Solution Structure of the Second RNA Recognition Motif (RRM) Domain of Murine T Cell Intracellular Antigen-1 (TIA-1) and Its RNA Recognition Mode[†,‡]

Kanako Kuwasako,[§,||] Mari Takahashi,[§,||] Naoya Tochio,[§] Chikage Abe,[§,⊥] Kengo Tsuda,[§] Makoto Inoue,[§] Takaho Terada,[§] Mikako Shirouzu,[§] Naohiro Kobayashi,[§] Takanori Kigawa,[§,#] Seiichi Taguchi,[◇] Akiko Tanaka,[§] Yoshihide Hayashizaki,[○] Peter Güntert,[§,Δ,▽] Yutaka Muto,[*,§] and Shigeyuki Yokoyama[*,§,+]

Protein Research Group, Genomic Sciences Center, Yokohama Institute, RIKEN, Tsurumi-ku, Yokohama 230-0045, Japan, Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama 226-8502, Japan, Division of Biotechnology and Macromolecular Chemistry, Graduate School of Engineering, Hokkaido University, Sapporo, Hokkaido 060-8628, Japan, Tatsuo Miyazawa Memorial Program, Yokohama Institute, RIKEN, Tsurumi-ku, Yokohama 230-0045, Japan, and Department of Biophysics and Biochemistry, Graduate School of Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

Received December 19, 2007; Revised Manuscript Received March 31, 2008

ABSTRACT: T cell intracellular antigen-1 (TIA-1), an apoptosis promoting factor, functions as a splicing regulator for the Fas pre-mRNA. TIA-1 possesses three RNA recognition motifs (RRMs) and a glutamine-rich domain. The second RRM (RRM2) is necessary and sufficient for tight, sequence-specific binding to the uridine-rich sequences buried around the 5′ splice sites. In the present study, we solved the solution structure of the murine TIA-1 RRM2 by heteronuclear−nuclear magnetic resonance spectroscopy. The TIA-1 RRM2 adopts the RRM fold ($\beta\alpha\beta\beta\alpha\beta$) and possesses an extra β-strand between β2 and β3, which forms an additional β-sheet with the C-terminal part of β2. We refer to this structure as the β2-β2′ β-loop. Interestingly, this characteristic β-loop structure is conserved among a number of RRMs, including the U2AF65 RRM2 and the Sex-lethal RRM1 and RRM2, which also bind to uridine-rich RNAs. Furthermore, we identified a new sequence motif in the β2-β2′ β-loop, the DxxT motif. Chemical shift perturbation analyses of both the main and side chains upon binding to the uridine pentamer RNA revealed that most of the β-sheet surface, including the β2-β2′ β-loop, is involved in the RNA binding. An investigation of the chemical shift perturbation revealed similarity in the RNA recognition modes between the TIA-1 and U2AF65 RRMs.

Pre-mRNA splicing is essential for gene expression. Regulation of splicing plays key roles in posttranscriptional gene expression control (*1*–*7*). Splicing occurs through two successive transesterification reactions: the cleavage of the 5′ splice site by a nucleophilic attack from the conserved branch site adenosine and the release of the intron through a nucleophilic attack to the 3′ splice site by the phosphate group of the 5′ splice site (*8*). These processes are accomplished by the spliceosome, a dynamic macromolecular complex of the U small nuclear ribonucleoproteins (U1, U2, U4, U5, and U6 snRNPs),[1] each of which is composed of proteins and snRNA (*3*). The U1 and U2 snRNAs base pair with the 5′ splice site and the branch point, respectively (*9*). In addition, various RNA-binding proteins play important roles in splicing regulation by binding to the 5′ and/or 3′ splice sites, thus preventing or promoting the recruitment of snRNPs to the relevant sites (*10*).

T cell intracellular antigen-1 (TIA-1) is an apoptosis promoting factor that functions as a splicing regulator for Fas pre-mRNA in the nucleus (*11*). TIA-1 binds to the uridine-rich sequences located downstream of the 5′ splice

[†] This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI) of the National Project on Protein Structural and Functional Analyses of the Ministry of Education, Culture, Sports, Science and Technology of Japan (S.Y.) and by grants from the Human Frontier Science Program (Y.M.).

[‡] The atomic coordinates for the ensemble of 20 energy-refined NMR conformers that represent the solution structure of TIA-1 RRM2 have been deposited in the Protein Data Bank, with the accession code 2RNE.

* Corresponding authors. S.Y.: telephone, +81 45 503 9196; fax, +81 45 503 9195; e-mail, yokoyama@biochem.s.u-tokyo.ac.jp. Y.M.: telephone, +81 45 503 9461; fax, +81 45 503 9460; e-mail, ymuto@gsc.riken.jp.

[§] Protein Research Group, Yokohama Institute, RIKEN.

[||] These authors contributed equally to this work.

[⊥] Present address: Analytical Research Core for Advanced Materials, Institute for Materials Research, Tohoku University, Sendai, Miyagi 980-8577, Japan.

[#] Tokyo Institute of Technology.

[◇] Hokkaido University.

[○] Genome Exploration Research Group, Yokohama Institute, RIKEN.

[Δ] Tatsuo Miyazawa Memorial Program, RIKEN Yokohama Institute.

[▽] Present address: Institute of Biophysical Chemistry, J. W. Goethe-University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany.

[+] The University of Tokyo.

[1] Abbreviations: COSY, correlated spectroscopy; *d*-DTT, 1,4-DL-dithiothreitol-d_{10}; HSQC, heteronuclear single-quantum coherence spectroscopy; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; rmsd, root-mean-square deviation; RNP1 motif, ribonucleoprotein 1 motif; RNP2 motif, ribonucleoprotein 2 motif; RRM, RNA recognition motif; snRNPs, small nuclear ribonucleoproteins; TALOS, torsion angle likelihood obtained from shift and sequence similarity; TIA-1, T cell intracellular antigen-1; TOCSY, total correlation spectroscopy; U1 snRNP, U1 small nuclear ribonucleoprotein.

10.1021/bi7024723 CCC: $40.75
Published on Web 05/24/2008

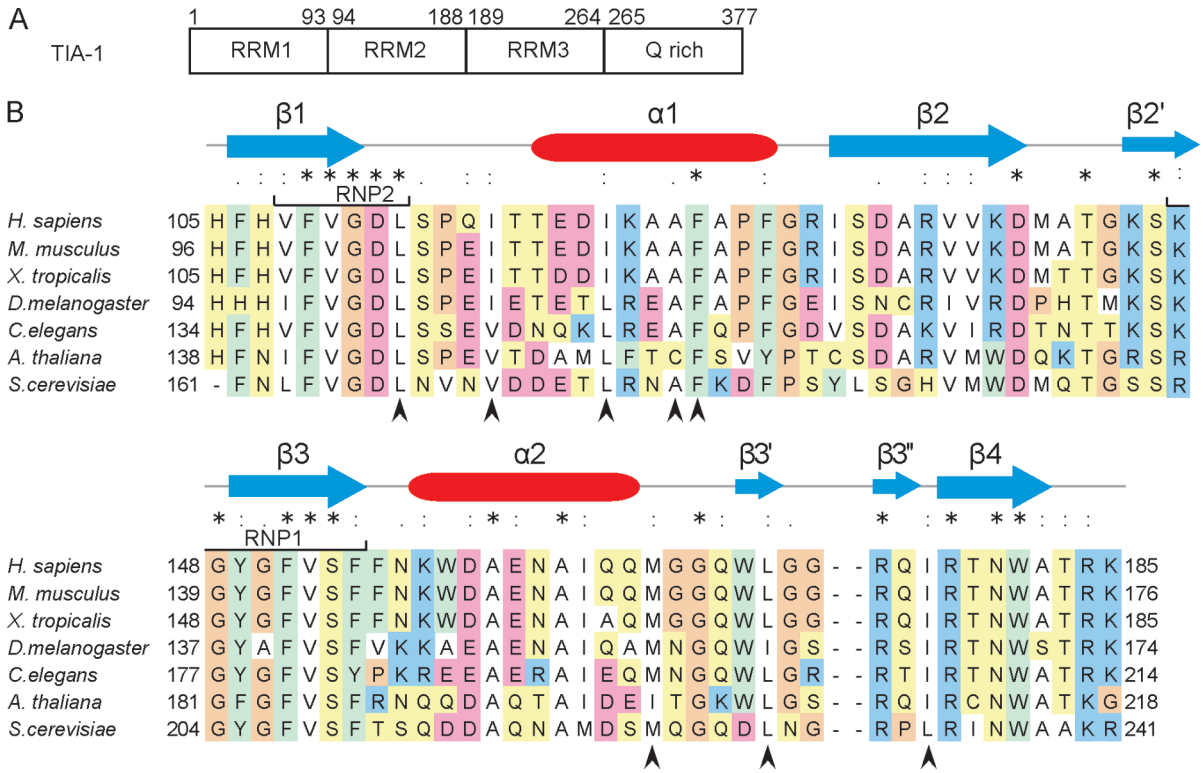

FIGURE 1: Primary structure of the TIA-1 RRM2. (A) Schematic diagram of the mouse TIA-1 protein. TIA-1 possesses three RNA recognition motifs (RRMs) and a glutamine-rich C-terminal domain (*13*). (B) Multiple sequence alignment of the TIA-1 RRM2. The TIA-1 sequences from *Homo sapiens* (accession code NP_071505), *Mus musculus* (AAH46812), *Xenopus tropicalis* (NP_989276), *Drosophila melanogaster* (NP_732942), *Caenorhabditis elegans* (NP_495121), *Arabidopsis thaliana* (ABH04584), and *Saccharomyces cerevisiae* (NP_014382) were aligned using ClustalX (*44*). Secondary structure elements are depicted with blue arrows (β-sheet) and red cylinders (α-helix) above the sequence alignment. The signature sequences of RNP1 and RNP2 are indicated by the black lines. The arrowheads indicate the residues forming the extensive network of hydrophobic interactions that rigidly hold the $\beta3'$-$\beta3''$ β-hairpin.

sites (e.g., intron 5 of Fas pre-mRNA) and recruits U1 snRNP by interacting with the U1C protein (*11*). During the splicing process of Fas pre-mRNA, TIA-1 plays an important role in including exon 6 and promoting the production of the full-length mRNA and, thereby, the production of the functional protein (*11*). Without exon 6, Fas mRNA encodes a soluble form of the receptor, which functions as an antagonist of Fas apoptosis signaling (*12*). TIA-1-mediated alternative splicing is therefore believed to be a critical regulator of cell death signals (*11*).

TIA-1 possesses three RNA recognition motifs (RRMs) and a glutamine-rich region (*13*) (Figure 1A). Previous studies revealed that the glutamine-rich region binds to U1C, with support from the first RRM (*14*). However, no RNA-binding activity has been reported for the first RRM, and the third RRM only weakly binds to RNA, with no sequence specificity (*13*). In contrast, the second RRM (RRM2) is required for the RNA-binding activity of TIA-1 to the uridine-rich sequence: when all of the uridine residues in the uridine-rich sequence are replaced with cytidine, TIA-1 no longer binds to the RNA sequence (*13*). Therefore, the RRM2 alone is necessary and sufficient for the strong, sequence-dependent binding to the RNA sequence (*13*). Consistent with its functional importance, RRM2 is highly conserved in a wide variety of organisms (Figure 1B). For better understanding of TIA-1-mediated apoptosis, it is important to elucidate how the TIA-1 RRM2 recognizes the uridine-rich sequences of some 5′ splice sites (such as intron 5 of the Fas pre-mRNA) to facilitate the recruitment of U1 snRNP.

