and 14 are inhibitor

Fig. 4. Comparison of UCH-L3 and papain-like active sites. Active site residues of UCH-L3. Gln89, Cys95, His169 and Asp184 are shown with thick lines. A representative collection of eight papain-like enzyme active sites is shown by thin lines following least-squares overlap on the active site residue C α atoms. The papain-like structures shown have PDB identifiers 9pap, 4pad, 1pop, 2act, 1aec, 1huc, 1csb and 1gec. Other papain-like structures used in structural comparisons in this paper are: 1the, 1cpj, 1pad, 2pad, 5pad, 6pad, 1stf, 1pip, 1ppp, 1pe6, 1ppd, 1ppn and 1ppo. Refer to the PDB for primary references to these structures, which are not included here because of space limits.

complexes; 13 are of papain, four cathepsin B, three actinidin, and one glycyl endopeptidase. Of the papain-like enzymes, cathepsin B has the structure with greatest overall similarity to UCH-L3, as indicated by a search performed with the Dali algorithm (Holm and Sander, 1993).

Overlap of the UCH-L3 active site triad (Cys95, His169, Asp184) with the active site Cys, His and Asn of the papain-like enzymes yields r.m.s. deviation values on the three C α atoms of between 0.07 and 0.32 Å for 21 papain-like structures in the Brookhaven protein database (see Figure 4). In addition, UCH-L3 Gln89 is structurally equivalent to Gln19 of papain, which participates in the formation of a catalytically important structure known as the oxyanion hole (Drenth *et al.*, 1976; Ménard *et al.*, 1991; Schröder *et al.*, 1993). Overlap of all four of these UCH-L3 active site residues on the papain-like enzymes yields r.m.s. deviation values that range from 0.59 to 0.79 Å for C α atoms, and from 0.84 to 1.2 Å for all atoms. Interestingly, the structural similarity extends to three buried water molecules of UCH-L3 that are located between the two lobes of the protein below the active site Cys and His. Two of these water molecules are also found in the papain-like enzymes, with the third site occupied by a serine side chain. It is possible that these conserved

water molecules serve simple architectural roles to allow juxtaposition of the two lobes of the enzyme. It is also possible that they function in catalysis, either by facilitating conformational change (Rashin *et al.*, 1986) or substrate binding (Meyer *et al.*, 1988).

Structural similarity at the active sites suggests that the catalytic mechanism of UCHs will resemble that of the papain-like enzymes (Storer and Ménard, 1994). Thus, it is likely that UCH-L3 Cys95 and His169 form a thiolate–imidazolium ion pair, Asp184 functions to orient the enzyme's active site and perhaps to stabilize the protonated form of His169, and Gln89 contributes to the oxyanion hole. These roles in catalysis are consistent with mutagenesis data for the Cys, His and Asp of UCH-L1 (Larsen *et al.*, 1996). In our unliganded structure, it appears unlikely that the Cys95 side chain is deprotonated because the carbonyl oxygen atom of Ser92 is positioned to form a linear 3.2 Å hydrogen bond with the Cys95 thiol. We propose that the thiolate ion will form after displacement of Ser92, which, as discussed below, is expected to undergo conformational change upon substrate binding.

Starting from overlap on the active site tetrad C α atoms, optimal C α superpositions of UCH-L3 with the papain-like enzymes were obtained using the program LSQMAN (Kleywegt and Jones, 1994). The best overlays were obtained with cathepsin B (Turk *et al.*, 1995) which shows 53 equivalent C α atoms with an r.m.s. deviation of 1.6 Å. Good agreement is also found with papain (Kamphuis *et al.*, 1984), which shows 39 equivalent C α atoms and an r.m.s. deviation of 1.2 Å. Superposition of UCH-L3 with papain on the 53 C α atoms of the optimal UCH-L3–cathepsin B overlap resulted in an r.m.s. deviation of 2.45 Å.

Segments of UCH-L3 that have structural equivalents in papain-like enzymes include most of the the central antiparallel β -sheet, helix 4 (which contains the active site Cys) and an extended β -like segment adjacent to helix 4 (see Figure 5). The major difference between these structures is that the active site helix precedes the β -sheet in papain, while the active site helix is formed from the sequence following the second β -strand of the sheet in UCH-L3. This may have important functional consequences because it allows the positioning of a disordered loop of 20 residues over the active site of UCH-L3. As discussed below, we hypothesize that this loop may play a role in substrate selection by the UCH enzymes.

