Crystal Structure of a Class I Ubiquitin Conjugating Enzyme (Ubc7) from *Saccharomyces cerevisiae* at 2.9 Å Resolution[†,‡]

William J. Cook[§]

Department of Pathology and Center for Macromolecular Crystallography, University of Alabama at Birmingham, Birmingham, Alabama 35253

Philip D. Martin,[§,||] Brian F. P. Edwards,[||] Russell K. Yamazaki,[⊥] and Vincent Chau*[,⊥]

Departments of Biochemistry and Pharmacology, Wayne State University School of Medicine, Detroit, Michigan 48201

Received October 21, 1996; Revised Manuscript Received December 3, 1996[⊗]

ABSTRACT: Ubiquitin-conjugating enzymes are a family of related proteins that participate in the ubiquitination of proteins. Previous studies on the crystal structures of *Saccharomyces cerevisiae* Ubc4 and *Arabidopsis thaliana* Ubc1 indicated that the smallest enzymes (class I), which consist entirely of the conserved core domain, share a common tertiary fold. Here we report the three-dimensional structure of the *S. cerevisiae* class I enzyme encoded by the **UBC7** gene. The crystal structure has been solved using molecular replacement techniques and refined by simulated annealing to an *R*-factor of 0.183 at 2.93 Å resolution. Bond lengths and angles in the molecule have root-mean-square deviations from ideal values of 0.016 Å and 2.3°, respectively. Ubc7 is an α/β protein with four α-helices and a four-stranded antiparallel β-sheet. With the exception of two regions where extra residues are present, the tertiary folding of Ubc7 is similar to those of the other two enzymes. The ubiquitin-accepting cysteine is located in a cleft between two loops. One of these loops is nonconserved, as this region of the Ubc7 molecule differs from the other two enzymes by having 13 extra residues. There is also a second single amino acid insertion that alters the orientation of the turn between the first two β-strands. Analysis of the 13 ubiquitin-conjugating enzyme sequences in *S. cerevisiae* indicates that there may be two other regions where extra residues could be inserted into the common tertiary fold. Both of these other regions exhibit significant deviations in the superposition of the three structures and, like the two insertion regions in Ubc7, may represent hypervariable regions within a common tertiary fold. As ubiquitin-conjugating enzymes interact with different substrates or other accessory proteins in the ubiquitination pathway, these variable surface regions may confer distinct specificity to individual enzymes.

Ubiquitin is a small, extremely conserved protein that is present in all eukaryotes. It functions in a variety of cellular processes but most prominently in specific proteolysis (Finley & Chau, 1991; Johnson *et al.*, 1992; Wilkinson, 1995; Hochstrasser, 1995). In ubiquitin-mediated proteolysis, multiple molecules of ubiquitin are attached to the target protein (Chau *et al.*, 1989; Gregori *et al.*, 1993), and these conjugates are then recognized and degraded by the 26S proteasome (Hochstrasser, 1995; Coux *et al.*, 1996). Most *in vivo* degradation of abnormal and naturally short-lived proteins is mediated by the ubiquitin-dependent pathway (Seufert & Jentsch, 1990; Finley & Chau, 1991; Hochstrasser, 1995, 1996; Murray, 1996).

The conjugation of ubiquitin to proteins requires ATP and takes place in a series of steps. In the first step, ubiquitin-activating enzyme (E1) forms a thiol ester linkage between its active site and the C-terminus of ubiquitin. This is an ATP-dependent step and proceeds through the formation of an intermediate with adenylate before being transferred to the thiol site. Ubiquitin is next transferred to a specific cysteine residue on an ubiquitin-conjugating enzyme, also called E2, where it is bound by a similar thiol ester linkage. Finally, ubiquitin−protein conjugates are formed by transferring ubiquitin from E2 to an acceptor protein, resulting in the formation of an isopeptide bond between the C-terminal carboxyl group of ubiquitin and an ϵ-amino group in the acceptor protein. This transfer may result directly from interaction of a specific E2 with its substrate protein or may require formation of a complex between the E2 enzyme and an additional substrate-binding protein known as E3.

Ubiquitin-conjugating enzymes are a family of closely related proteins (Jentsch, 1992). E2 gene families have been identified from many eukaryotic species; there are 13 unique E2s in the yeast *Saccharomyces cerevisiae*. All ubiquitin-conjugating enzymes apparently have a core domain of about 150 amino acids that show at least 25% sequence identity. Comparisons of their primary sequences suggest that the ubiquitin-accepting cysteine in all E2 enzymes is located within this core. E2 enzymes are grouped into four classes, depending on the presence and the location of unrelated sequences (Jentsch, 1992). Some of these enzymes contain extra C-terminal and/or N-terminal extensions from the core domain. Class I enzymes are the smallest E2 enzymes and

[†] Supported by NIH Grants GM-47604, GM-53136 (V.C.), and HL-57527 (B.F.P.E).

[‡] Coordinates for *S. cerevisiae* Ubc7 have been deposited with the Brookhaven Protein Data Bank, accession no. 2UCZ.

* Corresponding author: email, vchau@med.wayne.edu.

[§] These authors contributed equally to the work.

[||] Department of Biochemistry.

[⊥] Department of Pharmacology.

[⊗] Abstract published in *Advance ACS Abstracts,* February 1, 1997.

