Acta Cryst. (1998). D**54**, 891–898

Structure of Ubiquitin-Conjugating Enzyme 9 Displays Significant Differences with Other Ubiquitin-Conjugating Enzymes Which May Reflect its Specificity for Sumo Rather Than Ubiquitin

MARIE-FRANCE GIRAUD, JOANNA M. P. DESTERRO AND JAMES H. NAISMITH*

Centre for Biomolecular Science and School of Biomedical Sciences, Purdie Building, The University, St Andrews KY16 9ST, Scotland. E-mail: naismith@st-and.ac.uk

(*Received 26 November 1997; accepted 9 February 1998*)

Abstract

The three-dimensional structure of ubiquitin-conjugating enzyme 9 (Ubc9) has been obtained to a resolution of 2.8 Å by molecular replacement followed by a combination of automated refinement and graphical intervention. Diffraction data were recorded on a single crystal in space group $P4_3$ with cell dimensions $a = b = 73.9$, $c = 42.9$ Å. The final model has an R factor of 21.3% for all data to 2.8 Å. Only the N-terminal methionine, a two-residue N-terminal extension and a four-residue loop are not located by the final electron-density map. Ubc9 is now known to be the first sumo, a new ubiquitin-like protein, conjugating enzyme and does not conjugate ubiquitin. The structure of Ubc9 shows important differences compared with the structures of known ubiquitin-conjugating enzymes. At the N-terminal helix, the structural and sequence alignments are out of register by one amino acid giving Ubc9 a different recognition surface compared to ubiquitin-conjugating enzymes. This is coupled to a profound change in the electrostatic surface of the molecular face remote from the catalytic site. These differences may be important in recognition of other proteins in the Sumo conjugation pathway. The catalytic cysteine in Ubc9 has a positively charged lip and a negatively charged ridge nearby. Both these features seem confined to sumo-conjugating enzymes, and a sequence alignment of sumo and ubiquitin suggests how these might play a role in sumo/ubiquitin discrimination.

1. Introduction

Addition of ubiquitin to a variety of proteins is a common process in cells that leads either to the proteolytic degradation of ubiquinated proteins or to their functional modification (Finley & Chau, 1991; Rechtsteiner, 1991). Ubiquitin conjugation with target proteins involves three different steps. In the first step, which is ATP-dependent, a ubiquitin-activating enzyme (E_1) catalyzes the formation of a thioester linkage between its active-site cysteine and the terminal COOH group of ubiquitin. Ubiquitin is then transferred to the cysteine residue of an ubiquitin-conjugating enzyme (abbreviated as Ubc*n*, where *n* is an integer) also known as E_2. In the final step, ubiquitin is attached to ε-amino groups of lysine residues of the target protein. In most cases, this step requires the participation of a ubiquitin ligase (E_3), although some E_2's are able, *in vitro*, to ubiquitinate their specific substrate protein (King *et al.*, 1995). Thus, ubiquitin-conjugating enzymes play a key role in this process since they interact with at least three different proteins: E_1, ubiquitin and the acceptor protein.

Different isoforms of ubiquitin-conjugating enzymes have been characterized in a multitude of organisms and classified in subgroups according to their sequence homologies (Ubc1, Ubc4, Ubc7, Ubc9 ...). The crystal structures of three E_2's belonging to different subgroups (Ubc1 from plant *Arabidopsis thaliana*, Ubc4 and 7 from *Saccharomyces cerevisiae*) have been reported (Cook *et al.*, 1992, 1993; Martin *et al.*, 1997). Whatever the subgroup, all E_2's have a conserved core (>25% of identity) of roughly 150 residues, containing the ubiquitin-accepting cysteine. Some E_2's consist only of this core, while others have small internal insertions and/or COOH terminal extensions. E_2 constructs with deletion of the N-terminus show reduced activity, and it has been suggested that the N-terminus is important for E_1 recognition (Sullivan & Vierstra, 1991).

Yeast two hybrid screens have provided evidence that human Ubc9 binds to the human Rad51/Rad52 (Shen *et al.*, 1996), the protein encoded by the Wilm's tumor suppressor gene (Wang *et al.*, 1996), the Fas antigen (Wright *et al.*, 1996) and the adenovirus E1A oncoprotein (Hateboer *et al.*, 1996). The protein has also been shown to interact with proteins of the nuclear transport machinery (Saitoh *et al.*, 1997). Yeast Ubc9 is a nuclear protein that participates in cell-cycle regulation by targeting the degradation of cyclins (Seufert *et al.*, 1995), however attempts to demonstrate, *in vitro*, ubiquitination of cyclins failed. Desterro *et al.* (1997) and Schwarz *et al.* (1998) have recently demonstrated that Ubc9 does not conjugate ubiquitin but a ubiquitin-like protein. This protein has been called sumo (small ubiquitin-related modifier) in humans and is homologous to the yeast

Table 1. *Data collection*

Resolution range (Å)	Number of unique reflections	Average redundancy	R_{merge} (%)†	Completeness (%)
20.0–5.98	602	3.4	5.7	97.5
5.98–4.77	583	3.5	8.0	99.7
4.77–4.17	585	3.7	8.2	99.5
4.17–3.79	593	3.7	11.0	100
3.79–3.52	570	3.7	12.7	99.8
3.52–3.32	577	3.0	16.0	99.3
3.32–3.15	576	2.8	20.5	99.1
3.15–3.02	573	2.3	21.8	96.6
3.02–2.90	477	1.8	18.0	85.6
2.90–2.80	474	1.7	19.8	81.4
20.0–2.80	5610	3.0	10.0	95.9

† $R_{merge} = \sum\sum I(h)j - \langle I(h)\rangle / \sum\sum I(h)j$ where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

Smt3 protein (Johnson & Blobel, 1997). Desterro *et al.* (1997) have also shown that Ubc9 requires a distinct E_1 from the one used in the ubiquitination pathway for activation. Ubiquitin is not the only tag for protein modification and other 'ubiquitination-like' pathways involving other E_1 and other ubiquitin-like proteins like sumo could play an important role in protein processing (Desterro *et al.*, 1997; Haas *et al.*, 1987; Narasimhan, 1996; Shen *et al.*, 1996).