A likely mode of substrate binding to UCH-L3 is suggested by analogy with complexes of papain-like enzymes, in which bound inhibitors occupy either the S or S' site (Figure 6). (Substrate residues amino- and carboxy-terminal to the scissile bond are designated P and P' respectively, and the corresponding binding sites on the enzyme designated S and S'; Schechter and Berger, 1967.) The corresponding putative active site cleft of UCH-L3 is closed by two short segments of the enzyme which, as described below, we suggest will move to allow substrate binding. Our proposed location for the UCH active site cleft is supported by the clustering of invariant surface-exposed residues in the region of the S site inhibitors of papain-like enzymes (Figure 6C). This pattern of conserved residues is consistent with the very high specificity of UCH enzymes for ubiquitin, which is expected to bind to the proposed S sites, and the lack

Fig. 5. Comparison of UCH-L3 with cathepsin B. (A) UCH-L3 (upper) and cathepsin B (lower) shown in a similar orientation to Figure 3. Equivalent residues were defined by LSQMAN (Kleywegt and Jones, 1994). Pairs of C α atoms were included in the overlap if their separation is <3.0 Å and if they form a stretch of at least five contiguous residues. Equivalent residues, as defined by LSQMAN, are shown in the cyan ribbon representation, and listed here: residues 32–37 of UCH-L3::residues 152–157 of cathepsin B, 48–60::166–178, 84–90::18–24, 93–106::27–40, 167–174::197–204, 182–186::217–221. (B) Topology diagram of secondary structure for the β -sheet and helix 4 of UCH-L3 (upper) and structurally equivalent segments of cathepsin B (lower). Secondary structural elements are colored according to their order of occurrence along the amino acid sequence (red, orange, yellow, green, cyan, blue, magenta). The main topological difference is for the helix, which in papain-like enzymes is the first of these secondary structural elements in the sequence, while for UCH-L3 helix 4 is found between strands 2 and 3. The long disordered loop of UCH-L3 is indicated by a dotted line.

of selection for residues following ubiquitin, which are expected to bind in the proposed S' sites.

Further insight on substrate binding is provided by the observation that UCH-L3 binds to ubiquitin with a micromolar dissociation constant and that this interaction has a significant electrostatic component (Larsen *et al.*, 1996). We anticipate, therefore, that the positively charged basic face of ubiquitin (Wilkinson, 1988) will bind to UCH enzymes. Consistent with this idea, UCH-L3 has a molecular surface of almost entirely negative electrostatic potential (Nicholls *et al.*, 1991), including three invariant carboxylates (Glu10, Glu14 and Asp33) at the putative S sites. As shown in Figure 7, we have crudely docked ubiquitin against the proposed S sites of UCH-L3 so that electrostatic interactions appear favorable and the flexible C-terminal residues of ubiquitin are positioned analogously to the S site inhibitor of papain-like enzymes, with the ubiquitin C-terminus adjacent to the active site nucleophile,

Cys95. We do not propose specific residue interactions between UCH-L3 and ubiquitin because this docking exercise is only approximate and, as discussed below, it appears likely that UCH-L3 will undergo some conformational change upon binding to substrate. Hydrophobic surfaces on ubiquitin and UCH-L3 are also likely to contribute to the binding interaction.

Substrate-induced conformational changes

Comparison with ligand-bound complexes of papain-like enzymes suggests that the specificity of UCH enzymes for ubiquitin adducts may result, in part, from maintenance of an inactive enzyme conformation in the absence of a bound ubiquitin moiety. In the absence of a binding partner, the UCH-L3 active site cleft appears to be closed by two loops (see Figure 8). The first of these loops includes Leu9 and Glu10, which are in van der Waals contact with groups on the opposite side of the cleft, and

are in positions incompatible with the placement of papain-like enzyme inhibitors after least-squares overlap on active site residues. It also seems likely that residues 11 and 12 will have to move in order to accommodate substrate. Interestingly, Glu10 is one of the few surface-exposed UCH residues that is invariant, and it is possible that binding of positively charged groups on ubiquitin to Glu10 initiates opening of the UCH active site cleft.

The second loop that appears to block the active site, residues 90–94, spans the catalytic residues Gln89 and Cys95, and adopts a conformation that differs from the equivalent region of papain-like structures by displacements of >4 Å for the C^α atoms of residues 92 and 93. Consequently, the carbonyl oxygen of UCH-L3 Ser92 is buried in the oxyanion hole in a position analogous to the oxygen atom of inhibitors seen in the cysteine protease–inhibitor complex structures. Thus, Ser92 carbonyl oxygen forms hydrogen bonding interactions with both the thiol and main chain amide of Cys95. Because the adjacent residue, Asn93, is both highly exposed and invariant, we speculate that this side chain may participate in substrate binding, thereby providing a mechanism to open the active site. Conformational changes in both of the loops that appear to block the active site may be coupled since van der Waals contacts are observed from residue 9 to 93 and from 6 to 93 and 94.