S0006-2960(96)02639-6 CCC: $14.00

consist almost entirely of the conserved core domain. Those that contain either extra C-terminal or N-terminal extensions from the core domain are called class II and III enzymes, respectively. Class IV enzymes contain both N- and C-terminal extensions.

The structures of two ubiquitin conjugating enzymes have been determined (Cook *et al*., 1992, 1993). Both are class I enzymes and consist entirely of the conserved "core" domain. Comparison of these two structures suggested that class I ubiquitin-conjugating enzymes are highly conserved in their three-dimensional folding. However, significant variations are found between the surfaces of the two molecules, as most of the identical residues between the two enzymes are either buried or clustered on one surface that lies next to the ubiquitin-accepting cysteine. Previously, we speculated that this conserved surface functions in protein−protein binding during ubiquitin thiol ester formation and in fact may be the E1-binding site common to all E2 enzymes. The divergent surface regions may enable an individual class I enzyme to interact with its respective substrates or E3 proteins.

To extend our structural studies to ubiquitin conjugating enzymes that contain more than the minimum core domain, we undertook the structure of Ubc7, a yeast E2 that contains two insertions relative to the two E2s whose structures have been determined (Cook *et al*., 1992, 1993). Ubc7 from *S. cerevisiae* contains 165 amino acids and has a calculated molecular mass of 18 521 Da (Vassal *et al.*, 1992; Jungmann *et al*., 1993). The yeast Ubc7 protein is clearly a functional ubiquitin conjugating enzyme by the criteria of accepting ubiquitin from E1 *in vitro* (Yamazaki & Chau, 1996) and being rendered nonfunctional *in vivo* by mutation of the active site Cys89 to Ser (Hiller *et al.*, 1996).

The **UBC7** gene is not essential; deletion of the gene has been reported to confer a cadmium hypersensitivity phenotype (Jungmann *et al*., 1993). Functional studies at the genetic level indicate that Ubc7 protein is involved in the ubiquitin-dependent degradation of a diverse group of proteins including the repressor MATα2 protein (Chen *et al.*, 1993), a β-galactosidase construct with a nonapeptide insert (Sadis *et al.*, 1995), and mutated forms of the ER-membrane protein-translocating complex, Sec61p (Biederer *et al.*, 1996), and the soluble vacuolar carboxypeptidase, yscY (Hiller *et al.*, 1996).

We report here the determination of the three-dimensional structure of the yeast Ubc7 protein. The overall folding of this molecule is similar to the two other Class I E2 enzymes whose structures have been determined (Cook *et al*., 1992, 1993), although it shares only 38% sequence identity with *Arabidopsis* Ubc1 and 37% with yeast Ubc4. Comparison of these three structures reveals a surface region that may be the E1-binding site common to all E2 enzymes as well as divergent surface regions that may enable an individual class I enzyme to interact with its respective substrates or E3.

EXPERIMENTAL PROCEDURES

Expression of recombinant *S. cerevisiae* Ubc7 in *Escherichia coli* and purification have been described previously (Yamazaki & Chau, 1996). The purified protein is functional as judged by its ability to form the ubiquitin−thiol ester adduct. Crystals of Ubc7 were initially obtained by vapor-diffusion equilibration using a screening kit from Hampton Crystal (San Diego, CA) that is based on a modified sparse matrix sampling method (Jancarik & Kim, 1991). These drops consisted of 1 μL of a solution containing 10 mg of protein/mL of distilled water plus 1 μL of the reservoir solution. They were equilibrated against 0.7 mL of each trial solution at 23 and 4 °C. Within 24 h one set of conditions at 23 °C gave aggregates of hexagonal prisms. Further screening around these conditions produced a few small crystals suitable for X-ray diffraction studies. The best crystals were grown from 0.6 M sodium citrate at pH 7.4, and 0.1 M HEPES buffer.

Intensity data to 2.93 Å resolution were collected from a crystal at 22 °C with a Nicolet X-100A area detector using Cu Ka radiation from a Rigaku RU-300 rotating anode generator operating at 40 kV and 100 mA. Data sets were collected from two crystals. The detector-to-crystal distance was 14 cm, and the detector 2θ value was 10°. Oscillation frames covered 0.25° and were measured for 450 s. The crystals diffracted rather weakly and suffered significant decay within 24 h. Therefore, only 320 frames were collected from the first crystal and 280 from the second. A total of 25 491 reflections was processed and merged into 6557 unique reflections (93% complete at 2.93 Å resolution); the R_{sym} value (based on I) for these data was 15.4%. The highest resolution range (2.93−3.11 Å) contained 720 unique reflections (62% complete), but only 192 of these reflections (27%) had $I > 2\sigma(I)$. Indexing and integration of intensity data were carried out using the XENGEN processing programs (Howard *et al*., 1987).

Of the potential solutions generated by the indexing program, one had angles and axial lengths consistent with a hexagonal space group, and it was the only solution that successfully indexed all reflections. The cell parameters refined to $a = b = 106.5$ Å and $c = 49.3$ Å. The unit cell was tested for the two possible hexagonal Laue symmetry groups by comparing the integrated intensities of potentially equivalent reflections. This analysis confirmed the Laue group 6/*m*. The systematic absence of reflections 00*l* with $l \neq 6n$ indicated either space group $P6_1$ or $P6_5$. Based on a molecular mass of 18 521 Da as predicted from the cDNA, the calculated values of V_m (Matthews, 1968) for one or two molecules/crystallographic asymmetric unit are 4.36 and 2.18, respectively. These values correspond to solvent volume fractions of 72% and 44%, respectively, either of which would be reasonable.

