PROTEIN STRUCTURE REPORT

Crystal structure of the UBR-box from UBR6/FBXO11 reveals domain swapping mediated by zinc binding

Juliana Muñoz-Escobar, Guennadi Kozlov, and Kalle Gehring*

Department of Biochemistry, Groupe de Recherche Axé sur la Structure des Protéines, McGill University, Montreal, Quebec H3G0B1, Canada

Received 1 May 2017; Accepted 3 July 2017
DOI: 10.1002/pro.3227
Published online 10 July 2017 proteinscience.org

Abstract: The UBR-box is a 70-residue zinc finger domain present in the UBR family of E3 ubiquitin ligases that directly binds N-terminal degradation signals in substrate proteins. UBR6, also called FBXO11, is an UBR-box containing E3 ubiquitin ligase that does not bind N-terminal signals. Here, we present the crystal structure of the UBR-box domain from human UBR6. The dimeric crystal structure reveals a unique form of domain swapping mediated by zinc coordination, where three independent protein chains come together to regenerate the topology of the monomeric UBR-box fold. Analysis of the structure suggests that the absence of N-terminal residue binding arises from the lack of an amino acid binding pocket.

Keywords: zinc finger; UBR-box domain; domain swapping; zinc coordination; histidine; X-ray crystallography

Introduction

E3 ubiquitin ligases catalyze the transfer of ubiquitin to proteins targeted for proteasomal degradation. These enzymes recognize specific degradation signals termed degrons, located in substrate proteins.[1] The N-end rule pathway is a branch of the ubiquitin proteasome system that relates the *in vivo* half-life of a protein to the identity of its N-terminal residue.[2] The UBR-box is a 70-residue zinc finger domain present in the UBR family of E3 ubiquitin ligases. The mammalian genome encodes seven UBR-containing proteins named UBR1–UBR7. In mammals, UBR1, UBR2, UBR4, and UBR5 target proteins for proteasomal degradation in the N-end rule pathway by binding destabilizing N-degrons through the UBR-box domain.[3,4] Remarkably, the presence of the UBR-box does not guarantee the ability to recognize N-degrons. UBR3, UBR6, and UBR7 contain a UBR-box but do not bind destabilizing N-terminal residues.[3] In our previous work, we identified the molecular determinants for N-degron recognition in UBR1 and UBR2 by studying the crystal structure of the UBR-box and its

Statement The dimeric crystal structure reveals a unique form of domain swapping mediated by three zinc fingers. The zinc binding sites offer some insight into the significance of the structure and suggest that it arises from a unique form of domain swapping, where three independent protein chains come together to emulate the monomeric UBR-box fold.

Grant sponsor: Canadian Institutes of Health Research; Grant number: MOP-14219.

*Correspondence to: Kalle Gehring, Life Sciences Complex, Room 469, 3649 promenade Sir William Osler, Montreal, Quebec, H3G0B1 Canada. E-mail: kalle.gehring@mcgill.ca

ability to bind N-degron peptides. However, the role of the UBR-box domain in proteins that are not part of the N-end rule is still not understood.

UBR6, also called FBXO11, is an F-box subunit of the Skp1-Cullin-F-box ubiquitin ligase complex. As an F-box protein, UBR6 directly binds substrates for ubiquitylation and proteasomal degradation. It comprises an N-terminal F-box, three presumptive substrate-binding CASH domains, and a C-terminal UBR-box.[5] Despite the presence of a UBR domain, UBR6 does not bind N-degrons.[3] To understand the role of the UBR-box in proteins that are not part of the N-end rule pathway, we determined the crystal structure of the UBR-box domain from human UBR6. We analyze the distinct dimerization behavior observed in solution and in the crystal and propose an explanation for the inability of UBR6 to bind N-degrons.