Access to the active site appears to be restricted further by a 20 residue disordered loop consisting of residues 147–166 which spans the active site cleft. This loop may exist in several different conformations and, as discussed below, we propose that it may function in the definition of substrate specificity. The observation of van der Waals contact between residues 7 and 146 and a hydrogen bonding interaction between residues 5 and 146 in the UCH-L3 crystal structure suggest the possibility of a coordinated conformational change upon substrate binding that includes the disordered loop.

Masking of the UCH active site in the absence of bound substrate may function to limit non-specific cleavages by these cytoplasmic proteases. An analogous conformational change probably does not occur for the papain-like enzymes. Inspection of the liganded and unliganded structures in the Brookhaven database shows no significant conformational changes in the enzyme S sites upon binding inhibitor. The papain-like enzymes, which are generally secreted or lysosomal, employ an alternative strategy to limit inappropriate reactions. Inhibitory N-terminal propeptide extensions are cleaved only after import into the lysosome (Karrer *et al.*, 1993; Carmona *et al.*, 1996; Coulombe *et al.*, 1996; Cygler *et al.*, 1996; Turk *et al.*, 1996).

Substrate specificity

Although UCH-L3 has high specificity for ubiquitin N-terminal to the scissile bond, it is highly permissive for the residues following ubiquitin provided the adduct is small and unstructured (K.D.Wilkinson, unpublished). One possible rationale for the lack of activity against larger folded C-terminal ubiquitin fusions is that only highly

Fig. 6. Active site clefts of papain-like enzymes and UCH-L3. Orientation is the same as for Figure 3. (A) Glycyl endopeptidase complex with the inhibitor benzyloxycarbonyl-L-V-G-methylene, which occupies the S4, S3, S2 and S1 sites (O'Hara *et al.*, 1995). (B) Cathepsin B with the inhibitor CA030, which occupies the S2, S1, S1' and S2' sites (Turk *et al.*, 1995). Protein surfaces are colored gray/green according to curvature. Bound inhibitors are red. The active site Cys residue is yellow, other active site residues are magenta. (C) UCH-L3 molecular surface colored for the invariant residues of Figure 1. Active site residues are shown in magenta, basic residues blue, acidic residues red, polar residues cyan and hydrophobic residues green. This figure was prepared with the program GRASP (Nicholls *et al.*, 1991).

extended substrates can be accommodated in a deep narrow groove of UCH S' sites. The UCH-L3 crystal structure does not appear to possess such a groove, however, and thus the ordered protein visible in the crystal structure does not obviously explain the preference of UCH enzymes for small unfolded substrates. Although it is possible that a deep S' site substrate cleft could be formed by conformational change upon binding to substrate, the very low discrimination shown across a broad range of sequences that are cleaved from the ubiquitin C-terminus argues against this possibility.

We propose instead that exclusion of large ubiquitin

fusions from the UCH-L3 active site results from the 20 residue loop between Thr147 and Val166 that is disordered in our crystals. This loop is topologically distinct from the papain-like enzymes. The ends of the loop are anchored 20 Å apart on opposite sides of the active site Cys95, and three different classes of conformations can be envisaged for the loop with respect to the proposed UCH–substrate interaction geometry (see Figure 9).

(i) The loop may be sandwiched between the body of UCH-L3 and the ubiquitin moiety of a substrate (red conformation in Figure 9). This arrangement seems unlikely, however, in light of the probable ubiquitin-binding surface on UCH-L3 (see above). Furthermore, the

Fig. 7. Proposed orientation of UCH-L3–ubiquitin binding. This view is approximately perpendicular (from the left) to Figure 3. The crystal structure of UCH-L3 is shown with β -strands green, helix 4 cyan, and other structure yellow. The glycyl endopeptidase and cathepsin B S and S' site inhibitors of Figure 6A and B are shown in red and magenta respectively after least-squares overlap of the papain-like enzyme complexes on the UCH-L3 crystal structure. The structure of ubiquitin (Vijay-Kumar *et al.*, 1987), shown in gray, has been positioned with the basic face adjacent to UCH-L3, the C-terminal carboxylate adjacent to UCH-L3 Cys95, and with the flexible C-terminal residues following the path of the glycyl endopeptidase S site inhibitor.

Fig. 9. Possible orientations of the UCH-L3 disordered loop. The crystal structure of UCH-L3 is shown in the same color representation and orientation as Figure 7. The docked ubiquitin molecule has been moved slightly away from the UCH-L3 for clarity. Residues that follow ubiquitin in an α -linked substrate adduct have been included in dark gray. Three possible classes of conformation are shown in magenta, blue and red, for the disordered loop (residues 147–166) with respect to the substrate.