Here we report the X-ray crystal structure of recombinant human Ubc9 at a resolution of 2.8 Å. We compare it with the other known crystal structures of *A. thaliana* (plant Ubc1) (Cook *et al.*, 1992) and of yeast *S. cerevisiae* Ubc4 (Cook *et al.*, 1993) and Ubc7 (Martin *et al.*, 1997). The differences observed in the structure surrounding the reactive cysteine of human Ubc9 may reflect the specificity of Ubc9 for sumo and its E_1. During our study, the 2 Å structure of human Ubc9 with an N-terminal extension of eight residues was published by Tong *et al.* (1997) but with crystals belonging to different space groups than this study. The work of Tong *et al.* was published prior to the sumo-conjugating activity of Ubc9 being known.

2. Experimental procedures

2.1. *Overexpression and purification of Ubc9*

Ubc9 was cloned in pGEX-2T expression vector (Pharmacia) and DH5α cells were transformed with the resulting construct. Overexpression, glutathione Sepharose affinity chromatography and thrombin cleavage were performed as described (Jaffray *et al.*, 1995). GST and uncleaved fusion proteins were rebound onto the glutathione sepharose column. The unretained fraction was concentrated and dialysed against buffer *A* (40% ammonium sulfate, 20 m*M* sodium phosphate pH 7.2). Proteins were separated according their hydrophobic properties with an increasing gradient of buffer *B* (20 m*M* sodium phosphate, pH 7.2) on a POROS high-density phenyl HPLC column using the Biocad-Sprint system. The final yield of purified protein was 4.5 mg l^{-1} of LB culture. The recombinant protein contains a N-terminal extension of two residues (Gly-Ser) and no other modifications or deletions.

2.2. *Protein analysis*

Protein purity was judged on a denaturing electrophoresis gel (single band at an apparent molecular weight of 18 kDa) and on an iso-electric focusing gel (single band, p$I \simeq 9$). The molecular weight based on sequence is 18 152 Da with a pI of 8.7. Prior to crystallization, light scattering was performed on the protein in solution in 100 m*M* Tris–HCl, pH 7.6. The sample was monodispersed giving a Stokes radius of 24 Å (apparent molecular weight of 26 kDa).

2.3. *Crystallization and data collection*

HPLC fractions were pooled and concentrated with an amicon filter. The buffer was changed by a series of dilution and concentration steps. A total of 15 mg ml^{-1} protein was finally obtained in 10 m*M* Tris–HCl, 20 m*M* ammonium sulfate, 1 m*M* sodium phosphate. Crystals were grown by the hanging-drop vapour-diffusion method. A drop containing of 1.5 µl of protein solution and 1.5 µl of precipitant solution (30% polyethylene glycol 4000, 200 m*M* lithium sulfate, 100 m*M* Tris–HCl pH 8.25) was equilibrated against a well of precipitant solution at 293.5 K.

A crystal of 0.4 × 0.3 × 0.3 mm grew in 5 d, was mounted in a glass capillary and diffraction data were collected at room temperature using the Enraf–Nonius/MacScience DIP2000 image plate. X-rays were generated at a wavelength of 1.54 Å from an Enraf–Nonius FR591 rotating-anode generator and focused with mirrors. The crystal-to-detector distance was 140 mm. Data were recorded as 94 non-overlapping 28 min 1° oscillations and processed with *DENZO* and *SCALEPACK* (Otwinowski, 1993). The reflections were indexed in a tetragonal space group ($a = b = 73.9$, $c =$

42.9 Å). The V_m value (Matthews, 1968) was 3.25 Å^3 Da^{-1} for one molecule/asymmetric unit corresponding to a solvent fraction of 59%. The $l = 4n$ condition identified the space group as $P4_1$ or its enantiomorph $P4_3$. A summary of the data is given in Table 1.

2.4. *Structure determination*

The structure of Ubc9 was obtained by molecular replacement using *AMoRe* from the *CCP*4 suite (Navaza, 1994). Since plant Ubc1 (Cook *et al.*, 1992) displays 35% identity with human Ubc9, the atomic coordinates of Ubc1 were used for the search model. All the residues in Ubc9 differing from Ubc1 were changed to alanine. A radius of 26 Å was chosen for the rotation function with data between 12 and 3.4 Å. One maximum peak was obtained for $\theta_1 = 67.2°$, $\theta_2 = 35.6°$, $\theta_3 = 179.9°$ with a correlation coefficient of 0.14. Data between 10 and 3.8 Å were used for the translation function. One clear solution was obtained for space group $P4_3$ with a correlation coefficient of 0.4. After 20 cycles of rigid-body fitting using data between 10 and 3.6 Å the correlation coefficient was 0.43.

