5424–5437 *Nucleic Acids Research, 2016, Vol. 44, No. 11* *Published online 7 April 2016*
doi: 10.1093/nar/gkw218

Structural basis for recruiting and shuttling of the spliceosomal deubiquitinase USP4 by SART3

Joon Kyu Park[1], Tanuza Das[2], Eun Joo Song[2,*] and Eunice EunKyeong Kim[1,*]

[1]Biomedical Research Institute, Korea Institute of Science and Technology, Hwarangno 14-gil 5, Seongbuk-gu, Seoul 136-791, Republic of Korea and [2]Molecular Recognition Research Center, Korea Institute of Science and Technology, Hwarangno 14-gil 5, Seongbuk-gu, Seoul 136-791, Republic of Korea

Received October 27, 2015; Revised March 18, 2016; Accepted March 21, 2016

ABSTRACT

Squamous cell carcinoma antigen recognized by T-cells 3 (SART3) is a U4/U6 recycling factor as well as a targeting factor of USP4 and USP15. However, the details of how SART3 recognizes these deubiquitinases and how they get subsequently translocated into the nucleus are not known. Here, we present the crystal structures of the SART3 half-a-tetratricopeptide (HAT) repeat domain alone and in complex with the domain present in ubiquitin-specific protease (DUSP)-ubiquitin-like (UBL) domains of ubiquitin specific protease 4 (USP4). The 12 HAT repeats of SART3 are in two sub-domains (HAT-N and HAT-C) forming a dimer through HAT-C. USP4 binds SART3 at the opposite surface of the HAT-C dimer interface utilizing the β-structured linker between the DUSP and the UBL domains. The binding affinities of USP4 and USP15 to SART3 are 0.9 μM and 0.2 μM, respectively. The complex structure of SART3 nuclear localization signal (NLS) and importin-α reveals bipartite binding, and removal of SART3 NLS prevents the entry of USP4 (and USP15) into the nucleus and abrogates the subsequent deubiquitinase activity of USP4.

INTRODUCTION

Squamous cell carcinoma antigen recognized by T-cells 3 (SART3) was identified in human as a 110-kDa nuclear RNA-binding protein, designated p110^{nrb}, during attempts to isolate the human U6 small nuclear RNA capping enzyme. It consists of eight half-a-tetratricopeptide (HAT) repeats in the N-terminal followed by a nuclear localization signal (NLS) sequence, two RNA recognition motifs (RRMs) and a stretch of 10 highly conserved amino acids known as the Lsm-binding domain at the C-terminus (1–3). Around the same time, a different group identified SART3 as a tumor-rejection antigen which possesses tumor epitopes capable of inducing HLA-A24-restricted and tumor-specific cytotoxic T lymphocytes in cancer patients, thereby establishing it as a potential candidate for targeted immunotherapy (4,5). SART3 was also shown to influence HIV-1 gene expression and viral replication through its direct interaction with RNA polymerase II, and play a role in hematopoiesis through the transcriptional regulation of c-Myc (CMYC) (6,7).

Although SART3 is not detected in the spliceosome, a dynamic assembly of numerous proteins and five small nuclear RNAs (U1, U2, U4, U5 and U6 snRNAs) that serves to remove the intron sequences that are present in most eukaryotic pre-mRNAs, it has been shown to associate with U6 and U4/U6 di-snRNP during the recycling phase of the spliceosome cycle, and it is involved in the regulation of pre-mRNA splicing (3,8). Therefore, SART3 is analogous to Prp24 in *Schizosaccharomyces cerevisiae* (9). Recently, it was shown that SART3 functions as a targeting factor in the ubiquitin specific protease 4 (USP4)-mediated deubiquitination of K63-polyubiquitinated pre-mRNA-processing factor 3 (Prp3), one of the major components of the U4 snRNP, and this deubiquitination leads to disassembly of U4 and U6 components from tri-snRNP (10). SART3 is also reported as a targeting factor for another deubiquitinating enzyme (DUB), USP15, to histone H2B and histone deubiquitination regulates gene expression and/or DNA repair (11). In addition to these newly identified nuclear functions, both USP4 and USP15 are well known to function in the cytosol, i.e. USP4 modulates the Wnt/β-catenin, NF-κB, p53 and TGF-β signaling pathways (12–15) while USP15 performs functions in the TGF-β receptor and NF-κB signaling pathways (16–18). Interestingly, the two DUBs share domain structure, having domain specific for USP (DUSP) and Ub-like (UBL) domains at the N-terminus in addition to a catalytic domain with an additional UBL domain embedded, and the sequence identity between the two is about 58%.

There are other DUBs that function in both the cytosol and the nucleus (19). However, how they translocate between these two compartments is not well characterized ex-

*To whom correspondence should be addressed. Tel: +82 2 958 5937; Fax: +82 2 958 5909; Email: eunice@kist.re.kr
Correspondence may also be addressed to Eun Joo Song. Tel: +82 2 958 5086; Fax: +82 2 958 5170; Email: ejsong@kist.re.kr

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

cept for in a few cases, e.g. USP1. USP1 forms a complex with UAF1 in the cytoplasm, and then the complex translocates to the nucleus utilizing the two nuclear localization signals (NLSs) on USP1 (20). Although it has been suggested that USP4 also has a NLS sequence (21), it is not clear how it translocates between the cytosol and the nucleus. The additional domains and shorter structural motifs found in most of the DUBs identified are thought to contribute to the regulation of the DUB activity as well as govern specific sub-cellular localization or both (19,22). While the UBL domain has been associated with regulation of catalytic activity (23), the function of the DUSP domain is currently unknown.

Here, we report the crystal structures of the HAT repeat domain of human SART3 alone and in complex with the DUSP-UBL domains of USP4. In addition, we show how SART3 utilizes importin-α to translocate USP4 and presumably USP15 into the nucleus by showing the crystal structure of the SART3 NLS and importin-α complex. Biochemical analysis based on these structures provides a detailed understanding on how SART3 shuttles USP4 and USP15, but not USP11, into the nucleus for their functions.

MATERIALS AND METHODS

Protein expression and purification

The construct of *human* SART3 (SwissProt entry Q15020) HAT repeat domain (residues 94–611) was cloned into a pET-28a (Novagen) vector containing a Tobacco Etch Virus (TEV) protease-cleavable hexa-histidine tag at the N-terminal, and was expressed in *Escherichia coli* Rosetta (DE3) (Novagen). Cells were grown to an OD600 of 0.6 and induced with 0.5 mM IPTG at 18°C for ∼16 h. All subsequent steps were carried out at 4°C. Cells were harvested by centrifugation and lysed by sonication on the ice in a buffer containing 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) and a complete, EDTA-free protease inhibitor cocktail tablet (Roche). The lysates were loaded onto an nickle-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare), which was subsequently washed with lysis buffer supplemented with 20 mM imidazole. The protein was eluted in the same buffer with an imidazole gradient (20–500 mM imidazole), and TEV protease was added to fractions containing the target proteins for ∼12 h. Hexa-histidine tag removed fractions were further purified by Superdex 200S 26/60 size exclusion column (GE Healthcare) which was pre-equilibrated with 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 150 mM NaCl and 0.5 mM TCEP. The selenomethionine (SeMet) derivative was prepared by expression of SART3 HAT repeat domain in M9 minimal medium (Molecular Dimensions) containing 50 μg l^{-1} of SeMet (Sigma) which was purified using the scheme described above. All proteins were then concentrated to 20 mg ml^{-1} in same buffer using Vivaspin concentrator (Vivagen) as measured by Bradford assay, for crystallization trials.

SART3 HAT-C (residues 308–611) and HAT-C with Lα (residues 278–611) were cloned into a pET-28a vector and were overexpressed in *E. coli* Rosetta 2(DE3) cells at 18°C. Cells were harvested by centrifugation and lysed by sonication on ice in a buffer containing 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM TCEP and a complete, EDTA-free protease inhibitor cocktail tablet. The lysates were loaded onto the Ni-NTA column, which was subsequently washed with lysis buffer supplemented with 20 mM imidazole. The protein was eluted in the same buffer with an imidazole gradient (20–500 mM imidazole), and thrombin was added to the fractions for removal of hexa-histidine tag during ∼12 h. Hexa-histidine tag removed fractions were further purified by Superdex 200S 26/60 size exclusion column which was pre-equilibrated with 25 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM TCEP.

Wild-types of DUSP-UBL domains of *human* USP4 (SwissProt entry Q13107) comprising residues 1–230, *human* USP15 (SwissProt entry Q9Y4E8) comprising residues 1–226 and *human* USP11 (SwissProt entry P51784) comprising residues 73–288 were cloned into a pET-22b (Novagen) vector and were overexpressed in *E. coli* Rosetta 2(DE3) cells at 18°C. Harvested cells were disrupted in 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM TCEP and purified using the Ni-NTA affinity column. After several rounds of washing steps using the same buffer with 20 mM imidazole, the proteins were eluted with an imidazole gradient (20–500 mM imidazole). Pooled fractions were further purified using the Superdex 75S 26/60 size exclusion column (GE Healthcare) which was pre-equilibrated with 25 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM TCEP. USP15, USP11 and various mutants of USP4 ($USP4^{F127A}$, $USP4^{F127A/V128A}$, $USP4^{L11}$), USP15 ($USP15^{L11}$) and USP11 ($USP11^{L4}$ and $USP11^{L15}$) were overexpressed and purified using basically the same procedure.

For crystallization of USP4-SART3 complex, SART3 LαHAT-C and USP4 DUSP-UBL domains were expressed and purified individually as described above. The complex was formed by mixing the purified proteins in a 1:1.5 molar ratio and incubated for 1 h on the ice. The complex was then further purified on a Superdex 200S 26/60 size exclusion column to remove the excess proteins. Final protein complex was concentrated to 20 mg ml^{-1} in same buffer using Vivaspin concentrator (Vivagen) for crystallization trials.