The molecular replacement routines in X-PLOR (Brünger *et al*., 1987) were used to solve the crystal structure. To determine if there were local symmetry elements, a self-rotation function of the Patterson map was calculated in the resolution range of 12−4 Å, using a Patterson vector length of 25 Å. The product correlation function in direct space was calculated in the polar angle system, where κ is the rotational angle. The self-rotation function showed only symmetry characteristic of the space groups $P6_1$ or $P6_5$, i.e., 2-fold and 3-fold axes parallel to the *c* axis that had the same rotation function value as the identity operator. If there were two molecules in the asymmetric unit, the noncrystallographic local symmetry axis would have to be parallel to the crystallographic *c* axis and hidden under the crystallographic symmetry operators. There was, however, no indication that a second molecule was present from the self-rotation function.

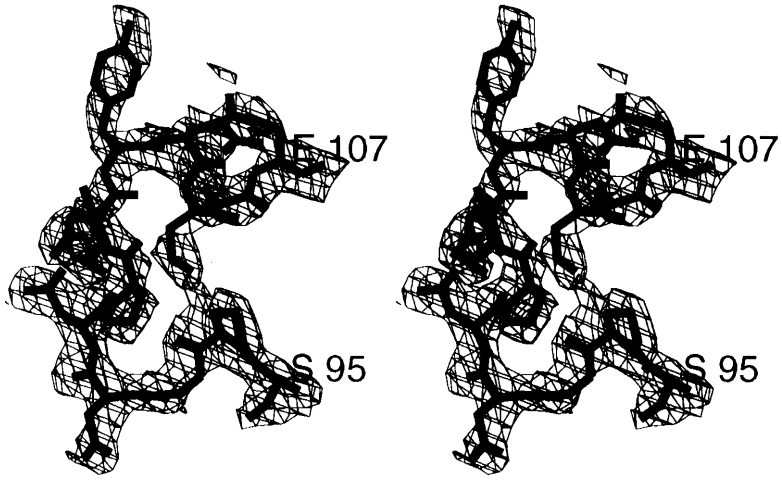

FIGURE 1: Stereoview of residues 95−107 in the Ubc7 structure with the associated electron density. The $2F_o - F_c$ map was calculated with phases based on the refined structure after these 13 residues were removed. Figures 1, 3, and 4 were prepared using CHAIN (Sack, 1988).

For the cross-rotation functions, the 2.4 Å crystal structure of Ubc1 from the plant *Arabidopsis thaliana* (Cook *et al*., 1992) was used as the search model. *Arabidopsis* Ubc1 has 152 amino acids and 38% sequence identity with yeast Ubc7. Compared to *Arabidopsis* Ubc1, yeast Ubc7 has two insertions. One is a single extra residue in the loop between the first and second strands of the β-sheet. The second is a 13-residue insertion that occurs two residues after the active site cysteine residue. The entire model (residues 1−150) was placed in a 120 Å orthogonal P1 unit cell with the center of mass at the origin, and a cross-rotation function was calculated using a radius of integration of 25 Å and data between 12 and 4 Å resolution. All of the peaks from the rotation search were subjected to Patterson correlation (PC) refinement (Brünger, 1990). This refinement produced a clear, single maximum peak at $\theta_1 = 180.6°$, $\theta_2 = 47.8°$, and $\theta_3 = 59.7°$.

The model was rotated by the angles corresponding to the rotation solution, and the translation function from X-PLOR was used with data from 12 to 4 Å resolution for each of the two enantiomorphic space groups. Comparison of the best translation function solutions for the two space groups showed that the solution using space group $P6_5$ gave the higher translation function value and the lower *R*-factor, using data from 8 to 4 Å resolution. The translation function solution had a correlation coefficient of 0.33, which was the maximum in the calculation, and was 10.1σ (0.023) above the mean (0.10) and 5.3σ above the next highest peak (0.21). On the basis of these results, the space group $P6_5$ was chosen. The translation vector for this solution was 0.145 in *x*, 0.550 in *y*, and 0.0 in *z* (since this is a polar space group, the relative *z* coordinate is arbitrary).

Prior to refinement and phase calculation, all residues in the model which differed from Ubc7 were truncated to alanine. The isotropic temperature factors for all the atoms were set to 30 Å^2 (from a Wilson plot). Refinement of this model for 25 cycles by rigid body refinement, using 2σ data from 7 to 2.9 Å resolution, gave an initial crystallographic *R*-factor of 44.4%. Data above 7 Å were not included in the refinement, since they would be especially influenced by disordered solvent. A $2F_o - F_c$ map revealed positions for the nonconserved side chains, which were modeled into the density. Additional density for the two insertions was also evident, and we were able to fit the single residue insertion but not the 13-residue stretch. Positional refinement (120 cycles, $R = 31.7\%$) followed by phase improvement with the program SQUASH (Zhang, 1993) allowed us to fit the long insertion. This model was subjected to a 4000° slow cool, positional refinement, and then several more rounds of simulated annealing alternating with examination of $2F_o - F_c$ to adjust manually residues with poor geometry. Using isotropic temperature factors of 15.0 Å^2 for all atoms, the *R*-factor at 2.93 Å resolution was 20.4%. Individual thermal parameters were then introduced with restraints on the differences of temperature factors between connected atoms; this reduced the *R*-factor to 18.3%, based on 4997 reflections in the range 2.93 Å $\leq d \leq$ 7.0 Å with $F > 2\sigma(F)$. Table 1 lists the statistical parameters describing the refinement.

