

## Structural Characterization of RNA Recognition Motif-2 Domain of SART3

Kyeong-Mi Bang,<sup>†,‡</sup> Na Youn Cho,<sup>†</sup> Won-Je Kim,<sup>†</sup> Ae-Ryung Kim,<sup>§</sup> Hyun Kyu Song,<sup>‡</sup>  
Eunice EunKyeong Kim,<sup>§</sup> and Nak-Kyoon Kim<sup>†,\*</sup>

<sup>†</sup>Advanced Analysis Center, Korea Institute of Science and Technology, Seoul 02792, Korea

\*E-mail: nkkim@kist.re.kr

<sup>‡</sup>Department of Life Sciences, Korea University, Seoul 02841, Korea

<sup>§</sup>Biomedical Research Institute, Korea Institute of Science and Technology, Seoul 02792, Korea

Received November 23, 2016, Accepted January 31, 2017, Published online February 28, 2017

Squamous cell carcinoma antigen recognized by T-cells 3 (SART3) is an essential recycling factor in pre-mRNA splicing, which is required for association of U4/U6 small nuclear ribonucleoprotein (snRNP). SART3 contains two RNA recognition motifs (RRMs), and they are responsible for the tertiary interaction with U6 small nuclear RNA. Despite the importance of structural studies for understanding complicated U4/U6 snRNP recycling mechanism, only a few of them have been performed for SART3. Here, the structure of SART3 RRM2 was characterized by heteronuclear multi-dimensional nuclear magnetic resonance experiments. Nearly complete <sup>1</sup>H, <sup>15</sup>N, and <sup>13</sup>C chemical shifts of the backbone residues of RRM2 were assigned. In addition, the secondary structure of RRM2 were predicted by the chemical shift index and TALOS+ analyses, and the results showed that RRM2 forms a “β<sub>1</sub>-α<sub>1</sub>-β<sub>2</sub>-β<sub>3</sub>-α<sub>2</sub>-β<sub>4</sub>-β<sub>5</sub>” structure, where β<sub>4</sub> is not common in the canonical RRM domain structures. Our results will provide structural basis for investigation of SART3-mediated U4/U6 snRNP complex formation.

**Keywords:** Squamous cell carcinoma antigen recognized by T-cells 3, Splicing, RNA recognition motifs, Nuclear magnetic resonance, Structure

### Introduction

RNA splicing is a series of process that removes introns and connects exons of the pre-mRNA. RNA splicing is catalyzed in a large RNA–protein complex called spliceosome, where more than 100 proteins and U1, U2, U4, U5, and U6 small nuclear RNAs (snRNAs) are associated.<sup>1,2</sup> After completion of a splicing, the protein- and RNA-components of the spliceosome are recycled for the initiation of a new splicing process. Among these components, the post-spliceosomal U4 and U6 small nuclear ribonucleo-proteins (snRNPs) are rejoined to form functional U4/U6 snRNPs, and squamous cell carcinoma antigen recognized by T-cells 3 (SART3) serves as a recycling factor for the formation of U4/U6 snRNP in human.

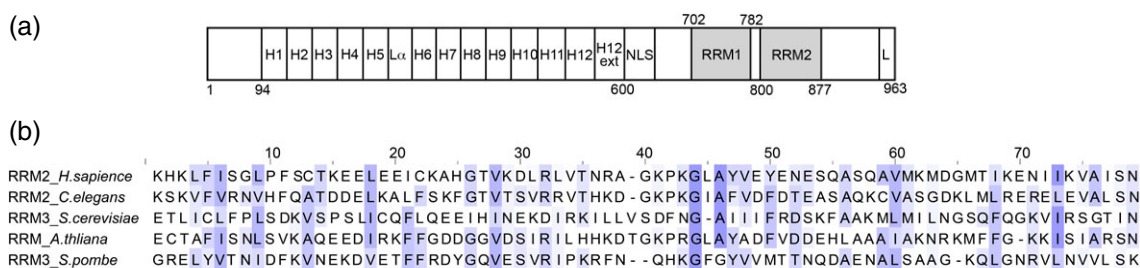
SART3 is a large (110 kDa) RNA-binding protein, which is involved in many biological processes including pre-mRNA splicing,<sup>3</sup> regulation of gene transcription,<sup>4</sup> and cell proliferation and differentiation.<sup>5</sup> SART3 consists of 12 half-a-tetratricopeptide (HAT) domains in the N-terminal half, a nuclear localization signal (NLS) sequence, two RNA recognition motifs 1 and 2 (RRM1 and RRM2), and an extension of 10 highly conserved amino acids residues in the C-terminal (C10) (Figure 1).<sup>6,7</sup>