A random subset of data (10%) was omitted from all refinement calculations to provide an assessment of the refinement (Brünger, 1993). The molecular replacement solution gave a free R factor of 47.5% and an R factor of 47.5% for all data to 2.8 Å. A bulk solvent correction was implemented according to the *X-PLOR* manual (Brünger, 1992). After restrained positional and thermal-factor refinements using *X-PLOR*, the free R and R-factor values were of 42.1 and 33.7%, respectively. Manual adjustment was performed with *O* using $2F_o - F_c$ and $F_o - F_c$ difference electron-density maps (Jones *et al.*, 1991). The $2F_o - F_c$ map revealed positions for the non-conserved side chains that were modelled into the density. Additional density for the small insertion (residues 100–101) was clear while density for the loop 28–37, the six C-terminal residues, the N-terminal 18 residues and Pro79 and Pro80 were either absent or ambiguous. After B-factor and bulk-solvent corrections with *X-PLOR*, the free R factor was 35.9% and R factor was 31.2% using all data to 2.8 Å. Using *CNS* (Brünger *et al.*, 1998) we performed torsional dynamics. This refinement allowed us to identify two *cis*-prolines (Pro69 and Pro79), re-fit the N-terminus, the C-terminus, residues 28–31 and 36. No density was observed for the N-terminal methionine, the two extension residues or residues 32–35 and these residues were excluded from calculation of the X-ray residual. The free R value and R value were 26.0 and 21.8%. After a further round of refinement the values were 25.1% for the free R factor and 20.9% for the R factor. The final R factor for all data is 21.3%. An electron-density map is shown in Fig. 1. Statistics of the final model are given in Table 2. Coordinates and structure factors have been deposited with the PDB (Bernstein *et al.*, 1977).†

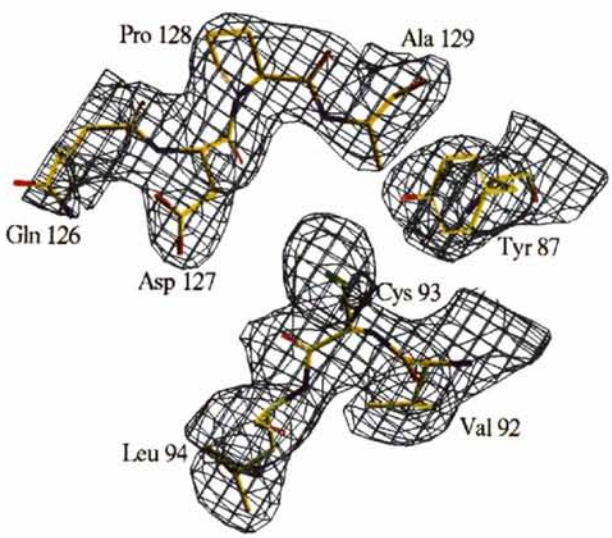

Fig. 1. The $2F_o - F_c$ electron density around the active-site residues. The phases for the fourier summation were derived from the coordinates of the final model. The map is contoured at 0.18 e Å^{-3}, 0.9σ.

Table 2. *Crystallographic data and refinement statistics*

Solvent content (%)	59
(1 molecule/asymmetric unit)	
Number of atoms	1268 (1233)†
Refinement	
Resolution range (Å)	26–2.8
Number of reflections	5604
Completeness (%)	95.9
R factor‡	20.9
Free R factor (%)	25.8
Thermal parameters (Å^2)	
Average B factor	40.9
R.m.s. $\langle B \rangle$ for bonded atoms	2.8
Stereochemistry	
R.m.s.d. bonds (Å)	0.01
R.m.s.d. angles (°)	1.73
Ramachandran plot	
Residues in most favoured regions (%)	92.7
Residues in allowed regions	7.3
Residues in disallowed regions	0

† The number of atoms included in a calculation of X-ray residual. ‡ R factor = $\sum_{hkl} |F_o(hkl) - F_c(hkl)| / \sum_{hkl} F_o(hkl)$.

2.5. *Sequence and structure analysis*

Multiple sequence alignments were performed with *DNAstar*. The r.m.s.d. distances between the different ubiquitin-conjugating enzymes were calculated with *LSQMAN* and B-factor analysis was performed with

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1A3S).

MOLEMAN (Kleywegt, unpublished programs). Ramachandran analysis was performed with *PROCHECK* (Laskowski *et al.*, 1993).

3. Results and discussion

3.1. *Overall structure of human Ubc9*

Human Ubc9 is an asymmetric protein with overall dimensions of approximately 52 × 36 × 25 Å. The protein appears to be a monomer in the crystal consistent with the light-scattering results. The protein consists of four α helices (*H*1, Ile4–Lys18; *H*2, Ile109–Asn121; *H*3, Ala131–Asn139; H4, Arg141–Lys154) and an antiparallel β-sheet composed of four strands (*S*1, Val25–Pro28; *S*2, Asn40–Pro46; *S*3, Leu57–Leu63; *S*4, Lys74–Phe77) (Fig. 2). The N-terminus is located on the opposite face from the C-terminus. Helices *H*1, *H*2 and *H*3 form one side of the protein, while the other side is composed mostly by one face of the β-sheet.

The reactive cysteine (Cys93) is located in a region of the molecule (residues 78–108) consisting of four turns and a 3_{10} helix (Ser98–Leu97). Residues Thr91 to Leu97 form a small depression in which Cys93 lies. Our structure is very similar to that determined independently by Tong *et al.* (1997), with an average r.m.s.d. for ordered C^{α} positions (150 residues) of 0.49 Å.