The glutathione S-transferase (GST)-tagged SART3 NLS, comprising residues 601–649, was overexpressed in *E. coli* strain Rossetta 2(DE3) at 18°C from pGEX-4T-1 (GE Healthcare), and hexa-histidine tagged *mouse* importin-α2 (SwissProt entry P52293) comprising residues 70–498 (ImpA) was overexpressed separately from the same strain at 18°C from pET-28a. After harvesting, the two sets of cells were mixed, suspended in 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.1 mM PMSF, and disrupted by sonication on the ice. After centrifugation, the lysates were loaded onto a GST affinity column (GE Healthcare), which was pre-equilibrated with 50 mM Tris, pH 8.0, 150 mM NaCl, 0.5 mM TCEP and eluted using the same buffer with 10 mM reduced glutathione. The GST-SART3 NLS and HIS_6-ImpA complex were cleaved by thrombin treatment for ∼12 h and further purified using a Superdex 200S 26/60 size exclusion column. The complex in 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM DTT was concentrated to 10 mg ml^{-1} using a Vivaspin concentrator for crystallization trials.

Crystallization and data collection

Diffraction quality crystals of both the native and the SeMet-substituted SART3 HAT repeat domain were obtained using hanging-drop vapor diffusion at 18°C, by mixing equal volumes of 20 mg ml^{-1} protein in 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 150 mM NaCl and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) with a reservoir solution containing 200 mM sodium formate, pH 7.5 and 30% PEG 4000. Crystals of SART3 LαHAT-C in complex with DUSP-UBL of USP4 were grown by hanging-drop vapor diffusion at 18°C. The protein solution was mixed in a 1:1 ratio with a reservoir solution containing 100 mM MES, pH 6.8, 200 mM $MgCl_2$ and 15% PEG 4000 Crystals of the SART3 NLS-ImpA complex were grown by hanging-drop vapor diffusion in a 1:1 protein/reservoir ratio, with the reservoir buffer containing 100 mM HEPES, pH 6.6, 1.6 M sodium citrate and 10 mM DTT at 18°C. All crystals were cryoprotected by adding ethyleneglycol to the reservoir solution to a final concentration of 20% prior to data collection.

All diffraction data sets were collected at a temperature of 100 K on the ADSC Quantum 315 CCD area detector at the 5C SB-II beamline of the Pohang Light Source in Pohang, Korea. Data sets for SeMet-substituted SART3 HAT repeat domain were collected at a wavelength of 0.9795 Å while data for the SART3 HAT-C and USP4 DUSP-UBL complex was collected at a wavelength of 1.0000 Å. The crystal of the SART3 HAT-C and USP4 DUSP-UBL complex diffracted to 3.0 Å resolution. Data set for the SART3 NLS and ImpA complex was collected at 1.7 Å resolution. All data sets were processed using HKL2000 (24). Statistics on the data collection are summarized in Table 1.

Structure solution and refinement

Initial selenium sites in the SART3 HAT repeat domain were identified by SAD with a 3.0 Å data sets from $P2_1$ crystal form (25) and these sites were used to obtain phases, and an initial backbone model was built using PHENIX AutoSol (26). After several rounds of refinement and manual model building using PHENIX (27) and Coot (28), the model was used for molecular replacement by Phaser (29) using the 2.6 Å native data. The structure was finalized by several cycles of manual model building in Coot and refinement in PHENIX. The final model for SART3 HAT repeat domain includes 94.3% of all residues in the favored regions of the Ramachandran plot.

The phasing information of SART3 LαHAT-C and DUSP-UBL of USP4 complex was initially obtained by the molecular replacement method using the LαHAT-C from this study as a search model by program MOLREP (30) from the CCP4 suite. Once the LαHAT-C domain of SART3 was located, the resulting electron density was good enough to orient the USP4 DUSP-UBL domains (PDB 3JYU unpublished data). The model was subsequently improved manually in Coot and refined with PHENIX. The final model has 90.3% of all residues in the favored regions of the Ramachandran plot.

Initial phase information of ImpA was obtained by molecular replacement with MOLREP using the crystal structure of mouse ImpA (PDB 1IAL (31)) as a search model. The initial electron density was improved by several rounds of refinements with PHENIX, and the NLS of SART3 were manually placed into the density map with Coot and refined to a final model which has 99.6% of all residues in the favored regions of Ramachandran plot. The stereochemical properties of all the structures were verified with MolProbity (32), and structure figures were generated in PyMOL (www.pymol.org). Refinement statistics for all three structures are summarized in Table 1.

Size exclusion chromatography with multi-angle light scattering (SEC-MALS) analysis

SEC-MALS experiments were performed using a high-performance liquid chromatography system (SHIMADZU coporation), equipped with a DAWN HELEOS II™ 18-angle MALS light scattering detector, an Optilab T-rEX refractive index detector, and ASTRA software for data analysis (all from Wyatt Technology). The 1 mg ml^{-1} of SART3 HAT repeat domain (residues 94–611) in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl and 1 mM DTT was injected onto a pre-equilibrated WTC-030S5 column (Wyatt Technology) at a flow rate of 0.5 ml min^{-1} at 25°C. The ASTRA 6.1 software was used to calculate the molecular mass and a BSA was used as a control protein.

Immunoprecipitation (IP) assay

Transfected HeLa cells were harvested and disrupted in 50 mM HEPES, pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 5 mM KCl, 0.1% Tween-20, 2 mM DTT, protease inhibitor cocktail (Roche). Cell lysates were centrifuged at 12 000 rpm for 30 min at 4°C and supernatants were incubated with anti-HA agarose (Sigma) bead for 4 h at 4°C. Bead were washed with buffer containing 50 mM HEPES, pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 5 mM KCl, 0.1% Tween-20, 2 mM DTT and eluted with 2x SDS–sample buffer. Samples were analyzed and detected by Western blotting.

Isothermal titration calorimetry (ITC)

ITC experiments were performed using ITC200 instrument (MicroCal, USA). All protein samples were dialyzed into 25 mM HEPES (pH 7.5), 150 mM NaCl and 0.5 mM TCEP before titration. Titrations were carried out by injecting consecutive aliquots of 300 μM protein samples into the ITC cell containing 15 μM protein samples except for the mutants of USP11 (using 600 μM) at 25°C. Binding stoichiometry (n), enthalpy (ΔH) and binding constants (K_D) were determined by fitting the data to a one-site binding model using Origin 7.0 (MicroCal, USA).

Immunofluorescence analysis

HeLa cells were seeded on coverslips and transfected with plasmids. After 48 h, cells were fixed with 4% formaldehyde, and incubated with anti-HA (Santa Cruz Biotechnology) or anti-Myc (Santa Cruz Biotechnology) antibodies, followed by secondary goat anti-rabbit antibody coupled to Alexa488 (Invitrogen) and goat anti-mouse antibody coupled to Alexa546 (Invitrogen). Nucleus was stained with

Table 1. Data collection and refinement statistics

	SART3 HAT repeat domain		SART3 HAT-C-USP4 DUSP-UBL	SART3 NLS-ImpA
	Native	Se-Met peak		
Data collection				
Space group	$P2_1$	$P2_1$	$P4_32_12$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c* (Å)	120.9, 80.5, 146.6	123.8, 81.0, 146.9	114.3, 114.3, 303.6	114.3, 114.3, 303.6
α, β, γ (°)	90.0, 99.3, 90.0	90.0, 100.1, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.0–2.6 (2.69–2.60)[a]	50.0–3.0 (3.11–3.00)	50.0–3.0 (3.08–3.00)	50.0–1.70 (1.76–1.70)
R_{sym} or R_{merge}	9.7 (39.5)	12.0 (33.1)	10.7 (35.6)	7.8 (45.3)
I / σI	14.5 (2.1)	12.1 (2.3)	11.7 (2.1)	19.8 (1.9)
Completeness (%)	95.1 (82.9)	97.3 (92.3)	98.1 (97.0)	99.0 (97.4)
Redundancy	2.5 (1.6)		4.0 (2.1)	2.8 (2.0)
Refinement				
Resolution (Å)	50.0–2.6		50.0–3.0	50.0–1.7
No. reflections	82 307		40 978	76 791
R_{work} / R_{free}	23.5 / 28.1		30.9 / 33.4	19.1 / 21.9
No. atoms				
Protein	16 629		8535	3385
Ligand/ion	0		0	0
Water	318		57	501
B-factors				
Protein	44.3		54.9	19.3
Ligand/ion	0		0	0
Water	54.3		31.2	37.6
R.m.s. deviations				
Bond lengths (Å)	0.008		0.005	0.017
Bond angles (°)	1.45		0.98	1.58
MolProbity Score[b]	1.90		2.98	1.54

All diffraction data were obtained from a single crystal.
[a]Values in parentheses are for highest-resolution shell.
[b]MolProbity score combines the clashcore, rotamer and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

DAPI (Invitrogen). Coverslips were mounted and fluorescence was visualized using 40X magnification on a confocal laser scanning microscope (Carl Zeiss, Inc.), and the images were analyzed using ZEN 2009 software.

Deubiquitination assay

HeLa cells were transfected with pCS2-His-ubiquitin, pCS2-USP4, pCS2-Myc-Prp3 and either pCS2-HA-SART3 or pCS2-HA-SART3$^{\Delta NLS}$ as indicated. After 24h, cells were resuspended in Buffer A (6 M guanidine-HCl, 0.1 M Na_2HPO_4/NaH_2PO_4, 10 mM imidazole, pH 8.0) and sonicated. Cell lysates were added to 50 μl of equilibrated Ni-NTA agarose followed by incubation for 4 h at room temperature. Beads were then washed twice with Buffer A, followed by two washes with Buffer A/TI (1 vol of Buffer A and 3 vol of Buffer TI (25 mM Tris-HCl, 20 mM imidazole at pH 6.8)), and finally one wash with Buffer TI. The protein conjugates were eluted in 30 μl 2X Laemmli/Imidazole (200 mM imidazole) and boiled at 95°C for 10 min. Elutes were analyzed by Western blotting.

