THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 23, Issue of June 10, pp. 22006–22011, 2005
Printed in U.S.A.
© 2005 by The American Society for Biochemistry and Molecular Biology, Inc.

Crystal Structure of a Fragment of Mouse Ubiquitin-activating Enzyme*

Received for publication, March 8, 2005
Published, JBC Papers in Press, March 16, 2005, DOI 10.1074/jbc.M502583200

Roman H. Szczepanowski‡§¶, Renata Filipek‡§¶, and Matthias Bochtler‡§‖

From the ‡International Institute of Molecular and Cell Biology, ul. Trojdena 4, 02-109 Warsaw, Poland and §Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01309 Dresden, Germany

Protein ubiquitination requires the sequential activity of three enzymes: a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2), and a ubiquitin-ligase (E3). The ubiquitin-transfer machinery is hierarchically organized; for every ubiquitin-activating enzyme, there are several ubiquitin-conjugating enzymes, and most ubiquitin-conjugating enzymes can in turn interact with multiple ubiquitin ligases. Despite the central role of ubiquitin-activating enzyme in this cascade, a crystal structure of a ubiquitin-activating enzyme is not available. The enzyme is thought to consist of an adenylation domain, a catalytic cysteine domain, a four-helix bundle, and possibly, a ubiquitin-like domain. Its adenylation domain can be modeled because it is clearly homologous to the structurally known adenylation domains of the activating enzymes for the small ubiquitin-like modifier (SUMO) and for the protein encoded by the neuronal precursor cell-expressed, developmentally down-regulated gene 8 (NEDD8). Low sequence similarity and vastly different domain lengths make modeling difficult for the catalytic cysteine domain that results from the juxtaposition of two catalytic cysteine half-domains. Here, we present a biochemical and crystallographic characterization of the two half-domains and the crystal structure of the larger, second catalytic cysteine half-domain of mouse ubiquitin-activating enzyme. We show that the domain is organized around a conserved folding motif that is also present in the NEDD8- and SUMO-activating enzymes, and we propose a tentative model for full-length ubiquitin-activating enzyme.

Ubiquitination is an important post-translational protein modification, with roles in a wide variety of cellular processes (1), including the cell cycle (2), the inflammatory response (3), the innate and adaptive immune response (4), DNA repair (5), and programmed cell death (6). In most cases, the ubiquitin pathway targets proteins for destruction by the 26 S proteasome (7), but other "non-traditional" effects of ubiquitination are known as well (8).

Ubiquitin-activating enzyme (Ubiquitin-E1)[1] catalyzes the first step of the ubiquitination pathway. The enzyme consumes ATP to attach ubiquitin to the active site cysteine residue of the enzyme in a labile thioester linkage, which allows the transfer of ubiquitin to various ubiquitin-conjugating enzymes. All available data are consistent with a three-step mechanism for the reaction. In the first step, ATP is consumed to convert ubiquitin to ubiquitin adenylate, and pyrophosphate is produced as a byproduct. In the second step, the catalytic cysteine residue of the enzyme attacks the adenylate to form a thioester and AMP. In the third step, ubiquitin is transferred to the cysteine residue of a ubiquitin-conjugating enzyme in a *trans*-thiolation reaction. Detailed kinetic studies have shown that ubiquitin activation proceeds by an ordered mechanism. ATP binding occurs first, followed by ubiquitin binding and finally adenylate formation. Formation of a new ubiquitin adenylate on the activating enzyme is thought to promote *trans*-thiolation of the thioester-linked ubiquitin to a conjugating enzyme (9).

Ubiquitin-E1 has not been crystallized yet, but structures for activating enzymes (E1s) of other ubiquitin-like modifiers (Ubls) are available. The first structurally characterized eukaryotic Ubl-E1 was the APPBP1·UBA3 complex, the NEDD8-E1, which was crystallized in the presence and absence of NEDD8 (10, 11) and also with a peptide and the core domain of the NEDD8-E2 (12, 13). Very recently, the structure of Sae1/Sae2, which acts as SUMO-E1, has been published (14). Based on weak sequence similarity to these proteins, the structure of ubiquitin-E1 is thought to consist primarily of the adenylation domain and the catalytic cysteine domain, plus a ubiquitin-like domain at the C terminus of the protein (11, 15). Residues in ubiquitin-E1 that are equivalent to a short four-helix bundle domain in NEDD8 have been included in the adenylation domain (Fig. 1).

Unlike the NEDD8- and SUMO-E1s, which are encoded on two separate polypeptide chains, ubiquitin-E1 is encoded by a single open reading frame with sequence similarity to ABBPB1 and Sae1 in its N-terminal part and to UBA3 and Sae2 in its C-terminal part (11, 14). The domain organization is not readily apparent from the amino acid sequence, because the catalytic cysteine domain is discontinuous and interspersed into the adenylation domain. Here, the two parts of the catalytic cysteine domain will be referred to as the first catalytic

* This work was supported by the Polish Ministry of Scientific Research and Information Technology (decision KO89/PO4/2004). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1Z7L) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ These authors contributed equally to this work.

‖ To whom correspondence should be addressed. Tel.: 48-22-5970732; Fax: 48-22-5970715; E-mail: MBochtler@iimcb.gov.pl.