3.2. *Comparison with ubiquitin-conjugating enzymes*

Human ubiquitin-conjugating enzyme has 35, 33 and 36% identity with plant Ubc1, yeast Ubc4 and yeast Ubc7 (Fig. 3). Sequence alignments show that human Ubc9 has two small insertions compared with the other ubiquitin-conjugating enzymes. These two insertions consist of residues Pro32–Met36 and Asp100–Lys101. The α-carbon backbone of human Ubc9 is very similar to those of the other ubiquitin-conjugating enzymes. The root-mean-square differences are 1.88 Å for 144 C^{α} equivalent atoms of plant Ubc1 (Fig. 4), 1.31 Å for 142 C^{α} equivalent atoms of yeast Ubc4 and 1.42 Å for 145 C^{α} equivalent atoms of yeast Ubc7.

3.3. *Main differences with the other Ubc enzymes*

The main differences with the other ubiquitin-conjugating enzymes are created by the two insertions. The two-residue insertion (Asp100 and Lys101) is located as a lip near to the active Cys93. This insertion forms a small accessible loop. Interestingly, insertions at this position are also found in yeast Ubc3 (Cdc34) and wheat Ubc7 (Fig. 3), but are much longer (12 and 13 residues, respectively). The structure of this insertion in Ubc7 is completely different to that in Ubc9. Wheat Ubc7 and Ubc3 have been shown to form poly-ubiquitin chains (Banerjee *et al.*, 1993; Van Nocker & Vierstra, 1993) and it was postulated that the 12 residues played a role in poly-ubiquitination. However, yeast Ubc7 contains a 13-residue insertion in this region but appears to lack poly-ubiquitination activity (Yamazaki & Chau, 1996). Furthermore, another ubiquitin-conjugating enzyme ($E2_{25k}$) with no insertion at this position can form poly-ubiquitin chains (Chen *et al.*, 1991). Poly-ubiquitination occurs on Lys48 of ubiquitin, the corresponding residue in sumo is Gln69. (There is some debate as to whether multi sumo chains exist at all.) Thus, it seems highly unlikely that the two-residue insertion seen in Ubc9 has any role in formation of putative multi-sumo chains. The five-residue insertion Pro32–Met36 is limited to and relatively conserved in all Ubc9's (Fig. 3). This region of structure is a flexible loop and disordered in our structure. The main chain of the loop must form a protruding region near the N-terminal α-helix [and in the structure of Tong *et al.* (1997) does so].

Residues 2–6 of Ubc9 have a different orientation than plant Ubc1 (Fig. 5) and yeast Ubc4. The superposition of plant Ubc1, yeast Ubc4, yeast Ubc7 and human Ubc9 structures, shows that C^{α} backbone of each

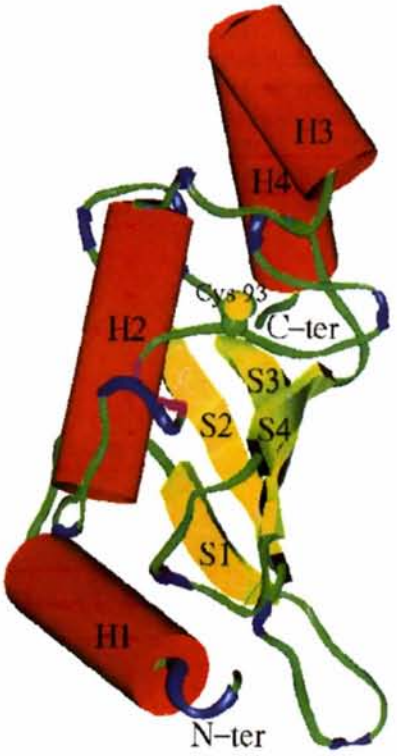

Fig. 2. Schematic representation of human Ubc9. The four α-helices are represented in red (*H*1, *H*2, *H*3, *H*4) and the four β-strands (*S*1, *S*2, *S*3, *S*4) in yellow. The 3_{10} helix is symbolized by a pink ribbon. Turns are indicated in blue. The N- and C-terminus and the reactive cysteine are labelled. Figure created with *Insight* (Molecular Simulations Inc.).