Table 1: Crystallographic Data and Refinement Statistics

space group	$P6_5$
unit cell	$a = b = 106.5$ Å, $c = 49.3$ Å
residues in model	3−165
refinement	
resolution range	7.0−2.93 Å
no. of reflections ($F > 2\sigma$)	4997
completeness	77.1%
R-factor[a]	18.3%
thermal parameters	
$<B>$ for main chain atoms	10.2 Å^2
$<B>$ for side chain atoms	12.5 Å^2
stereochemistry	
rms nonideality bond lengths	0.016 Å
rms nonideality bond angles	2.3°
rms dihedral angle	26.0°
rms improper angles	1.9°
Ramachandran plot	
residues in most favored regions	86.7%
residues in additional allowed regions	11.1%
residues in generously allowed regions	2.2%
residues in nonallowed regions	0.0%

[a] $R\text{-factor} = \sum_{hkl}|F_o(hkl) - F_c(hkl)|/\sum_{hkl}F_o(hkl)$.

The final $2F_o - F_c$ electron density map is generally of good quality (Figure 1), although the electron density for some of the side chains of hydrophilic residues on the surface is poor, presumably due to large thermal motion or disorder. The final model includes 1289 protein atoms from 163

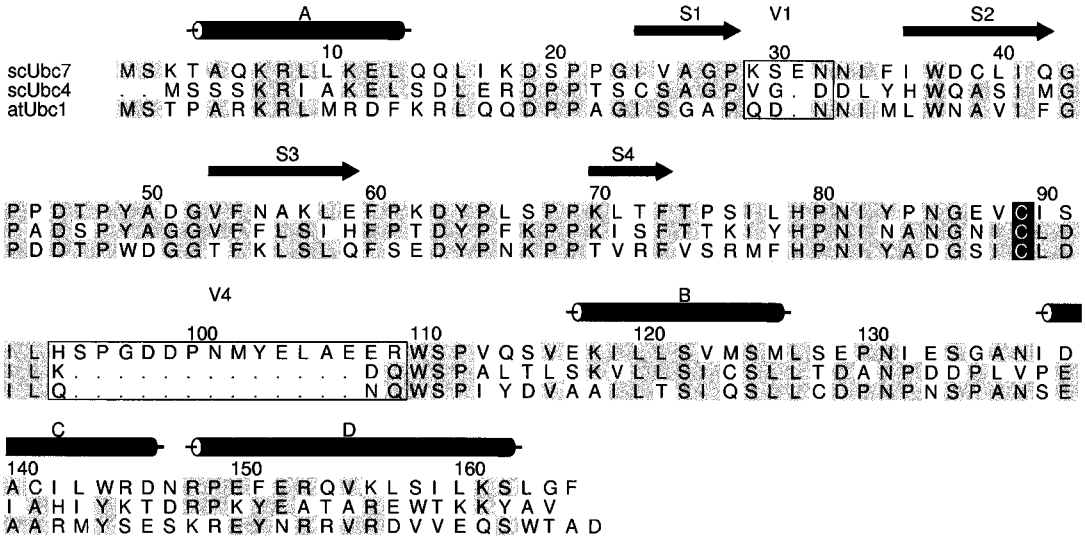

FIGURE 2: Alignment of the sequences for Ubc7, Ubc4, and AtUbc1. Secondary structural features and position numbers shown are those for Ubc7. Identities in at least two of the three sequences are shaded. The active site cysteine residues are shown with a black background.

residues; there was no significant electron density for the first two residues, presumably due to large thermal motion or disorder. Superposition of the *R*-factor curve with theoretical curves for different mean positional errors gives an estimated error of 0.27 Å in the atomic coordinates (Luzzati, 1952). The final coordinates of Ubc7 deviate from ideal bond lengths and angles by 0.016 Å and 2.3°, respectively.

Comparison of the stereochemistry of the structure with other structures solved at the same resolution using PROCHECK (Laskowski *et al*., 1993) showed that three of the five main chain parameters were better than average, and the other two parameters (ζ angle and H-bond energy standard deviations) were less than the mean. A Ramachandran plot showed that 87% of the residues were in most favored regions, and none were in disallowed regions. For the five side chain parameters, one was better than average, and the remainder were less than the mean. Averaged temperature factors for residues range from 2 to 44 Å^2, with an average *B*-factor of 11 Å^2 for all protein atoms. Coordinates for *S. cerevisiae* Ubc7 have been deposited with the Brookhaven Protein Data Bank, accession no. 2UCZ.