In the present study, we solved the solution structure of murine TIA-1 RRM2, which revealed the characteristic loop structure between the $\beta2$ and $\beta3$ strands. The $\beta2$-$\beta3$ loop structure was quite similar to those of the U2AF65 RRM2 and the Sex-lethal RRM1 and RRM2, which bind to polypyrimidine tracts at the 3′ splice site but lack strong amino acid sequence similarity with the TIA-1 RRM2, except for the $\beta2$-$\beta3$ loop region. Therefore, we refer to this characteristic structure as the $\beta2$-$\beta2'$ β-loop. Chemical shift perturbation analyses using a uridine pentamer RNA revealed that this β-loop structure is involved in the RNA-binding activity of the TIA-1 RRM2, together with the β-sheet surface. The similarity between the RNA-binding modes of the TIA-1 RRM2 and the U2AF65 RRM2 provides a basis for understanding the RNA recognition manner of TIA-1.

MATERIALS AND METHODS

Protein Expression and Purification. For the protein production, we utilized *in vitro* cell-free and *in vivo* protein production systems. In the *in vitro* cell-free system, the DNA fragment encoding the second RRM domain (Gln89−Gln190) of mouse TIA-1 (SwissProt accession no. P52912) was amplified by PCR from the murine full-length cDNA clone. This DNA fragment was cloned into the expression vector pCR2.1 (Invitrogen), as a fusion with an N-terminal native His affinity tag and a TEV protease cleavage site. The ^{13}C/^{15}N-labeled fusion protein was synthesized by the cell-free protein expression system (*15, 16*). The lysate was clarified by centrifugation at 16000*g* for 20 min and filtration with a

0.45 mm membrane (Millipore). The clarified lysate was applied to a 5 mL His-trap column (GE Healthcare Biosciences), which was eluted with an imidazole gradient from 20 to 500 mM, and the tag was removed by an incubation with TEV protease for 1 h at 30 °C. The cleaved His tag was segregated from the cleaved protein by rechromatograpy using a His-trap nickel column.

For the *in vivo* protein production system, the amplified DNA fragment was cloned into the expression vector pGEX6P-1 (GE Healthcare), as a fusion with an N-terminal glutathione *S*-transferase (GST) tag and a TEV protease cleavage site. The $^{13}C/^{15}N$-labeled fusion protein was overexpressed in *Escherichia coli* strain BL21(DE3) cells using modified minimum medium supplemented with 50 mg/L ampicillin (*17*). The harvested culture was lysed by sonication in phosphate buffer, containing 1 mM DTT, 1 mM PMSF, and protease inhibitor cocktail for general use (Nacalai Tesque). The lysate was applied to a glutathione-Sepharose 4 Fast Flow (GE Healthcare) column and was eluted by 10 mM reduced glutathione. The tag was removed by an incubation with TEV protease overnight at room temperature. The tag-free TIA-1 RRM2 was further purified by RESOURCE S column chromatography (GE Healthcare). The NMR samples were concentrated to approximately 1.1 mM in 20 mM *d*-Tris-HCl buffer (pH 7.0), containing 100 mM NaCl, 1 mM 1,4-DL-dithiothreitol-d_{10} (*d*-DTT), and 0.02% NaN_3 (in 90% ^1H_2O/10% ^2H_2O), using an Amicon Ultra-15 concentrator (5000 MWCO; Millipore).

NMR Spectroscopy. NMR experiments were performed at 25 °C on 800 MHz spectrometers (Bruker AV800). The ^1H, ^{15}N, and ^{13}C chemical shifts were referenced relative to the frequency of the ^2H lock resonance of water. Backbone and side-chain assignments were obtained by using a combination of standard triple resonance experiments (*18*). 2D [^1H,^{15}N]-HSQC, and 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH spectra were used for the ^1H, ^{15}N, and ^{13}C assignments of the protein backbone. Side-chain ^1H and ^{13}C assignments of the nonaromatic side chains, including all Pro residues, were obtained using 2D [^1H,^{13}C]-HSQC, and 3D HBHA(CO)NH, H(CCCO)NH, (H)CC(CO)-NH, HCCH-COSY, HCCH-TOCSY, and (H)CCH-TOCSY spectra. Assignments were checked for consistency with 3D ^{15}N-edited [^1H,^1H]-NOESY and ^{13}C-edited [^1H,^1H]-NOESY spectra. The ^1H and ^{13}C spin systems of the aromatic rings of the Phe, Trp, His, and Tyr residues were identified using 3D HCCH-COSY and HCCH-TOCSY spectra, and the 3D ^{13}C-edited [^1H,^1H]-NOESY spectrum was used for the sequence-specific resonance assignment of the aromatic side chains. NOESY spectra were recorded with mixing times of 80 ms. The 2D and 3D spectra were processed using NMRPipe (*19*). Analyses of the processed data were performed with the programs NMRView (*20*) and KUJIRA (*21*).

For the amide chemical shift titration experiments, the U/C-pentamer RNA (5′-[UUUUU]-3′ and 5′-[CCCCC]-3′) (Dharmacon) was dissolved in 20 mM *d*-Tris-HCl buffer (pH 7.0) containing 100 mM NaCl and 1 mM *d*-DTT to make a 6 mM solution. 2D [^1H,^{15}N]- and [^1H,^{13}C]-HSQC spectra were recorded while increasing the concentration of the RNA relative to that of the TIA-1 RRM2 solution [0.1 mM (Figure 8) and 0.53 mM (Figure 7)] to a final 1:2.0 ratio of TIA-1 RRM2/RNA.

The measurements of the nitrogen relaxation times, T_1 and T_2, and the proton−nitrogen heteronuclear NOEs were performed on a 600 MHz spectrometer with a cryoprobe (Bruker AV 600) at 25 °C, using the ^{15}N,^{13}C-labeled TIA-1 RRM2 at 0.53 mM (*22*). Eight different relaxation delays were recorded for the ^{15}N T_1 (T_1 delays = 5, 65, 145, 246, 366, 527, 757, and 1148 ms) and ^{15}N T_2 (T_2 delays = 32, 48, 64, 80, 96, 112, 128, and 144 ms) relaxation experiments. The ^{15}N T_1 and ^{15}N T_2 values were extracted using a curve-fitting subroutine included in the Sparky program (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). The proton−nitrogen heteronuclear NOE values were calculated as the ratio between the cross-peak intensities with (I) and without (I_0) ^1H saturation (I/I_0). The errors were estimated from the root mean square of the baseline noise in the two spectra (*22*). Using the TENSOR2 software (*23*), the correlation time was analyzed using the T_1/T_2/NOE values of residues Gln89−Gln190, except for Phe97, Val101, Phe146, Ser182, and proline residues.

Structure Calculations. The three-dimensional structures of the complex were determined by combined automated NOESY cross-peak assignment (*18*) and structure calculations with torsion angle dynamics (*24*) implemented in the program CYANA, version 2.0.17 (*25*). Dihedral angle constraints for φ and ψ were obtained from the main-chain and $^{13}C^{\beta}$ chemical shift values using the program TALOS (*26*) and by analyzing the NOESY spectra. Structure calculations started from 100 randomized conformers and used the standard CYANA simulated annealing schedule (*24*) with 20000 torsion angle dynamic steps per conformer. The 20 conformers with the lowest final CYANA target function values were subjected to restrained energy minimization in a water shell with the program OPALp (*27*), using the AMBER force field (*28*). PROCHECK-NMR (*29*) and MOLMOL (*30*) were used to validate and to visualize the final structures, respectively. Using MOLMOL (*30*), the hydrogen bonds were analyzed on the basis of the following criteria: the distance between the donor and acceptor atoms is smaller than 2.4 Å, and the supplementary angle between the line formed by the donor and hydrogen atom and the line formed by the acceptor and hydrogen atom is smaller than 35°. If the hydrogen bonds were found in more than 4 of the 20 structures (20%) by MOLMOL, we approved of the presence of the hydrogen bonds, according to Auweter et al. (*31*). The electrostatics were calculated by MOLMOL, using the Poisson−Boltzmann equation (*32*).

Structure Analyses. The 106 RRMs from the PDB were used for the structural comparison. The lengths and amino acid compositions of the β1-α1, β2-β3, and α2-β4 loops were examined. A total of 364 human RRMs were identified in the SMART sequence database (*33*), based on the definition of RRMs provided by the Pfam database (*34*). Their amino acid sequences were aligned using the ClustalX software (*35*). Similarity in the amino acid composition was assessed on the basis of the categorization of the amino acid residues as aliphatic (Ala, Val, Leu, Ile, and Met), aromatic (Phe, Tyr, and Trp), basic (Lys and Arg), acidic (Asp and Glu), neutral (Ser, Thr, Cys, Asn, Gln, and His), or small (Gly and Pro).

For the examination of the structural similarity of the β2-β3 loop, we selected 25 RRMs with β2-β3 loops of the same

Table 1: Summary of Conformational Restraints and Structural Statistics for the 20 Energy-Refined Conformers of the Murine TIA-1 RRM2 Structure

NOE upper distance restraints	
intraresidual ($\|i - j\| = 0$)	417
sequential ($\|i - j\| = 1$)	485
medium range ($1 < \|i - j\| < 5$)	323
long range ($\|i - j\| \geq 5$)	795
total NOE	2020
dihedral angle restraints (including 89 φ/ψ dihedral angle restraints from TALOS)[a]	111
CYANA target function value (Å^2)	0.065 ± 0.017
distance restraint violations >0.2 Å	0
dihedral angle restraint violations >5°	0
AMBER energies[b] (kcal/mol)	
total	-4181 ± 93.5
van der Waals	-262 ± 11
electrostatic	-4688 ± 94
Ramachandran plot statistics[c,d] (%)	
residues in most favored regions	88.0
residues in additionally allowed regions	12.0
residues in generously allowed regions	0.0
residues in disallowed regions	0.0
average rmsd from mean coordinates[d] (Å)	
backbone	0.29
heavy atoms	0.64

[a] See ref *26*. [b] See ref *28*. [c] See ref *29*. [d] For the structured regions comprising residues 97−173 of TIA-1 RRM2.

length as that of the TIA-1 RRM2 and calculated the rmsd values for the residues corresponding to residues 139−143 of the TIA-1 RRM2. Simultaneously, the differences in the φ and ψ dihedral angles for the corresponding residues were evaluated among these RRMs. Namely, the φ and ψ dihedral angles for each residue were classified into three categories, based on the Ramachandran diagram: α-helix, β-strand, and others. The "dihedral angle difference score" of each RRM was then defined as the number of amino acid residues that were classified into different categories from those of the corresponding residues in the TIA-1 RRM2.