Accession codes

The atomic coordinates and structure factors for the SART3 HAT repeat domain alone and in complex with USP4 DUSP-UBL domains as well as the complex of SART3 NLS peptide and importin-α have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 5CTQ, 5CTR and 5CTT, respectively).

RESULTS

SART3 has 12 HAT repeats

The crystal structure of SART3 HAT repeat domain (residues 94–611) was determined by the selenomethionyl single anomalous diffraction method (25). The statistics on data collection and refinement are summarized in Table 1. There are four molecules of SART3 HAT (labeled as molecules A, B, C and D), and the following description applies to all four molecules since the root-mean-square deviation (r.m.s.d.) for the four molecules varies from 0.84 to 1.41 Å for 485–491 Cα atoms (Supplementary Figure S1). The overall domain structure of SART3 with its HAT repeat domain redefined based on our crystal structure is shown in Figure 1. The crystal structure reveals that residues 94–600 form 12 HAT repeats as shown in Figure 1A and B, each HAT repeat being a helix-loop-helix (denoted as αA and αB), on the contrary to 8 HAT repeats predicted based on amino acid sequence analysis (2). The entire HAT repeat domain can be divided into two sub-domains: the 'N-terminal HAT repeat domain' consisting of five HAT repeats corresponding to residues 94–277 (HAT-N domain), and the 'C-terminal HAT repeat domain' consisting of the remaining seven HAT repeats corresponding to residues 308–600 (HAT-C domain), and a linker helix between the two sub-domains formed by residues 278–307 (Lα). They are colored in grey, purple and green, respectively. The HAT-N domain is almost perpendicular to the HAT-C domain, and while the HAT-C domain forms a curved struc-

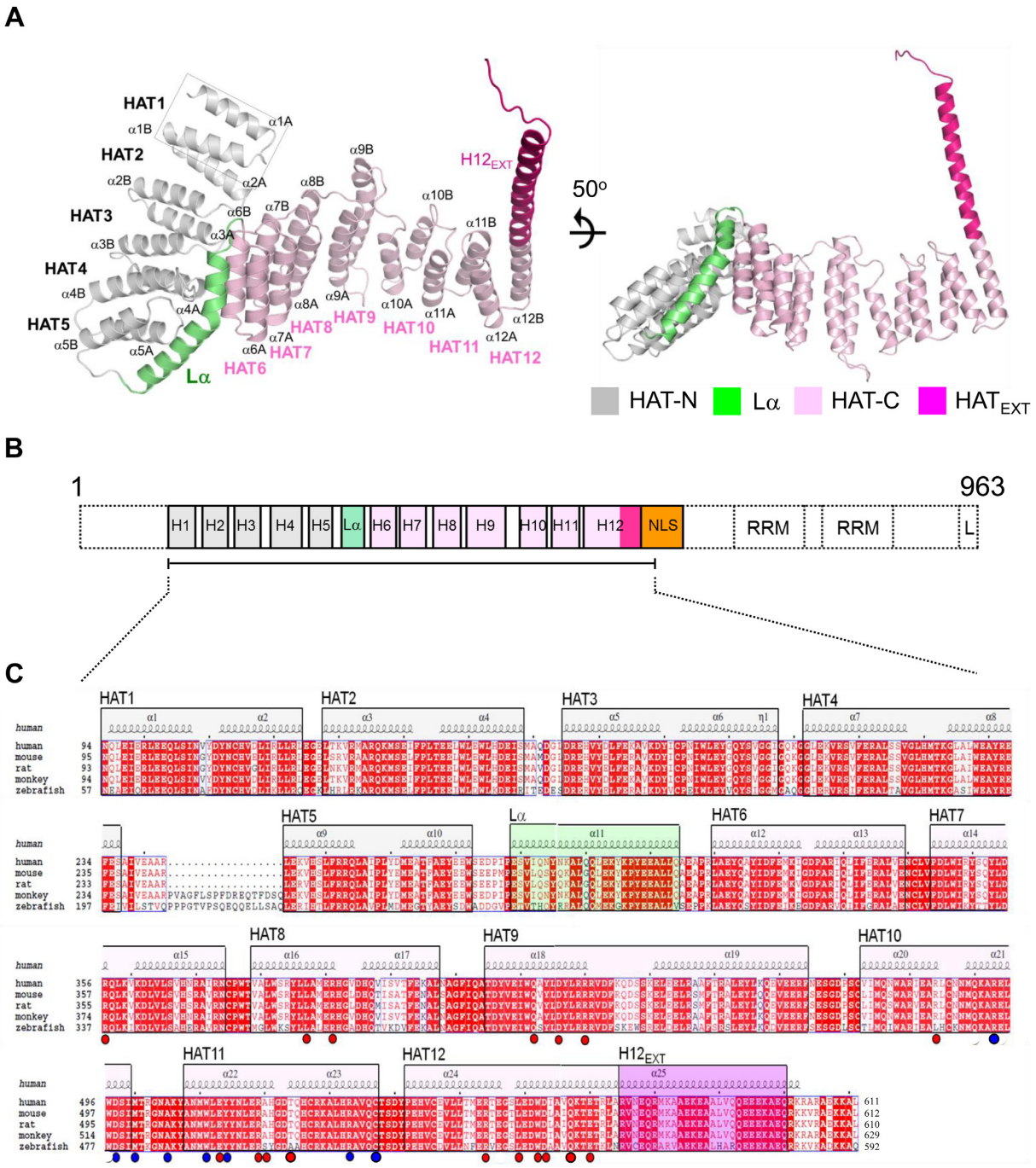

Figure 1. Crystal structure of the SART3 HAT repeat domain. (**A**) Ribbon representation of the SART3 HAT repeat domain. Five HAT repeats at the N-terminal (HAT-N), seven HAT repeats at the C-terminal (HAT-C) and the linker helix (Lα) are shown in grey, pink and green, respectively. The region after HAT12, $H12_{EXT}$, is highlighted in dark pink. (**B**) Domain organization of SART3 showing 12 HAT repeats, NLS, two RNA recognition motifs (RRMs) and one Sm-like (LSm) binding region. The construct used in the structural study is indicated below. (**C**) Structure based sequence alignment of the SART3 HAT repeat domain with the residues at dimer interface and the USP4 binding surface marked by red and blue circles, respectively. Secondary structural elements are shown above the sequence. Figure was generated by ESPript 3.0.

ture, the HAT-N domain together with Lα is rather flat as seen in Figure 1A. The extended long helix of HAT12 is formed by residues 574–600 ($H12_{EXT}$) as seen in the figure. The $H12_{EXT}$ in molecules B, C and D are shorter. The SART3 HAT repeat domain is strictly conserved among vertebrates (Figure 1C).

SART3 forms a tail-to-tail dimer

There are four molecules in the asymmetric unit in this crystal form as mentioned above, and interestingly molecules A and B are oriented in tail-to-tail fashion forming a homo-dimer (Figure 2A) with approximate dimensions of 160 Å × 95 Å × 65 Å. Molecules C and D also form a dimer in the same manner (Supplementary Figure S1A). The r.m.s.d. between AB and CD dimers is 1.02 Å for 994 Cα atoms. The two monomers in the pair are oriented in 'tail-to-tail' anti-parallel fashion with the concave faces of the HAT-C domain juxtaposing each other and the respective HAT-N domains facing away from the dimer interface. The dimeric nature of the SART3 HAT repeat domain seen in the crystal structure is consistent with the results from the solution. The elution volume of SART3 HAT repeat domain in solution is close to the expected molecular mass of SART3 HAT dimer as seen in Figure 2B. Also, to confirm the self-assembly of SART3 *in vivo*, HeLa cells were transfected with Myc- and HA-tagged SART3. As shown in Figure 2C, the purified HA-SART3 co-precipitated with Myc-SART3 suggesting that SART3 forms and functions as a dimer.

A total of 1200 $Å^2$ of the surface area of each monomer is buried at the interface, and this corresponds to 5.1% of the total surface area of a monomer, and 8.3% of the HAT-C, i.e. leaving out the HAT-N and Lα which is not involved in dimerization. The dimer interface is mostly formed by the residues from α12A and α12B (HAT12) of molecule A interacting with α7A, α8A and α9A of molecule B, and further interactions are made by residues from α10A, α11A and α11B from molecule A and α11B, α11A and α10A of the molecule B (Figure 2A and D). As seen in Figure 2D, the residues at the interface are both polar and hydrophobic in nature; interestingly they are conserved across species (Figure 1C). There are six arginine residues, namely R356, R391, R430, R484, R519 and R554, that are found at the interface, mostly located in the lower half (Supplementary Figure S2A), and all six of them make extensive interactions with the other monomer, e.g. R356, R391 and R519 make salt bridges with E570', D563' and E513', respectively, and R554 and R484 make several hydrogen bonds with Q422, D426 and T524' (Figure 2D). The interactions that occur at the dimer interface will be discussed later. The electrostatic surface diagram shows a highly charged surface (Figure 2E); in particular, the surface of HAT-N is relatively flat in nature and is highly negatively charged, with an estimated isoelectric point (pI) value of 4.7, and the implications of this will be discussed in depth later.