Multiple sequence alignments were performed with either the PILEUP program (Genetics Computer Group, Madison, WI) using the FASTA matrix or the CLUSTAL W program (Thompson *et al*., 1994) using the BLOSUM matrix. Alignments were formatted using ALSCRIPT, Version 2.0 (Barton, 1993). Accession codes for the yeast sequences are as follows: Ubc1, Swiss-Protein (SP) P21734; Ubc2, SP P06104; Ubc3, SP P14682; Ubc4, SP P15731; Ubc5, SP P15732; Ubc6, SP P33296; Ubc7, SP Q02159; Ubc8, SP P28263; Ubc9, PIR S52414; Ubc10, SP P29340; Ubc11, Saccharomyces Genome Database (SGD) Clone YOR339C; Ubc12, SGD CloneYLR306W; Ubc13, SGD Clone YDR092W.

RESULTS AND DISCUSSION

S. cerevisiae Ubc7 contains 165 amino acids and belongs to the class of E2s that consists entirely of the relatively conserved globular core domain (class I) (Jentsch, 1992). Ubc7 shares 37% residue identity with the 148-residue *S. cerevisiae* Ubc4 and the 150-residue *A. thaliana* Ubc1 (AtUbc1) whose crystal structures had previously been determined (Cook *et al*., 1992, 1993). These two previously determined structures suggested that class I enzymes are conserved in tertiary folding. However, the alignment of the Ubc7 sequence with those of Ubc4 and AtUbc1 required placing 14 of the extra residues in Ubc7 in two separate variable regions (V1 and V4, Figure 2). One of these, V1, is a single amino acid (Glu31), and the second is a 13-residue insertion (Ser95−Glu107) just distal to the ubiquitin-accepting cysteine at position 89. The final model of the Ubc7 molecule revealed a structure with an overall folding that is remarkably similar to those of Ubc4 and AtUbc1. The effect of the two predicted insertions altered two surface regions but did not significantly change the overall folding of pattern of class I enzymes seen with Ubc4 and AtUbc1.

The Ubc7 molecule is asymmetric with overall dimensions of approximately 24 Å × 41 Å × 49 Å. Figure 3 shows an α-carbon backbone of the Ubc7 molecule. Like Ubc4 and AtUbc1, there is one antiparallel β-sheet with four strands (labeled S1 to S4), bounded on each end and on one side by four helices (labeled A, B, C, and D). One face of the β-sheet is on the surface of the protein. The four β-strands include residues 24−27, 36−42, 53−59, and 71−73. The torsion angles for residues Glu87 are also consistent with β-strand geometry, and there are two hydrogen bonds to strand S4: Phe73 N−Gly86 O and Leu71 O−Val88 N. The four α-helices consist of residues 7−13, 116−128, 138−146, and 148−162. There are three short stretches of 3_{10} helix in Ubc7 (residues 91−93, 104−108 in the 13-residue insertion, and 132−134), which are not present in Ubc4 or AtUbc1.

The ubiquitin-accepting cysteine in Ubc7 is located in a long extended stretch between the fourth strand (S4) of the β-sheet and the second α-helix (B) (Figure 3). A stereoview of the environment around the ubiquitin-accepting cysteine is shown in Figure 4. Even in Ubc7, where there is a large insertion close to Cys89, the cysteine side chain is exposed and sits in a slight depression on the surface. It appears that ubiquitin can approach the crucial cysteine from only one direction. There are three loops surrounding Cys89; the closest residues are Tyr83, His94, and Ser134. The loop containing Ser134 is several ångstroms closer to the ubiq-

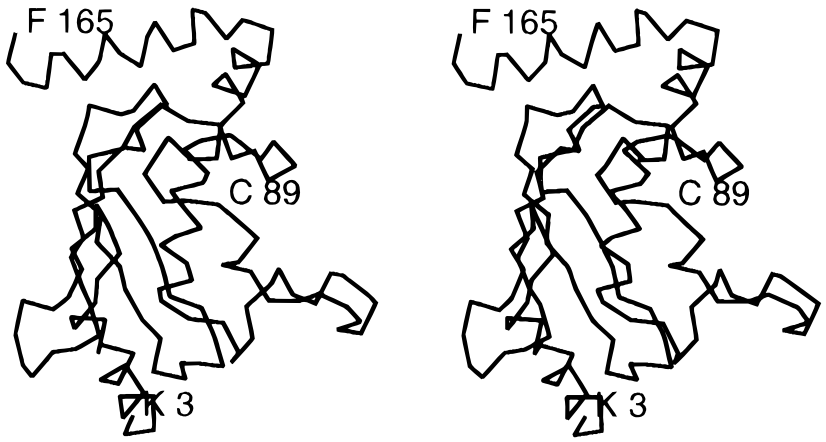

FIGURE 3: Stereo drawing of Ubc7 based on the α-carbon positions. The N- and C-termini, as well as the ubiquitin-accepting cysteine, are labeled.