Isothermal Titration Calorimetry (ITC) Measurements. ITC measurements were performed at 25 °C by using a Microcal (Amherst, MA) VP-ITC calorimeter. Samples were buffered with 20 mM Tris-HCl (pH 7.0), 100 mM NaCl, and 1 mM DTT and were thoroughly degassed before use. At first, a 2.0 mL aliquot of 10 μM TIA-1 RRM2 was placed in the cell chamber. A 20-fold concentration of the four kinds of pentamer RNAs (5′-[UUUUU]-3′, 5′-[CCCCC]-3′, 5′-[AAAAA]-3′, and 5′-[GGGGG]-3′) was then injected into it. Hereafter, these RNAs are referred as to U5, C5, A5, and G5 RNAs, respectively. The heat generated due to dilution of the titrants was very small and was ignored for the analysis. The data were analyzed with Microcal ORIGIN software, using a binding model that assumes a single site of interaction.

RESULTS AND DISCUSSION

Solution Structure of the Second RRM Domain of Murine TIA-1. Using standard multidimensional heteronuclear NMR spectroscopy, we assigned 99.4% of the main-chain resonances and 88.3% of the side-chain resonances of residues 96−176 of the murine TIA-1 RRM2 (see Materials and Methods). We determined the solution structure of the TIA-1 RRM2 on the basis of 2020 ^{1}H−^{1}H distance constraints from nuclear Overhauser effect spectroscopy (NOESY) and 111 torsion angle restraints (Table 1). Out of 100 independently calculated structures, 20 with the lowest CYANA target function values were subjected to restrained energy minimization (*27*). The resulting RRM structures (residues 96−176) were well determined and highly converged (0.52 Å rmsd) (Figure 2A and Table 1), while the N- and C-terminal extensions were disordered. The overall structure of the TIA-1 RRM2 adopts a canonical RRM fold (βαββαβ) (Figure 2B). Residues 97−102 (β1), 123−131 (β2), 140−145 (β3), and 169−173 (β4) constitute the four-stranded antiparallel β-sheet (Figure 2B, left panel). Helix 1 (α1, 110−120) and helix 2 (α2, 148−157) are formed between the β1- and β2-strands and between the β3- and β4-strands, respectively, and underpin the β-sheet (Figure 2B, right panel).

The β3′-β3″ β-Hairpin. In addition to the β1, β2, β3, and β4 strands, a short β-hairpin was identified between α2 and β4 (β3′, Trp162-Leu163; β3″, Arg166-Gln167) (Figures 2B and 3A). The β3′-β3″ β-hairpin intimately contacts the β1-α1 loop and the α1 helix and forms hydrophobic interactions involving Leu104 and Ile108 from the β1-α1 loop, Ile113, Ala116, and Phe117 from the α1 helix, Met158 from the α2-β3′ loop, Leu163 from the β3′-β3″ β-hairpin, and Ile168 from the β3″-β4 loop (Figure 2E). Thus, the β3′-β3″ β-hairpin is highly ordered in the 20 conformers that represent the solution structure (Figure 2A). The eight hydrophobic residues are conserved in a wide variety of organisms (Figure 1B) and form a hydrophobic patch on the α-helical surface (Figure 2D, right panel).

Characteristic Conformation of the β2-β2′ β-Loop. The β2 strand is longer than those of other RRMs, as it is extended on the C-terminal side by two residues (Lys130 and Asp131) with a slight kink at Val129 (Figures 2B and 3B). The Val129-Lys130-Asp131 strand hydrogen bonds with an extra β-strand formed between β2 and β3 (β2′, Lys136-Ser137-Lys138) (Figures 1B, 2B, and 3B). Between the β2 and β2′ strands, residues 132−135 form a type I β-turn (Figure 2F). The O^{δ} atom of Asp131 is located close to the H^{γ} atom of Thr134 and the H^{N} atoms of Ala133 and Thr134. The O^{γ} atom of Thr134 is located close to the H^{N} atom of Lys136. The Asp and Thr residues may contribute to the characteristic loop conformation. On the other hand, a single bulged-out residue, Gly139, intervenes between the β2′ and β3 strands (Figure 3B). In the 20 conformers (Figure 2A), the Gly139 residue exhibits φ and ψ values of $+122 \pm 29°$ and $-22 \pm 9°$, respectively, which are different from those of typical β-strands. This particular conformation, characterized by the β2:β2′ pairing and the β-turn with the putative hydrogen bonds (Figure 2F), is shared by all 20 of the conformers (Figure 2A). Hereafter, we refer to the β2-β3 loop (Lys130-Gly139) as the β2-β2′ β-loop. The β2-β2′ β-loop protrudes from the β-sheet surface and has a negative charge contributed by Asp131, located at the C-terminus of the β2 strand (Figure 2C, left panel).

Positively Charged Surface. As indicated in Figure 2C, the solvent-accessible surface of the β-sheet is highly positively charged, with five basic residues, Arg127 (the β2 strand), Lys138 (the β2′-β3 loop), Arg169 (the β4 strand), and Arg175 and Lys176 (the C-terminal extension). In addition, the top and bottom sides of the β-sheet surface are also positively charged, by His96 and His98 (the N-terminal region of the β1 strand) as well as Lys136 (the β2′ strand) and Arg166 (the β3″ strand). These positively charged

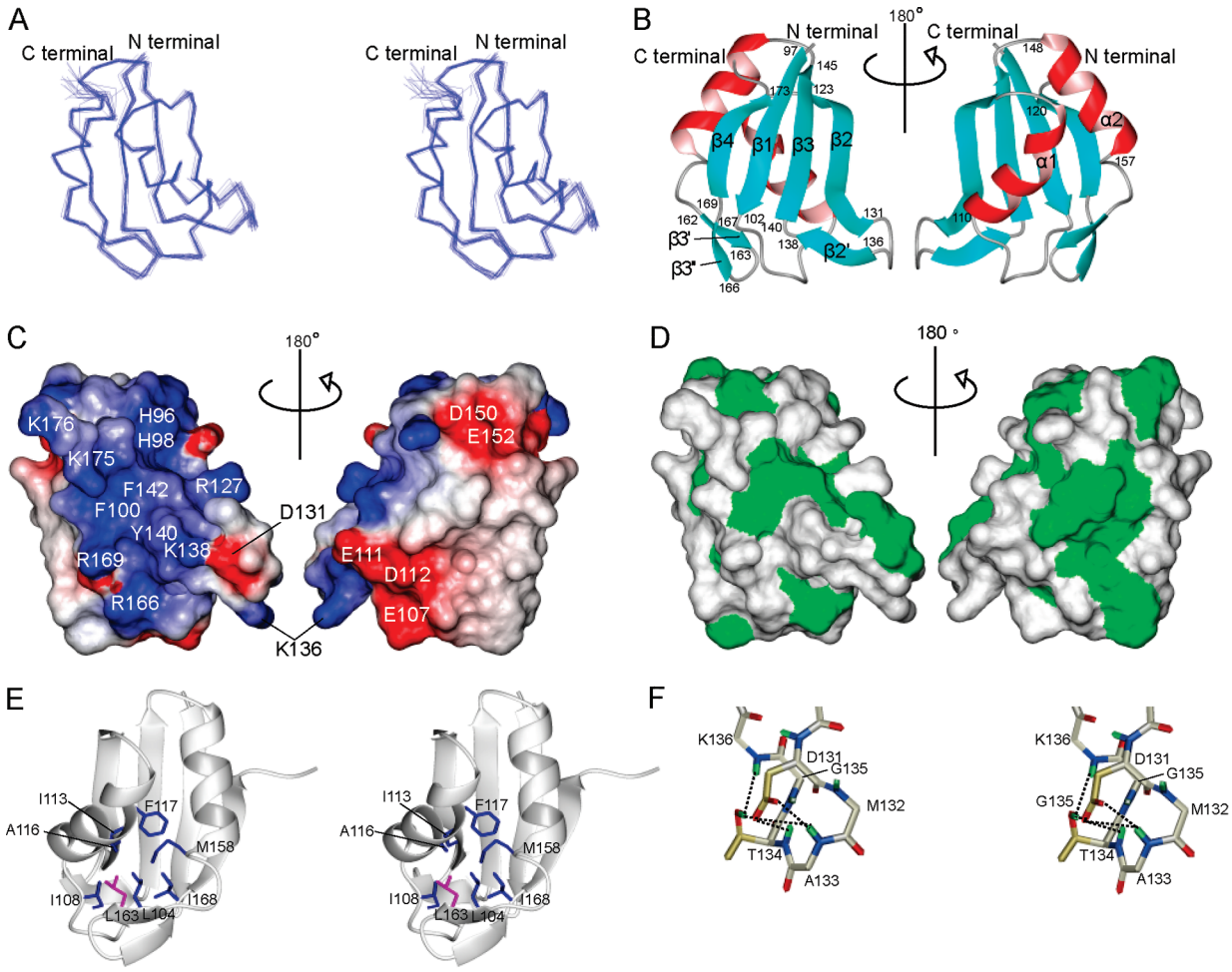

FIGURE 2: Solution structure of the second RRM domain of TIA-1. (A) Superimposition of the 20 lowest energy conformers of the TIA-1 RRM2 (His96−Lys176). Blue lines represent C^{α} traces. (B) Ribbon representation of the TIA-1 RRM2. The secondary structure elements and the sequence numbering are indicated (left panel, β1-β4, β2′, β3′, and β3″; right panel, α1 and α2). (C) Electrostatic surface potential of the TIA-1 RRM2. Positive and negative potentials are colored blue and red, respectively. The molecular orientation is the same as in (B). (D) Surface properties of the TIA-1 RRM2. The hydrophobic residues (Ala, Ile, Leu, Met, Phe, Pro, Trp, and Val) are colored green. The molecular orientation is the same as in (B). (E) Hydrophobic core formation by the β3′-β3″ loop. The side chain of the Leu163 residue is colored magenta. The side chains of the residues that form the hydrophobic core with the β3′-β3″ loop are colored blue. (F) Potential hydrogen bonds formed within the β2-β2′ β-loop. The hydrogen bonds were calculated by MOLMOL and are represented by broken lines (see also Materials and Methods). The main chains and the side chains of the Asp131 and Thr134 residues are shown by gray and pale yellow sticks, respectively (red, oxygen; blue, nitrogen; green, amide proton).

residues surround the aromatic residues, Phe100, Tyr140, and Phe142, within the RNP motifs on the β1 and β3 strands. On the other hand, the α-helical surface is negatively charged, by Glu107 (the β1-α1 loop), Glu111 and Asp112 (the N-terminal region of the α1 helix), and Asp150 and Glu152 (the α2 helix) (Figure 2C).

Comparison of the β1-α1 and α2-β4 Loops of the TIA-1 RRM2 with Those of Other RRMs. In almost all of the RRM domains, the aromatic residues of the RNP motifs on β1 and β3 directly contact the RNA bases by stacking interactions (*36*). In addition, the β1-α1, β2-β3, and α2-β4 regions play important roles in the RNA recognition (*36*). Therefore, we compared the structures of these loops in the TIA-1 RRM2 with those in other RRMs. We examined the sequences of 106 RRMs with solved structures and found that about 70−80% of them have β1-α1 and α2-β4 loops with the same lengths as those in the TIA-1 RRM2 (Figure 4A,B). The two loop regions in most of these RRMs are structurally similar to those of the TIA-1 RRM2 (Supporting Information Figure S1A−D), although their RNA-binding specificities are diverse. Furthermore, a sequence analysis of the two loop regions identified Leu104, Ile108, Leu163, and Ile168 as the most conserved residues (Figure 4A,B). These hydrophobic amino acid residues are buried inside the molecule (Supporting Information Figure S1A,C), indicating that they are involved in the formation of the tertiary structure. On the other hand, the RNA-binding surface has various types of amino acid residues, which show a negligible correlation with their positions in the amino acid sequence (Figure 4A,B).