Fig. 8. The active site cleft of UCH-L3 is blocked. Stereoview of the UCH-L3 active site in approximately the same orientation as Figure 3. The active site residues Gln89, Cys95, His169 and Asp184 are labeled Q, C, H and D, respectively. UCH-L3 residues Leu9, Glu10, Ala11 and Ser92 are labeled. UCH-L3 is colored cyan, with the two segments proposed to move upon binding of substrate colored green (residues 9–12 and 90–94). The S4–S1 site inhibitor of glycyl endopeptidase (Figure 6A) is shown in red after superposition on the UCH active site residue C $^{\alpha}$ atoms.

loop sequence is not well conserved, and thus seems poorly suited to mediate interactions with ubiquitin, for which all UCH enzymes that have been characterized exhibit high specificity.

(ii) A second possible conformation places the loop over the active site, with residues C-terminal to the scissile bond passing through the loop (blue in Figure 9). When modeled in a maximally open conformation, the loop has an internal diameter of ~ 15 Å, which is suitable for passage of an unfolded extended polypeptide chain, although it is expected to limit passage of even a small folded structure such as an α -helix. A problem with this model is that the *D.melanogaster* UCH is able to cleave ubiquitin from conjugates with the large substrate IkB α (Roff *et al.*, 1996), and that the *S.cerevisiae* UCH cleaves conjugates from cytochrome c (R.E.Cohen, personal communication).

(iii) Alternatively, the disordered loop may fold completely away from the proposed ubiquitin-binding surface (magenta in Figure 9). This conformation would be analogous to the occluding loop of cathepsin B, which is also located along the S' sites and defines the exopeptidase specificity of cathepsin B by making specific interactions with the substrate carboxy-terminus two residues beyond the scissile bond (Turk *et al.*, 1995). An important topological distinction is that, unlike the disordered loop of UCH-L3, the cathepsin B occluding loop does not straddle the active site cleft (in Figure 5A the occluding loop partially obscures the active site Gln, Cys and His of cathepsin B).

It is possible that upon binding of ubiquitin adducts, the disordered loop will remain mobile, fluctuating between the extreme magenta and blue conformations of Figure 9. Thus, the loop will impede active site access for a wide range of larger substrates, which may eventually attain a productive complex by using either the blue or magenta conformations. It is also possible that the disordered loop plays a more active role in the selection of substrates *in vivo*, perhaps even becoming ordered and contributing directly to binding of some physiological substrates. This model suggests the intriguing possibility that the disordered loops of the different UCH enzymes, which are of similar length but relatively dissimilar sequence identities, function as modular units to confer different substrate specificities on the various UCH isozymes.

Materials and methods

Crystallization

The recombinant human UCH-L3 used in these studies was purified as described (Larsen *et al.*, 1996). The protein solution used in crystallization trials was 12 mg/ml UCH-L3 in 50 mM Tris-HCl, pH 7.6, 15 mM β -mercaptoethanol, 1 mM EDTA. This solution was stored in aliquots at -70°C . Crystallization was performed at 4°C in sitting drops. The reservoir solution was 26% (w/w) polyethyleneglycol 4000, 200 mM sodium acetate, 100 mM PIPES pH 6.7 and 10 mM dithiothreitol. The drop solution was 3 μl of protein solution mixed with 3 μl of reservoir solution. These conditions produced crystalline aggregates after 4–5 days.

Single crystals were obtained by microseeding. One of the initial aggregates was ground up with a needle, and the needle streaked through a drop that was identical to that described above, but which had equilibrated for 3–5 days. Small single crystals appeared after several days.

Large crystals were obtained by macroseeding. Using a rayon loop, a small single crystal was transferred into reservoir solution, allowed to wash for several minutes, and then transferred into another drop that

had been equilibrated for 3–5 days. The same reservoir and drop conditions used to obtain the initial aggregates were also used for the subsequent micro and macroseeding. The crystals attain their maximum size in 5–10 days following macroseeding. Typical crystal dimensions are 0.3 mm $\times$ 0.3 mm $\times$ 0.6 mm.

For generation of selenomethionine-substituted UCH-L3 (SeUCH-L3), the *gal*[−], *met*[−] auxotroph B834(DE3) of the BL21 strain (Studier and Moffatt, 1986) harboring pRSL3 (Larsen *et al.*, 1996) was grown on LB agar as colonies. A single colony was inoculated into 50 ml of LB medium and grown overnight, followed by dilution into 6 l of modified M9 medium. Solutions O, P, S and V (Weber *et al.*, 1992), uracil (final concentration of 1 mM) and selenomethionine (final concentration of 50 g/l) were sterile filtered and added to M9 media.

At an OD_{600 nm} of 0.6, the cells were induced with 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG) for 3 h before harvesting by centrifugation. Purification of SeUCH-L3 was the same as for wild-type. Ion electrospray mass spectrometry showed an incorporation of $>98\%$ Se at each Met codon. SeUCH-L3 and wild-type UCH-L3 have comparable specific activities. SeUCH-L3 crystals were grown under the same conditions as native protein, although in this case the seeding steps proved unnecessary and growth time from the initial set up was 5–10 days.