[1] The abbreviations used are: E1, activating enzyme; E2, conjugating enzyme; E3, ligase; FCCH, first catalytic cysteine half-domain; FH, first half-domain; SCCH, second catalytic cysteine half-domain; FOM, figure of merit; SUMO, small ubiquitin-like modifier; NEDD8, neuronal precursor cell-expressed developmentally downregulated gene 8 (encoded protein); Sae1, SUMO-activating enzyme, subunit 1; Sae2, SUMO-activating enzyme, subunit 2; APPBP1, amyloid β precursor protein binding protein 1; UBA3, ubiquitin-activating enzyme 3; MAD, multiple wavelength anomalous diffraction; NCS, noncrystallographic symmetry; Ubl, ubiquitin-like modifier.

This paper is available on line at http://www.jbc.org

cysteine half-domain (FCCH, Fig. 1, *2*) and the second catalytic cysteine half-domain (SCCH, Fig. 1, *4* and *5*), even though the two halves differ in molecular weight in both the ubiquitin-E1 and the homologues.

Sequence conservation among Ubl-E1s is highest for the adenylation half-domains, which are also homologous to each other. Therefore, it is very likely that the adenylation domain of ubiquitin-E1 resembles the pseudodimeric adenylation domains of the NEDD8-E1 and SUMO-E1, and also the bacterial, dimeric MoeB protein (16). In contrast, no confident homology model can be built for the catalytic cysteine half-domains. Both half-domains differ significantly between E1s for different ubiquitin-like modifiers (Fig. 2).

The FCCH consists of two antiparallel β-strands and a disordered region in SUMO-E1. It is larger (~100 residues) in ubiquitin-E1 (~100 residues) and largest in the NEDD8-E1 (~230 residues), where this half-domain is almost entirely helical. The FCCH does not contain the catalytic cysteine residue, and its function in the various Ubl-activator proteins is not known (Fig. 2).

The SCCH is built around a short core motif (~80 residues), which is present in all Ubl-activator proteins and includes the active site catalytic cysteine residue. In NEDD8-E1, this core region represents the entire SCCH. In SUMO-activating enzyme, the SCCH is expanded by an ~140-residue insertion, which exceeds the core region in size. An even larger, unrelated insertion is present in ubiquitin-E1. The function of the insertions in the SCCH is presently unclear.

In this work, we present a biochemical characterization of the two catalytic cysteine half-domains and the crystal structure of the SCCH, and we propose a tentative model for the structure of ubiquitin-activating enzyme.

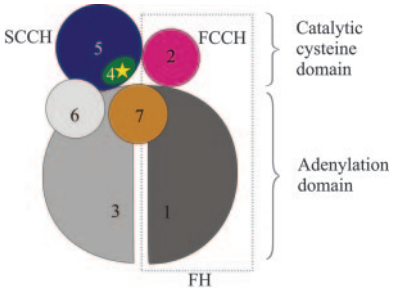

FIG. 1. **Schematic diagram of the likely domain structure of ubiquitin-E1 (*1–6*) in complex with ubiquitin (7).** The fragments of ubiquitin-E1 are first adenylation half-domain (*1*), FCCH (*2*), second adenylation half-domain (*3*), SCCH conserved region (*4*), SCCH poorly conserved region (*5*), and Ubl-domain (*6*). The *dotted rectangle* marks the extent of the FH, which consists of the first adenylation half-domain and the FCCH. The *yellow star* marks the location of the catalytic cysteine residue.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Protein Purification—The clone for mouse ubiquitin-activating enzyme was a kind gift from Prof. H. Seino, National Institute of Genetics, Mishima, Japan. Standard PCR techniques were used to amplify fragments corresponding to residues 202–312 (FCCH construct), residues 626–891 (SCCH construct), and residues 1–439 (FH construct), according to Swiss Protein Data Base entry Q02053 numbering. These fragments were cloned via EcoRI and XhoI into pET15b(+) mod, which was created from pET15b(+) (Novagen) by a deletion of the original EcoRI site and by a modification of the multiple cloning site, which eliminated the thrombin cleavage site and placed an EcoRI site immediately downstream of a vector-encoded histidine tag. The amino acid sequence of all constructs was MGHHHHHHEF, which was directly followed by the sequence of the ubiquitin-E1 fragments. For protein expression, the plasmids were transformed into *Escherichia coli* BL21(DE3) cells, which were grown at 37 °C to an A_{600} of 0.7–1.0, shifted to 28 °C for 0.5 h, induced with 0.5 mM isopropyl β-D-thiogalactopyranoside, and harvested 4 h after induction. After affinity chromatography on nitrilotriacetic acid-agarose loaded with Ni^{2+} (Qiagen), the proteins were subjected to a gel filtration step in 5 mM Tris, pH 7.5, 1 mM EDTA on Sephacryl S-200 HR (Amersham Biosciences) and concentrated by ultrafiltration.

CD Measurements—CD measurements were performed with a Jasco J-810 spectrometer equipped with a thermostated cell holder using a 0.02-cm cell. Three scans at 20 °C between 250 and 190 nm were recorded and averaged. Spectra were corrected for buffer CD signal and normalized for protein concentration based on mean residue molecular weights: 109.07 for the FCCH, 108.89 for the FH, and 112.96 for SCCH. Spectra were analyzed with SELCON3, CDSSTR, and CONTIN of the CDPro software package (17).