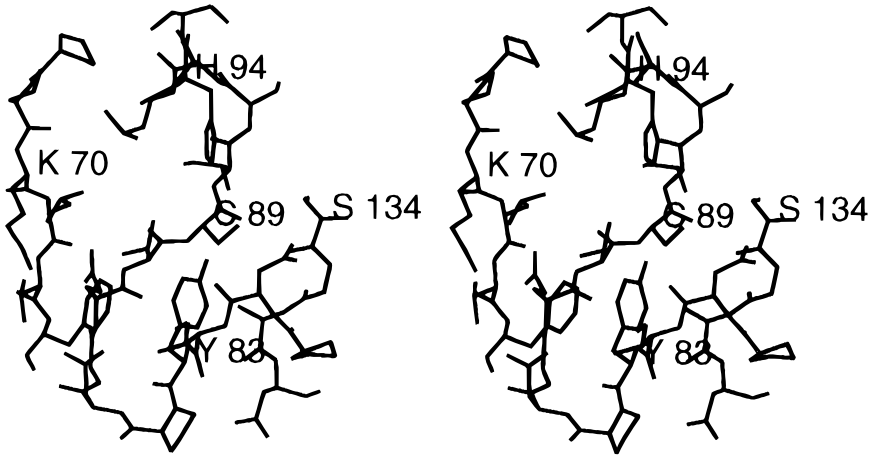

FIGURE 4: Stereo drawing of the environment around the ubiquitin-accepting cysteine in yeast Ubc7. Residues 68−73, 80−95, and 134−137 are shown.

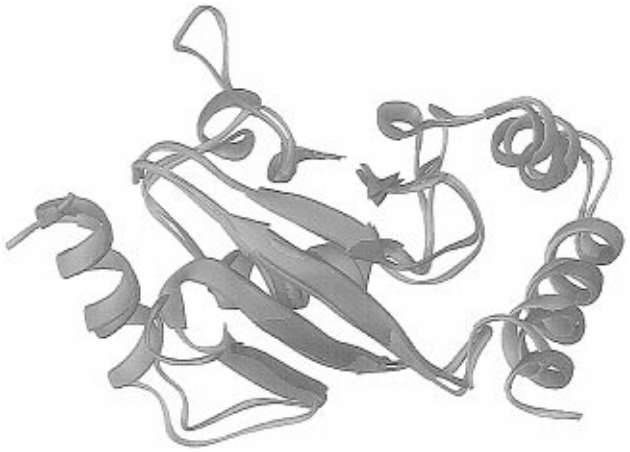

FIGURE 5: Ribbon drawing of the yeast Ubc7 molecule (red) superimposed on the yeast Ubc4 molecule (green). The superposition is based on the Cα positions for residues 3−31, 33−94, and 108−164 from Ubc7 and residues 1−29, 30−91, and 92−148 from Ubc4. The ubiquitin-accepting cysteine in each protein is yellow.

uitin-accepting cysteine in Ubc7 compared to the corresponding loop in yeast Ubc4.

The conserved tertiary folding of class I enzymes is evident from comparing the ribbon diagrams of yeast Ubc4 and Ubc7 (Figure 5). Excluding residues in the two insertions, the remaining α-carbon atoms of Ubc7 can be superimposed with corresponding atoms in Ubc4 with a root-mean-square deviation of 1.8 Å, using the superposition method of Kabsch (1976). A similar value of 1.8 Å was found for the superposition of Ubc7 with AtUbc1 (data not shown). These values are not significantly different from the value of 1.6 Å obtained for the superposition of Ubc4 and AtUbc1 (Cook *et al.*, 1993). Most differences occur at the loops and are fairly small. Compared to Ubc4, all four strands of the β-sheet in Ubc7 are slightly shorter; there are a total of 21 residues in the four strands of Ubc7 compared to 27 residues in Ubc4. Judging by the deviations in the superposition, helix A and the loop linking it to the first β-strand in Ubc7 are more similar to corresponding regions in AtUbc1 than in Ubc4. Significant deviations are also found in the superposition at the C-terminal end of helix D compared to both Ubc4 and AtUbc1. Pro65 of Ubc7 has the *cis* conformation and aligns with the previously found *cis*-Pro62 of Ubc4 and *cis*-Pro64 of AtUbc1 (Figure 2).

The most obvious differences between Ubc7 and the other two E2s are regions containing the two insertions (V1 and V4 of Figure 2). Comparison of these regions in the three crystal structures suggests that they represent hypervariable surfaces in a common E2 tertiary fold. The 13-residue insertion in the variable region V4 has the effect of introducing a much larger surface loop flanking one side of the ubiquitin-accepting cysteine. In Ubc4, this loop is comprised of Lys91−Gln93, while this region in Ubc7 is comprised of 16 residues (His94−Arg109) and contains a short stretch of 3_{10} helix and two tight turns (Figure 3).