Linker between DUSP and UBL domains of USP4 mediates SART3 binding

Next, to understand the molecular basis for SART3 and DUB interaction, we co-purified various forms of the SART3 HAT repeat domain together with the DUSP-UBL domains of USP4 and USP15 (Figure 3A). As a result we obtained the crystal structure of the LαHAT-C domain in complex with the DUSP-UBL domains of USP4. The structure was solved by molecular replacement method using the structures of the LαHAT-C domain of SART3 from this study and the DUSP-UBL domains of *mouse* USP4 (PDB 3JYU unpublished data) at 3.0 Å resolution. As shown in Figure 3B, the SART3 LαHAT-C domain also forms a homo-dimer with the pair oriented tail-to-tail in an anti-parallel fashion in the same manner as seen in the SART3 HAT repeat domain alone, and USP4 DUSP-UBL domains bound at the hydrophobic region between HAT10 and HAT11 of the SART3 HAT-C domain (Figure 3B and C). The overall structure of LαHAT-C domain of SART3 in complex with USP4 is similar to the SART3 HAT repeat domain alone with the r.m.s.d. of 0.95 Å. The DUSP domain adopts an α-tripod fold while the UBL domain adopts a β-grasp fold as reported previously (33–35). The linker between the two domains forms a well-defined β-hairpin conformation as seen in the Figure 3C. Interestingly, the highly conserved linker region, namely the residues ^{125}GLFVKH^{130}, is the major interaction site for the SART3 HAT-C domain as shown in the figure (Figure 3C and D), i.e. the UBL domain does not make direct interaction with SART3. The side chain atoms of F127 and V128 of USP4 are embedded into the hydrophobic pocket created by R493, M500, W511, Y514, R533 and C537 of SART3 as shown in Figure 3E. In addition, there are several hydrogen bonds formed between USP4 and SART3, e.g. the side-chain of USP4 H130 and the side chain of SART3 D497; the main-chain carbonyl oxygen of USP4 L126 and the side-chain amide of SART3 R533. Additionally the side-chain atoms of USP4 W54 make contacts with the side chain atoms of SART3 K440. All the residues of SART3 that are involved in the interaction with USP4 are absolutely conserved (Figure 1C).

Interaction between SART3 HAT repeat domain and DUSP-UBL domains of USP4 and USP15

To obtain the binding stoichiometry (n), enthalpy change (ΔH) and affinity between the DUSP-UBL domains of USP4 and the SART3 HAT repeat domain, ITC experiments were carried out and the results are shown in Figure 4A. Also included in the analysis are USP15 as well as USP11, since they have the same domain structure as USP4 (34), and the sequence identities for the DUSP-UBL domains of these with USP4 are 71% and 39%, respectively. The result shows that the binding between the DUSP-UBL domains of USP4 and the SART3 HAT repeat domain is exothermic, with an enthalpy change (ΔH) of −21.6 ± 0.6 kcal mol^{-1} and the best fit was obtained with a one-site binding model with the binding stoichiometry (n) of 1.07 ± 0.02 and dissociation constant (K_D) of 0.88 ± 0.13 μM. Similarly, USP15 shows the same binding pattern and binding ratio ($n = 0.99 \pm 0.01$) but with a slightly lower K_D value of 0.19 ± 0.04 μM and a ΔH of −16.1 ± 0.1 kcal mol^{-1}, while USP11 shows no detectable binding. ITC measurements were also carried out using the LαHAT-C domain of SART3 under the same conditions.

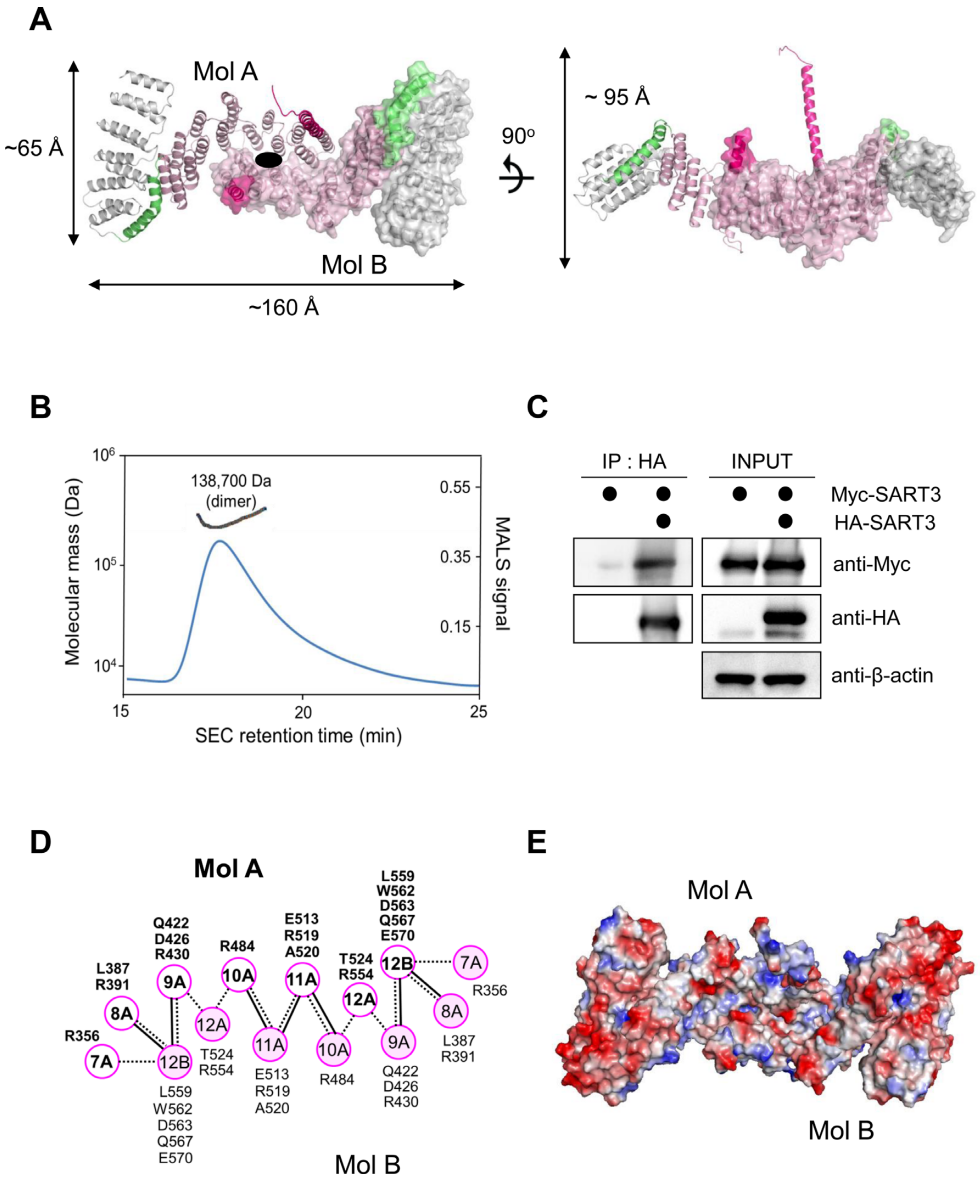

Figure 2. Self-association of the SART3 HAT repeat domain. (**A**) SART3 homo-dimer shown in the ribbon (Mol A) and surface (Mol B) representations. Residues of $H12_{EXT}$ (583–611) in Mol B are left out since they are not well-defined in the electron density maps. (**B**) Size exclusion chromatography with multi-angle light scattering (SEC-MALS) elution profile of SART3 HAT dimer. Molecular mass is estimated as 138.7 kDa while the value predicted for a monomer is 61.7 kDa. Protein concentration at the peak maxima evaluated from refractive index is 1 mg ml^{-1}. (**C**) Self-assembly of Myc- and HA-SART3 *in vivo.* HeLa cells were transfected with Myc-SART3 and HA-SART3. After HA-SART3 was purified on HA-agarose, co-precipitated Myc-SART3 was detected using anti-Myc antibodies. (**D**) Interactions at the dimer interface. Both the polar and hydrophobic interactions at the dimer interface are indicated by dashed and solid lines, respectively. (**E**) Electrostatic surface representation of the SART3 HAT dimer. Structural comparison of SART3 with CstF-77 is in Supplementary Figure S2.

Again, both USP4 and USP15 show the same binding stoichiometry ($n = 1.05 \pm 0.01$ and 1.09 ± 0.01, respectively), affinity ($K_D = 0.58 \pm 0.03$ μM and 0.14 ± 0.06 μM, respectively), and enthalpy change ($\Delta H = -21.1 \pm 0.4$ and -17.7 ± 0.5 kcal mol^{-1}, respectively) (Figure 4B). In order to verify the importance of the β-hairpin linker between the DUSP and UBL domains as seen in the crystal structure of SART3-USP4 complex, we generated a series of mutants, e.g. the linkers of USP4 (^{124}HGLFVKHC^{131}) and USP15 (^{124}QGMFVKHC^{131}) replaced by the USP11 linker (^{186}LPNIQ^{190}) and vice versa, denoted as USP4^{L11} and USP15^{L11}, USP11^{L4} and USP11^{L15}. As expected, both USP4^{L11} and USP15^{L11} lost the ability to bind the SART3 HAT repeat domain (Figure 4C). On the contrary USP11^{L4} and USP11^{L15}showed binding to SART3 at a 1:1 molar ratio ($n = 0.89 \pm 0.15$ and 1.06 ± 0.09, respectively), but with significant higher K_D ($K_D = 20.1 \pm 4.36$ μM and 18.7 ± 2.25 μM, respectively) and ΔH values ($\Delta H = -9.1 \pm 0.4$ and -5.7 ± 0.7 kcal mol^{-1}, respectively) (Figure 4D).