Size Exclusion Chromatography—Gel filtration runs were performed at 24 °C using the Ettan LC fast protein liquid chromatography system (Amersham Biosciences) equipped with a Superose 12 HR 10/30 column. The running buffer was 10 mM Tris, pH 7.5, 50 mM NaCl (buffer A), and the flow rate was 0.4 ml/min. Individual proteins and equimolar protein mixtures were prepared in a final volume of 120 μl, adjusted with buffer A, and incubated at 25 °C for 10 min. Vitamin B_{12} (molecular mass 1.35 kDa) myoglobin (17 kDa), ovalbumin (44 kDa), bovine γ-globulin (158 kDa), and thyroglobulin (670 kDa) (all from Bio-Rad) were used for calibration.

Crystallization—For crystallization, the SCCH was concentrated by ultrafiltration to 42 mg/ml and supplemented with 2 mM of the reducing agent tris(2-carboxyethyl)phosphine hydrochloride. Crystallization trials were set up at room temperature (21 °C) as sitting drop vapor diffusion experiments with 0.5 ml of reservoir buffer, by mixing 2 μl of reservoir buffer with 2 μl of protein solution. Crystals were obtained from a range of reservoir buffers, 0.075 M Tris, pH 8.3–8.7, 1.8–2.0 M ammonium sulfate, and 30–33% glycerol, and belonged to space groups H3 and H32. Choosing low ammonium sulfate concentration and a pH value at the extremes of the crystallization condition appeared to favor the growth of the H3 form, but sometimes both forms grew in the same drop. H32 crystals routinely appeared overnight, H3 crystals grew somewhat slower and were distinctly different morphologically. Both crystal forms could be flash-cryocooled in mother liquor.

Structure Determination—In-house screening identified H32 crystals soaked with the tantalum bromide cluster as promising derivatives, and thus a three-wavelength MAD dataset was collected at the tantalum edge at BW6, Deutsches Elektronen-Synchrotron (absorption wavelength 1.2546 Å, inflection wavelength 1.2551 Å, remote wave-

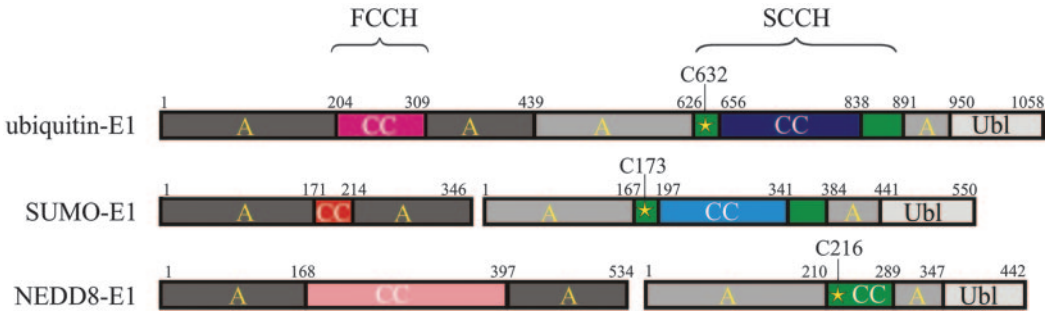

FIG. 2. **Domain organization mapped to linear sequence for the ubiquitin-E1 (*top*), SUMO-E1 (*middle*), and NEDD8-E1 (*bottom*).** The adenylation half-domains are shown in *dark* and *light gray*, the FCCH in different *shades* of *red*, the SCCH in *green* and *blue* (representing the well conserved and poorly conserved parts, respectively), and the Ubl domain in *very light gray*. *A*, adenylation domain; *CC*, cysteine catalytic domain. The catalytic cysteine residue is marked by a *yellow star*. Color coding is consistent with that in Fig. 1.