Data collection and processing

The native and SeUCH-L3 crystals are isomorphous; space group P2₁2₁2₁, cell dimensions: $a = 48.6$ Å, $b = 60.8$ Å, $c = 81.4$ Å. There is one molecule in the asymmetric unit, and the Matthews's parameter, V_m , is 2.37 Å³/Da, which corresponds to a solvent content of 48% (Matthews, 1968).

All data were collected at 100 K. Prior to cryocooling, the crystals were transferred to the reservoir solution, and then to a series of solutions that were identical except for 2% increments in glycerol concentration up to a final concentration of 18% glycerol. The cryoprotected crystals were suspended in a rayon loop and cooled by plunging into liquid nitrogen.

Multiwavelength data were collected from a single SeUCH-L3 crystal on an MAR imaging plate detector at beamline X12C of the National Synchrotron Light Source, Brookhaven National Laboratory. The three wavelengths collected were selected from the fluorescence spectrum; λ_1 (0.9796 Å) was chosen as the inflection, or rise, corresponding to the minimum value of f' ; λ_2 (0.9793 Å) was taken as the peak, corresponding to the maximum in f' ; λ_3 (0.9300 Å) was chosen for the remote wavelength, corresponding to the maximum in f' . Data from each wavelength were indexed and integrated independently, and data from all three wavelengths were scaled together from 6.0 to 2.2 Å. The resulting scale factors were then applied separately to each individual wavelength for data from 30 to 2.35 Å. Data from a native crystal were collected to 1.8 Å resolution on an MAR imaging plate detector at beamline 7-1 of the Stanford Synchrotron Radiation Laboratory, Palo Alto. All data were processed with the programs DENZO and SCALEPAK (Otwinowski, 1993). See Table I for data statistics.

Structure determination and refinement

Crystallographic computing was performed using programs from the CCP4 suite (CCP4, 1994), unless otherwise stated. Of the seven methionine residues in UCH-L3, all except the amino-terminal Met are ordered. The six selenium sites were identified from difference Patterson and Fourier functions using the program XtalView (McRee, 1992). Selenium parameters were refined in MLPHARE (Otwinowski, 1991), treating λ_1 as the native data of a conventional multiple isomorphous phase determination (Ramakrishnan and Biou, 1997). The mean figure of merit calculated by MLPHARE was 0.42.

Phases computed with MLPHARE were refined by solvent flattening and histogram shifting with the program DM (Cowtan, 1994) to a mean figure of merit of 0.77. The resulting electron density map was readily interpretable for the majority of the UCH-L3 sequence (see Figure 2). Rounds of refinement with XPLOR (Brünger, 1992b) were interspersed with model building (Jones *et al.*, 1991). λ_1 amplitudes from 10.0 to 2.35 Å resolution were used in the refinement, with phase restraints also applied. At this stage, the R -value against 10.0–2.35 Å data was 24.3% and the free R -value was 30.4% (Brünger, 1992a). No sigma cuts were applied to refinement or R -value calculations.

Refinement was continued against 6.0–1.8 Å data collected from a native crystal (see Table II). Because of a slight deviation from true isomorphism between the native and SeUCH-L3 crystals, phase restraints were not employed for the high resolution refinement. The final model includes 104 water molecules and 206 of the total 230 UCH-L3 residues.

Table I. Data processing statistics

	Native SSRL ^a	λ1 NSLS ^a	λ2 NSLS ^a	λ3 NSLS ^a
Wavelength (Å)	1.080	0.9796	0.9793	0.930
Resolution limit (Å)	1.80	2.35	2.35	2.35
High resolution shell (Å)	(1.83–1.80)	(2.43–2.35)	(2.43–2.35)	(2.43–2.35)
No. of unique reflections	23 334	11 282	11 217	11 108
Completeness (%)	98 (93)	96 (86)	90 (84)	89 (91)
$\langle I/\sigma(I) \rangle$	20 (5)	15 (5)	15 (5)	15 (6)
Redundancy ^b	4.5 (3)	2 (1)	2 (1)	2 (1.5)
R_{sym} (%) ^c	4.0 (19)	4.9 (13.2)	4.4 (12.1)	4.8 (13.7)
Mosaicity (°)	1.18	0.42	0.42	0.42

^aData were collected on MAR imaging plate detectors on beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) or beamline X12C of the National Synchrotron Light Source (NSLS).

^bRedundancy is defined as the ratio of observed/unique structure factor amplitudes.

^c $R_{\text{sym}} = 100 * \sum_{\text{hkl}} \sum_i |I_i - \langle I \rangle| / \sum \langle I \rangle$.

Data were processed with DENZO and SCALEPACK (Otwinowski, 1993).