Comparison of the β2-β3 Loops. In contrast to the β1-α1 and α2-β4 loops, both the length and conformation of the β2-β3 loop are diverse among the RRMs (*37*). As described above, the β2-β3 loop in the TIA-1 RRM2 assumes a characteristic conformation, designated as the β2-β2′ β-loop. To determine whether the β2-β3 loops in other RRMs exhibit the same structural characteristics as those of the TIA-1 RRM2 β2-β2′ β-loop, we examined the sequences of the 106 RRMs with solved structures and found that 25 of them have a β2-β3 loop with the same length as that of the murine TIA-1 RRM2 (10 amino acid residues, corresponding to Lys130−Gly139). In 19 of the 25 RRM structures, the β2-β3 loops are well ordered

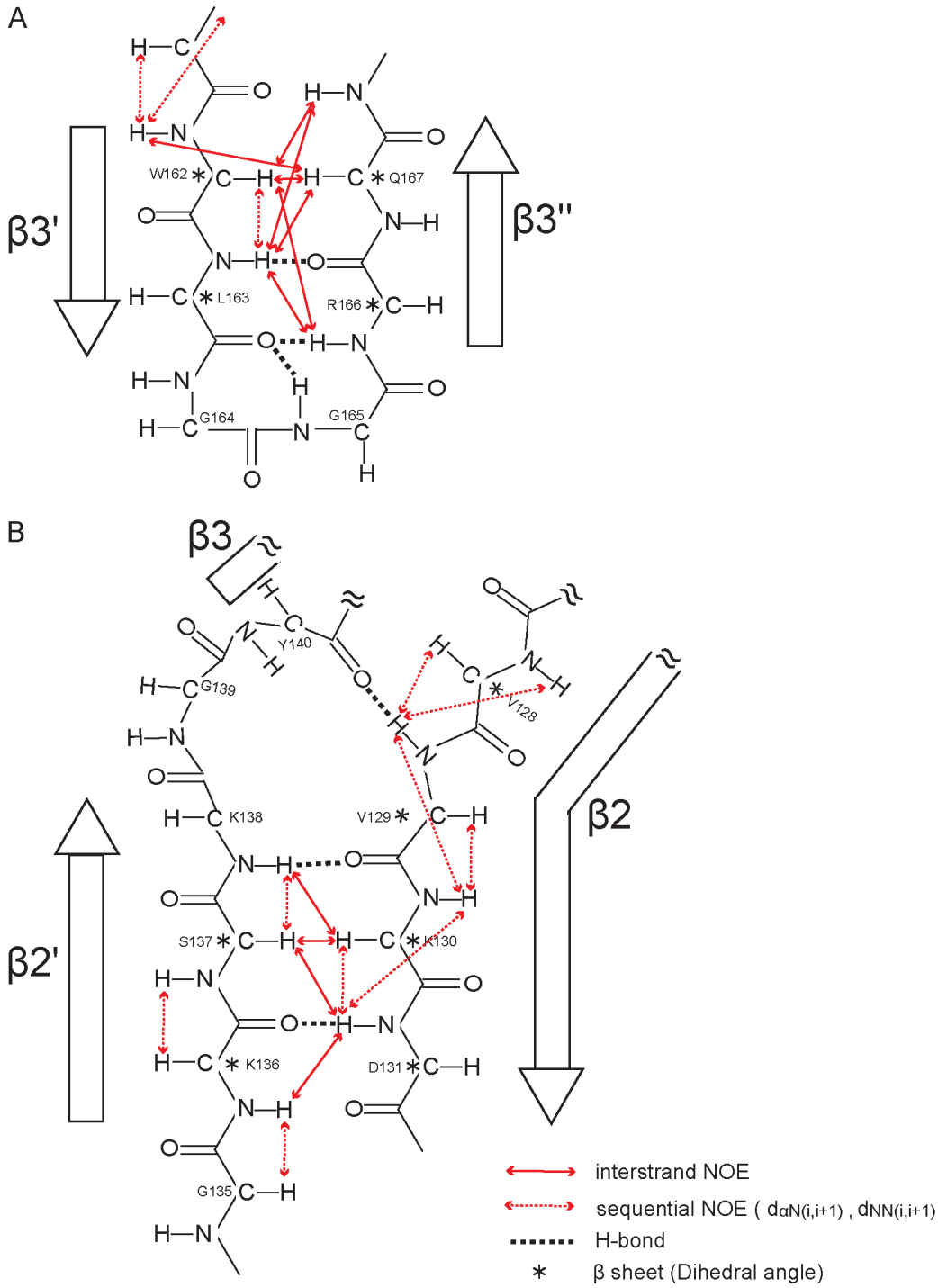

FIGURE 3: Schematic representation of the β-hairpin and the $\beta2$-$\beta2'$ β-loop structures. (A) Schematic representation of the β-hairpin structure. Interstrand and sequential NOEs are represented by double-ended continuous and broken arrows, respectively. The dihedral angle values of four residues (Trp162, Leu163, Arg166, and Gln167) are included in the β-sheet region of the Ramachandran plot (Trp162, $-118.8 < \varphi < -104.5°$, $158.8 < \psi < 99.2°$; Leu163, $-158.9 < \varphi < -85.1°$, $130.7 < \psi < 112.0°$; Arg166, $-148.6 < \varphi < -129.2°$, $166.0 < \psi < 147.1°$; Gln167, $-77.9 < \varphi < -66.1°$, $174.9 < \psi < 136.2°$). The hydrogen bonds were calculated by MOLMOL and are represented by broken lines. These results indicated that the β-hairpin is formed by residues 162−163 ($\beta3'$) and 166−167 ($\beta3''$) in the $\beta1$-$\beta4$ loop region. (B) Schematic representation of the $\beta2$-$\beta2'$ β-loop structure. Interstrand and sequential NOEs are represented by double-ended continuous and broken arrows, respectively. The dihedral angle values of five residues (Val129, Lys130, Asp131, Lys136, and Ser137) are included in the β-sheet region of the Ramachandran plot (Val129, $-86.2 < \varphi < -57.9°$, $146.8 < \psi < 133.1°$; Lys130, $-140.9 < \varphi < -123.6°$, $164.3 < \psi < 145.8°$; Asp131, $-75.3 < \varphi < -59.0°$, $158.4 < \psi < 107.6°$; Lys136, $-100.9 < \varphi < -79.0°$, $163.4 < \psi < 141.5°$; Ser137, $-66.1 < \varphi < -46.3°$, $130.1 < \psi < 117.3°$). In addition, the φ value of Lys138 is included in the β-sheet region of the Ramachandran plot. The hydrogen bonds were calculated by MOLMOL and are represented by broken lines. These results indicated that the additional β-strand, formed between $\beta2$ and $\beta3$ (residues 136−138: $\beta2'$), pairs with the C-terminal region of $\beta2$ (residues 129−131).

and their conformations are strikingly similar to that of the TIA-1 RRM2, with rmsd values smaller than 1.0 Å (Figure 4D, left panel). Some of the 19 RRM structures were determined as the RRM·RNA complex structures, for example, U2AF65 with poly(U) (*38*), Sex-lethal with a U-rich RNA (*39*), HuD with an AU-rich element (*40*), and poly(A)-binding protein with poly(A) (*41*) (Supporting Information Figure S1E,F). In the other six RRM struc-