length 1.0500 Å; see Table I). Bijvoet differences that were measured at the peak and inflection wavelengths were strongly correlated up to a resolution of 5 Å (correlation >0.6 for the range 8.0–5.0 Å; correlation <0.3 for the range 4.7–3.3 Å), suggesting that it would not be possible to resolve individual tantalum sites. The SHELXD program (18) identified three cluster sites (weights 1.0, 0.89, 0.80 for the correct sites, 0.18 for the first noise site). Phasing with the SHELXE program (19) showed a clear preference for one hand (after 20 cycles, contrast 0.559 for the correct enantiomer *versus* 0.108 for the wrong enantiomer). The MLPHARE program (20) was then used for the calculation of an optimized electron density map. As the orientation of the clusters was not resolved, they were represented by single atoms in their centers with very high B-factors (refined to values between 150 and 250 $Å^2$), and phasing was truncated at 4.5 Å (FOM 0.62 for the range 20.0–4.5 Å). The SIGMAA program (20) was then used to combine the MAD phases with the SIR phases that were obtained from the in-house Ta_6Br_{14} data and a native dataset collected at BW6. After solvent flattening and histogram matching with DM (Density Modification) software (20) to improve the phases and extend them to higher resolution (FOM 0.51 for the range 20.0–2.8 Å), the resulting map was still of insufficient quality for manual model building but could be used to derive an averaging mask and approximate NCS symmetry. Refinement of the local symmetry operators with DM and cyclic 3-fold averaging then resulted in a map of sufficient quality (FOM 0.61 for the phases in the range 20.0–2.8 Å; average correlation between NCS regions 0.80) to manually build an approximate model from secondary structure templates. As the free *R*-factor was still very poor, the model was annealed with CNS (Crystallography and NMR System) software, with diffraction data in the resolution range 10.0–3.5 Å. The resulting model, with a free *R*-factor of 49.2% for this resolution range, turned out to be a very tight, noncrystallographic trimer. Thus, it appeared probable that the H3 crystal form could contain the same trimer. Indeed, MOLREP (molecular replacement) software (20) yielded clear signals in both the rotation and translation search prodedures, when the trimer was used as a search model. The values of the normalized rotation function were 11.05, 10.11, and 9.89 for three correct solutions, and 5.17 for the highest scoring incorrect solution. The translation function was also easily interpreted. The correlation was 47.8% for the best correct solution *versus* 31.3% for the highest-scoring incorrect solution. Multicrystal averaging with DMMULTI (Density Modification for Multiple Crystals) software (20) yielded a much improved map (average correlation between NCS-related molecules, 0.92). This map (FOM 0.81 for phases in the range 20.0–2.8 Å) was of sufficient quality to allow confident tracing of nearly the complete protein. Throughout the trace, most of the side chains were visible, and the sequence could be assigned starting from the easily identified fragment WGDCVTWACHHW, with three characteristic tryptophan residues. Because of the slightly better diffraction and the availability of experimental phases, the H32 crystal form was chosen for further refinement, which was carried out with CNS software (21) and NCS restraints with standard weights. The final model comprises 255 residues, with a gap of six residues in a disordered region (residues 815–820). Three bound tantalum bromide clusters that were used for phasing have been modeled as well. The orientations of these clusters could not be resolved and have been chosen arbitrarily in the final model. The refinement statistics for the final model are summarized in Table II.

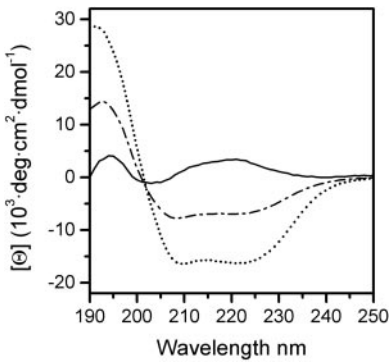

FIG. 3. **Circular dichroism spectra of the FCCH (*continuous line*), the FH (*dash-dotted line*), and the SCCH (*dotted line*).**

RESULTS AND DISCUSSION

Expression of the Two Cysteine Catalytic Half-domains—The determination of crystal structures of Ubl-E1s has helped to elucidate the complex domain structure of ubiquitin-E1, which would have been difficult to deduce from the amino acid sequence alone. We took advantage of the structural information to delineate the domain boundaries of mouse ubiquitin-E1 and expressed the two cysteine catalytic half-domains recombinantly in *E. coli*. As we were skeptical about the ability of the short FCCH to fold autonomously, we expressed this fragment both alone and as part of a larger fragment (residues 1–439) that comprised the first adenylation and catalytic cysteine half-domains. As this fragment represents roughly the first half of mouse ubiquitin-activating enzyme, it will be referred to as the FH fragment. All proteins were produced as fusion proteins with an N-terminal histidine tag and were purified in mg amounts by standard affinity chromatography techniques (see "Experimental Procedures").

The Two Cysteine Catalytic Half-domains Can Fold Autonomously—To assess the folding of the recombinantly expressed proteins, CD spectra were collected (Fig. 3). The CD spectrum for the FCCH indicated that the protein was ~40% β-structure and contained either very little or no helix, in agreement with sequence-based secondary structure predictions but contrary to prior speculation about similar folds of FCCH and its counterpart in the NEDD8 activator. As β-proteins can be difficult to distinguish from unfolded proteins by circular dichroism, we next checked protein folding by NMR. The spectra indicate a slight tendency of the FCCH to aggregate, particularly in low salt, but peak dispersions were clearly incompatible with a fully unfolded protein (data not shown). In sizing chromatography, the FCCH migrates with an apparent molecular mass of 18.4 ± 4.7 kDa, slightly higher than the calculated mass of 13.2 kDa (Fig. 4*A*).

The FH was less prone to aggregation than the FCCH. This fragment (roughly equivalent to APPBP1 and Sae1) migrated with an apparent molecular mass of 59.4 ± 14.5 kDa, in acceptable agreement with the calculated molecular mass of 49.0 kDa (Fig. 4*B*). The helical features in its CD spectrum (~25%