Table II. Refinement statistics

Resolution range (Å)	6.0–1.8
High resolution shell (Å)	(1.88–1.80)
R -value (%) ^a	22.9 (35.9)
R_{free} (%) ^b	29.1 (36.1)
R.m.s.d. (bonds) (Å) ^c	0.009
R.m.s.d. (angles) (°) ^c	1.609
No. of residues included (total)	206 (230)
No. of atoms with occupancy = 0.0 ^d	16
No. of water molecules	104
$\langle B \rangle$ (Å ²) protein/water	28.1/35.1
No. of ϕ/ψ angles (%): most favored	92.9
additional	7.1

^a R -value = $100 * \sum (|F_o(\text{obs})| - |F_o(\text{calc})|) / \sum |F_o(\text{obs})|$.

^b R_{free} = R -value for a randomly selected subset (5%) of the data that were not used for minimization of the crystallographic residual (Brünger, 1992a).

^cStereochemistry was analyzed with PROCHECK (Laskowski *et al.*, 1993).

^dNon-hydrogen atoms only. Atoms of the Arg136, Arg145 and Glu199 side chains were assigned an occupancy of zero because they lack defined electron density.

The current R -value is 22.9% and the free R -value is 29.1%. The first four residues at the amino-terminus are disordered, as are residues 147–166. The model has good stereochemistry as judged by PROCHECK (Laskowski *et al.*, 1993). Coordinates and diffraction data will be deposited with the Brookhaven Protein Database. Until coordinates are available through the database, they may be obtained from C.Phill by e-mail (chris@msscc.med.utah.edu).

Acknowledgements

We thank Robert Sweet, Mike Soltis and Harmut Luecke for assistance with data collection; Frank Whitby for help with the program XtalView; Venki Ramakrishnan for advice on MAD phasing; Kevin Cowtan for advice on the use of DM; and Robert Cohen, Venki Ramakrishnan, Martin Rechsteiner, Wes Sundquist and members of our laboratories for critical reading of the manuscript. This work has been supported by grants from the National Institutes of Health GM50163 (C.P.H.), GM30308 (K.D.W.), and by a grant from the Lucille P. Markey Charitable Trust. S.C.J. was an NIH predoctoral trainee (5-T32-GM08573). C.N.L. was an NIH predoctoral trainee (5-T32-GM08367). C.P.H. is the recipient of an American Cancer Society Junior Faculty Research Award (JFRA-513).

References

Bacon, D. and Anderson, W.F. (1988) A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics*, **6**, 219–220.

- Baker, R.T., Tobias, J.W. and Varshavsky, A. (1992) Ubiquitin-specific proteases of *Saccharomyces cerevisiae*: cloning of UBP2 and UBP3, and functional analysis of the UBP gene family. *J. Biol. Chem.*, **267**, 23364–23375.
- Bradbury, E.M. (1992) Reversible histone modifications and the chromosome cell cycle. *BioEssays*, **14**, 9–16.
- Brünger, A.T. (1992a) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
- Brünger, A.T. (1992b) *X-PLOR Version 3.1, A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
- Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J.S. and Menard, R. (1996) Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry*, **35**, 8149–8157.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K. and Varshavsky, A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.
- Chen, Z.J., Parent, L. and Maniatis, T. (1996) Site-specific phosphorylation of IκBα by a novel ubiquitination-dependent protein kinase activity. *Cell*, **84**, 853–862.
- Ciechanover, A. and Schwartz, A.L. (1994) The ubiquitin-mediated proteolytic pathway: mechanisms of recognition of the proteolytic substrate and involvement in the degradation of native cellular proteins. *FASEB J.*, **8**, 182–191.
- Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J.S. and Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.*, **15**, 5492–5503.
- Cowtan, K.D. (1994) 'dm': an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsl. Protein Crystallogr.*, **31**, 34–38.
- Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C. and Mort, J.S. (1996) Structure of rat procathepsin B. Model for inhibition of cysteine protease activity by the proregion. *Structure*, **4**, 405–416.
- Drenth, J., Kalk, K.H. and Swen, H.M. (1976) Binding of chloromethyl ketone substrate analogues to crystalline papain. *Biochemistry*, **15**, 3731–3738.
- Gregori, L., Poosch, M.S., Cousins, G. and Chau, V. (1990) A uniform isopeptide-linked multiubiquitin chain is sufficient to target substrate for degradation in ubiquitin-mediated proteolysis. *J. Biol. Chem.*, **265**, 8354–8357.
- Haas, A.L., Katzung, D.J., Reback, P.M. and Guarino, L.A. (1996) Functional characterization of the ubiquitin variant encoded by the baculovirus *Autographa californica*. *Biochemistry*, **35**, 5385–5394.
- Henchoz, S., De Rubertis, F., Pauli, D. and Spierer, P. (1996) The dose of a putative ubiquitin-specific protease affects position-effect variegation in *Drosophila melanogaster*. *Mol. Cell. Biol.*, **16**, 5717–5725.
- Hershko, A. and Ciechanover, A. (1992) The ubiquitin system for protein degradation. *Annu. Rev. Biochem.*, **61**, 761–807.
- Hicke, L. and Riezman, H. (1996) Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. *Cell*, **84**, 277–287.
- Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