The recycling of U4/U6 snRNP is associated with SART3 binding to U6 snRNA.<sup>3</sup> Both the HAT and two

RRM domains are required for U4/U6 snRNP recycling *in vitro*. However, for U6 RNA binding, the HAT domain is not necessary, but the two RRM domains are critical.<sup>6</sup> The conserved C10 domain is not essential for recycling of U4/U6 snRNP *in vitro*,<sup>6</sup> but the interaction of C10 with seven LSm2–LSm8 proteins largely enhances the efficient interaction of SART3 with U6 snRNA compared to the interaction of SART3 itself to U6 snRNA.<sup>8</sup>

Prp24 is a yeast ortholog of SART3, which is composed of four RRM1–4 and the C-terminal conserved sequence.<sup>6</sup> Compared to SART3, Prp24 lacks the counterpart of HAT of SART3. The tertiary interactions between the RRM domains of Prp24 and U6 snRNA were revealed recently by the structural investigation of the Prp24–U6 snRNA complex.<sup>9</sup> In the structure, RRM1 does not have any specific contact with U6 snRNA, but RRM2 and RRM3 significantly interact with the bulge region of U6 snRNA. This interaction of RRM with single-stranded RNA is a canonical RRM–RNA binding pattern as observed in many other RRM–RNA complexes.<sup>10</sup> RRM4 plays a role in forming the overall conformation of the RNA–protein complex, in that the RRM4 interaction with two double stranded RNA regions, internal stem loop and telestem, caused a significant RNA bend.

In comparison with Prp24, there are only two RRM1–2 in SART3. Therefore, it is important to reveal how the two



**Figure 1.** (a) Domain structure of human SART3. Twelve HAT repeats, NLS, two RRM, and a Sm-like (LSm) binding region are shown. The boundaries of the domains are indicated by amino acid residue number. (b) Sequence alignment of the SART3 RRM2 domain with the RRM domains of homologous proteins. The SART3 RRM2 sequence (NP\_055521) is aligned with *Caenorhabditis elegans* (CAA97405), *Saccharomyces cerevisiae* (P49960), *Arabidopsis thaliana* (CAB45062), and *Schizosaccharomyces pombe* (CAB52740). The identity of amino acids is shown by the purple shading, where the more conserved amino acids are shown in darker color.

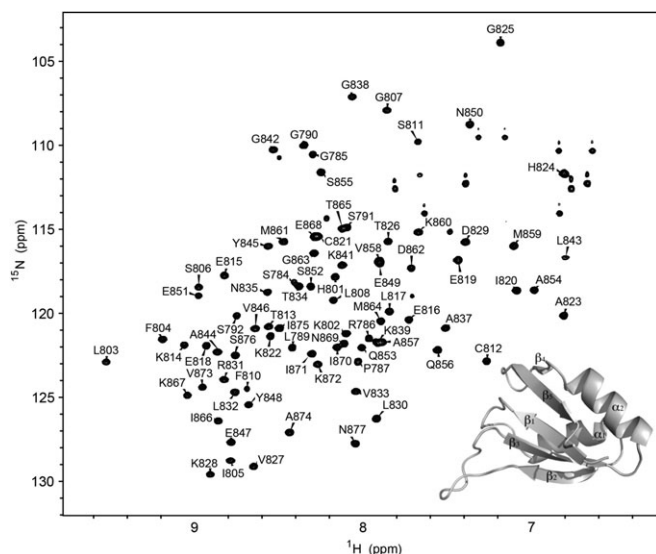
RRMs of human SART3 recognize U6 snRNA, and to compare the U6 snRNA-binding mode of SART3 with Prp24. The RRM1 and 2 of SART3 would recognize the U6 snRNA in a similar way to the RRM2 and 3 of yeast Prp24, because RRM2 and 3 are the functionally and structurally important domains of Prp24.<sup>9,11</sup> Despite this importance, no detailed structural investigation of the SART3 RRM interaction with U6 snRNA has been performed. The only available tertiary structure of SART3 is the RRM2 structure alone (Figure 2, inset, PDB ID: 2DO4), yet the deposition was incomplete in that no nuclear magnetic resonance (NMR) restraints have been reported. The SART3 RRM2 sequence was compared with the RRM domains of its orthologs, which have some sequence similarities (Figure 1(b)).