Results and Discussion

The UBR-box domain of UBR6 is a mixture of monomers and dimers in solution

UBR6 eluted in two peaks during size exclusion chromatography (SEC), suggesting the protein forms both dimers and monomers in solution [Fig. 1(A)]. No exchange between the two forms was observed in chromatography experiments. We used multiangle light scattering coupled with SEC (SEC-MALS) to confirm the molecular mass of the purified fractions [Fig. 1(B)]. Dimerization has been observed in other zinc fingers[6,7] and F-box proteins[8,9] but the UBR-boxes of UBR1 and UBR2 do not form dimers.[10–12] It is unclear if the UBR6 UBR-box is a dimer in the context of the full-length protein and if it affects its function.

Domain swapping in the UBR-box domain from UBR6

Crystallization experiments were set up for both monomer and dimer fractions of UBR6. Crystals grew after approximately 6 months for both samples. All crystals had the same space group and crystal form regardless of the input sample. Solving the structure by phasing with anomalous diffraction of the zinc atoms showed that the crystals had four identical molecules in the asymmetric unit arranged as two dimers. Each chain consists of three α-helices, two antiparallel β-strands, and two long loops [Fig. 1(C)]. Dimerization is mediated by zinc fingers located in the extremes of the dimer that coordinate three zinc atoms each [Fig. 1(C,E)]. In addition, a central zinc is tetrahedrally coordinated by histidine residues 848 and 883 from two chains.

The tetrahedral topology of the central zinc-binding site comprising only histidine residues is uncommon. Banaszak *et al*. report a zinc-coordination site of only histidine residues[13] [Fig. 1(C)]. The most common coordination topology of tetrahedral zinc sites is Cys_2-Cys/His-Cys/His (with positional variations). Around 30% of structural zinc coordination spheres in the protein data bank (PDB) have a Cys_4 topology while only 4% have a His_3-Asp sphere.[14] The histidines composing the central zinc-binding site are not conserved in UBR-box domains in other UBR proteins [Fig. 2(A)].

To test the effect of this zinc-binding site on dimerization, we mutated the histidines to alanine and probed the mutant proteins by SEC. Dimer and monomer peaks were observed for both the His848Ala and His883Ala mutants but the His848Ala mutant showed an increase in the fraction of monomers [Fig. 1(D)]. A His848/883Ala mutant was designed but the protein did not express. These results suggest that the dimerization observed in solution is not dependent on the central zinc coordination site. We believe this zinc coordination site may be an artifact of crystal packing as zinc was present in the crystallization drop at 10 μ*M*.

The remaining zinc binding sites offer some insight into the significance of the structure and suggest that it arises from a unique form of domain swapping. The second zinc finger has an unusual topology also observed in UBR1 and UBR2,[11,12] consisting of two zinc ions each tetrahedrally coordinated by the shared cysteine 868 [Fig. 1(E)]. All but one of the zinc-coordinating residues are conserved across the UBR family [Fig. 2(A)]. In UBR1 and UBR2, a histidine residue forms a Cys_3-His coordination sphere while, in UBR6, this residue is replaced by Cys899 to generate a Cys_4 site. A remaining zinc atom in each zinc finger is bound to a Cys_2-His_2 site found in UBR1/2 structures [Fig. 1(E)].

Even though most of the zinc-binding residues are conserved in UBR6 compared to UBR1 and UBR2 [Fig. 2(A)], the structure of UBR6 is strikingly different. Analysis of the UBR6 structure suggests that it is the result of domain swapping where three independent protein chains come together to regenerate the topology observed in the monomeric UBR-box fold [Fig. 2(B,D)]. The first zinc ion is coordinated by Cys868 and Cys872 from chain one (green), and Cys890 and Cys899 from chain two (cyan). The second zinc atom is coordinated by Cys865 and Cys868 from chain one (green), Cys888 from chain two (cyan), and Cys835 from an adjacent chain (yellow) [Figs. 1(E) and 2(B)]. This zinc finger is observed twice in the structure as two protein chains coordinate each zinc-binding site in opposite sides of the dimer [Fig. 1(C)]. The third zinc finger has a typical Cys_2-His_2 topology as observed in UBR1 and UBR2 [Fig. 1(E)]. This zinc-binding site is the only one formed by one protein chain. In UBR1 and UBR2, these zinc fingers form a rigid scaffold that frames the N-degron binding site. Given the conservation of the zinc coordination spheres, we believe the reconstructed monomeric structure shown in Figure 2(B) is probably a close