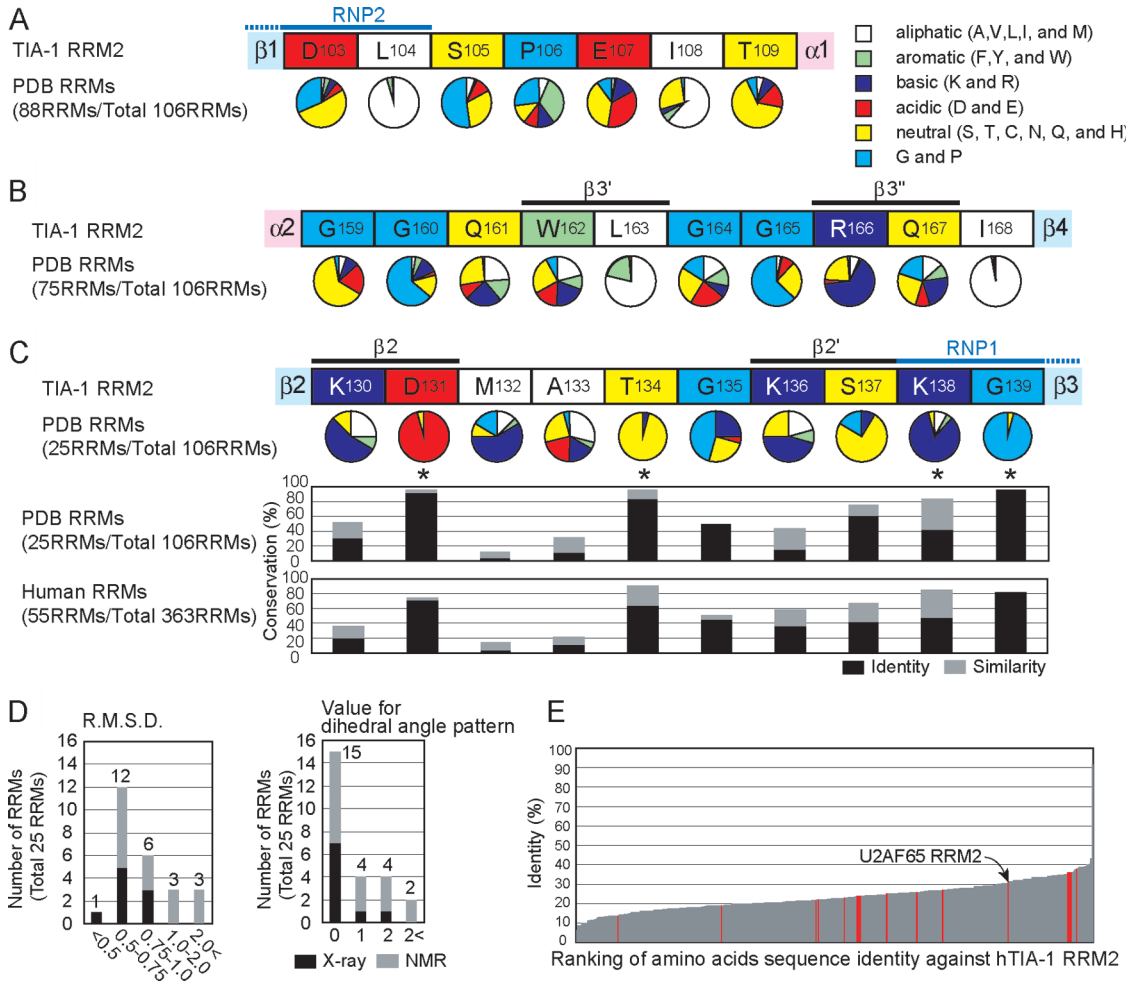

FIGURE 4: Structure and sequence comparisons of the β1-α1, β2-β3, and α2-β4 loops between the TIA-1 RRM2 and other RRMs. (A) Comparison of the amino acid compositions of the β1-α1 loop between the TIA-1 RRM2 and other RRMs. The top panel shows the sequence of the β1-α1 loop in the TIA-1 RRM2. The bottom panel shows the frequency of the amino acid type at each position, using the RRMs with structures that have been solved and with the same loop length as that of the TIA-1 RRM2 (88 RRMs among a total of 106 RRMs, i.e., 83%). (B) Comparison of the amino acid compositions of the α2-β4 loop between the TIA-1 RRM2 and other RRMs, as described in (A) (75 RRMs among a total of 106 RRMs, i.e., 71%). (C) Comparison of the amino acid compositions and sequences of the β2-β3 loop in the RRM domains. The top and second panels show the comparison of the amino acid compositions of the β2-β3 loop between the TIA-1 RRM2 and other RRMs, as described in (A) (25 RRMs among a total of 106 RRMs, i.e., 24%). The third and bottom panels depict sequence identities and similarities of RRMs from PDB and SMART (55 RRMs among a total of 363 RRMs, i.e., 15%), respectively, at each residue in the β2-β3 loop. The identities and similarities were calculated against the TIA-1 RRM2. Black bars represent identity, and gray bars represent the difference between similarity and identity. (D) Distribution of structural similarity. The left and right panels show the frequency distributions of the RRMs from the PDB, with the indicated rmsd values and dihedral angle difference scores, respectively. Black and gray bars represent structures solved by X-ray and NMR, respectively. (E) Distribution of RRM sequence identities against the murine TIA-1 RRM2. The RRM sequences were obtained from the SMART database. The sequence identities of the entire RRMs are sorted in ascending order. RRMs with a β2-β2′ β-loop highly similar to that of the murine TIA-1 RRM2 (rmsd < 1.0 Å) are indicated in red.

tures, all of which were solved in solution, the β2-β3 loops are highly disordered and exhibit much larger rmsd values. These observations suggest that the β2-β3 loop might adopt essentially the same conformation when it is ordered.

A Structural Subfamily of RRMs Characterized by the β2-β2′ β-Loop Conformation. In addition to the rmsd values, we estimated the structural similarity between the β2-β3 loops by examining the dihedral angles φ and ψ as follows. Each residue in the β2-β3 loop was classified into three categories (α-helix, β-strand, and others) on the basis of the φ/ψ pattern and was compared between the murine TIA-1 RRM2 and other RRMs. The number of amino acid residues in different categories from those of the corresponding residues in the TIA-1 RRM2 provides a score for the local structural difference. For example, the "dihedral angle difference score" is 0 when there is no difference in the categories of all of the 10 amino acid residues constituting the β2-β3 loop or when the φ/ψ patterns are the same as those of the TIA-1 RRM2 throughout the β2-β3 loop (Figure 4D, right panel). Among the above-mentioned 25 RRMs with the 10-residue β2-β3 loop, the scores of 15 RRMs (60%) were 0, and only two RRMs had scores higher than 2 (Figure 4D, right panel). Furthermore, five of the above-mentioned six RRMs with larger rmsd values also had scores lower than 2. Therefore, even when the β2-β3 loops were disordered, their main-chain conformations did not differ much from that of the TIA-1 RRM2. This implies that the β2-β2′ β-loop conformation could define a structural subfamily in the RRM family.

The Novel Sequence Motif, DxxT, in the β2-β2′ β-Loop. The sequence analysis of the 25 RRMs identified the well-conserved Asp and Thr residues (corresponding to Asp131 and Thr134 of murine TIA-1) in the β2-β2′ β-loop (Figure 4C), while the β1-α1 and α2-β4 loops conserve some hydrophobic residues. Therefore, we examined not only the frequency of the amino acid type but also the identity at each position and found that the amino acid identities at these two positions are higher than 80% and even higher than that of Lys138 within the RNP1 motif (Figure 4C). In contrast, the two residues following Asp131 (Met132 and Ala133) show much lower similarity among the 25 RRMs (identities of 4% and 12%, respectively) (Figure 4C). Therefore, we will refer to this sequence motif as the "DxxT" motif hereafter. The DxxT motif is particularly well conserved in the 19 RRMs with β2-β3 loops that are well ordered and highly similar to that of the TIA-1 RRM2 (rmsd < 1.0 Å). Moreover, the three residues between the DxxT and RNP1 motifs are relatively well conserved (identities/similarities of 48%/50%, 15%/44%, and 62%/76%, respectively) (Figure 4C).

The Subfamily of RRMs with the DxxT Motif in the 10-Residue β2-β2′ β-Loop. We searched the SMART database (*33*) for all RRMs encoded in the human genome, using the definition of RRMs provided by the Pfam database (*34*), and thus identified 363 RRMs (excluding the human TIA-1 RRM2). As shown in Figure 4E, we examined the identity between the entire sequence of the murine TIA-1 RRM2 and those of the 363 RRMs. The U2AF65 RRM2 shows 30% identity, which is similar to those of many other RRMs and does not seem to be outstanding. Moreover, the RRMs with β2-β2′ β-loops that are very similar to that of the TIA-1 RRM2 (rmsd < 1.0 Å) showed sequence identities from 14% to 38% and were widely distributed in the ranking of the RRMs based on the sequence identity against the murine TIA-1 RRM2 (Figure 4E). In addition, the fruit fly Sex-lethal RRM1 and RRM2 showed 29% and 20% identities, respectively, against the murine TIA-1 RRM2. On the other hand, as many as 55 of the 363 human RRMs have 10-residue β2-β2′ β-loops, and interestingly, as shown in Figure 4C, the DxxT motif is particularly well conserved in these 55 RRMs. Therefore, RRMs that have the 10-residue β2-β2′ β-loop with the DxxT motif appear to constitute a unique "subfamily" in which the entire amino acid sequences of the members do not particularly show high homology. This "subfamily" defined by the structural feature may be a consequence of either convergent or divergent evolution.

Dynamic Properties of the Murine TIA-1 RRM2. To define the flexible and rigid parts of the TIA-1 RRM2, we analyzed the dynamics by measuring the backbone nitrogen relaxation times T_1 and T_2 and the proton−nitrogen heteronuclear NOEs (Figure 5). In addition to the RRM core (residues 97−173), we assigned the N- and C-terminal extensions (residues 89−96 and 175−190, respectively), except for a few residues (see the legend of Figure 5). The T_1 values for the majority of the residues in the RRM core (average of 0.62 s for residues 97−173) are significantly larger than those for the N- and C-terminal extensions (average of 0.36 s for residues 89−96 and 175−190) (p value of t-test < 0.01). The T_2 values for the RRM core are also distinct from those for the N- and C-terminal extensions (66 and 120 ms, respectively, on average) (p value of t-test < 0.01). Moreover, the

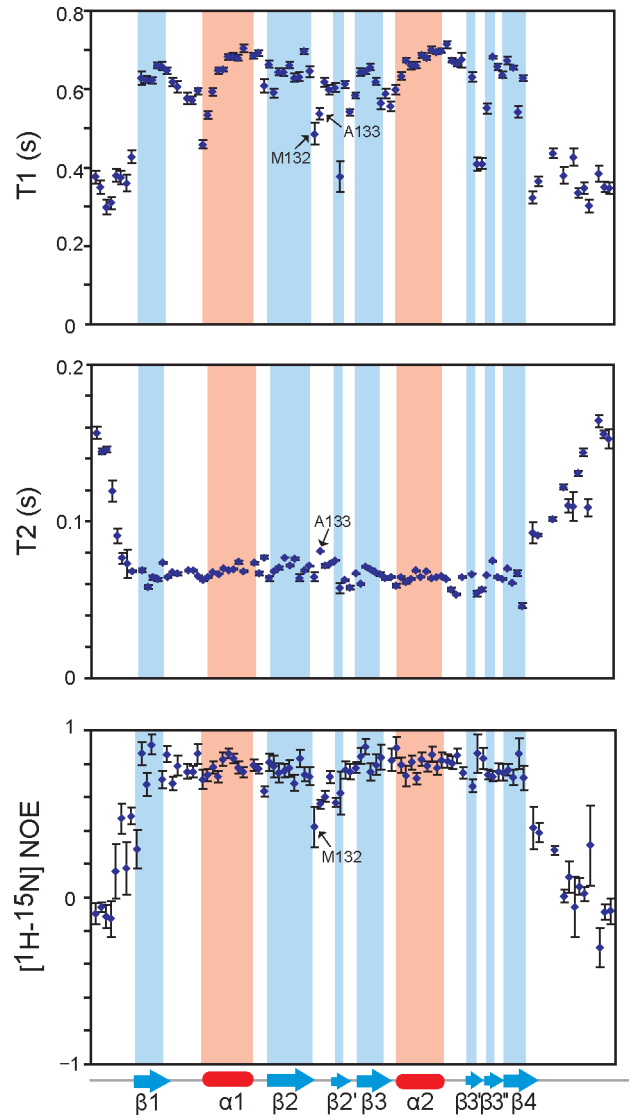

FIGURE 5: Dynamics of the TIA-1 RRM2. The resonance peaks of Phe97 and Ser182, which overlapped each other, are not shown. In the proton−nitrogen heteronuclear NOEs, residues with error bars larger than 1 are also not shown (Val101 and Phe146), because the values of proton−nitrogen heteronuclear NOEs are smaller than 1, in theory (*22*).

heteronuclear NOE values for the RRM core (average 0.76) are also much larger than those of the N- and C-terminal extensions (average ≈0.0 for residues 89−96 and 175−190) (p value of t-test < 0.01). Therefore, the disordered N- and C-terminal extensions are more flexible than the rest of the RRM in solution. Intriguingly, the NOE value for the Met132 residue is different from those for the RRM core. In addition, the T_1 values for the Met132 and Ala133 residues and the T_2 value for the Ala133 residue differ from those for the RRM core. These results indicate that the two residues, which are located between the β2 and β2′ strands and correspond to the "xx" in the DxxT motif, are slightly more flexible in solution (*42*). From the T_1/T_2/NOE analysis, the correlation time of TIA-1 RRM2 was estimated to be 5.5×10^{-9} (*22*). On the basis of the correlation time, the effective hydrodynamic radius of the TIA-1 RRM2 molecule was calculated as 18 Å, indicating a monomeric protein (*43*).