Here, as a first step to structurally understand the SART3–U6 snRNA tertiary interaction, we characterized the RRM2 domain using NMR spectroscopy, the most conserved

domain of the human SART3.<sup>3</sup> Nearly complete NMR resonances of the backbone atoms of the RRM2 domain of SART3 were assigned. Interestingly, the CSI (Chemical Shift Index) and TALOS+ (Torsion Angle Likelihood Obtained from Shift and Sequence Similarity +) analyses showed that RRM2 formed a “ $\beta_1$ - $\alpha_1$ - $\beta_2$ - $\beta_3$ - $\alpha_2$ - $\beta_4$ - $\beta_5$ ” secondary structure, where  $\beta_4$  was an extra sheet structure that was not observed in the canonical RRM domain structure. Our results will provide important structural basis for investigation of the tertiary interaction between SART3 and U6-snRNA and for further studies of SART3-mediated U4/U6 recycling pathway.

## Experimental

**Protein Purification.** The DNA encoding RRM2 domain (a.a. 800–877) of SART3 was amplified by polymerase chain reaction and expressed from the pET28a plasmid, modifying the protein coding sequences with a short N-terminal His-tag. The plasmid was transformed into *Escherichia coli* (*E. coli*) cells, BL21(DE3). To acquire  $^{15}\text{N}$  and  $^{13}\text{C}$ -labeled protein, transformed *E. coli* cells containing the plasmids, pET28a were grown in minimal (M9) media, which is enriched with  $^{13}\text{C}$ -D-glucose and  $^{15}\text{NH}_4\text{Cl}$ . For unlabeled RRM2 growth, Luria broth media was used. Protein production was induced with 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside for 16 h at 18°C. Cells were harvested by centrifugation at 4,000 rpm, 4°C and stored at –80°C until further purification. The cell pellets were gently resuspended in 25 mM HEPES pH 7.5, 150 mM NaCl, 2 mM  $\beta$ -mercaptoethanol, 0.2 mM phenylmethane sulfonyl fluoride, and lysed by sonication. The cell lysates were centrifuged at 18,000 rpm for 30 min at 4°C, and the supernatant was filtered to remove the cell debris. The protein was purified by Ni-affinity chromatography followed by the gel filtration step using a Hi-load Superdex75 26/60 column, and the purified protein samples were then analyzed by SDS-PAGE. For NMR studies, the buffer was exchanged into 20 mM sodium phosphate pH 6.5, 100 mM NaCl, 0.5 mM dithiothreitol, and concentrated using a centrifugal filter device (Amicon, Merck Millipore, Germany).



**Figure 2.** 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum and assignments of the amide resonances of RRM2. The construct contains 78 residues from a.a. 800–877. The observable  $^1\text{H}$ - $^{15}\text{N}$  pairs are labeled by their residue numbers. The solution NMR structure of RRM2 domain (PDB ID: 2DO4) is shown within the spectrum.

**NMR Spectroscopy.** The NMR data were collected on the Bruker 800MHz NMR spectrometer with a cryogenic z-gradient triple resonance probe. All NMR experiments were uniformly executed at 25°C. The obtained NMR data were processed using Topspin software (Bruker, Germany), and analyzed using Sparky (University of California, San Francisco, CA, USA). The 3D CBCA(CO)NH, HNCA, HNCACB, HN(CA)CO, HNCO, HN(CO)CA, HBHA(CO)NH and 2D  $^1\text{H}$ - $^{15}\text{N}$ -TROSY HSQC experiments were used to complete sequential backbone assignments of RRM2.