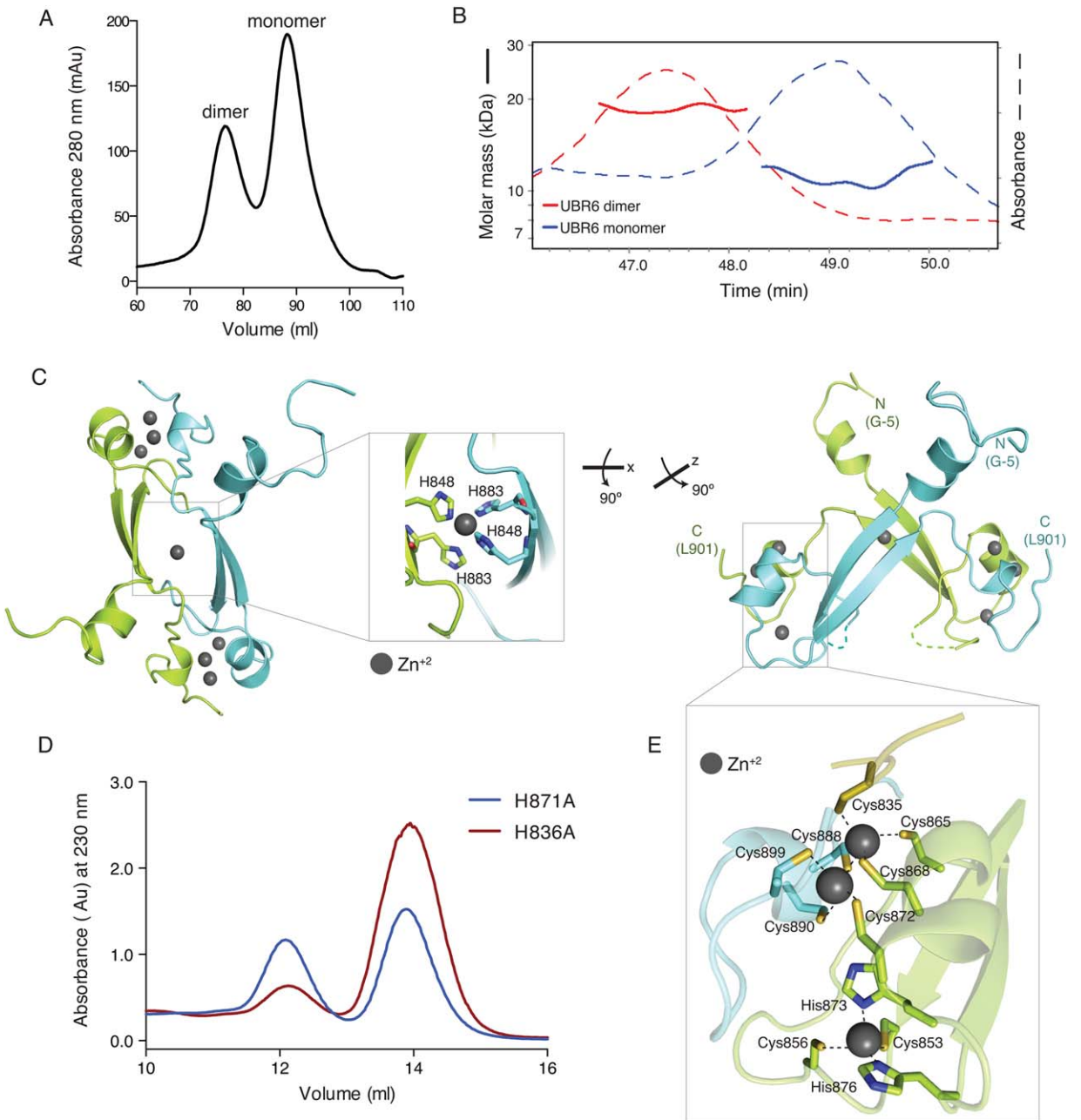

Figure 1. UBR-box from UBR6 forms different dimers in solution and in the crystal structure. (A) SEC of the UBR-box from UBR6 shows monomer and dimer fractions. (B) Molecular weight profiles from SEC-MALS for monomer (11 kDa) and dimer (18 kDa) fractions. (C) Crystal structure of the UBR-box from UBR6. The dimer structure is stabilized by seven zinc atoms. The central zinc binding site has an unusual topology with four histidine residues in a tetrahedral coordination sphere. Secondary structural elements are α1 (Ile839-Tyr845), α2 (Val866-Cys872), α3 (Asp889-Ala892), β1 (Met847-Cys853), and β2 (Val878-Asp884). (D) Size exclusion experiment shows that loss of the central zinc coordination sphere alters but does not prevent dimerization in solution. (E) Intermolecular zinc coordination in the UBR-box structure of UBR6. Zinc binding residues are conserved in the UBR family.

representation of the physiological structure of the UBR6 UBR-box.

One question that remains is why the UBR-box in UBR6 does not bind N-degrons. We were unable to detect binding of peptides with N-terminal arginine residues to the UBR-box of UBR6 by nuclear magnetic resonance. We labeled the UBR6 domain with 15N and acquired ^{1}H-^{15}N correlation spectra adding increasing concentrations of unlabeled Arg-Ile-Phe-Ser peptide, which binds strongly to UBR1 and UBR2. UBR6 signals did not shift or disappear indicating no binding. This result is in agreement with a published study with the full-length protein.[3]

In UBR1 and UBR2, N-degron recognition occurs through a negatively charged pocket that recognizes the positive N-terminal residue, and a secondary