In order to confirm the interactions seen between USP4 and SART3 in the complex structure (Figure 3C and E) several mutations at the binding interface were tested. First,

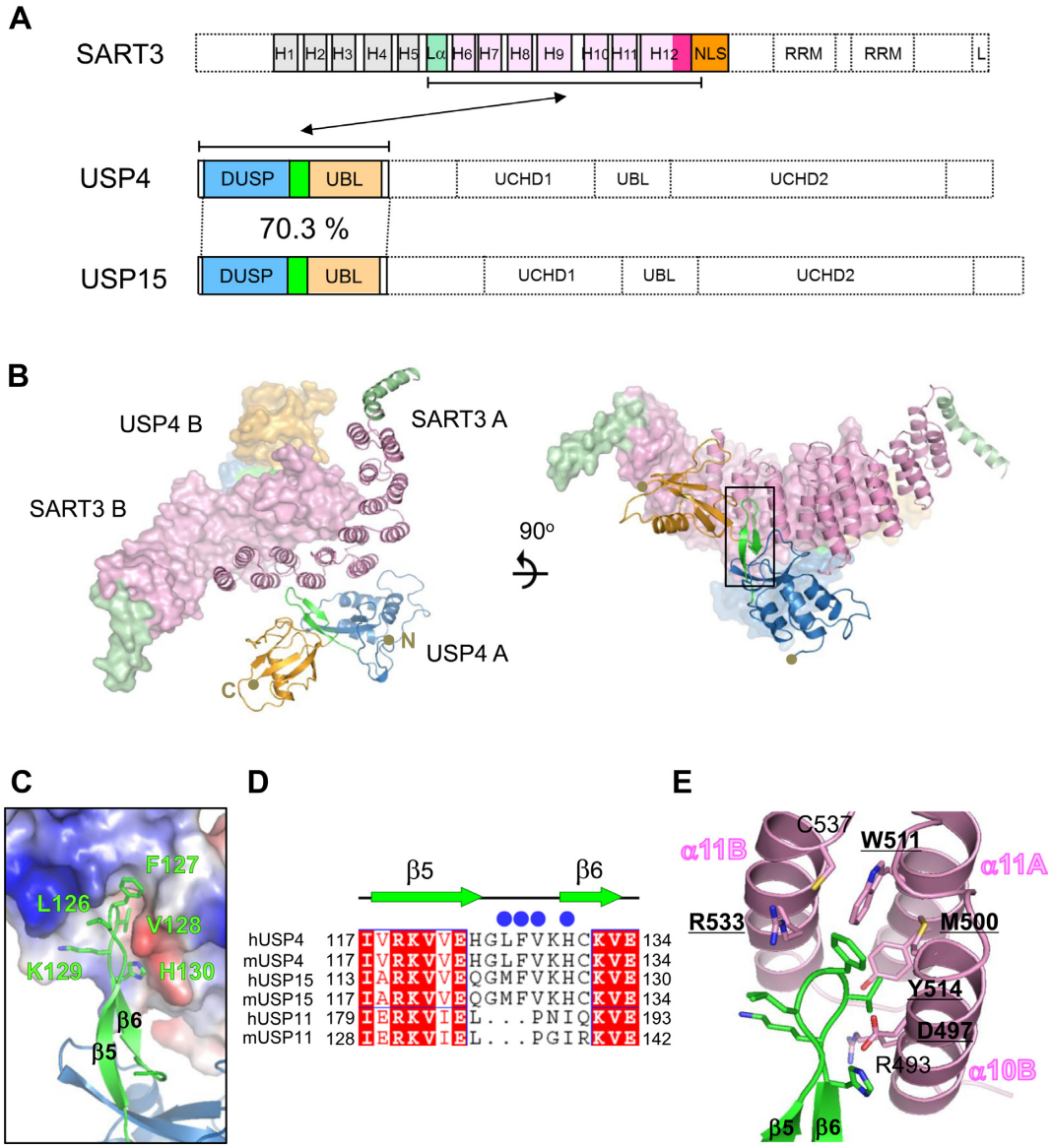

Figure 3. Crystal structure of SART3 LαHAT-C and USP4 DUSP-UBL complex. (**A**) Domain organization of USP4 and USP15. The constructs used in the complex structure are indicated below. (**B**) Crystal structure of the SART3 LαHAT-C domain complexed with the DUSP-UBL domains of USP4. They also form a homo-dimer and the two are shown in the ribbon and surface representations for clarity. (**C**) Close up view at the interface between USP4 and SART3 shown in electrostatic surface. (**D**) Structure based sequence alignment of linker region between the DUSP and UBL domains of USP4, USP15 and USP11. (**E**) Interactions at the interface with the key residues highlighted. The β-hairpin formed by the linker between the DUSP and UBL domains of USP4 binds on the interface between HAT10 and HAT11 of SART3. (Comparison with the structures of the DUSP-UBL domains alone is in Supplementary Figure S3).

two mutants of USP4, namely F127A and F127A/V128A, were generated. As shown in Figure 4E, neither F127A nor F127A/V128A showed any affinity for SART3, confirming that both Phe127 and Val128 of USP4 are critical for SART3 binding. Second, three mutants of SART3, namely D497A/R533A (DR), D497A/M500A/R533A (DMR) and D497A/M500A/W511A/Y514A/R533A (DMWYR) were generated and tested in cell. These are the residues lining the binding surface for USP4 (Figure 3E) and presumably USP15. As shown in Figure 4F, the DMR and the DMWYR mutants display reduced affinity for USP4 compared to wild-type while the DM mutant does not. Taken together, the β-hairpin of the linker region of the DUSP and UBL domains of both USP4 and USP15 are the major determinant for the SART3 binding, and only the SART3 HAT-C domain appears to be involved in binding, with no apparent contribution from the HAT-N domain.

SART3 NLS binds to importin-α in a bipartite manner

The residues 601–648 have been annotated as the NLS for SART3 based on sequence analysis using the PSORT II (36) and cNLS Mapper (38) programs. In order to identify the exact NLS required for the nuclear import of SART3 as well as the SART3-USP4 complex, we have co-purified NLS

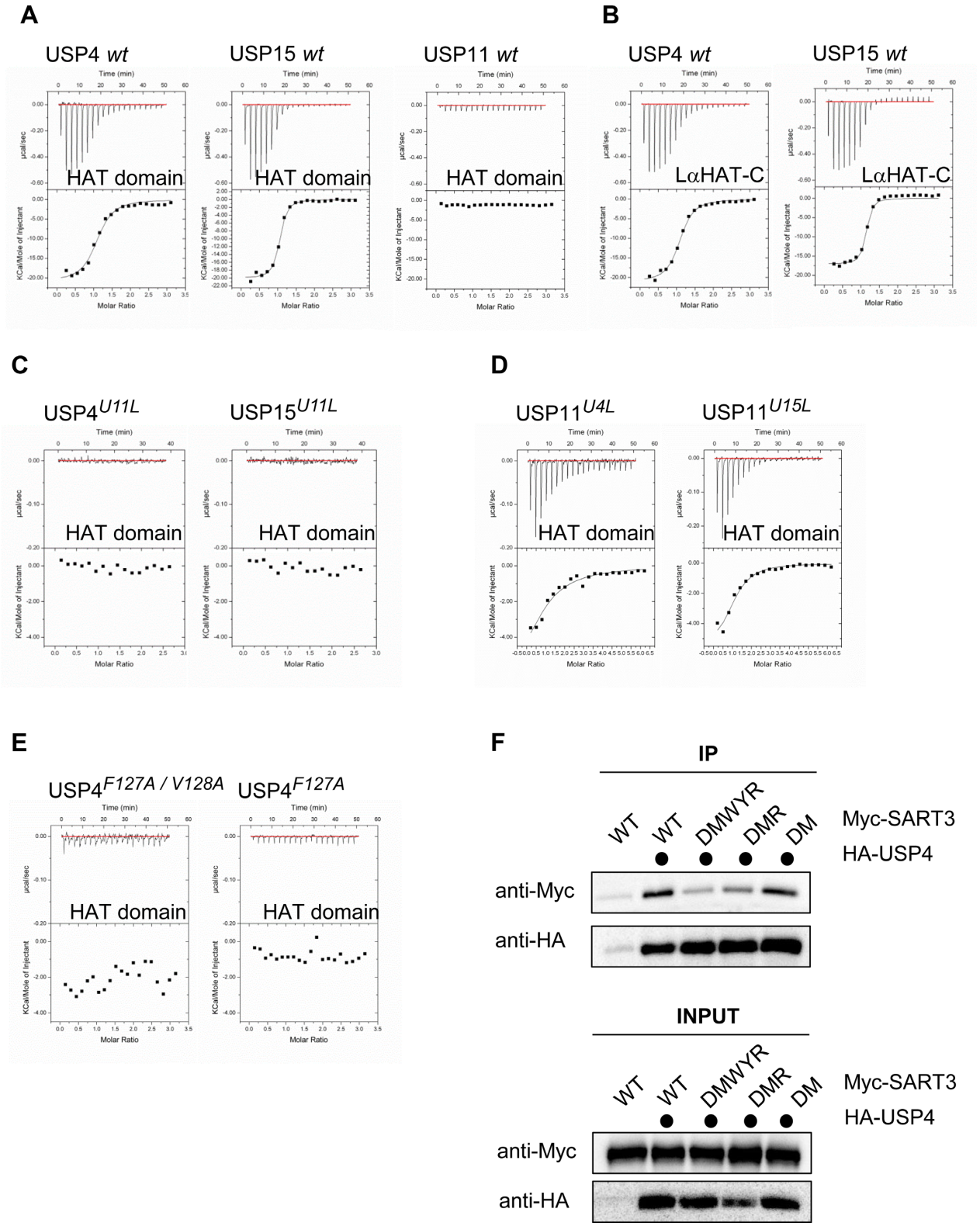

Figure 4. Interactions of SART3 with USPs. (**A**–**E**) ITC analysis on SART3 interactions with USP4, USP15 and USP11. The top panel shows the heat effects associated with the injection, while the bottom panel shows the binding isotherm corresponding to the data in the top panel and the best fitted curve with a one-site binding model. (**A**) DUSP-UBL domains of USP4 (left), USP15 (middle) and USP11 (right) added to the HAT repeat domain of SART3. (**B**) USP4 (left) and USP15 (right) to LαHAT-C domain of SART3. (**C**) Mutants of USP4^{U11L} (left) and USP15^{U11L} (right) to SART3 HAT repeat domain. (**D**) Mutants of USP11^{U4L} (left) and USP11^{U15L} (right) to SART3 HAT repeat domain. (**E**) USP4 F127A/V128A (left) and USP4 F127A (right) to SART3 HAT repeat domain. (**F**) Interaction between SART3 mutants and USP4 *in vivo.* HeLa cells were transfected with Myc-tagged mutants of SART3 and HA-USP4. HA-USP4 was purified on HA-agarose, and co-precipitated mutants of Myc-SART3 were detected using anti-Myc antibodies.