- Huang, Y., Baker, R.T. and Fischer-Vize, J.A. (1995) Control of cell fate by a deubiquitinating enzyme encoded by the *fat facets* gene. *Science*, **270**, 1828–1831.
- Hutchinson, E.G. and Thornton, J.M. (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.*, **5**, 212–220.
- Jentsch, S. (1992) Ubiquitin-dependent protein degradation: a cellular perspective. *Trends Cell Biol.*, **2**, 98–103.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
- Kamphuis, I.G., Kalk, K.H., Swarte, M.B.A. and Drenth, J. (1984) Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol.*, **179**, 233–256.
- Karrer, K.M., Peiffer, S.L. and DiTomas, M.E. (1993) Two distinct gene subfamilies within the family of cysteine protease genes. *Proc. Natl Acad. Sci. USA*, **90**, 3063–3067.
- Kleywegt, G.J. and Jones, T.A. (1994) A super position. *Joint CCP4 and ESF-EACBM Newslett. Protein Crystallogr.*, **31**, 9–14.
- Kraulis, P.J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Larsen, C.N., Price, J.S. and Wilkinson, K.D. (1996) Substrate binding and catalysis by ubiquitin C-terminal hydrolases: identification of two active site residues. *Biochemistry*, **35**, 6735–6744.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Liu, C.-C., Miller, H.I., Kohr, W.J., Jr and Silber, J.I. (1989) Purification of a ubiquitin protein peptidase from yeast with efficient *in vitro* assays. *J. Biol. Chem.*, **264**, 20331–20338.
- Mahajan, R., Delphin, C., Guan, T., Gerace, L. and Melchior, F. (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell*, **88**, 97–107.
- Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
- Matunis, M.J., Coutavas, E. and Blobel, G. (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell. Biol.*, **135**, 1457–1470.
- McRee, D.E. (1992) A visual protein crystallographic software system for X11/XView. *J. Mol. Graph.*, **10**, 44–46.
- Ménard, R., Carrière, J., Laflamme, P., Plouffe, C., Khouri, H.E., Vernet, T., Tessier, D.C., Thomas, D.Y. and Storer, A.C. (1991) Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain. *Biochemistry*, **30**, 8924–8928.
- Meyer, E., Cole, G., Radhakrishnan, R. and Epp, O. (1988) Structure of native porcine pancreatic elastase at 1.65 Å resolution. *Acta Crystallogr.*, **B44**, 26–55.
- Miller, H.I., Henzel, W.J., Ridgway, J.B., Kuang, W.-J., Chisholm, V. and Liu, C.-C. (1989) Cloning and expression of a yeast ubiquitin-protein cleaving activity in *Escherichia coli*. *BioTechnology*, **7**, 698–704.
- Moazed, D. and Johnson, A.D. (1996) A deubiquitinating enzyme interacts with SIR4 and regulates silencing in *S.cerevisiae*. *Cell*, **86**, 667–677.
- Narasimhan, J., Potter, J.L. and Haas, A.L. (1996) Conjugation of the 15-kDa interferon-induced ubiquitin homolog is distinct from that of ubiquitin. *J. Biol. Chem.*, **271**, 324–330.
- Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins, Struct. Funct. Genet.*, **11**, 281–296.
- O'Hara, B.P., Hemmings, A.M., Buttle, D.J. and Pearl, L.H. (1995) Crystal structure of glycyl endopeptidase from carica papaya: a cysteine endopeptidase of unusual substrate specificity. *Biochemistry*, **34**, 13190–13195.
- Olvera, J. and Wool, I.G. (1993) The carboxyl extension of a ubiquitin-like protein is rat ribosomal protein S30. *J. Biol. Chem.*, **268**, 17967–17974.
- Otwinski, Z. (1991) Maximum likelihood refinement of heavy atom parameters. In Wolf, Evans, P.R. and Leslie, A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering*. Daresbury Laboratory, Warrington, UK, pp. 80–86.
- Otwinski, Z. (1993) Oscillation data reduction program. In Sawyer, L., Isaacs, N. and Bailey, S. (eds), *Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.
- Özkaynak, E., Finley, D., Solomon, M.J. and Varshavsky, A. (1987) The yeast ubiquitin genes: a family of natural gene fusions. *EMBO J.*, **6**, 1429–1439.
- Papa, F.R. and Hochstrasser, M. (1993) The yeast *DOA4* gene encodes a deubiquitinating enzyme related to a product of the human *tre-2* oncogene. *Nature*, **366**, 313–319.