molecule's N-terminal α-helix (residues 6–18) is related to each other by small rigid-body translations, due to crystal packing and/or small changes in surrounding structure. However, despite the superposition of backbone atoms, the sequence and structural alignments do not correspond. Ubc9 is out of register by one amino acid in residues 6–13 with all other Ubc proteins (residues 7–13 in Ubc9 superimpose with 8–14 in all other Ubc's) (Figs. 3 and 5). Although Lys14 of Ubc9 could be viewed as an insertion, as it does not superimpose with any residue in any ubiquitin-conjugating enzyme structure, it is part of regular helix. A distortion is present in the helices of other structures. At residue 15 the sequence and structural alignments coincide.

```
                  10        20        30        40        50        60        70        80
                   |         |         |         |         |         |         |         |
1 MSG------IALSRLAQERKAWRKDHP-FGFVAVPTKNPDGTMNLMNWECAIPG-KKGTPWEGGLFKLRMLFKDDYPSSPPKCKFEPPLFH  83
2 MSG------IALSRLAQERKAWRKDHP-FGFVAVPTKNPDGTMNLMNWECAIPG-KKGTPWEGGLFKLRMLFKDDYPSSPPKCKFEPPLFH  83
3 MSS------LCLQRLQEERKKWRKDHP-FGFYAKPVKKADGSMDLQKWEAGIPG-KEGTNWAGGVYPITVEYPNEYPSKPPKVKFPAGFYH  83
4 MST------PARKRLMRDFKRLQQDPP-AGISGAPQDN-----NIMLWNAVIFG-PDDTPWDGGTFKLSLQFSEDYPNKPPTVRFVSRMFH  78
5 MSS------S--KRIAKELSDLERDPP-TSCSAGPVGD-----DLYHWQASIMG-PADSPYAGGVFFLSIHFPTDYPFKPPKISFTTKIYH  76
6 MAS---------KRILKELKDLQKDPP-TSCSAGPVAE-----DMFHWQATIMG-PSDSPYSGGVFLVTIHFPPDYPFKPPKVAFRTKVFH  75
7 MS--KTAQKRLL----KELQQLIKDSPP-GIVAGPKSEN----NIFIWDCLIQG-PPDTPYADGVFNAKLEFPKDYPLSPPKLTFTPSILH  79
8 MA--TAPARRASSS--RSSSEISRTTPSMGFQLGFVDDS----NVFEWQVTIIG-PPETLYDGGYFNAIMSFPQNYPNSPPTVRFTSEMWH  82
9 MSSRKSTASSLLLRQYRELTDPKKAIPS--FHIELEDDS----NIFTYNIGVMVLNEDSIYHGGFFKAQMRFPEDFPFSPPQFRFTPAIYH  85

                  90        100                 110       120       130       140       150
                   |         |                   |         |         |         |         |
1 PNVYPSGTVCLSILEEDK-----------DWRPAITIKQILLGIQELLNEPNIQDPAQAEAYTIYCQNRVEYEKRVRA-QAKKFAPS      158
2 PNVYPSGTVCLSILEEDK-----------DWRPAITINQLFIGIQELLNEPNIQEPAQAEAYTIYCQNRVEYEKRFRA-QAKKFCPS      158
3 PNVYPSGTICLSILNEDQ-----------DWRPAITLKQIVLGVQDLLDSPNPNSPAQEPAWRSFSRNKAEYDKKVLL-QAKQYS-K      157
4 PNIYADGSICLDILQN-------------QWSPIYDVAAILTSIQSLLCDPNPNSPANSEAARMYSESKREYNRRVRDVVEQSWTAD      152
5 PNINANGNICLDILKD-------------QWSPALTLSKVLLSICSLLTDANPDDPLVPEIAHIYKTDRPKYEATAREWT-KKYAV       148
6 PNINSNGSICLDILKE-------------QWSPALTISKVLLSICSLLTDPNPDDPLVPEIAHMYKTDKNKYESTARTWT-QKYAMG      148
7 PNIYPNGEVCISILHSPGDDPNMYELAEERWSPVQSVEKILLSVMSMLSEPNIESGANIDACILWRDNRPEFERQVKLSILKSLG-----F 165
8 PNVYPDGRVCISI-HPPGDDPNGYELASERWTPVHTVESIVLSIISMLSSPNDESPANIEAAKDWREKQDEFKKKVRRAVRKSQEM----L 168
9 PNVYRDGRLCISILHQSGD-PMTDEPDAETWSPVQTVESVLISIVSLLEDPNINSPANVDAAVDYRKNPEQYKQRVKMEVERSKQDIPKGF 175
```

(*a*)

```
                  10        20        30        40        50        60        70        80        90
                   |         |         |         |         |         |         |         |         |
1 MSD----QEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGG  97
2 MSDSEVNQEAKPEVKP--EVKPETHINLKV-SDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG 101
3 M-----------------------QIFVKTLT--GKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG  76
4 M-----------------------QIFVKTLT--GKTITLEVESSDTIDNVKSKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG  76
5 M-----------------------QIFVKTLT--GKTITLEVESSDTIDNVKSKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG  76
                                      **   * *     **  **    *          * ** *          ** *  *   * *
```

(*b*)

Fig. 3. (*a*) Alignments of amino-acid sequence from 1: human, murine and xenopus Ubc9 (the three proteins have 100% identity); 2: *Mesocricetus auratus* Ubc9; 3: *Saccharomyces cerevisiae* Ubc9; 4: plant *Arabidopsis thaliana* Ubc1 (used as a model for molecular replacement); 5: *S. cerevisiae* Ubc4; 6: *A. thaliana* Ubc9 (more related to Ubc4 enzymes than Ubc9 enzymes); 7: *S. cerevisiae* Ubc7; 8: wheat Ubc7; 9: *S. cerevisiae* Ubc3 (C-terminal extension truncated to fit figure). The insertions in human Ubc9 are highlighted in yellow. Amino acids identical to human Ubc9 residues are shown in blue and the catalytic cysteine is labeled in red. The numbering refers to the human Ubc9 sequence. (*b*) Alignments of amino-acid sequence from 1: human sumo; 2: *S. cerevisiae* Smt3; 3: human ubiquitin; 4: *S. cerevisiae* ubiquitin; 5: *Neurospora crassa* ubiquitin. Residues identical to sumo residues are shown in blue and the reactive terminal glycine in red. Charge differences between ubiquitin-like proteins [sumo (1) and Smt3 (2)] and ubiquitins (sequences 3, 4 and 5) are indicated by green stars. The numbering refers to sumo sequence. Protein sequences were obtained from Swissprot data bank and alignments performed with *DNAstar*, using the default parameters of the clustal method.