Chemical Shift Perturbation Analysis of the TIA-1 RRM2 upon RNA Binding. The TIA-1 RRM2 reportedly binds strongly and stably to a uridine-rich RNA sequence (i.e., five

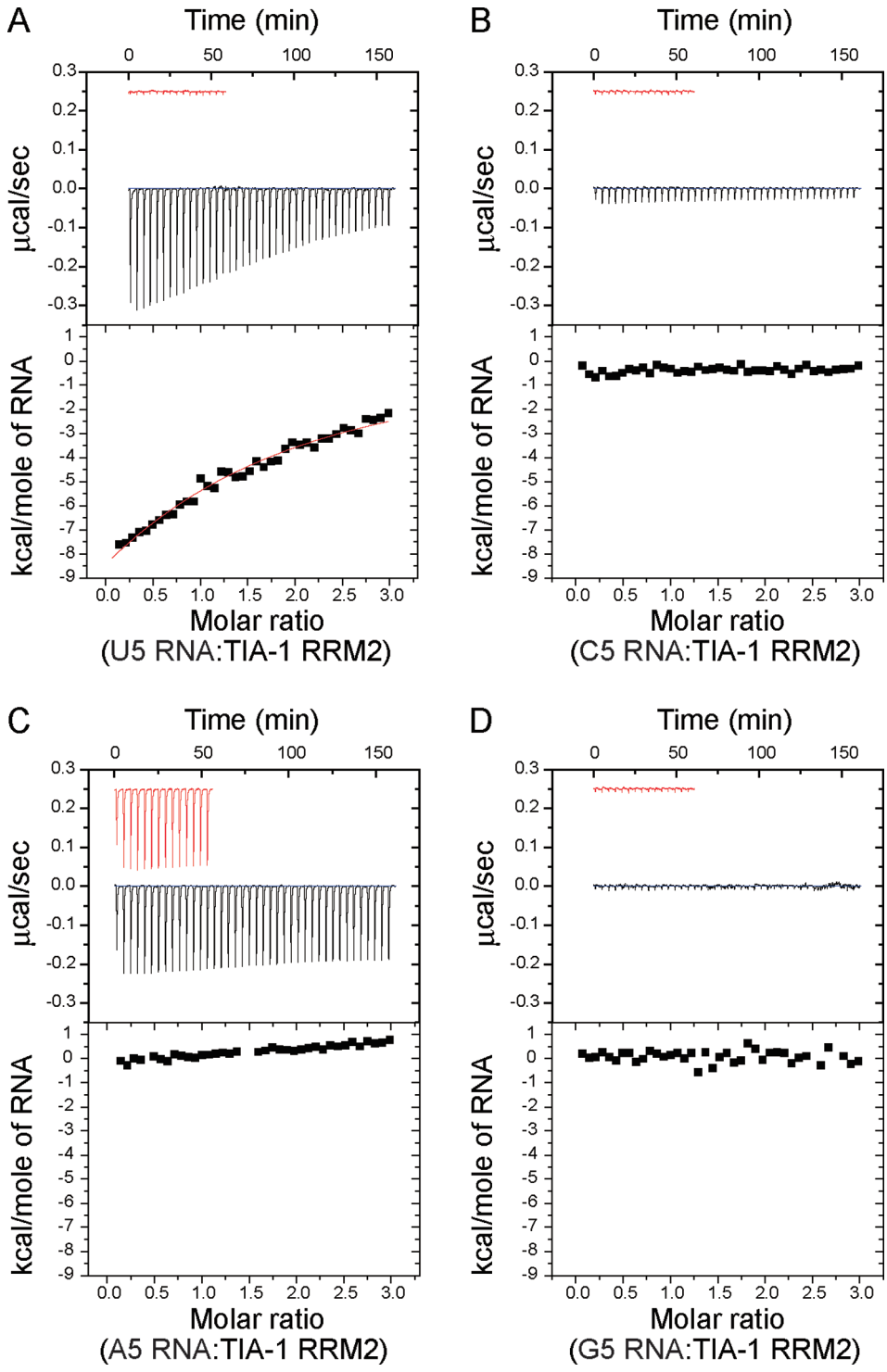

FIGURE 6: ITC measurements of TIA-1 RRM2 binding to four kinds of RNA pentamers: (A) U5 RNA, (B) C5 RNA, (C) A5 RNA, and (D) G5 RNA. Raw data as a function of time are shown in the top panel, and a plot of the total heat released as a function of the molar ratio of RNAs is shown in the bottom panel. The continuous line represents the nonlinear, least-squares best fit to the experimental data using a one-site model. The values of the fitted parameters are shown in the box.

consecutive uridines) (*13*). Consistent with the previous study, the TIA1 RRM2 strongly bound to U5 RNA (K_d approximately 30 μM), but not to A5/C5/G5 RNAs, in ITC measurements (Figure 6). Therefore, we examined the binding of U5 RNA to this RRM by a chemical shift perturbation analysis (Figure 7 and Supporting Information Figure S2; see also Materials and Methods). We mixed the RNA with the TIA-1 RRM2 at protein:RNA ratios from 1:0 to 1:1.9. Some of the main-chain $^{1}H-^{15}N$ resonances shifted in a continuous manner with the increasing RNA concentration (Supporting Information Figure S2). We assigned all of the backbone amide resonances of residues 89−190 by analyzing the NOESY spectra, except for Asp103, Thr174, and Lys176 (Figure 7A). In particular, the cross-peak of Gly139 in the $\beta2$-$\beta2'$ β-loop, which belongs to the RNP1 motif and is proximal to the DxxT motif, was strongly affected upon RNA binding: it shifted to a higher magnetic field and was simultaneously broadened until the RNA concentration was about 0.4-fold over the TIA-1 RRM2 concentration (Supporting Information Figure S2C). When

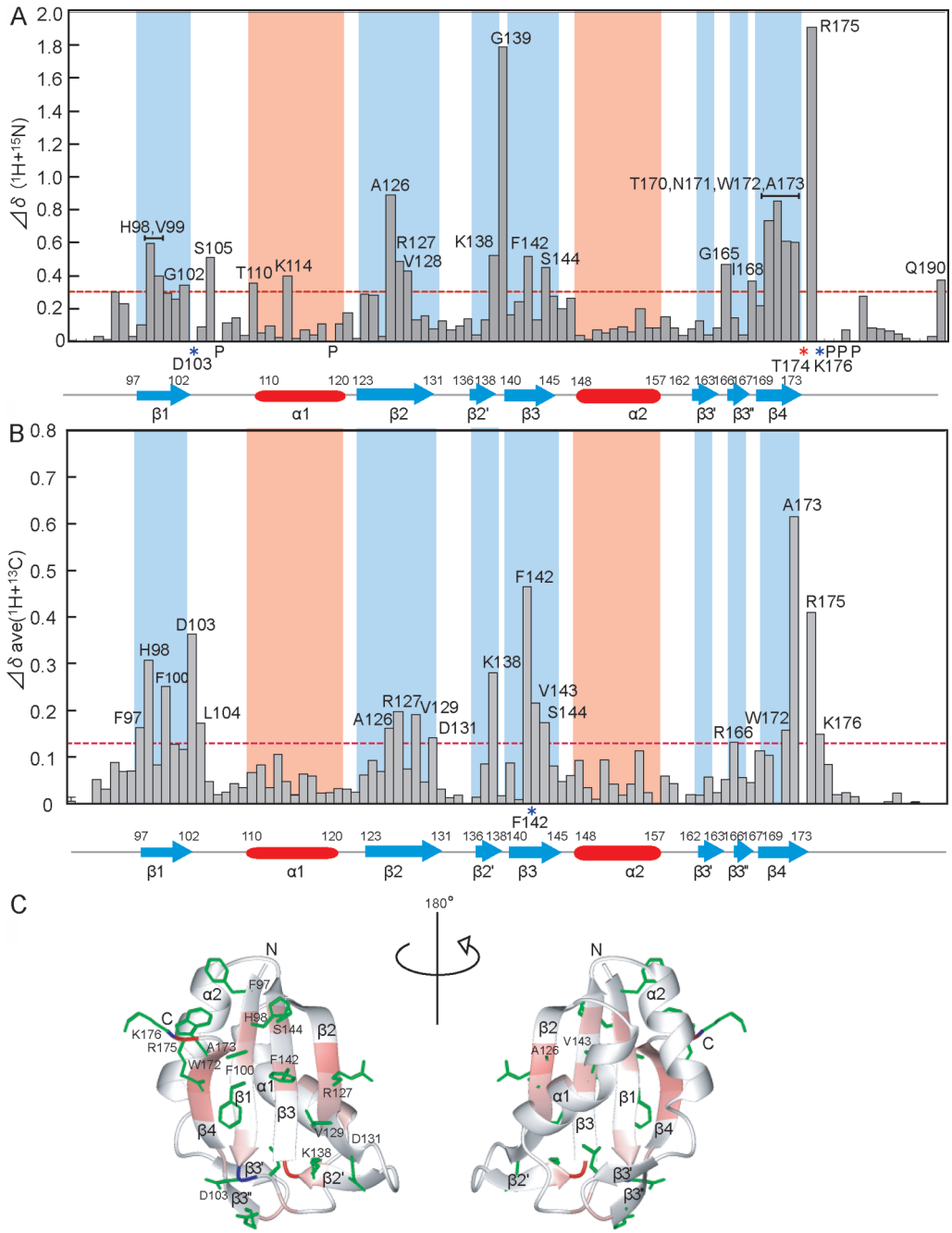

FIGURE 7: Quantification of the chemical shift perturbation values of the TIA-1 RRM2 upon RNA binding (ratio of protein:RNA = 1:1.1, protein concentration 0.53 mM). (A) Main-chain amide protons and nitrogens. The perturbation values were obtained from the [^{1}H,^{15}N]-HSQC spectrum. The absolute values of the chemical shift change $\Delta\delta$ (^{15}N + ^{1}H$_N$) are shown. $\Delta\delta$ (^{15}N + ^{1}H$_N$) was calculated as follows: $\Delta\delta$ (^{15}N + ^{1}H$_N$) = $(\delta_{^{15}N}^2 + \delta_{^1H}^2)^{1/2}$. Perturbation values greater than the 75% baseline (0.1070 ppm) plus three times the standard deviation of the baseline (3 × 0.0793 ppm) were defined as significant perturbations (i.e., the significance level is 0.34 ppm). Black letters indicate amino acid residues with significant chemical shift changes. Asp103 and Lys176, with resonances that disappeared after the addition of the RNA, are indicated by blue asterisks. We could not assign the resonances of the main-chain amide protons and nitrogen of Thr174, as is indicated by red asterisks. (B) Side-chain chemical shift perturbation values in the [^{1}H,^{13}C]-HSQC spectrum. $\Delta\delta$ (^{13}C + ^{1}H) was calculated as follows: $\Delta\delta$ (^{13}C + ^{1}H$_N$) = $(\delta_{^{13}C}^2 + \delta_{^1H}^2)^{1/2}$. The average perturbation value of the atoms assigned in each residue is shown. The definition of a significant perturbation is the same as in (A) (0.1 ppm). Residues with significant chemical shift changes are indicated. The resonances of the C$^\zeta$ atom of Phe142 disappeared upon RNA binding. (C) Ribbon representation of the TIA-1 RRM2. The chemical shift perturbation values of the main-chain amides are color coded from white (0 ppm) to red (1.9 ppm). Asp103 and Lys176, with resonances that disappeared after the addition of the RNA, are colored blue. The side chains of residues with significant side-chain chemical shift perturbations upon RNA binding are shown in green.

the RNA concentration reached 1.2-fold over the protein concentration, the cross-peak of the main-chain amide resonance of Gly139 was sharpened and remained at the same chemical shift value (Supporting Information Figure S2C). This indicates that the exchange rate between the RNA-bound and RNA-free states of the protein is fast in comparison with the NMR time scale.