of SART3 with ImpA (residues 70–498 of *mouse* importin-α) and the complex was crystallized. The structure was determined by the molecular replacement method using the armadillo repeat (ARM) of ImpA crystal structure (PDB ID: 1IAL (31)) at 1.7 Å resolution. A difference map including ImpA gave clear conformation of the SART3 NLS bound. The final model (Figure 5A), composed of an elongated gently curved domain consisting of 10 ARM repeats, each repeat containing 3 α-helices (H1, H2 and H3) connected by loop, is practically identical to the starting model as indicated by a r.m.s.d. of 0.2 Å for 429 Cα atoms. SART3 NLS bound to both the major and the minor NLS-binding sites that are located within a shallow groove on the concave face of ImpA, namely the regions in ARM repeats 2–4 and ARM repeats 7–8, respectively. Both major and minor pockets are formed by solvent-exposed, conserved tryptophans together with a set of invariant asparagines four residues downstream. The residues 601–605 and 639–648 of SART3 are clearly bound to the minor and the major sites, respectively, while the linker between the two is not well defined in the electron density maps (Figure 5A). The interaction between the two proteins at both the major and minor sites involves both hydrophobic and electrostatic interactions.

As shown in Figure 5A, K602 and R603 are positioned at the P1' and P2' in the 'minor site'. The side chain of R603 forms a salt-bridge with E396 and a hydrogen bond with S360. In addition, the main chain nitrogen and oxygen atoms of K602, R603, A604 and R605 form hydrogen bonds and salt bridges with ImpA for a stable interaction. In the 'major' site, the hydrophobic portions of the basic side chains of residues R644 and R646 in P3 and P5 are sandwiched between the indole side chains of residues W231, W184 and W142 of ImpA. The side chains of K643 and R644 make hydrogen bonds with T155 and N228, respectively. And K643 and R646 form a salt bridge with D192 and Q181. Q640 forms hydrogen bonds with the side chains of S234 and Y277. The main chain carbonyl oxygens of Q640, P641, S642, R644, R646 and E648 forms hydrogen bonds with the side chains of R238, W231, N235, W184, N188, W142, N146 and S105. ITC analysis shows that the binding affinity (K_D) between SART3 NLS and ImpA is $K_D = 0.21 \pm 0.01$ μM with $n = 1.01 \pm 0.01$ and $\Delta H = -8.6 \pm 0.2$ kcal mol^{-1} (Figure 5B).

SART3 NLS is essential for the spliceosomal activity of USP4

SART3 is known to function as a recruiting as well as a targeting factor for USP4 in pre-mRNA processing, and USP4 does not appear to contain an NLS of its own, it is plausible that USP4 interaction with SART3 allows for USP4 nuclear translocation. USP15 was included in the study. To confirm whether SART3 is required for the nuclear translocation of USP4 and USP15, we first examined the localization of SART3 following transfection of HeLa cells with SART3 or a mutant that lacks residues 601–648, which encode for the NLS of SART3 (referred to as SART3$^{\Delta NLS}$ hereafter) based on the complex structure in this study. As shown in Figure 5C, the wild-type of SART3 is completely localized in the nucleus but SART3$^{\Delta NLS}$ is dispersed throughout the cytoplasm and excluded from the nucleus. These data indicate that the SART3 NLS plays an essential role in the localization of SART3. We further examined whether SART3 NLS affects the localization of USP4 and USP15. Consistent with previous studies, USP4 and USP15 localize to the nucleus when co-transfected with SART3 (10,37), while co-transfection with SART3$^{\Delta NLS}$ was not sufficient to drive the nuclear localization of USP4 and USP15 (Figure 5D). Collectively, these results suggest that SART3 NLS is essential for the translocation of USP4 and USP15 in addition to its own nuclear localization. Finally, we examined whether the SART3 NLS has an effect on the deubiquitination of USP4 nuclear substrates. The result shows that USP4 no longer deubiquitinates Prp3 in the presence of SART3$^{\Delta NLS}$ (Figure 5E). Since SART3 appears to be a targeting factor of USP4 for substrate Prp3, we examined whether USP4 co-localizes with Prp3 in the presence of SART3$^{\Delta NLS}$, and as expected, USP4 does not translocate to the nucleus, and therefore no longer co-localizes with Prp3 (Figure 5F). Taken together, these results indicate that SART3 NLS is essential for the deubiquitination of substrate of USP4.

DISCUSSION

In order to search for structurally homologous proteins of SART3, the DALI server was used (39). Initial attempts using the monomeric SART3, leaving out the H12$_{EXT}$, gave *mouse* CstF-77 (*m*CstF-77; PDB 2OOE (40)), *yeast* Rna14 (PDB 4E85 (41)) and *fungi* CstF-77 (*f*CstF-77; PDB 2UY1 (42)) as hits, with Z-scores of 22.2, 20.5 and 17.7 and r.m.s.d. values of 6.1, 6.0 and 5.9 Å, respectively. CstF-77 is a 77-kDa cleavage stimulation factor that is a member of the heterotrimeric protein complex that is required for the cleavage of the 3' signaling region from newly synthesized pre-mRNA molecules, and Rna14 is the *yeast* homologue of CstF-77. They share a sequence identity of only about 16.2% with SART3. While *m*CstF-77 and *y*CstF-77 consist of 12 HAT repeats, *f*CstF-77 has 11 HAT repeats. Interestingly, they also form two domain structures, namely HAT-N and HAT-C. Furthermore, all three form a tight homodimers, using the HAT-C domain with seven HAT repeats in a fashion similar to SART3. However, there are significant differences between SART3 and the CstF-77s. The orientation of the between HAT-N and HAT-C in SART3, e.g. the angle between HAT5 α2 and HAT6 α1, is about 80° while the corresponding values are 130°, 175° and 150° in *mouse*, *yeast* and *fungi* CstF-77s, respectively, and these structural differences between SART3 and CstF-77s affect not only the shape but also the dimensions of the molecules (Supplementary Figure S2).

One of the most striking differences is the region following the HAT12 repeat, H12$_{EXT}$. In SART3, this region (residues 574–600) was predicted to form a coiled-coil structure, but instead it forms an extended long helix as seen in Figures 1A and 2A. H12$_{EXT}$ does not make significant contacts with either the body of the HAT repeat domain or the H12$_{EXT}$ of the second molecule in the dimer. However, in both *m*CstF-77 and *y*CstF-77, the residues following HAT12 (residues 514–549 and 553–582, respectively) form two short helices and two short anti-parallel β-strands, and they are embedded at the dimeric interface, making con-