Fig. 4. Stereo drawing of the α-carbon backbone of human Ubc9 (line) and plant Ubc1 (dashed line). The N-terminus is labelled.

3.4. *Active site*

The sulfydryl group of Cys93 is close to the carboxyl groups of Glu98 and Asp127 and in the vicinity of the hydroxyl group of Tyr87 (Fig. 6). Asn85 forms a hydrogen bond between its carbonyl group and the N atom of Cys93. This region is also rigidified by a hydrogen bond between the N atom of Tyr87 and the carbonyl group of Thr91. The r.m.s.d. for 14 C^{α} atoms around the catalytic site (Tyr87–Leu94, Glu98–Lys101, Gln126–Asp127) is 0.28 Å between our structure and that from Tong *et al.* (1997).

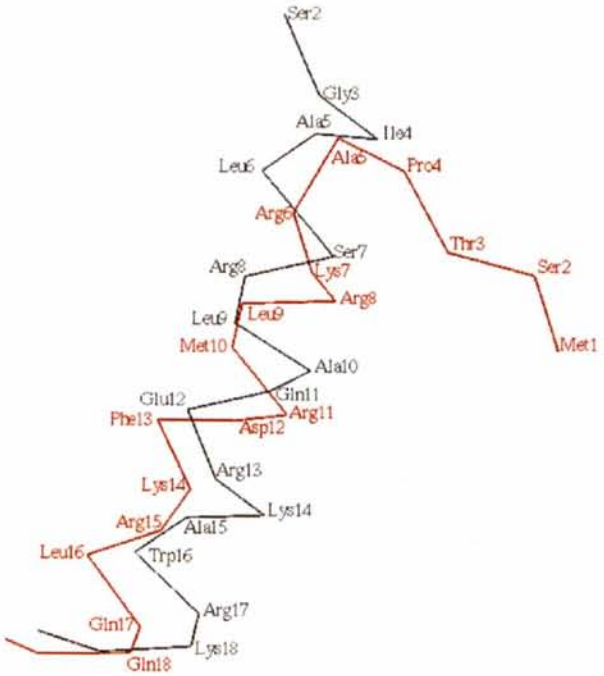

Fig. 5. The N-terminus of Ubc9 (black) and Ubc1p (red). The orientation of the first six residues is very different in both structures, although this region is found to be poorly ordered in all Ubc structures. The most striking feature is that the helix (residues 7–14) are out of register by one residue (*e.g.* Ser7 in Ubc9 superimposes with Arg8 of Ubc1p). The structures of Ubc7 and Ubc4 show the same change in register of the helix compared to Ubc9. The structure of this N-terminal helix in Ubc9 is very similar to that found by Tong *et al.* (1997) in their high-resolution structure of Ubc9.

Surprisingly, residues near the reactive cysteine (residues 91, 92, 94, 95) are not especially well conserved between plant Ubc1, yeast Ubc7, Ubc4 and Ubc3 (Fig. 3). Other residues such as His83, Pro84, Asn85, Ile96 and Leu97 are highly conserved suggesting that those residues may be important to maintain the cysteine in appropriate conformation. This is reinforced by the observation that a single mutation Leu99→Ser in yeast Ubc3 (equivalent of Leu97 in human Ubc9) is sufficient to generate an inactive enzyme (Banerjee *et al.*, 1995). A noticeable feature of the active site of Ubc9 is the presence of Lys101, which is located close to the S atom of Cys93 (Figs. 6 and 7).

3.5. *Implications for E1 and sumo recognition*

The N-terminus is known to be important for activity (Sullivan & Vierstra, 1991) and given its remoteness from the active site presumed to play a role in E_1 binding. The non-catalytic face of Ubc9 is much more positively charged than other Ubc enzymes and is a reflection of the much higher p*I* of Ubc9 compared with ubiquitin conjugating enzymes (8.7 *versus* < 6.5) (Fig. 7). A combination of this positively charged face and the dramatically altered recognition surface of the N-terminal helix, due to the register change, may allow Ubc9 to identify its specific E_1.

Ubiquitin is a small globular protein with a C-terminal extension that contains the reactive glycine (Gly76) (Vijay-Kumar *et al.*, 1987). One side of the protein is

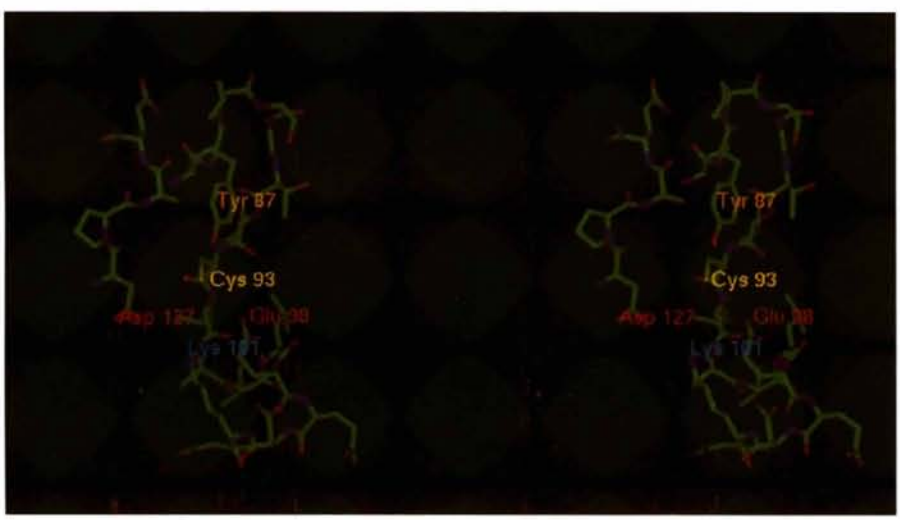

Fig. 6. Stereoview of the catalytic site of human Ubc9. Lys101 and the negatively charged amino acids are clearly visible. O atoms are red, N atoms blue, C atoms green and S atoms yellow. Figure created with *Insight* (Molecular Simulations Inc.).