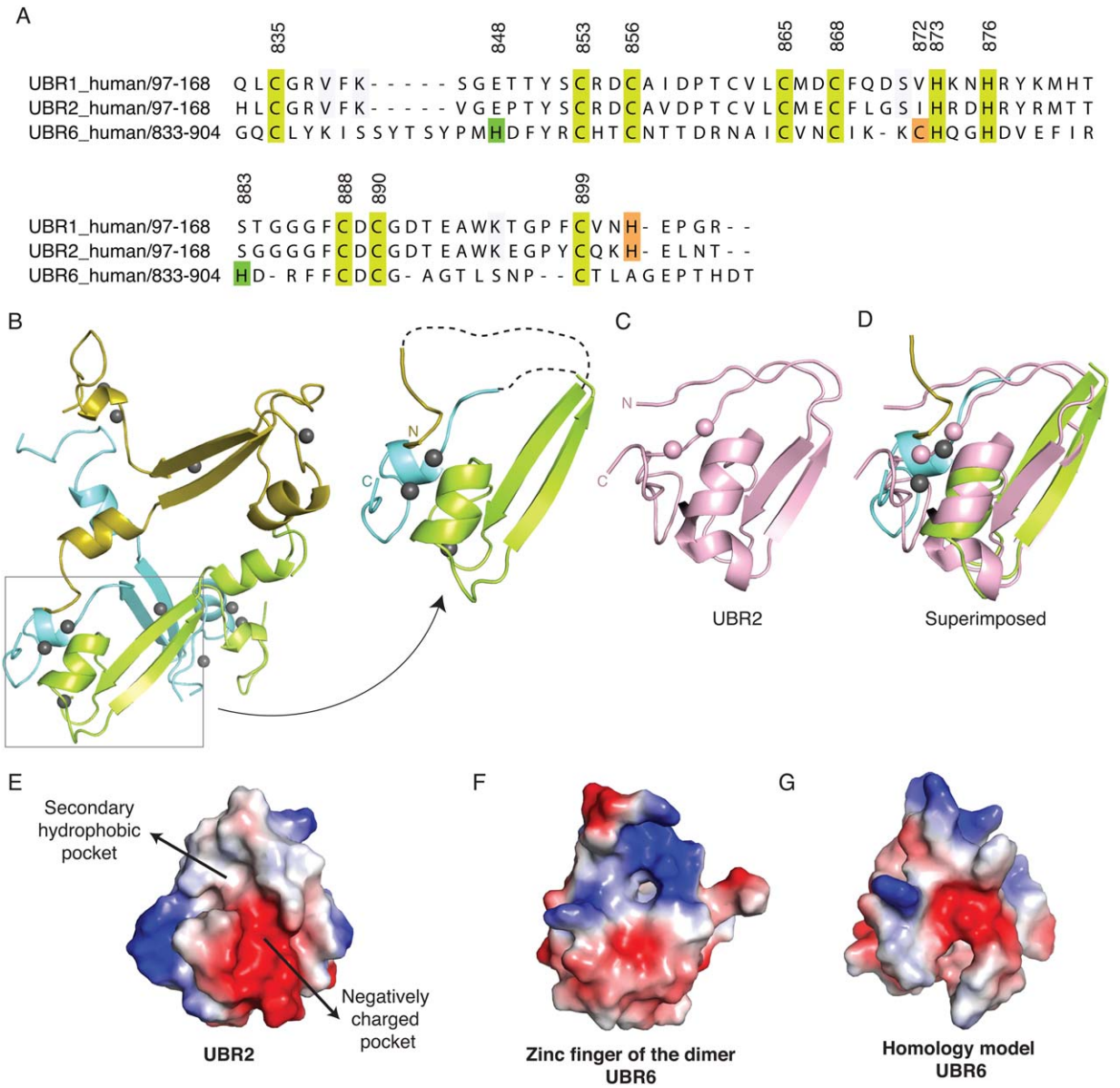

Figure 2. Dimer crystal structure emulates the monomeric UBR-box fold. (A) Sequence alignment of the UBR-box domains from human UBR1, UBR2 and UBR6 highlights the conservation of zinc-coordinating residues (yellow). The C-terminal histidine residue in UBR1 and UBR2 is not conserved in UBR6. Instead, Cys872 completes a tetrahedral coordination site (orange). The histidine residues coordinating the central zinc ion in UBR6 are not conserved in the UBR family (green). (B) Three of the four protein chains in the asymmetric unit generate a topology that resembles the UBR-box fold. (C) Crystal structure of the UBR-box from UBR2 (PDB: 3NY3). (D) Superimposed structures of UBR2 and the zinc finger dimer of UBR6. (E) Electrostatics potential surface representation of UBR2 (PDB: 3NY3) shows N-degron binding pockets. (F) Monomer model of UBR6. Zinc finger of the dimer (panel B, right). Absence of pockets could explain the lack of N-degron binding in UBR6. (G) Homology model shows lack of interacting pockets. Structure model was generated based on human and yeast UBR-box structures using SWISS-MODEL.

hydrophobic pocket, which binds the second residue of the substrate [Fig. 2(E)]. An electrostatic potential surface representation of the UBR6 fragment containing the conserved zinc fingers lacks any grooves in the area where N-degron binding is expected [Fig. 2(F)]. A homology model generated by the SWISS-MODEL server based on the available UBR-box structures (human and yeast) [Fig. 2(G)] also shows the absence of grooves that would support N-degron binding.