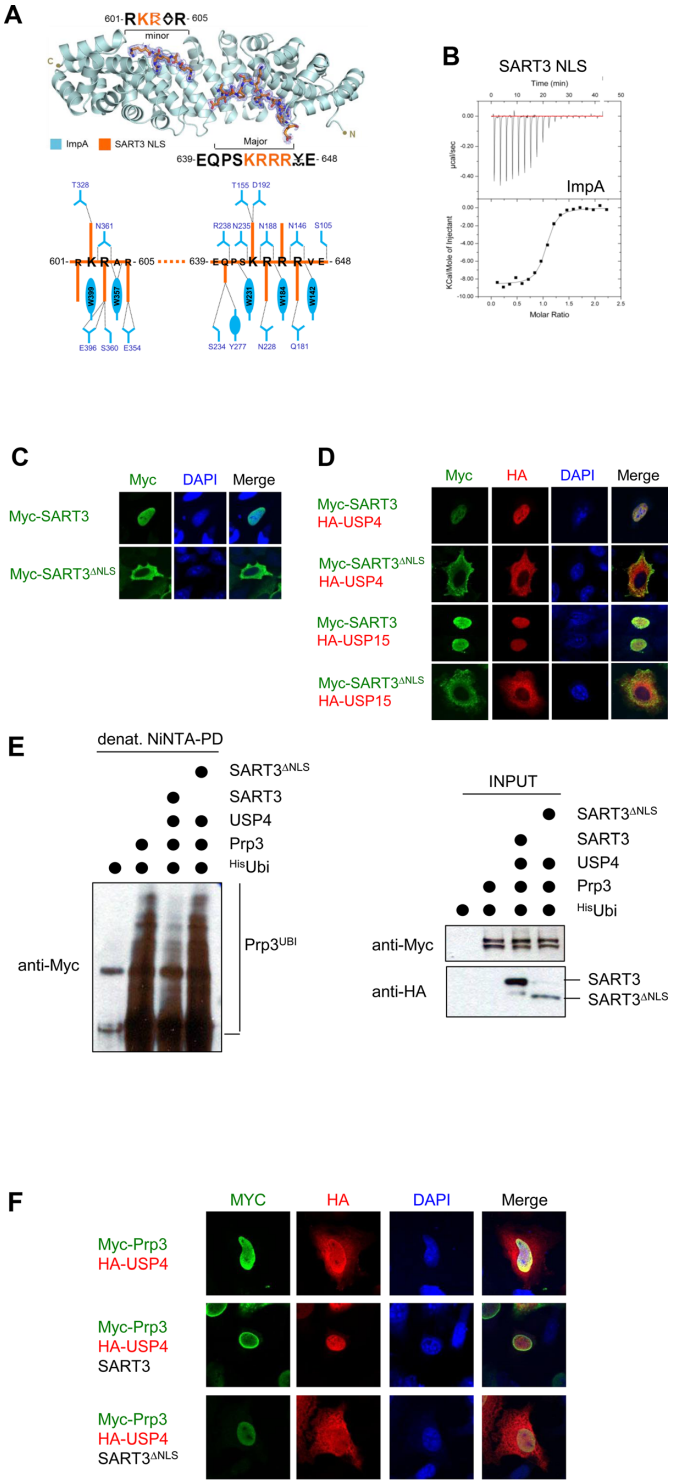

Figure 5. Translocation of USP4 and USP15 into the nucleus and SART3. (**A**) Crystal structure of SART3 NLS complexed to importin-α. ImpA (cyan) is shown ribbon drawing while and SART3 NLS (orange) is shown in stick representation, and the electron density map (2*Fo-Fc* map contoured at 1σ level) shown around the major and minor sites of ImpA. Sequence of SART3 NLS is shown for the two sites with the height of the amino acid reflecting the conservation across the species. Interactions between SART3 and ImpA are shown below. (**B**) ITC analysis of SART3 NLS binding to ImpA. (**C**) Localization of SART3 and SART3$^{\Delta NLS}$. HeLa cells transfected with Myc-SART3 or Myc-SART3$^{\Delta NLS}$ were stained with anti-Myc antibodies. Images were acquired by confocal microscope. (**D**) Intracellular localization of USP4 or USP15 in the presence of SART3$^{\Delta NLS}$. The localization of HA-USP4 or HA-USP15 in HeLa cells was analyzed by confocal microscopy in the presence of co-expressed SART3 or SART3$^{\Delta NLS}$. (**E**) Prp3 is not deubiquitinated by USP4 in the presence of SART3$^{\Delta NLS}$. Myc-Prp3 was expressed with His-ubiquitin, USP4 and either HA-SART3 or HA-SART3$^{\Delta NLS}$ and covalently modified proteins were purified on NiNTA-agarose under denaturing conditions. Ubiquitinated Prp3 was detected by anti-Myc antibodies. (**F**) USP4 does not co-localize with Prp3 in the presence of SART3$^{\Delta NLS}$. The localization of HA-USP4 or Myc-Prp3in HeLa cells was analyzed by confocal microscopy in the presence of co-expressed SART3 or SART3$^{\Delta NLS}$.

tacts with HAT7-10 of the other monomer. In *f*CstF-77, the corresponding segment (residues 417–454) forms an α-helix, again located at the dimeric interface interacting with HAT6-8 of the other monomer. Another notable difference is the arrangement of the HAT-N and HAT-C domains. Another interesting point is the nature of the electrostatic surface (Supplementary Figure S2). This $H12_{EXT}$ helix is composed of highly charged residues, both negative and positive. The surface consisting of E578, E585, E587, E594, E595, E596 and E599 forms a negatively charged patch while the surface comprising R580, K582, K586 and K597 forms positively charged streaks along the helix. Considering that $H12_{EXT}$ is followed by the two RRM domains, these positive streaks on the $H12_{EXT}$ may play a role in binding the other components of the U4 or U6 snRNPs.

The dimer interfaces of both SART3 and CstF-77s involve only the HAT-C domains, yet they differ from one another. First, the two HAT-C domains of SART3 are almost parallel, while in the CstF-77s they are more open at the region corresponding to $H12_{EXT}$ and therefore shaped more like a boat (Figure 2A). Second, the charged residues are clustered at the bottom region in SART3 while they are more evenly distributed in CstF-77s. As mentioned earlier, in the case of SART3 the residues involved in the dimer interactions are located at the bottom half of the HAT-C domain. Furthermore, the residues involved in the dimerization of CstF-77s are spread out (Figure 2B and Supplementary Figure S2B). The surface area buried on the interfaces are 2500 $Å^2$, 2500 $Å^2$ and 2000 $Å^2$ for *m*CstF-77, *y*CstF-77 and *f*CstF-77, respectively, and these correspond to 8.8% to 9.5% of the total area for a monomer. These values are somewhat lower than what is typically reported for a dimer (43), but if one considers only the HAT-C that is involved in dimerization, these values become 12.3% to 14.5%. These values are significantly larger than those of SART3. However, if the $H12_{EXT}$ and the corresponding regions are excluded in the calculation, the dimeric interfaces are 960 $Å^2$, 700 $Å^2$ and 1100 $Å^2$, respectively, corresponding to 2.5 to 5.1% of the total area of each monomer. These values then become comparable or even less than that of SART3, suggesting that the region corresponding to $HAT12_{EXT}$ of CstF-77s contributes significantly to dimer stability (Supplementary Figure S2B). In the case of *m*CstF-77 light-scattering, yeast two-hybrid as well as analytical centrifugation, measurements all suggested self-association (40).

The crystal structure of USP4 and SART3 complex in this study reveals that the linker between the DUSP and UBL domains of USP4 forms a β-structure, and the residues forming the β-hairpin loop, namely L126, F127, V128 and H130, are the binding determinant for SART3 binding. Both the structure of the linker and the orientation of the DUSP and UBL domains in the complex structure are similar to the 'closed' conformation of the three reported structures for DUSP-UBL domains (PDBs 3JYU, 4A3O (34) and 3T9L (35)) (Supplementary Figure S3). All the residues of both USP4 and SART3 involved in the interaction are absolutely conserved (Figure 1C). The linker region between the DUSP and UBL domain of USP15 is almost identical to that of USP4 as shown in Figure 3D except leucine is replaced by methionine which is quite similar in volume (44). On the other hand, this linker is shorter by three residues in USP11. Although it was shown that this region also adopts a similar β–hairpin structure (45), it does not have the key residues for SART3 binding. Therefore, no binding is expected (Figure 3C and D) and this was confirmed as shown in Figure 4. Furthermore, the ITC results on the USP11 mutants with the β-hairpin loop replaced by those of USP4 and USP15 exhibit SART3 binding, while the USP4 and USP15 mutants with the shorter loop do not bind SART3. Taken together, both the length and the composition of the β-hairpin region in the linker between the DUSP and UBL domains are critical in binding SART3.

The crystal structure of the SART3 NLS-ImpA complex shows that K602-R603 residues occupy the P1'-P2' pockets (minor site) while K643-R644-R645-R646 residues occupy the P1-P2-P3-P4 pockets (major site), and this is in good agreement with the bipartite cNLS consensus sequence, KR-$X_{10\text{-}12}$-K-K/R-X-K/R (Figure 5) (46,47). The estimated K_D between SART3 NLS and ImpA is comparable to those reported for bipartite cNLSs (10–1 μM). The linker between the major and the minor sites in SART3 is 39 residues long and the residues 606–638 are not well defined in the electron density maps. In other bipartite binding proteins, the linker varies in both length and amino acid composition and it has been suggested that it plays a role in the specific recognition of the cargo protein by ImpA. Confocal microscopy analysis clearly demonstrates that the NLS is required for SART3 to relocate to the nucleus and both USP4 and USP15 are recruited to the nucleus in the presence of SART3. Furthermore, when the NLS is deleted from SART3, deubiquitination activity of USP4 is diminished although the interaction between SART3 and USPs are intact (Figure 5E and Supplementary Figure S4). Therefore, SART3 NLS is essential for the proper relocalization of USP4 and USP15 to the nucleus and their subsequent nuclear deubiquitination activities.

In addition to the recruiting role of SART3, it functions as a targeting factor at the spliceosome during the recycling phase. It is required for proper recycling of U4 and U6 components through dissociation from U4/U6.U5 tri-snRNP. For this, SART3 together with USP4 needs to recognize the K63-linked polyubiquitinated Prp3 in the tri-snRNP (10). Although Prp3 is a substrate of USP4, the two do not make a direct interaction rather SART3 mediates the interaction between the two. Furthermore, the highly conserved C-terminal region of Prp3 was identified to interact with the N-terminal region of SART3 (10,48). Interestingly, the 416–550 region of Prp3 is highly basic with an estimated pI value of 10.3. It is tempting to suggest that Prp3 binds to the highly negatively charged (pI ≈ 4.7) surface of HAT-N surface (Figure 2E).

The details of how SART3 initiates the recycling functions at the spliceosome are largely unknown. However, recent the structural report on the complex of core of U6 snRNP and Prp24 sheds some light on the role and mechanism (49). Prp24, which has four RRMs and a LSm sequence at the C-terminus with a pI value of 9.3, revealed a unique interlocked architecture of RNA and protein (9,49). Three of the four RRMs (RRM2, RRM3 and RRM4) as well as the region immediately preceding RRM1 in Prp24 form extensive interactions with RNA. Although the sequence identity between Prp24 and the C-terminal end of

SART3 is only 22%, the pI value for the C-terminal region of SART3, i.e. beyond the NLS, is 9.4. Also, it is well established that SART3 binds specific U6 snRNA (3). Based on the fact that multiple RRMs in Prp24 are required for the proper recognition of U6 snRNA, one may contemplate how SART3, which has only two RRMs facilitates the necessary recognition of U6 snRNA. It may well be that SART3 also functions as a dimer in the pre-mRNA processing, i.e. four RRM domains.

The functions of USP4 and USP15 are unraveling rapidly, and some appear to be somewhat parallel. It is quite possible that USP15 functions as another spliceosomal DUB as suggested by reciprocal proteomic analysis (22), and SART3 serves as a versatile platform for dynamic and accurate deubiquitination actions in spliceosome. We hope that the information reported in this study aids in understanding the complex and dynamic actions of the spliceosome.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Dr. Kook Han Kim and Mr. Sang Chul Shin for the helpful discussions and the staff at Pohang Light Source for assistance during data collection.

FUNDING

Global Research Laboratory program of the Ministry of Science, ICT and Future Planning [NRF-2011-0021713]; Institutional grant from Korea Institute of Science and Technology. Funding for open access charge: Global Research Laboratory program of the Ministry of Science, ICT and Future Planning [NRF-2011-0021713]; Institutional grant from Korea Institute of Science and Technology.
Conflict of interest statement. None declared.