positive, while the other is negative. However, even with the three-dimensional structure of ubiquitin, no unambiguous model for a ubiquitin/Ubc complex exists. Sequence alignments between different ubiquitins, sumo and the yeast sumo homolog (Smt3) reveal that conserved charges in ubiquitins are changed in sumo (and Smt3) (Fig. 3). Sumo has a p*I* of 5.4, which is more negatively charged than ubiquitin (p*I* 7.2). This correlates with the unusually high p*I* for Ubc9 suggesting an electrostatic basis for Ubc9 sumo recognition. However, the charge difference between sumo and ubiquitin may play a role in the downstream processing of modified proteins. Overall the C-terminus of sumo is much more negatively charged than ubiquitin: the conserved residues Arg74, Arg72, His68 and Lys63 of ubiquitin proteins are all found as negative or neutral residues in sumo or Smt3 (Fig. 3). In particular we note that Arg74, near the reactive Gly76, in ubiquitin is replaced by Thr95 in sumo and Ile99 in Smt3 (Fig. 3).

In contrast to the positive non-catalytic face, human Ubc9 has a negatively charged ridge close to the reactive cysteine (Fig. 7). The ridge is composed of Asp67, Glu98, Asp100, Asp102 and Asp127. Except for yeast Ubc9, the other Ubc9's (Fig. 3) have negatively charged residues at equivalent positions. In other ubiquitin-conjugating enzymes, this negatively charged patch is not found due to non-conservative substitution of acidic residues (Fig. 3). Presumably this negatively charged ridge in Ubc9 would repel the negatively charged C-terminus of sumo, this does not eliminate ambiguity in the sumo Ubc9 complex but it does substantially delimit possible complex structures. Ubc9 diplays a positively charged region on the same face as the active site, that may interact with the negatively charged C-terminus of

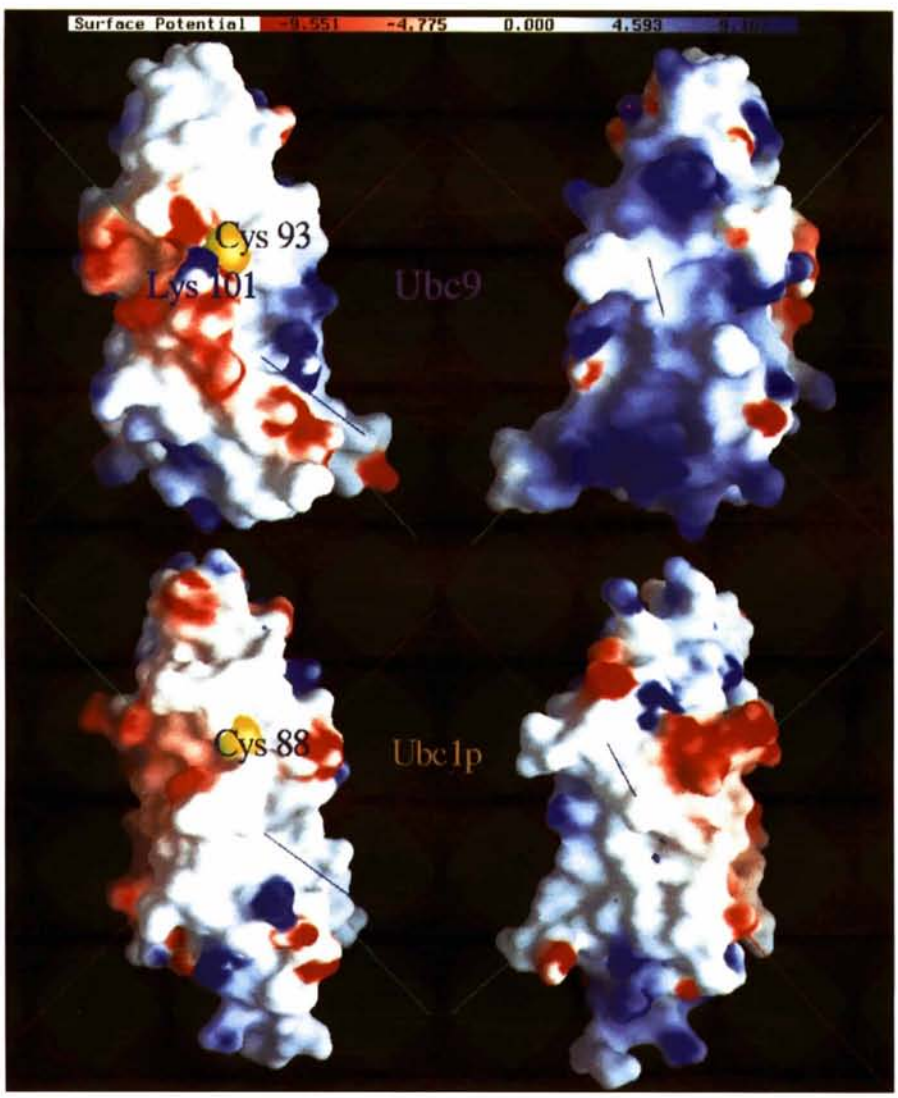

Fig. 7. Electric surface potentials of human Ubc9 and plant Ubc1 (Ubc1p). The SG atom of the reactive cysteine is shown in yellow. The right pictures correspond to the back face of the left pictures (180° rotation about an vertical axis in the plane of the paper). The *x*, *y* and *z* axes are shown. Figure created with *GRASP* (Nicholls *et al.*, 1991).

sumo (Fig. 7). A model of the complex awaits the structure of sumo.