A physiological structure of the UBR-box domain from a UBR protein that does not bind N-degrons will clarify the structural differences that prevent type-1 binding. Moreover, such structure may elucidate alternative functions of the UBR-box that are independent of the N-end rule. Among the possibilities for UBR-box function are those of zinc finger proteins. Tertiary structured zinc fingers confer specific binding activities to various molecules

Table I. *Data Processing and Refinement Statistics*

	UBR6 (833–904)
Wavelength	1.2822
Resolution range	41.72–2.202 (2.281–2.202)
Space group	P 21 21 21
Unit cell	67.57 69.2 82.59 90 90 90
Unique reflections	20,183 (1952)
Redundancy	4.3 (4.1)
Completeness (%)	99.81 (99)
Mean I/sigma(I)	14.67 (2.88)
Wilson *B*-factor	36.59
R-meas	0.094 (0.608)
CC1/2	0.851
Reflections used in refinement	20,172 (1949)
Reflections used for *R*-free	1010 (98)
R-work	0.1948 (0.2940)
R-free	0.2474 (0.3414)
Number of non-hydrogen atoms	2377
Macromolecules	2210
Ligands	50
Protein residues	117
RMS (bonds)	0.005
RMS (angles)	0.69
Ramachandran favored (%)	98.87
Ramachandran allowed (%)	1.13
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.41
Clashscore	5.54
Average *B*-factor	50.03
Macromolecules	49.69
Ligands	67.83
Solvent	48.78
Number of TLS groups	13