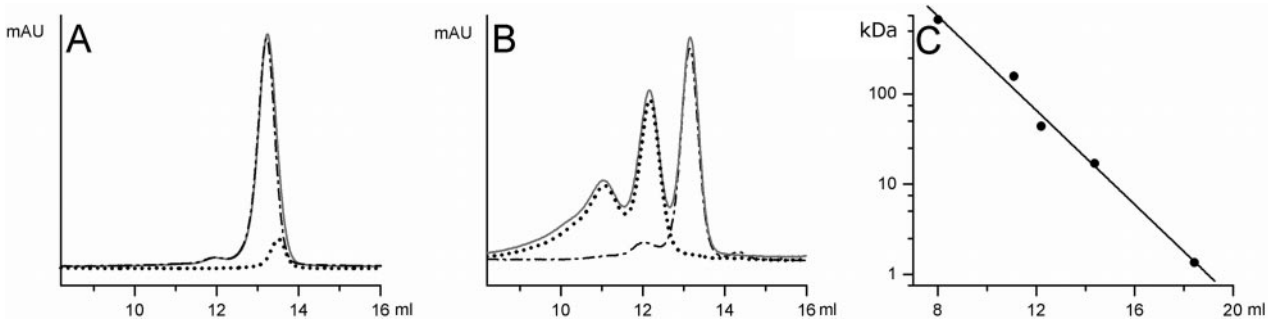

FIG. 4. *A*, lack of comigration between the FCCH (*dotted trace*) and the SCCH (*dashed trace*). The *continuously drawn trace* was obtained when the two components were co-injected. *B*, lack of comigration between the FH (*dotted trace*) and SCCH (*dash-dotted trace*). The *continuously drawn trace* was obtained for the equimolar mixture. *C*, calibration curve for molecular mass estimation.

TABLE I
Data collection and phasing statistics

The SIR[a] data contribute little to phasing (FOM improvement from 0.62 to 0.64), because the phase information in the SIR data is redundant with the phase information in the dispersive MAD differences.

Space group	H32					H3
Phasing protocol	SIR		MAD			
Crystal	Native	Ta_6Br_{14}		Ta_6Br_{14}		Native
a, b (Å)	216.5	215.8		216.1		212.9
c (Å)	195.2	196.1		196.1		98.6
Wavelength (Å)	1.0500	1.54	1.0500	1.2546	1.2551	1.0500
Total measured reflections	64,257	34,264	446,787	288,354	225,040	105,533
Unique reflections (Bijvoet separate)			82,774	50,738	50,712	
Unique reflections (Bijvoet merged)	25,999	9,413	43,074	26,389	26,319	36,644
Resolution	15.0–3.3	15.0–4.5	20.0–2.8	20.0–3.3	20.0–3.3	40.0–2.9
Completeness (%)	98.5	89.5[b]	98.6	99.2	99.4	99.4
I/σ (all/last shell)	8.5 (3.4)	5.0 (2.7)	30 (3.5)	28 (7.1)	25 (6.7)	21 (4.0)
R_{sym} (all/last shell) (%)	6.7 (21.0)	14.5 (31.5)	7.1 (23.0)	6.9 (23.0)	6.4 (21.6)	5.5 (28.5)
Number of clusters	0	3		3		0
FOM	0.31 (15–4.5 Å)			0.62 (20.0–4.5 Å)		NA[c]
FOM (combined)	0.64 (20.0–4.5 Å, SIGMAA)					NA
FOM (after 3-fold averaging solvent)	0.61 (20.0–2.8 Å, DM)					NA
FOM (after 2-crystal averaging)	0.81 (20.0–2.8 Å, DMMULTI)					

[a] SIR, single isomorphous replacement.
[b] Thin resolution shells had to be excluded from processing, because fine Ta_6Br_{14} precipitate gave rise to Debye-Scherrer rings.
[c] NA, not applicable.

helix, 25% β-sheet, 20% turn) are likely because of the helices in the first adenylation half-domain.

The SCCH is predominantly helical (~50% helix, 7% β-sheet, 15% turn). In sizing chromatography, the fragment migrated as a monomer, with an apparent molecular mass of 31.8 ± 7.7 kDa, in excellent agreement with the 31.2 kDa calculated mass (Fig. 4*B*).

The Two Cysteine Catalytic Domains Do Not Comigrate—The possibility of producing folded cysteine catalytic half-domains suggested that the half-domains may act as independent units, rather than as building blocks for a tight complex. To distinguish between the two models, approximately equimolar amounts of the FCCH and the SCCH were injected either separately or together into a Superose 12 HR 10/30 column. No tendency for comigration was observed (Fig. 4*A*). As this negative result could have been due to partial folding or the slight aggregation of the FCCH, we replaced the FCCH with the larger FH, which was less prone to aggregation. As in the previous experiment, the two fragments did not appear to interact (Fig. 4*B*). We conclude that the association between the FCCH and the SCCH in ubiquitin-activating enzyme is primarily mediated by the covalent links to the adenylation domain, and only to a lesser extent, if at all, by direct noncovalent interactions between the two domains.

Structure Determination—The FCCH did not crystallize, but crystals of the SCCH could be grown in two different space groups, H32 and H3. Both crystal forms contained three monomers in the asymmetric unit that were assembled into trimers in an essentially identical manner. As full-length ubiquitin-activating enzyme does not trimerize, the arrangement is likely a crystal packing effect and will therefore not be discussed further. The two crystal forms were solved by a combination of MAD methods and multicrystal averaging (see "Experimental Procedures" and Table I). The H32 crystal form was chosen for refinement, both because MAD phases were collected for this crystal form and because it diffracted to a slightly higher resolution than the H3 form. The final model for the H32 form comprises 257 residues in each of the three subunits in the asymmetric unit, plus three tantalum bromide clusters that were used for phasing and have no physiological meaning. Stereochemical and refinement parameters of the final model reflect the comparatively poor ordering of the molecules in the crystal (Table II).