To further investigate the RNA binding to the TIA-1 RRM2, we quantified the effects of the U5 RNA on the chemical shift values of the main-chain amide resonances,

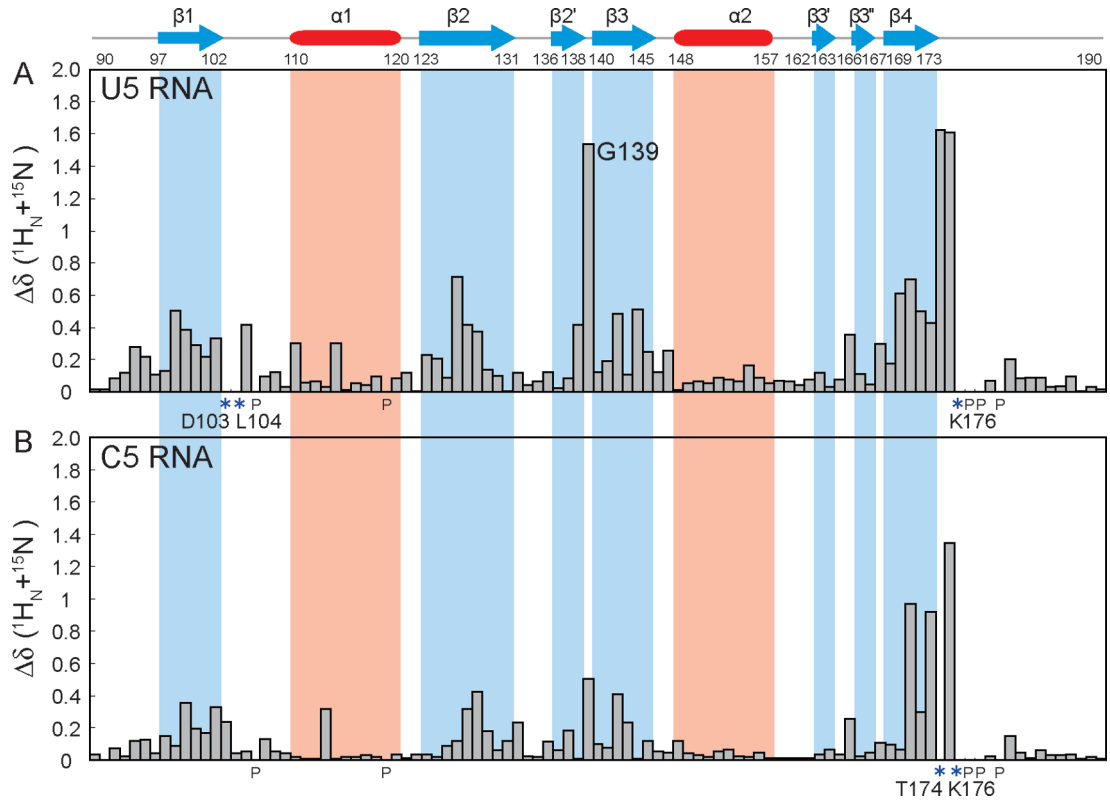

FIGURE 8: Quantification of the chemical shift perturbation values of the TIA-1 RRM2 upon U5/C5 RNA binding (ratio of protein:RNA = 1:2, protein concentration 0.1 mM): (A) U5 RNA and (B) C5 RNA. The perturbation values were obtained as described in Figure 7A. The blue asterisks indicate the residues with resonances that either overlapped or disappeared after the addition of the RNA.

using the [^{1}H,^{15}N]-HSQC spectra (Figure 7A). The cross-peaks of the residues located in β1 (His98, Val99, and Gly102), β2 (Ala126, Arg127, and Val128), β3 (Phe142 and Ser144), and β4 (Thr170, Asn171, Trp172, and Ala173) were significantly affected (Figure 7A). In addition, the cross-peaks of the residues in the N-terminal region (Ser94), the β1-α1 loop (Ser105 and Thr110), α1 (Lys114), the β2-β2′ β-loop (Lys138 and Gly139), the β3′-β3″ β-hairpin (Gly165 and Ile168), and the C-terminal region (Arg175 and Glu190) were significantly affected (Figure 7A).

In the case of U5 RNA binding, we also investigated the chemical shift perturbations of the side-chain resonances in the [^{1}H,^{13}C]-HSQC spectra, using ^{15}N,^{13}C-labeled TIA-1 RRM2 (Figure 7B). The residues with side-chain cross-peaks that were affected by RNA binding correspond to those with affected main-chain amide resonances: the side-chain cross-peaks of the residues located in β1 (Phe97, His98, and Phe100), β2 (Ala126, Arg127, Val129, and Asp131), β3 (Phe142, Val143, and Ser144), and β4 (Trp172 and Ala173) were significantly affected (Figure 7B). In addition, the side-chain cross-peaks corresponding to the β1-α1 loop (Asp103 and Leu104), the β2-β2′ β-loop (Lys138), the β3′-β3″ β-hairpin (Arg166), and the C-terminal region (Arg175 and Lys176) were strongly affected (Figure 7B).

Therefore, the residues on the β-sheet surface and those proximal to the β-sheet are likely to contribute to RNA binding by the TIA-1 RRM2 (Figure 7C). On the other hand, although Glu190 also displayed a significant main-chain chemical shift perturbation, no significant side-chain perturbation was observed for this residue, suggesting that Glu190 is less likely to be involved directly in the RNA recognition by the TIA-1 RRM2. Importantly, the chemical shifts of the amide resonances from both aromatic residues (Phe100 and Phe142) in the RNP were significantly affected by the RNA. In this context, large chemical shift perturbations were reportedly observed for the amide resonances from the RNP residues in the canonical RRMs when they directly contact the RNA bases through stacking interactions (*31*). Therefore, our results suggest that both of the aromatic RNP residues in the TIA-1 RRM2 also play important roles in the RNA recognition, probably through stacking interactions with the RNA bases. All of the residues involved in the interactions with the RNA are located in the positively charged patches, suggesting that electrostatic interactions contribute to the RNA-binding activity of the TIA-1 RRM2 (Figures 2C and 7C). In summary, the TIA-1 RRM2 interacts with the U5 RNA via its β-sheet surface, including the β1-α1 loop, the β2-β2′ β-loop, and the C-terminal region slightly downstream of the β4 strand.

For comparison, we selected another pyrimidine stretch, C5 RNA, according to a previous report (*8*) and examined the binding of the C5 RNA to the TIA-1 RRM2 by a chemical shift perturbation analysis (Figure 8; see also Materials and Methods). Although the ITC experiments indicated that the binding between C5 RNA and TIA-1 RRM2 is not very strong, the C5 RNA showed a chemical shift perturbation pattern relatively similar to that of the U5 RNA. However, Gly139 in the β2-β2′ β-loop, which is proximal to the DxxT motif, was not strongly affected by the C5 RNA, as compared to the U5 RNA binding (Figure 8). The difference may reflect the sequence specificity of TIA-1 RRM2.

Comparison of the RNA-Binding Residues of the TIA-1 RRM2 with Those of RRM1 and RRM3. Among the three

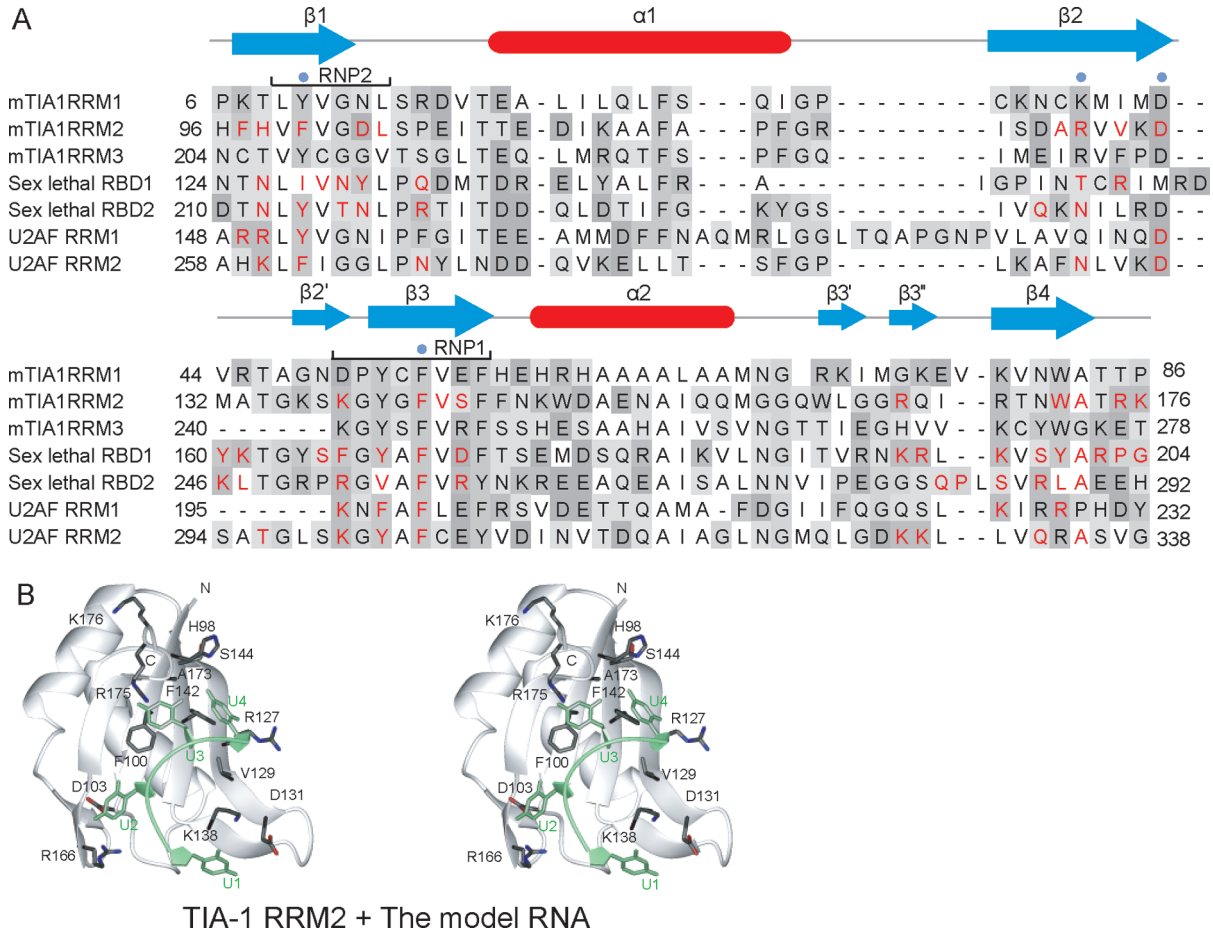

FIGURE 9: Comparison of the RNA recognition modes between the TIA-1 RRM2 and other RRMs. (A) Multiple sequence alignment of the TIA-1 and other RRMs. The RRMs from mouse TIA-1 (AAH46812), the fruit fly Sex-lethal protein (AAA28922), and the human U2AF65 protein (CAA45409) were aligned using ClustalX (*44*). Secondary structure elements are depicted with blue arrows (β-sheet) and red cylinders (α-helix) above the sequence alignment. The signature sequences of RNP1 and RNP2 are depicted by the black lines. The red letters indicate the residues that contribute to RNA recognition (*38, 39*). The blue circles indicate the residues that appear to play important roles in the RNA recognition and are conserved among the three RRMs of TIA-1. (B) Superposition of the RNA molecule from the protein−RNA complex structures of U2AF65 RRM2 on the TIA-1 RRM2. The protein−RNA complex structures were superposed on the TIA-1 RRM2 with reference to the β1 and β3 strands. Amino acid residues with significant side-chain chemical shift changes upon RNA binding on the β-sheet surface are represented by stick models.