Lys101 in human Ubc9 generatcs a positively charged lip close to the reactive cysteine in the midst of the negatively charged ridge (Fig. 7). This lip may discriminate between sumo and ubiquitin, by repelling Arg74 of ubiquitin. We feel this is important as without Lys101, the cysteine of Ubc9 sits in a very negative pocket which could favour ubiquitin binding. The two-residue insertion (residues 100–101) is conserved in all Ubc9's and Lys101 is conserved in all with the exception of yeast Ubc9 (changed to Gln). However, the yeast Ubc9 has an additional accompanying mutation (Glu98 to Asn98). This would significantly lower the negative potential around the active-site cysteine, possibly eliminating the need for the positive lip.

3.6. *Summary*

We have determined the structure of human Ubc9 the first sumo-conjugating protein to be characterized. Ubc9 shows several important differences compared to classical ubiquitin-conjugating enzymes. The principal differences occur at the N-terminus, the catalytic site and on the face of the protein opposite the catalytic site. We have shown how these differences in structure may be responsible for sumo/ubiquitin discrimination and E_1 selection.

We thank Ron Hay for encouragement and the Wellcome Trust and BBSRC for support. MJPD was supported by the JNICT-Praxis XXI (Portugal). We thank Axel Brünger and the authors of *CNS* for a pre-release version of *CNS*. We thank the reviewers and editor for advice in revising the manuscript.

References

Banerjee, A., Deshaies, R. J. & Chau, V. (1995). *J. Biol. Chem.* **270**, 26209–26215.

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

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Biol. Chem.* **112**, 535–542

Brünger, A. T. (1992). *X-PLOR, Version 3-1, Manual*, Yale University, New Haven, CT, USA.

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.

Brünger, 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. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Chen, Z. J., 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. P. & Chau, V. (1993). *Biochemistry*, **32**, 13809–13817.

Desterro, J. M. P., Thomson, J. & Hay, R. T. (1997). *FEBS Lett.* **417**, 297–300.

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

Haas, A. L., Arhens, P., Bright, P. M. & Ankel, H. (1987). *J. Biol. Chem.* **262**, 11315–11323.

Hateboer, G., Hijmans, E. M., Nooil, J. B. D., Schlenker, S., Jentsch, S. & Bernards, R. (1996). *J. Biol. Chem.* **271**, 25906–25911.

Jaffray, E., Wood, K. M. & Hay, R. T. (1995). *Mol. Cell. Biol.* **15**, 2166–2172.

Johnson, E. S. & Blobel, G. (1997). *J. Biol. Chem.* **272**, 26799–26802.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

King, R. W., Peters, J. M., Tugendreich, S., Rolfe, M., Hieter, P. & Kirschner, M. W. (1995). *Cell*, **81**, 279–288.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 548–558.

Martin, P. D., Edwards, B. F. P., Yamazaki, R. K. & Chau, V. (1997). *Biochemistry*, **36**, 1621–1627.

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

Narasimhan, J., Potter, J. L. & Haas, A. L. (1996). *J. Biol. Chem.* **271**, 324–330.

Navaza, J. (1994). *Acta Cryst.* D**50**, 157–163.

Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend, 29–30 January, 1993*, edited by L. Sawyer, N. Isaacs & S. W. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Rechsteiner, M. (1991). *Cell*, **66**, 615–618.

Saitoh, H., Pu, R., Cavenagh, M. & Dasso, M. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 3736–3741.

Schwartz, S. E, Matuschewski, K., Liakopoulos, D., Scheffner, M. & Jentsch, S. (1998). *Proc. Natl Acad. Sci USA*, **95**, 560–564.

Seufert, W., Futcher, B. & Jentsch, S. (1995). *Nature (London)*, **373**, 78–81.

Shen, Z., Pardington-Purtymun, P. E., Comeaux, J. C., Moyzis, R. K. & Chen, D. J. (1996). *Genomics*, **37**, 183–186.

Sullivan, M. L. & Vierstra, R. D. (1991). *J. Biol. Chem.* **266**, 23878–23885.

Tong, H., Hateboer, G., Perrakis, A., Bernards, R. & Sixma, T. K. (1997). *J. Biol. Chem.* **272**, 21381–21387.

Van Nocker, S. & Vierstra, R. D. (1993). *J. Biol. Chem.* **268**, 24766–24773.

Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). *J. Mol. Biol.* **194**, 531–544.

Wang, Z.-Y, Qiu, Q.-Q, Seufert, W., Taguchi, T., Testa, J. R., Whitmore, S. A., Callen, D. F., Welsh, D., Shenk, T. & Deuel, T. F. (1996). *J. Biol. Chem.* **271**, 24811–24816.

Wright, D. A., Futcher, B., Ghosh, P. & Geha, R. S. (1996). *J. Biol. Chem.* **271**, 31037–31043.

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