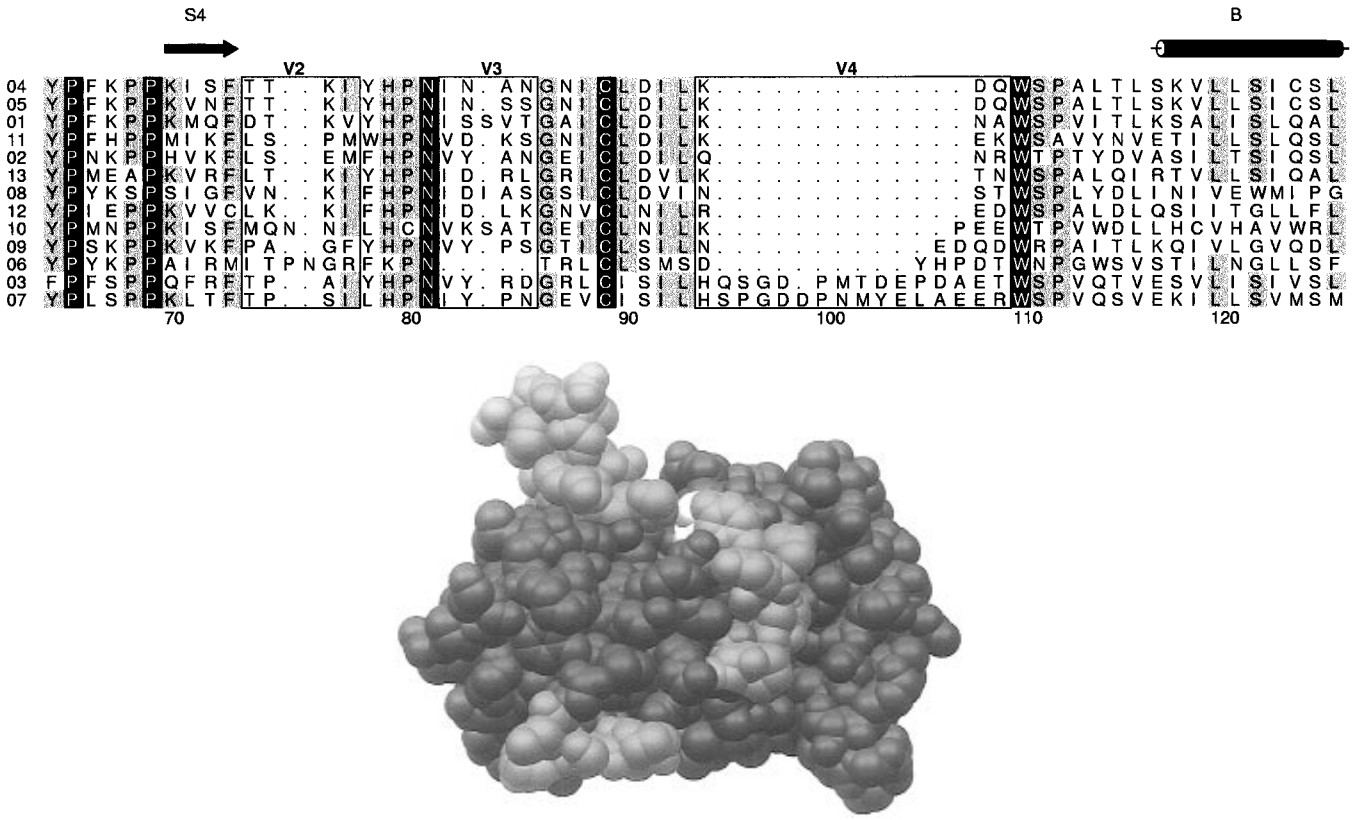

FIGURE 6: (A, top) Alignment of sequences around the active site cysteine of 13 *S. cerevisiae* ubiquitin-conjugating enzymes. Invariant residues are highlighted with a black background. Identities in seven or more sequences are highlighted in gray. The secondary structural features shown above the alignment and the numbers below refer to the Ubc7 sequence. (B, bottom) Space-filling model of yeast Ubc7. The orientation of the model is the same as in Figure 5. The four segments in Ubc7 that correspond to the variable regions (V1−V4) in Figures 2 and 6 (residues 29−32, 74−77, 82−85, 94−109) are in green, the ubiquitin-accepting Cys89 is in yellow, and all others are in red.

An alignment of the amino acid sequences surrounding the active site cysteine of the 13 E2s from *S. cerevisiae* is shown in Figure 6A. This region contains the five invariant amino acids (Pro65, Pro69, Asn81, the active site Cys89, and Trp110 according to the Ubc7 numbering) found in full-length alignments of the yeast sequences. Overall the level of similarity is much greater in this region than the N-terminal and C-terminal regions (not shown), even when those sequences with extensions are excluded from the alignment.

Of the 13 E2 sequences, seven are similar to Ubc4 in that the loop within V4 is comprised of three residues. In the remaining five sequences, this loop varies in length from 4 to 16 residues. Variations in these regions are also found in E2s from other species. It is significant that, even with 13 extra residues, the regions adjoining this loop in Ubc7 are remarkably similar to those in Ubc4. In Ubc7 this loop is flanked on the N-terminal side by a highly conserved leucine (Leu93) and on the C-terminal side by a tryptophan (Trp110), which is one of the five invariant residues in *S. cerevisiae* E2s. In the superposition of Ubc7 and Ubc4, the α-carbon atoms at the positions of Leu93 and Trp110 deviate between the two molecules with values of 0.14 and 0.30 Å, respectively, well below the RMS deviation value of 1.8 Å.

The functions of the variable insertions in V4 in general and the 13 amino acid insert in Ubc7 in particular are unknown. Yeast Ubc7 appears to be a member of a family consisting of at least six Ubc7-like proteins, all containing 12- or 13 residue inserts in the V4 region. A homolog of the yeast **UBC7** gene whose predicted protein sequence contains a similar insert has been identified in *Caenorhabditis elegans* (Swiss-Protein UBC7_CAEEL, P34477), and homologs have been cloned, expressed, and functionally studied from plant sources (van Nocker & Vierstra, 1991; van Nocker *et al*., 1996). The plant sequences are of particular interest since these proteins have been demonstrated to form multiubiquitin chains through Lys48-specific linkages. A similar 12-residue insert is present in yeast Ubc3, which also catalyzes the formation of Lys48-specific ubiquitin−ubiquitin linkages (Banerjee *et al*., 1993). However, the V4 insert does not play an obligatory role in multiubiquitin chain formation since the bovine $E2_{25K}$ protein that is very active in chain formation does not contain an insert (Chen *et al*., 1991). Further, activity for multiubiquitin chain formation cannot be detected with yeast Ubc7 (Yamazaki & Chau, 1996).