TABLE II
Refinement statistics

Resolution (Å)	10.0–2.8
Reflections work/test	39,768/2,109
Protein atoms (excluding H)	6093
Water molecules	0
R-factor (%)	27.7
R_{free} (%)	30.8
Root mean square deviation bond distance (Å)	0.01
Root mean square deviation angles (°)	1.4
Average B-factor (Å^2)	77
B(iso) from Wilson statistics (Å^2)	87
Ramachandran core region (%)	87.3
Ramachandran allowed region (%)	12.7
Ramachandran additionally allowed region (%)	0
Ramachandran disallowed region (%)	0

Structure of the Large, Second Catalytic Cysteine Half-domain—The crystal structure of the SCCH domain of mouse ubiquitin-E1 is presented in *ribbon* representation in Fig. 5*A*. In this orientation, the shape of the domain can be described as a distorted "U" with a large, central cleft in the middle. The cleft is bridged by a long and poorly structured region of the protein that lacks electron density for four residues altogether (Fig. 5*A*, *pink*). The topology of the SCCH is rather complex. Neither of the two "arms" of the "U" (subdomains) is built up from an uninterrupted stretch of amino acids. The rather complicated fold places the N- and C-terminal ends of the half-domain in close proximity. The active site cysteine (Fig. 5*A*, *green ball*) is located near the N terminus of the domain, just upstream of a very short helix. The location of cysteine residues at the N terminus of helices is thought to enhance their nucleophilicity, but in the present case, the helix is so short and irregular that any influence of the helix dipole moment appears questionable.

Comparison with the SCCHs of SUMO- and NEDD8-activating Enzyme—The SCCHs of the Ubl-E1s differ greatly in sequence, except at the N and C termini of the half-domains, where homology is clearly recognizable. The crystal structure places these conserved regions close to each other, in the region around the catalytic cysteine residue in the active site (drawn in Fig. 5*A*). Quantitative structure comparisons with the DALI (distance matrix alignment) program showed that the fold similarities were statistically significant. The scores in standard deviations above average for the superpositions of the SCCH of ubiquitin-E1 with

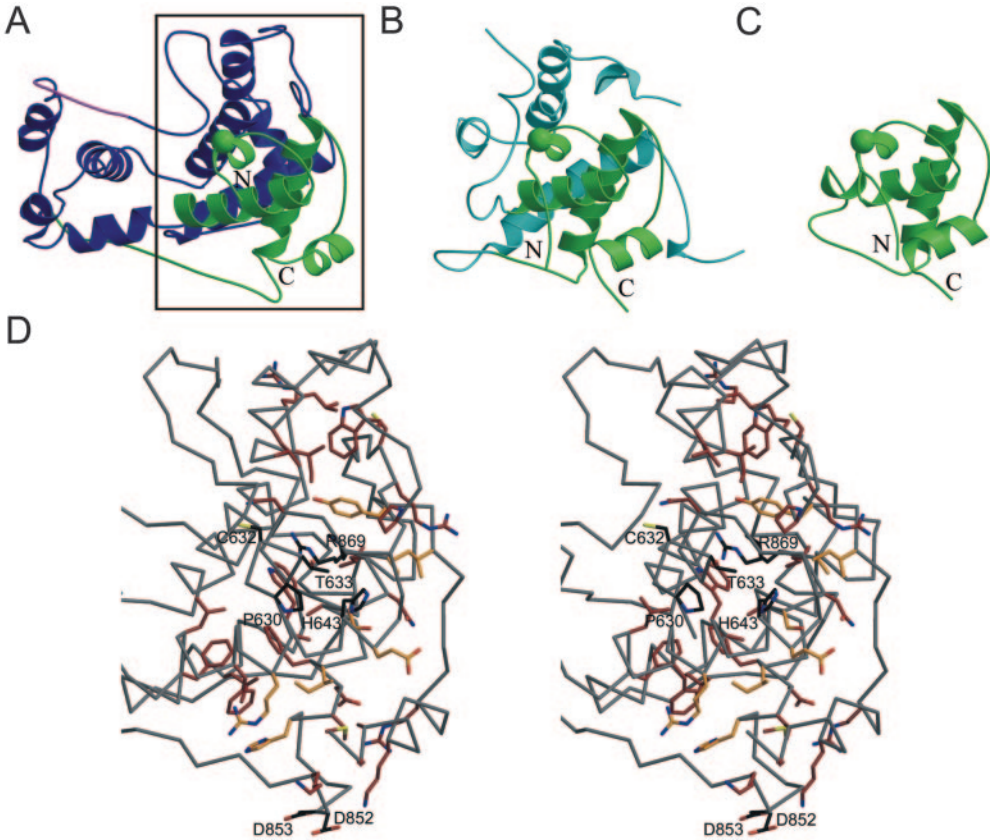

FIG. 5. Comparison of the structures of the SCCHs of ubiquitin-E1 (*A*), SUMO-E1 (*B*), and NEDD8-E1 (*C*). The coordinates for *B* and *C* were taken from Protein Data Bank entries 1R4N (10) and 1Y8R (14), respectively. The common core of the fold is shown in *green*, and the divergent regions in *blue*. *Pink* residues in *A* are disordered in the crystal structure. The *green ball* marks the location of the catalytic cysteine residue. *D*, stereo representation of the indicated region of the SCCH of ubiquitin-E1. Residues that are conserved between ubiquitin-E1 and SUMO-E1 are shown in *brown*, and residues that are conserved between ubiquitin-E1 and NEDD8-E1 are presented in *yellow*. Residues that are present in all three enzymes are drawn in *black*. The numbering of conserved residues is according to the mouse ubiquitin-E1 sequence.