RRMs in TIA-1, only RRM2 has sequence-specific RNA-binding activity, and the TIA-1 RRM3 reportedly binds to cellular RNA only weakly in a non-sequence-specific manner (*13*). Therefore, we compared the TIA-1 RRM2 and the other two RRMs of TIA-1 (RRM1 and RRM3) (Figure 9A). We focused on the solvent-accessible residues on the β-sheet surface that showed chemical shift perturbations for the side-chain resonances: His98, Phe100, Asp103, Arg127, Val129, Asp131, Lys138, Phe142, Ser144, Arg166, Ala173, Arg175, and Lys176. As shown in Figure 9A, the conservation of these 13 residues among the three RRMs in murine TIA-1 is only about 30%; four residues (Phe100, Arg127, Asp131, and Phe142), of which two are the RNP motif residues, are essentially conserved in the three RRMs. The four conserved residues are located at the center of the β-sheet surface and at the C-terminus of the β2 strand. On the other hand, the positively charged residues surrounding the aromatic residues in the RNP are not conserved. In addition, although the β2-β3 loop of RRM2 is characteristic ("the β2-β2′ β-loop"), those of the other RRMs are different; RRM1 does not have a positively charged residue immediately upstream of the β3 strand, and the length of the β2-β3 loop in RRM3 is shorter than 10 amino acid residues (Figure 9A). These differences in the surface properties could confer the distinctive sequence-specific RNA-binding property of RRM2.

Similarity in the RNA-Binding Mode between the TIA-1 RRM2 and the U2AF65 RRM2. We compared the TIA-1 RRM2 with other RRMs that bind to U-rich RNA sequences from U2AF65 and Sex-lethal (Figure 9A). Both of them bind to U-rich RNAs by two tandemly repeated RRMs (RRM1 and RRM2) (*38, 39*). All of them hold the RNAs on their β-sheet surface through stacking interactions with the aromatic rings in the RNP (Supporting Information Figure S3A). In addition to the β-sheet, these RRMs utilize the loops connecting the secondary structures for RNA binding: the β2-β3 loop (corresponding to the β2-β2′ β-loop in the TIA-1 RRM2), the α2-β4 loop (corresponding to the β3′-β3″ β-hairpin in the TIA-1 RRM2), and the positively charged residues immediately following the β4 strand. These loops are intimately associated with the β-sheet and recognize the RNAs collectively. Among these four RRMs, the loop structures are conserved in the U2AF65 RRM2 and the Sex-lethal RRM1 (Supporting Information Figure S1A,C,E). Furthermore, the positions of the aromatic rings in the RNP motifs, which are responsible for RNA binding, are identical between the U2AF65 RRM2 and the TIA-1 RRM2 (Figure

9A). Therefore, insights into the RNA recognition mode of the TIA-1 RRM2 can be obtained from that of the U2AF65 RRM2·RNA complex. When we superimposed the RNA structure from the U2AF65 RRM2·RNA complex on the TIA-1 RRM2 structure, the RNA was accommodated well by the β-sheet surface of the TIA-1 RRM2 (Figure 9B; see also Supporting Information Figure S3B): the third and fourth uridines (U3 and U4, respectively) stack on the RNP aromatic residues, Phe100 and Phe142, respectively. In this model, the first uridine, U1, is located close to Asp131 and Lys138, and the second uridine, U2, is surrounded by Asp103 and Arg166. Moreover, U3 is sandwiched between Phe100 and Arg175. The stacking interaction between U4 and Phe142 is probably supported by Arg127 and Val129. His98 may also change the χ^1 angle from $-60°$ to $-180°$ to support the stacking interaction. Most of the residues with chemical shifts significantly affected by the U5 RNA are located close to the modeled RNA (Figure 9B), except for Ser144 and Lys176. Therefore, the RNA recognition mode by the TIA-1 RRM2 is likely to be similar to that of the U2AF65 RRM2.

On the other hand, we observed that the residues located on the β-sheet surface, which are not involved in RNA binding in the U2AF65 RRM2, were affected by the RNA binding (Asp103, Val129, Ser144, Arg175, and Lys176). Consequently, these parts of the β-sheet surface are also likely to contribute to the RNA recognition. These residues may be important for the strong RNA-binding activity of the TIA-1 RRM2. The complex structure of the TIA-1 RRM2 and the poly(U) RNA will clarify the RNA recognition mode of the TIA-1 RRM2.

ACKNOWLEDGMENT

We thank Dr. T. Yabuki, Dr. M. Aoki, Dr. T. Matsuda, Dr. T. Harada, Dr. S. Watanabe, and Ms. E. Seki for sample preparation. We also thank Dr. H. Li, Dr. S. Koshiba, Dr. S. Ohnishi, Ms. S. Suzuki, Dr. T. Nagata, Dr. T. Someya, and Mr. T. Tomizawa for help with the NMR measurement, data analysis, structure calculations, and structure refinement. We are grateful to Dr. K. Kurimoto for helpful discussions about the manuscript. Finally we thank Ms. A. Ishii and Ms. T. Nakayama for assistance in the manuscript preparation.

SUPPORTING INFORMATION AVAILABLE

Figures showing the structure comparison of the β1-α1, β2-β3, and α2-β4 loops between the TIA-1 RRM2 and other RRMs (Figure S1), the amide $^1H-^{15}N$ HSQC spectra of chemical shift perturbations of the TIA-1 RRM2 upon the addition of U5 RNA (Figure S2), and superposition of the RNA molecule from the protein−RNA complex structures of U2AF65 and Sex-lethal on the TIA-1 RRM2 (Figure S3). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Dziembowski, A., Ventura, A. P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprevote, O., and Seraphin, B. (2004) Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J. 23*, 4847–4856.
2. Hastings, M. L., and Krainer, A. R. (2001) Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol. 13*, 302–309.
3. Jurica, M. S., and Moore, M. J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell 12*, 5–14.
4. Nilsen, T. W. (2003) The spliceosome: the most complex macromolecular machine in the cell? *BioEssays 25*, 1147–1149.
5. Varani, G., and Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct. 27*, 407–445.
6. Will, C. L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Luhrmann, R. (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J. 21*, 4978–4988.
7. Zhu, J., and Krainer, A. R. (2000) Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev. 14*, 3166–3178.
8. Query, C. C., Moore, M. J., and Sharp, P. A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev. 8*, 587–597.
9. Izquierdo, J. M., and Valcarcel, J. (2006) A simple principle to explain the evolution of pre-mRNA splicing. *Genes Dev. 20*, 1679–1684.
10. Blencowe, B. J. (2006) Alternative splicing: new insights from global analyses. *Cell 126*, 37–47.
11. Izquierdo, J. M., Majos, N., Bonnal, S., Martinez, C., Castelo, R., Guigo, R., Bilbao, D., and Valcarcel, J. (2005) Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol. Cell 19*, 475–484.
12. Shin, C., and Manley, J. L. (2004) Cell signalling and the control of pre-mRNA splicing. *Nat. Rev. 5*, 727–738.
13. Dember, L. M., Kim, N. D., Liu, K. Q., and Anderson, P. (1996) Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. *J. Biol. Chem. 271*, 2783–2788.
14. Forch, P., Puig, O., Martinez, C., Seraphin, B., and Valcarcel, J. (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites. *EMBO J. 21*, 6882–6892.
15. Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A., and Yokoyama, S. (2004) Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics 5*, 63–68.
16. Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett. 442*, 15–19.
17. Kuwasako, K., He, F., Inoue, M., Tanaka, A., Sugano, S., Güntert, P., Muto, Y., and Yokoyama, S. (2006) Solution structures of the SURP domains and the subunit-assembly mechanism within the splicing factor SF3a complex in 17S U2 snRNP. *Structure 14*, 1677–1689.
18. Bax, A. (1994) Multidimensional nuclear magnetic resonance methods for protein studies. *Curr. Opin. Struct. Biol. 4*, 738–744.
19. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.
20. Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol. 278*, 313–352.
21. Kobayashi, N., Iwahara, J., Koshiba, S., Tomizawa, T., Tochio, N., Güntert, P., Kigawa, T., and Yokoyama, S. (2007) KUJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to high-throughput NMR structure studies. *J. Biomol. NMR 39*, 31–52.
22. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry 33*, 5984–6003.
23. Cordier, F., Caffrey, M., Brutscher, B., Cusanovich, M. A., Marion, D., and Blackledge, M. (1998) Solution structure, rotational diffusion anisotropy and local backbone dynamics of *Rhodobacter capsulatus* cytochrome c_2. *J. Mol. Biol. 281*, 341–361.
24. Güntert, P., Mumenthaler, C., and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol. 273*, 283–298.
25. Güntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol. Biol. 278*, 353–378.
26. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR 13*, 289–302.
27. Koradi, R., Billeter, M., and Güntert, P. (2000) Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun. 124*, 139–147.

28. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, J. K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Cajdwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc. 117*, 5179–5197.
29. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR 8*, 477–486.
30. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics 14*, 51−55, 29−32.
31. Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L., Pitsch, S., and Allain, F. H. (2006) Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J. 25*, 163–173.
32. Nicholls, A., and Honig, B. (1990) A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Bolzmann equation. *J. Comput. Chem. 12*, 435–445.
33. Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res. 34*, D257–D260.
34. Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res. 34*, D247–D251.
35. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res. 25*, 4876–4882.
36. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J. 272*, 2118–2131.
37. Fleming, K., Ghuman, J., Yuan, X., Simpson, P., Szendroi, A., Matthews, S., and Curry, S. (2003) Solution structure and RNA interactions of the RNA recognition motif from eukaryotic translation initiation factor 4B. *Biochemistry 42*, 8966–8975.
38. Sickmier, E. A., Frato, K. E., Shen, H., Paranawithana, S. R., Green, M. R., and Kielkopf, C. L. (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell 23*, 49–59.
39. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature 398*, 579–585.
40. Wang, X., and Tanaka Hall, T. M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol. 8*, 141–145.
41. Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell 98*, 835–845.
42. Renisio, J. G., Perez, J., Czisch, M., Guenneugues, M., Bornet, O., Frenken, L., Cambillau, C., and Darbon, H. (2002) Solution structure and backbone dynamics of an antigen-free heavy chain variable domain (VHH) from Llama. *Proteins 47*, 546–555.
43. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (2007) *Protein NMR Spectroscopy, Principles and Practice*, Academic Press, San Diego.
44. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem. Sci. 23*, 403–405.

BI7024723