Values in parenthesis are for the highest resolution shell.

beyond proteins such as DNA, RNA, and small molecules. It remains to be discovered if the UBR-box is a protein–protein interaction domain or if it serves other macromolecules.

In conclusion, this study exposes the ability of the UBR-box domain from FBXO11 to form monomers and dimers in solution. We present a domain-swapped crystal structure mediated by three zinc coordination sites that emulate the known UBR-box fold and propose an explanation for the lack of N-degron binding.

Methods

Protein expression and purification

Human UBR6 (833–904) and H848A, H883A, and H848/883A mutants were cloned into pGEX-6p-1 vector and expressed in BL21(DE3) *E. coli* cells. Cultures were grown at 37°C until O.D. at 600 nm reached 0.7. At this point, 100 μ*M* $ZnCl_2$ was added, and protein was induced with 0.5 m*M* Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were grown for 20 h at 16°C. Lysis of pellets was done in 50 m*M* HEPES pH 7.5, 200 m*M* NaCl, 0.1 m*M* phenylmethylsulfonyl fluoride (PMSF), 10 μM $ZnCl_2$, 5% glycerol, and 5 m*M* β-mercaptoethanol buffer. GST-tagged protein was purified using glutathione *S*-transferase (GST) sepharose beads and eluted with 20 m*M* reduced glutathione. GST tag was cleaved using PreScission Protease from GE life sciences. Proteins were further purified by SEC using a HiLoad Superdex 75 16/600 column and 20 m*M* HEPES pH 7.5, 100 m*M* NaCl, 2 m*M* β-mercaptoethanol and 10 μ*M* $ZnCl_2$ buffer. Gel filtration experiments for UBR6 mutants were performed in a Superdex 75 10/300 GL column.

SEC-MALS

SEC-MALS experiments were done in with the Wyatt miniDawn TREOS and Optilab rex units using a Superose 6 10/300 GL column and 8 mg/mL of protein. Data analysis was performed with the ASTRA software.

Crystallization, data collection, and structure determination

Crystals of 10 mg/mL UBR6 (monomer and dimer fractions) were grown at 22°C by sitting drop vapor diffusion against 0.2 *M* $NaNO_3$, 0.1 M Bis-Tris propane pH 6.5, and 25% polyethylene glycol 3350. Crystals were cryoprotected with 20% ethylene glycol and flash cooled with liquid nitrogen. X-ray diffraction datasets were collected at λ 1.28 at the Canadian Light Source facility beam line 08B1–1 using a RAYONIX MX300HE detector. Data were processed using HKL2000.[15] Protein structure was solved using AutoSol program from PHENIX.[16] Model was finalized manually using COOT.[17] Structure refinement was done with CCP4[18] and PHENIX[16] suites. Translation/libration/screw (TLS) vibration motion was applied at the last stage of refinement.[19] Refinement statistics are given in Table I. Figures were produced using PyMol.[20] The UBR6 homology model was generated by SWISS-MODEL server.[21]

Accession number

The coordinates and structure factors of the UBR-box domain from UBR6 have been deposited in the PDB with accession code 5VMD.

Acknowledgements

X-ray diffraction data were acquired at the Canadian Light Source, which is supported by the Canada Foundation for Innovation, Natural Sciences and Engineering Research Council of Canada, the University of Saskatchewan, the Government of Saskatchewan, Western Economic Diversification Canada, the National Research Council Canada, and the Canadian Institutes of Health Research.