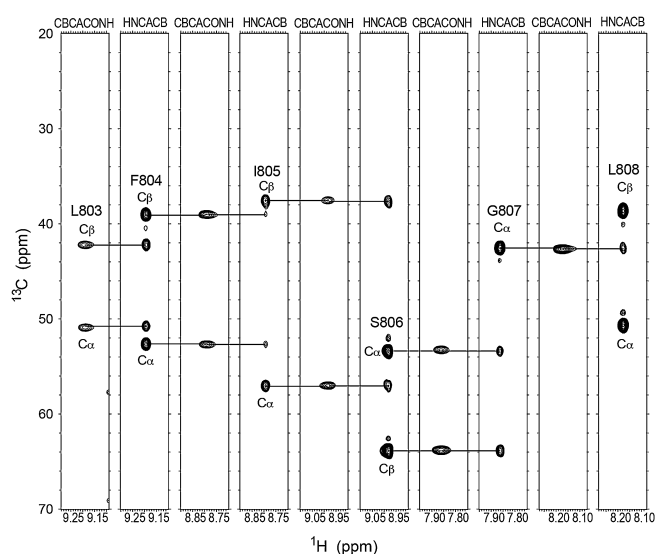
**Secondary Structure Analysis.** The secondary structure was determined from the H and C chemical shift values, using the CSI 3.0 program.<sup>12</sup> The RRM2 sequence and chemical shift values of  $\text{H}_\alpha$ , HN, N,  $\text{C}_\alpha$ ,  $\text{C}_\beta$ , and  $\text{C}'$  were used for the secondary structure prediction. The CSI values are written as “1”, “0”, or “-1” based on the chemical shift value of a residue. For the up-field shifted resonances compared to the random coil residues, the CSI values are “1”s. For the down-field shifted residues, they are “-1”s. For the residues with chemical shifts within the reference value range, they are marked as “0”s. For  $\text{H}_\alpha$  and  $\text{C}_\beta$ , a group with repeating four or more “-1” without interruption of a “1” means an  $\alpha$ -helix. Also a group with three or more “1” without interruption of a “-1” means a  $\beta$ -sheet. The chemical shift of  $\text{C}_\beta$  can only be used to confirm  $\beta$ -sheet. In contrast, for  $\text{C}_\alpha$  and  $\text{C}'$ , a group with repeating four or more “1” without interruption of “-1” means an  $\alpha$ -helix. Also a group with repeating three or more “-1” without interruption of a “1” means a  $\beta$ -sheet. All other regions are indicated as a random coil.<sup>13</sup> To estimate backbone dihedral angles, TALOS+<sup>14</sup> method was used, based on the chemical shifts.

## Results and Discussion

A good dispersion of the peaks in the 2D  $^1\text{H}$ - $^{15}\text{N}$ -HSQC TROSY spectrum implies that RRM2 has a well-folded tertiary structure (Figure 2). The backbone assignment of uniformly  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled RRM2 protein was obtained from the analyses of a suite of 2D and 3D NMR spectra shown in Experimental section. Briefly, the correlation between the chemical shift of an amide residue and  $\text{C}_\alpha$  and  $\text{C}_\beta$  of its preceding residue was determined by comparing the CBCA(CO)NH and HNCACB spectra (Figure 3).

The HNCO and HN(CA)CO spectra provided the chemical shift correlation between amide and carbonyl carbon ( $\text{C}'$ ), and the HNCA and HN(CO)CA spectra provided the correlation between amide and  $\text{C}_\alpha$ . Also, the chemical shifts of  $\text{H}_\alpha$  and  $\text{H}_\beta$  were determined by using HBHA(CO)NH spectra as well as using the amide—carbon correlation spectra.