In comparison to Ubc4, the addition in region V1 of an extra amino acid in the turn between the first two β-strands of Ubc7 has the effect of shifting the tight turn by two residues. This turn lies at the surface of the molecule and protrudes slightly. Like the region of the 13-residue insertion, region V1 also seems to vary in length among *S. cerevisiae* E2 sequences. Seven of these sequences contain insertions that range from one to five residues (data not shown), and it is likely that this region also represents a hypervariable surface. There are at least two other regions in *S. cerevisiae* E2 sequences that may contain variable length insertions. These two regions are located between the last β-strand and the ubiquitin-accepting cysteine (V2 and V3 in Figure 6A). Although insertions are absent in

Ubc4 and Ubc7, these regions exhibit significant deviations in the superposed structures, suggesting that these may also be hypervariable surfaces in the conserved E2 tertiary fold.

When the four variable regions among yeast E2s are highlighted in the structure of Ubc7, it is clear that the four regions are not randomly distributed but are located on one surface (Figure 6B). Our analysis suggests that this surface is highly variable among different E2 enzymes. One possible function for such a hypervariable surface is to provide specificity for binding different E3 proteins or substrates. Experiments to test this hypothesis are currently underway in our laboratory.

REFERENCES

Banerjee, A., Gregori, L., Xu, Y., & Chau, V. (1993) *J. Biol. Chem. 268*, 5668−5675.

Barton, G. J. (1993) *Protein Eng. 6*, 37−40.

Biederer, T., Volkwein, C., & Sommer, T. (1996) *EMBO J. 15*, 2069−2076.

Brünger, A. T. (1990) *Acta Crystallogr., Sect. A 46*, 46−57.

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458−460.

Chau, V., Tobias, J. W., Bachmair, A., Marriot, D., Ecker, D. J., Gonda, D. K., & Varshavsky, A. (1989) *Science 243*, 1576−1583.

Chen, P., Johnson, P., Sommer, T., Jentsch, S., & Hochstrasser, M. (1993) *Cell 74*, 357−369.

Chen, Z., Niles, E. G., & Pickart, C. M. (1991) *J. Biol. Chem. 266*, 15698−15704.

Cook, W. J., Jeffrey, L. C., Sullivan, M. L., & Vierstra, R. D. (1992) *J. Biol. Chem. 267*, 15116−15121.

Cook, W. J., Jeffrey, L. C., Xu, Y., & Chau, V. (1993) *Biochemistry 32*, 13809−13817.

Coux, O., Tanaka, K., & Goldberg, A. L. (1996) *Annu. Rev. Biochem. 65*, 801−847.

Finley, D., & Chau, V. (1991) *Annu. Rev. Cell Biol. 7*, 25−69.

Gregori, L., Poosch, M., Cousins, G., & Chau, V. (1990) *J. Biol. Chem. 265*, 8354−8357.

Hiller, M. M., Finger, A., Schweiger, M., & Wolf, D. H. (1996) *Science 273*, 1725−1728.

Hochstrasser, M. (1995) *Curr. Opin. Cell Biol. 7*, 215−223.

Hochstrasser, M. (1996) *Cell 84*, 813−815.

Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H., & Salemme, F. R. (1987) *J. Appl. Crystallogr. 20*, 383−387.

Jancarik, J., & Kim, S.-H. (1991) *J. Appl. Crystallogr. 24*, 409−411.

Jentsch, S. (1992) *Annu. Rev. Genet. 26*, 179−207.

Johnson, E. S., Bartel, B., Seufert, W., & Varshavsky, A. (1992) *EMBO J. 11*, 497−505.

Jungmann, J., Reins, H-A., Schobert, C., & Jentsch, S. (1993) *Nature 361*, 369−371.

Kabsch, W. (1976) *Acta Crystallogr., Sect. A 32*, 922−923.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.

Luzzati, V. (1952) *Acta Crystallogr. 5*, 802−810.

Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.

Murray, A. (1995) *Cell 81*, 149−152.

Sack, J. S. (1988) *J. Mol. Graphics 6*, 224−225.

Sadis, S., Atienza, C., Jr., & Finley, D. (1995) *Mol. Cell. Biol. 15*, 4086−4094.

Seufert, W., & Jentsch, S. (1990) *EMBO J. 9*, 543−550.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.

van Nocker, S., & Vierstra, R. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10297−10301.

van Nocker, S., Walker, J. M., & Vierstra, R. D. (1996) *J. Biol. Chem. 271*, 12150−12158.

Vassal, A., Boulet, A., Decoster, E., & Faye, G. (1992) *Biochim. Biophys. Acta 1132*, 211−213.

Wilkinson, K. D. (1995) *Annu. Rev. Nutr. 15*, 161−189.

Yamazaki, R. K., & Chau, V. (1996) *Protein Expression Purif. 7*, 122−127.

Zhang, K. Y. J. (1993) *Acta Crystallogr., Sect. D 49*, 213−222.

BI962639E