REFERENCES

1. Gu,J., Shimba,S., Nomura,N. and Reddy,R. (1998) Isolation and characterization of a new 110 kDa human nuclear RNA-binding protein (p110nrb). *Biochim. Biophys. Acta*, **1399**, 1–9.
2. Preker,P.J. and Keller,W. (1998) The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem. Sci.*, **23**, 15–16.
3. Bell,M., Schreiner,S., Damianov,A., Reddy,R. and Bindereif,A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.*, **21**, 2724–2735.
4. Yang,D., Nakao,M., Shichijo,S., Sasatomi,T., Takasu,H., Matsumoto,H., Mori,K., Hayashi,A., Yamana,H., Shirouzu,K. *et al.* (1999) Identification of a gene coding for a protein possessing shared tumor epitopes capable of inducing HLA-A24-restricted cytotoxic T lymphocytes in cancer patients. *Cancer Res.*, **59**, 4056–4063.
5. Iseki,K., Matsunaga,H., Komatsu,N., Suekane,S., Noguchi,M., Itoh,K. and Yamada,A. (2010) Evaluation of a new oil adjuvant for use in peptide-based cancer vaccination. *Cancer Sci.*, **101**, 2110–2114.
6. Liu,Y., Timani,K., Mantel,C., Fan,Y., Hangoc,G., Cooper,S., He,J.J. and Broxmeyer,H.E. (2011) TIP110/p110nrb/SART3/p110 regulation of hematopoiesis through CMYC. *Blood*, **117**, 5643–5651.
7. Zhao,W., Liu,Y., Timani,K.A. and He,J.J. (2014) Tip110 protein binds to unphosphorylated RNA polymerase II and promotes its phosphorylation and HIV-1 long terminal repeat transcription. *J. Biol. Chem.*, **289**, 190–202.
8. Chen,W. and Moore,M.J. (2014) The spliceosome: disorder and dynamics defined. *Curr. Opin. Struct. Biol.*, **24**, 141–149.
9. Rader,S.D. and Guthrie,C. (2002) A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA*, **8**, 1378–1392.
10. Song,E.J., Werner,S.L., Neubauer,J., Stegmeier,F., Aspden,J., Rio,D., Harper,J.W., Elledge,S.J., Kirschner,M.W. and Rape,M. (2010) The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev.*, **24**, 1434–1447.
11. Long,L., Thelen,J.P., Furgason,M., Haj-Yahya,M., Brik,A., Cheng,D., Peng,J. and Yao,T. (2014) The U4/U6 recycling factor SART3 has histone chaperone activity and associates with USP15 to regulate H2B deubiquitination. *J. Biol. Chem.*, **289**, 8916–8930.
12. Zhao,B., Schlesiger,C., Masucci,M.G. and Lindsten,K. (2009) The ubiquitin specific protease 4 (USP4) is a new player in the Wnt signalling pathway. *J. Cell. Mol. Med.*, **13**, 1886–1895.
13. Fan,Y.H., Yu,Y., Mao,R.F., Tan,X.J., Xu,G.F., Zhang,H., Lu,X.B., Fu,S.B. and Yang,J. (2011) USP4 targets TAK1 to downregulate TNFalpha-induced NF-kappaB activation. *Cell Death Differ.*, **18**, 1547–1560.
14. Zhang,X., Berger,F.G., Yang,J. and Lu,X. (2011) USP4 inhibits p53 through deubiquitinating and stabilizing ARF-BP1. *EMBO J.*, **30**, 2177–2189.
15. Zhang,L., Zhou,F., Drabsch,Y., Gao,R., Snaar-Jagalska,B.E., Mickanin,C., Huang,H., Sheppard,K.A., Porter,J.A., Lu,C.X. *et al.* (2012) USP4 is regulated by AKT phosphorylation and directly deubiquitylates TGF-beta type I receptor. *Nat. Cell Biol.*, **14**, 717–726.
16. Schweitzer,K., Bozko,P.M., Dubiel,W. and Naumann,M. (2007) CSN controls NF-kappaB by deubiquitinylation of IkappaBalpha. *EMBO J.*, **26**, 1532–1541.
17. Inui,M., Manfrin,A., Mamidi,A., Martello,G., Morsut,L., Soligo,S., Enzo,E., Moro,S., Polo,S., Dupont,S. *et al.* (2011) USP15 is a deubiquitylating enzyme for receptor-activated SMADs. *Nat. Cell Biol.*, **13**, 1368–1375.
18. Eichhorn,P.J., Rodon,L., Gonzalez-Junca,A., Dirac,A., Gili,M., Martinez-Saez,E., Aura,C., Barba,I., Peg,V., Prat,A. *et al.* (2012) USP15 stabilizes TGF-beta receptor I and promotes oncogenesis through the activation of TGF-beta signaling in glioblastoma. *Nat. Med.*, **18**, 429–435.
19. Urbe,S., Liu,H., Hayes,S.D., Heride,C., Rigden,D.J. and Clague,M.J. (2012) Systematic survey of deubiquitinase localization identifies USP21 as a regulator of centrosome- and microtubule-associated functions. *Mol. Biol. Cell*, **23**, 1095–1103.
20. Garcia-Santisteban,I., Zorroza,K. and Rodriguez,J.A. (2012) Two nuclear localization signals in USP1 mediate nuclear import of the USP1/UAF1 complex. *PloS One*, **7**, e38570.
21. Soboleva,T.A., Jans,D.A., Johnson-Saliba,M. and Baker,R.T. (2005) Nuclear-cytoplasmic shuttling of the oncogenic mouse UNP/USP4 deubiquitylating enzyme. *J. Biol. Chem.*, **280**, 745–752.
22. Sowa,M.E., Bennett,E.J., Gygi,S.P. and Harper,J.W. (2009) Defining the human deubiquitinating enzyme interaction landscape. *Cell*, **138**, 389–403.
23. Faesen,A.C., Luna-Vargas,M.P. and Sixma,T.K. (2012) The role of UBL domains in ubiquitin-specific proteases. *Biochem. Soc. Trans.*, **40**, 539–545.
24. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.*, **276**, 307–326.
25. Hendrickson,W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.
26. Terwilliger,T.C., Adams,P.D., Read,R.J., McCoy,A.J., Moriarty,N.W., Grosse-Kunstleve,R.W., Afonine,P.V., Zwart,P.H. and Hung,L.W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.*, **65**, 582–601.
27. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

28. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
29. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
30. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 22–25.
31. Kobe,B. (1999) Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. *Nat. Struct. Biol.*, **6**, 388–397.
32. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
33. de Jong,R.N., Ab,E., Diercks,T., Truffault,V., Daniels,M., Kaptein,R. and Folkers,G.E. (2006) Solution structure of the human ubiquitin-specific protease 15 DUSP domain. *J. Biol. Chem.*, **281**, 5026–5031.
34. Elliott,P.R., Liu,H., Pastok,M.W., Grossmann,G.J., Rigden,D.J., Clague,M.J., Urbe,S. and Barsukov,I.L. (2011) Structural variability of the ubiquitin specific protease DUSP-UBL double domains. *FEBS Lett.*, **585**, 3385–3390.
35. Harper,S., Besong,T.M., Emsley,J., Scott,D.J. and Dreveny,I. (2011) Structure of the USP15 N-terminal domains: a beta-hairpin mediates close association between the DUSP and UBL domains. *Biochemistry*, **50**, 7995–8004.
36. Nakai,K. and Horton,P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.*, **24**, 34–36.
37. Timani,K.A., Liu,Y., Suvannasankha,A. and He,J.J. (2014) Regulation of ubiquitin-proteasome system-mediated Tip110 protein degradation by USP15. *Int. J. Biochem. Cell Biol.*, **54**, 10–19.
38. Kosugi,S., Hasebe,M., Tomita,M. and Yanagawa,H. (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 10171–10176.
39. Holm,L. and Rosenstrom,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
40. Bai,Y., Auperin,T.C., Chou,C.Y., Chang,G.G., Manley,J.L. and Tong,L. (2007) Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. *Mol. Cell*, **25**, 863–875.
41. Paulson,A.R. and Tong,L. (2012) Crystal structure of the Rna14-Rna15 complex. *RNA*, **18**, 1154–1162.
42. Legrand,P., Pinaud,N., Minvielle-Sebastia,L. and Fribourg,S. (2007) The structure of the CstF-77 homodimer provides insights into CstF assembly. *Nucleic Acids Res.*, **35**, 4515–4522.
43. Janin,J., Miller,S. and Chothia,C. (1988) Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.*, **204**, 155–164.
44. Richards,F.M. (1977) Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.*, **6**, 151–176.
45. Harper,S., Gratton,H.E., Cornaciu,I., Oberer,M., Scott,D.J., Emsley,J. and Dreveny,I. (2014) Structure and catalytic regulatory function of ubiquitin specific protease 11 N-terminal and ubiquitin-like domains. *Biochemistry*, **53**, 2966–2978.
46. Kosugi,S., Hasebe,M., Matsumura,N., Takashima,H., Miyamoto-Sato,E., Tomita,M. and Yanagawa,H. (2009) Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.*, **284**, 478–485.
47. Marfori,M., Lonhienne,T.G., Forwood,J.K. and Kobe,B. (2012) Structural basis of high-affinity nuclear localization signal interactions with importin-alpha. *Traffic*, **13**, 532–548.
48. Medenbach,J., Schreiner,S., Liu,S., Luhrmann,R. and Bindereif,A. (2004) Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling. *Mol. Cell. Biol.*, **24**, 7392–7401.
49. Montemayor,E.J., Curran,E.C., Liao,H.H., Andrews,K.L., Treba,C.N., Butcher,S.E. and Brow,D.A. (2014) Core structure of the U6 small nuclear ribonucleoprotein at 1.7-A resolution. *Nat. Struct. Mol. Biol.*, **21**, 544–551.