Thus, we completed the backbone  $\text{H}_\alpha$ ,  $\text{H}_\beta$ , HN, N,  $\text{C}_\alpha$ ,  $\text{C}_\beta$ , and  $\text{C}'$  assignments of the residues 800–877 of RRM2. The chemical shifts of  $^{15}\text{N}$  of two prolines P809 and P840 could not be determined by this sequential assignment, and

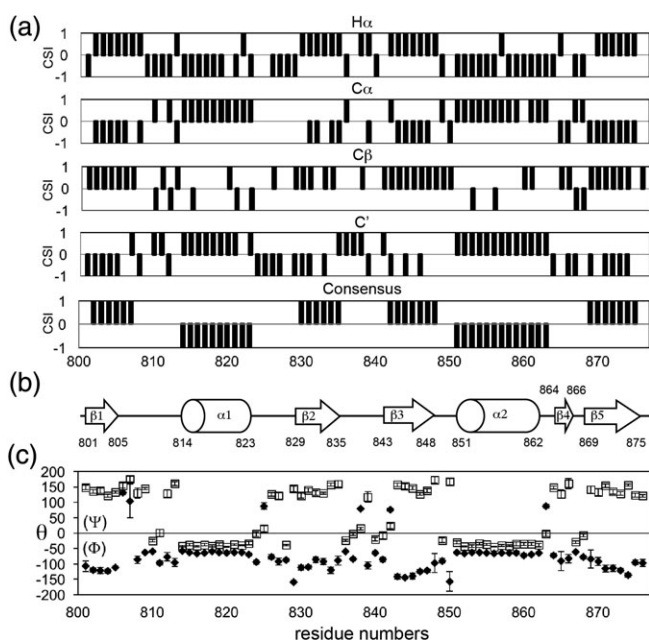


**Figure 3.** Strip plots from the CBCA(CO)NH and HNCACB spectra used for sequential assignment of SART3 RRM2. Sequential  $\text{C}_\alpha$  and  $\text{C}_\beta$  connectivity for the residues L803–L808 is shown with solid lines.

the HN and  $^{15}\text{N}$  of K800 and R836 were not determined due to the significant peak-overlaps. One hundred percent of  $\text{C}_\alpha$ , 93% of  $\text{C}_\beta$ , and 100% of  $\text{C}'$  resonances, including two prolines were assigned, respectively.

For the chemical shifts of alpha and beta protons, 93% of  $\text{H}_\alpha$  and 85% of  $\text{H}_\beta$  were assigned. All available  $^{15}\text{N}$ ,  $^{13}\text{C}$ , and  $^1\text{H}$  chemical shifts of backbone atoms of the RRM2 are listed in Supporting Information (Table S1), and have been deposited in BioMagResBank (<http://www.bmrb.wisc.edu>) database under access code of 26948.

Next, the assigned backbone chemical shifts were used to predict the secondary structure of RRM2.<sup>12</sup> The CSI results of RRM2 showed a patterned distribution of the marks -1, 0, and 1 (Figure 4(a)), which indicated that RRM2 consists of two alpha helices and four beta sheets. In the consensus CSI, the helical residues were assigned as K814–A823 (helix 1) and E851–G863 (helix 2) and the sheet residues as K802–G807 (sheet 1), L830–N835 (sheet 2), G842–Y848 (sheet 3), and N869–I875 (sheet 4), respectively. The boundaries of the helices and sheets of RRM2 in the CSI structure were almost the same as seen in the solution structure with only one- or two-residue differences, but the solution structure included an additional short beta sheet after helix 2. Therefore, to further characterize the secondary structure of RRM2, the backbone torsion angles of  $\Phi$  and  $\Psi$  were calculated using TALOS+ (Figure 4 (b) and (c)). The TALOS+ predicted secondary structure was almost consistent with the solution structure including the additional  $\beta_4$  that was not observed in the CSI results. The average torsional angle values ( $\Phi$  and  $\Psi$ ) of the helices and sheets were  $-116.4 \pm 7.0$ ,  $-135.4 \pm 9.4$  ( $\beta_1$ ),  $-63.3 \pm 3.1$ ,  $-39.4 \pm 3.5$  ( $\alpha_1$ ),  $-109.7 \pm 25.3$ ,  $-140.7 \pm 14.4$  ( $\beta_2$ ),  $-127.8 \pm 17.7$ ,  $-149.0 \pm 15.8$  ( $\beta_3$ ),  $-65.2 \pm 2.9$ ,  $-38.0 \pm 4.4$  ( $\alpha_2$ ),  $-81.7 \pm 9.1$ ,  $-145.3 \pm 17.2$  ( $\beta_4$ ), and  $-106.8 \pm 17.5$ ,