References

1. Hershko A, Ciechanover A (1998) The ubiquitin system. Annu Rev Biochem 67:425–479.

2. Bachmair A, Finley D, Varshavsky A (1986) In vivo half-life of a protein is a function of its amino-terminal residue. Science 234:179–186.
3. Tasaki T, Zakrzewska A, Dudgeon DD, Jiang Y, Lazo JS, Kwon YT (2009) The substrate recognition domains of the N-end rule pathway. J Biol Chem 284:1884–1895.
4. Tasaki T, Mulder LC, Iwamatsu A, Lee MJ, Davydov IV, Varshavsky A, Muesing M, Kwon YT (2005) A family of mammalian e3 ubiquitin ligases that contain the UBR box motif and recognize n-degrons. Mol Cell Biol 25:7120–7136.
5. Duan S, Cermak L, Pagan JK, Rossi M, Martinengo C, di Celle PF, Chapuy B, Shipp M, Chiarle R, Pagano M (2012) Fbxo11 targets bcl6 for degradation and is inactivated in diffuse large B-cell lymphomas. Nature 481: 90–93.
6. McCarty AS, Kleiger G, Eisenberg D, Smale ST (2003) Selective dimerization of a c2h2 zinc finger subfamily. Mol Cell 11:459–470.
7. Aras S, Singh G, Johnston K, Foster T, Aiyar A (2009) Zinc coordination is required for and regulates transcription activation by epstein-barr nuclear antigen 1. PLoS Pathog 5:e1000469.
8. Chew EH, Poobalasingam T, Hawkey CJ, Hagen T (2007) Characterization of cullin-based e3 ubiquitin ligases in intact mammalian cells–evidence for cullin dimerization. Cell Signal 19:1071–1080.
9. Zhuang M, Calabrese MF, Liu J, Waddell MB, Nourse A, Hammel M, Miller DJ, Walden H, Duda DM, Seyedin SN, Hoggard T ,Harper JW ,White KP ,Schulman BA (2009) Structures of spop-substrate complexes: insights into molecular architectures of btb-cul3 ubiquitin ligases. Mol Cell 36:39–50.
10. Munoz-Escobar J, Matta-Camacho E, Cho C, Kozlov G, Gehring K (2017) Bound waters mediate binding of diverse substrates to a ubiquitin ligase. Structure 25: 719–729.
11. Matta-Camacho E, Kozlov G, Li FF, Gehring K (2010) Structural basis of substrate recognition and specificity in the N-end rule pathway. Nat Struct Mol Biol 17: 1182–1187.
12. Choi WS, Jeong BC, Joo YJ, Lee MR, Kim J, Eck MJ, Song HK (2010) Structural basis for the recognition of N-end rule substrates by the ubr box of ubiquitin ligases. Nat Struct Mol Biol 17:1175–1181.
13. Banaszak K, Martin-Diaconescu V, Bellucci M, Zambelli B, Rypniewski W, Maroney MJ, Ciurli S (2012) Crystallographic and X-ray absorption spectroscopic characterization of *Helicobacter pylori* UreE bound to Ni(2)(+) and Zn(2)(+) reveals a role for the disordered C-terminal arm in metal trafficking. Biochem J 441:1017–1026.
14. Laitaoja M, Valjakka J, Janis J (2013) Zinc coordination spheres in protein structures. Inorg Chem 52: 10983–10991.
15. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326.
16. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ ,Moriarty NW ,Oeffner R ,Read RJ ,Richardson DC ,Richardson JS ,Terwilliger TC ,Zwart PH (2010) Phenix: a comprehensive python-based system for macromolecular structure solution. Acta Cryst D66:213–221.
17. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of coot. Acta Cryst D66:486–501.
18. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ ,Murshudov GN ,Pannu NS ,Potterton EA ,Powell HR ,Read RJ ,Vagin A ,Wilson KS (2011) Overview of the CCP4 suite and current developments. Acta Cryst D67:235–242.
19. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing tls motion. Acta Cryst D62:439–450.
20. Schrodinger (2015) The pymol molecular graphics system, version~1.8. New York: Schrodinger LLC.
21. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schewe T. (2014) Swiss-model: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258.