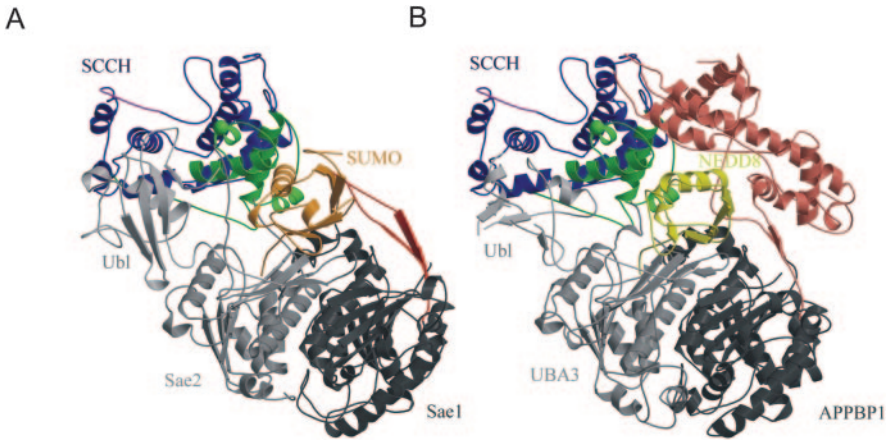

FIG. 6. **"Grafts" of ubiquitin-E1 SCCH onto the SCCH domains of SUMO-E1 (*A*) and NEDD8-E1 (*B*).** The genuine SCCH domains of SUMO-E1 and NEDD8-E1 are not shown. The location of the grafted SCCH was fixed by a superposition of the conserved core (shown in *green*) onto the equivalent core in the original SCCHs. The color coding is consistent with Figs. 1 and 2. The SUMO-E1/SUMO coordinates were taken from Protein Data Bank entry 1R4N (10) and the NEDD8-E1/NEDD8 coordinates from Protein Data Bank entry 1Y8R (14).

the equivalent half-domains of SUMO-E1 (Fig. 5*B*) and NEDD8-E1 (Fig. 5*C*) were 14.0 and 6.4, respectively, well above the threshold for random fold similarities (22).

Remarkably, the core folding motif of the SCCHs is present also in the FCCH of NEDD8-E1 (DALI score 3.1, not shown). It is not present in the very compact FCCH of SUMO-E1 and is unlikely to be found in the FCCH of ubiquitin-E1, which appears to be almost devoid of α-helix. Sequence comparisons show that residues that are strictly conserved in the SCCHs, including the active site cysteine, are not conserved in the FCCH of NEDD8-E1, suggesting that, in this case, the similarity is purely structural.

Accessory Catalytic Residues?—In all Ubl-E1s, the catalytic cysteine residue is thought to carry out a nucleophilic attack on the terminal carbonyl carbon atom of the ubiquitin-like modifier, displacing the leaving group AMP. As AMP is a good leaving group (by comparison with the more usual alcohols and amines), it is unclear whether the catalytic cysteine residue Cys-632 in mouse ubiquitin-E1 requires assistance from accessory catalytic residues. The basic residue that comes closest (5–6 Å) to the active site cysteine in the mouse ubiquitin-E1 SCCH is His-808, but this residue is not conserved among ubiquitin-E1s from different species and lacks equivalents in SUMO- and NEDD8-E1s. In addition to His-808, Lys-635 and Lys-806 of the mouse ubiquitin-E1 SCCH lie within a 10-Å shell around the sulfur atom of the active site cysteine, but in

the crystal structure, the side chains of both residues are disordered. Only Lys-806 is conserved among ubiquitin-E1s from different species, and only Lys-635 (but not Lys-806) superimposes with a spatially equivalent lysine in NEDD8-E1. Both lysines are replaced by arginines in the SUMO-E1 structure (not shown).

To identify conserved residues, we superimposed all available SCCH structures (Fig. 5*D*). Only two basic residues, His-643 and Arg-869, are present in all three structures, and both seem too far away for a direct involvement in cysteine activation. Moreover, His-643 is separated from the active site cysteine by Thr-633, a strictly conserved residue, which has been shown to be important for function in NEDD8-E1 (11). Thr-633 is unlikely to act as a general base, both chemically and structurally. In the present model, the cysteine sulfur points away from the threonine, but this rotamer assignment is unreliable because of the limited resolution of the x-ray data and because the catalytic cysteine is uncomfortably close to the domain boundary, so that packing artifacts cannot be excluded. Of course, the lack of a convincing general base residue in the SCCH does not preclude the presence of such a residue in the rest of the enzyme.