**Figure 4.** Secondary structural information (a) CSI values of each residue for H $\alpha$ , C $\alpha$ , C $\beta$ , and C' are shown along the residue numbers. The consensus CSI diagram shows the overall secondary structure of RRM2, where the negative and positive values indicate  $\alpha$ -helical and  $\beta$ -sheeted regions, respectively. (b) The schematic representation of the secondary structure of the RRM2 predicted by TALOS+. The arrows and cylinders represent the sheets and the helices, and the residue numbers are shown. (c)  $\Phi$  and  $\Psi$  angle values calculated by TALOS+. The filled diamonds represent the backbone  $\Phi$  angles and the open rectangles the backbone  $\Psi$  angles. The standard deviation of each angle is represented by a vertical line.

$-136.5 \pm 13.2$  ( $\beta_5$ ), respectively, which were well consistent with the typical sheet- and helical propensity.

### Conclusion

Overall, the results of CSI and TALOS+ suggest that the SART3 RRM2 domain adopts the typical RRM domain structure with an extra beta sheet after the second alpha helix. However, since  $\beta_4$  is composed of only two residues M864 and T865, this extra sheet may not exist after further refinement of the NMR structure. Our NMR chemical shift

values and the calculated secondary structure of the SART3 RRM2 domain will be used to predict the structure of the full RRM1–RRM2 domain and to obtain insights into the characteristic binding properties of RRM2 to U6 snRNA.

**Acknowledgments.** This study was supported financially by the National Research Foundation of Korea (NRF) (2015R1A2A2A04005596 to N.-K.K.), and Korea Institute of Science and Technology (2V04611 to N.-K.K.). We thank to Korea Basic Science Institute (KBSI) for using their NMR spectroscopy.

**Supporting Information.** Additional supporting information is available in the online version of this article.

### References

1. C. L. Will, R. Luhrmann, *Cold Spring Harb. Perspect. Biol.* **2011**, 3, 1–23. doi:10.1101/cshperspect.a003707.
2. A. G. Matera, Z. Wang, *Nat. Rev. Mol. Cell Biol.* **2014**, 15, 108.
3. M. Bell, S. Schreiner, A. Damianov, R. Reddy, A. Bindereif, *EMBO J.* **2002**, 21, 2724.
4. W. Zhao, Y. Liu, K. A. Timani, J. J. He, *J. Biol. Chem.* **2014**, 289, 190.
5. A. Whitmill, K. A. Timani, Y. Liu, J. J. He, *Life Sci.* **2016**, 149, 79.
6. J. Medenbach, S. Schreiner, S. Liu, R. Luhrmann, A. Bindereif, *Mol. Cell Biol.* **2004**, 24, 7392.
7. J. K. Park, T. Das, E. J. Song, E. E. Kim, *Nucleic Acids Res.* **2016**, 44, 5424.
8. K. Licht, J. Medenbach, R. Luhrmann, C. Kambach, A. Bindereif, *RNA* **2008**, 14, 1532.
9. E. J. Montemayor, E. C. Curran, H. H. Liao, K. L. Andrews, C. N. Treba, S. E. Butcher, D. A. Brow, *Nat. Struct. Mol. Biol.* **2014**, 21, 544.
10. G. M. Daubner, A. Clery, F. H. Allain, *Curr. Opin. Struct. Biol.* **2013**, 23, 100.
11. K. W. Shannon, C. Guthrie, *Genes Dev.* **1991**, 5, 773.
12. N. E. Hafsa, D. Arndt, D. S. Wishart, *Nucleic Acids Res.* **2015**, 43, W370.
13. D. S. Wishart, B. D. Sykes, *J. Biomol. NMR* **1994**, 4, 171.
14. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, 44, 213.