- Pickart, C.M. and Rose, I.A. (1985) Ubiquitin carboxyl-terminal hydrolase acts on ubiquitin carboxyl-terminal amides. *J. Biol. Chem.*, **260**, 7903–7910.
- Pickart, C.M. and Rose, I.A. (1986) Mechanism of ubiquitin carboxyl-terminal hydrolase: borohydride and hydroxylamine inactivate in the presence of ubiquitin. *J. Biol. Chem.*, **261**, 10210–10217.
- Ramakrishnan, V. and Biou, V. (1997) Treatment of multiwavelength anomalous diffraction as a special case of multiple isomorphous replacement. *Methods Enzymol.*, **276**, 538–557.
- Rashin, A.A., Iofin, M. and Honig, B. (1986) Internal cavities and buried waters in globular proteins. *Biochemistry*, **25**, 3619–3625.
- Rawlings, N.D. and Barrett, A.J. (1994) Families of cysteine peptidases. *Methods Enzymol.*, **244**, 461–486.
- Roff, M., Thompson, J., Rodriguez, M.S., Jacque, J.-M., Baleux, F., Arenzana-Seisdedos, F. and Hay, R.T. (1996) Role of IκBα ubiquitination in signal-induced activation of NF-κB *in vivo*. *J. Biol. Chem.*, **271**, 7844–7850.
- Rost, B. and Sander, C. (1993) Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl Acad. Sci. USA*, **90**, 7558–7562.
- Schechter, I. and Berger, A. (1967) On the size of the active site of proteases. I. Papain. *Biochem. Biophys. Res. Commun.*, **27**, 157–162.
- Schröder, E., Phillips, C., Garmen, E., Harlos, K. and Crawford, C. (1993) X-ray crystallographic structure of a papain-leupeptin complex. *FEBS Lett.*, **315**, 38–42.
- Storer, A.C. and Ménard, R. (1994) Catalytic mechanism in papain family of cysteine peptidases. *Methods Enzymol.*, **244**, 486–500.
- Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, **189**, 113–130.
- Tobias, J.W. and Varshavsky, A. (1991) Cloning and functional analysis of the ubiquitin-specific protease gene Ubp1 of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **266**, 12021–12028.
- Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R. and Turk, V. (1995) Crystal structure of cathepsin B inhibited with CA030 at 2.0 Å resolution: a basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry*, **34**, 4791–4797.
- Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M. and Turk, V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett.*, **384**, 211–214.
- Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.*, **194**, 531–544.
- Weber, D.J., Gittis, A.G., Mullen, G.P., Abeygunawardana, C., Latmann, E.E. and Mildvan, A.S. (1992) NMR docking of a substrate into the X-ray structure of staphylococcal nuclease. *Proteins: Struct., Funct. Genet.*, **13**, 275–287.
- Wilkinson, K.D. (1988) Purification and structural properties of ubiquitin. In Rechsteiner, M. (ed.), *Ubiquitin*. Plenum Press, New York, pp. 5–38.
- Wilkinson, K.D. (1997) Ubiquitin carboxyl-terminal hydrolase. In Barrett, A. and Rawlings, N. (eds), *A Handbook of Proteolytic Enzymes*. Academic Press, in press.
- Wilkinson, K.D., Cox, M.J., Mayer, A.N. and Frey, T. (1986) Synthesis and characterization of ubiquitin ethyl ester, a new substrate for ubiquitin carboxyl-terminal hydrolase. *Biochemistry*, **25**, 6644–6649.
- Wilkinson, K.D., Lee, K., Deshpande, S., Duerksen-Hughes, P., Boss, J.M. and Pohl, J. (1989) The neuron-specific protein PGP9.5 is a ubiquitin carboxyl-terminal hydrolase. *Science*, **246**, 670–673.
- Wilkinson, K.D., Deshpande, S. and Larsen, C.N. (1992) Comparisons of neuronal (PGP 9.5) and non-neuronal ubiquitin C-terminal hydrolases. *Biochem. Soc. Trans.*, **20**, 631–637.
- Wilkinson, K.D., Tashayev, V.L., O'Connor, L.B., Larsen, C.N., Kasperek, E. and Pickart, C.M. (1995) Metabolism of the polyubiquitin degradation signal: structure, mechanism, and role of isopeptidase T. *Biochemistry*, **34**, 14535–14546.
- Zhu, Y., Carroll, M., Papa, F.R., Hochstrasser, M. and D'Andrea, A.D. (1996) DUB-1, a deubiquitinating enzyme with growth-suppressing activity. *Proc. Natl Acad. Sci. USA*, **93**, 3275–3279.

Received on January 30, 1997; revised on March 19, 1997