A Model for Ubiquitin-activating Enzyme—The SUMO-E1/SUMO and NEDD8-E1/NEDD8 crystal structures allow the building of tentative models of the ubiquitin-E1 ubiquitin complex by grafting the SCCH from the present crystal structure onto these structures (Fig. 6), so that the overlap between the conserved regions of the ubiquitin-E1 and SUMO-E1/NEDD8-E1 SCCHs is maximal. Some justification for this procedure can be derived from the significant sequence similarity of Ubl-E1s and from the similar orientation of the conserved region of the SCCH relative to the adenylation domain in the SUMO-E1/SUMO and NEDD8-E1/NEDD8 crystal structures. Moreover, detailed sequence comparisons strongly suggest that ubiquitin binds to ubiquitin-E1 similar to how SUMO and NEDD8 bind to their respective E1s (10, 11, 23), so that the hybrid models are predictive to some extent.

The "true" model of ubiquitin-E1/ubiquitin is likely to differ from the displayed models primarily in the FCCH. In ubiquitin-E1, this domain is intermediate in size between the FCCHs of SUMO-E1 and NEDD8-E1 and probably the predominantly β-structure. The SUMO-E1-based model for ubiquitin-E1 places the FCCH relatively far away from the SCCH, in agreement with the experimental result that the FCCH and SCCH of ubiquitin-E1 do not comigrate in sizing chromatography experiments. The model is also consistent with the lack of interaction between the SCCH and the FH. The NEDD8-E1-based model of ubiquitin-activating enzyme indicates clashes between the NEDD8-E1 FCCH and the ubiquitin-E1 SCCH in the hybrid model, but these are likely irrelevant, because the FCCH of ubiquitin-E1 is ~120 residues smaller than the NEDD8-E1.

As in the template structures (10), the hybrid models place the catalytic cysteine residue in the SCCH >30 Å away from the C terminus of ubiquitin or the ubiquitin-like modifier. In the case of the ubiquitin-activating enzyme, it remains to be seen whether this aspect of the model simply requires correction or whether movements of the catalytic cysteine half-domains, the adenylation domain, and the small modifier occur as part of the catalytic cycle.

Acknowledgments—We are grateful to Prof. Hideaki Seino, National Institute of Genetics, Mishima, Japan, for the gift of the E1 mouse c-DNA clone, to Marek Wojciechowski for help during a summer practicum, and to Prof. Robert Huber, Max Planck Institute for Biochemistry, Martinsried, Germany, for the tantalum bromide cluster. We thank Prof. Hans Bartunik Max-Planck-Arbeitsgruppen für Strukturelle Molekularbiologie, Hamburg, Germany, for generous allocation of beam time at beamline BW6 at the Deutsche Elektronensynchrotron Hamburg (DESY) and Dr. Gleb Bourenkov, Max-Planck-Arbeitsgruppen für Strukturelle Molekularbiologie, Hamburg, Germany, for assistance during data collection.

REFERENCES

1. Hershko, A., and Ciechanover, A. (1998) *Annu. Rev. Biochem.* **67,** 425–479
2. Hershko, A. (1997) *Curr. Opin. Cell Biol.* **9,** 788–799
3. Alkalay, I., Yaron, A., Hatzubai, A., Orian, A., Ciechanover, A., and Ben-Neriah, Y. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 10599–10603
4. Liu, Y. C. (2004) *Annu. Rev. Immunol.* **22,** 81–127
5. Michael, D., and Oren, M. (2003) *Semin. Cancer Biol.* **13,** 49–58
6. Jesenberger, V., and Jentsch, S. (2002) *Nat. Rev. Mol. Cell Biol.* **3,** 112–121
7. Ciechanover, A. (1998) *EMBO J.* **17,** 7151–7160
8. Schnell, J. D., and Hicke, L. (2003) *J. Biol. Chem.* **278,** 35857–35860
9. Haas, A. L., Warms, J. V., Hershko, A., and Rose, I. A. (1982) *J. Biol. Chem.* **257,** 2543–2548
10. Walden, H., Podgorski, M. S., Huang, D. T., Miller, D. W., Howard, R. J., Minor, D. L., Jr., Holton, J. M., and Schulman, B. A. (2003) *Mol. Cell* **12,** 1427–1437
11. Walden, H., Podgorski, M. S., and Schulman, B. A. (2003) *Nature* **422,** 330–334
12. Huang, D. T., Miller, D. W., Mathew, R., Cassell, R., Holton, J. M., Roussel, M. F., and Schulman, B. A. (2004) *Nat. Struct. Mol. Biol.* **11,** 927–935
13. Huang, D. T., Paydar, A., Zhuang, M., Waddell, M. B., Holton, J. M., and Schulman, B. A. (2005) *Mol. Cell* **17,** 341–350
14. Lois, L. M., and Lima, C. D. (2005) *EMBO J.* **24,** 439–451
15. Huang, D. T., Walden, H., Duda, D., and Schulman, B. A. (2004) *Oncogene* **23,** 1958–1971
16. Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V., and Schindelin, H. (2001) *Nature* **414,** 325–329
17. Sreerama, N., and Woody, R. W. (2000) *Anal. Biochem.* **287,** 252–260
18. Schneider, T. A., and Sheldrick, G. M. (2002) *Acta Crystallogr.* **D58,** 1772–1779
19. Sheldrick, G. M. (2002) *Zeitschrift für Kristallographie* **217,** 644–650
20. Collaborative Computational Project Number 4. (1994) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **50,** 760–763
21. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **54,** 905–921
22. Holm, L., and Park, J. (2000) *Bioinformatics* **16,** 566–567
23. VanDemark, A. P., and Hill, C. P. (2003) *Nat. Struct. Biol.* **10,** 244–246