HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2015 December 23.

Published in final edited form as:

Mol Cell. 2006 January 6; 21(1): 75–85. doi:10.1016/j.molcel.2005.10.027.

Structural Basis for Recognition and Sequestration of UUU_{OH} 3′ Temini of Nascent RNA Polymerase III Transcripts by La, a Rheumatic Disease Autoantigen

Marianna Teplova[1], **Yu-Ren Yuan**[1], **Anh Tuân Phan**[1], **Lucy Malinina**[1], **Serge Ilin**[1], **Alexei Teplov**[1], and **Dinshaw J. Patel**[1,*]

[1]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

Summary

The nuclear phosphoprotein La was identified as an autoantigen in patients with systemic lupus erythematosus and Sjogren's syndrome. La binds to and protects the UUU_{OH} 3′ terminii of nascent RNA polymerase III transcripts from exonuclease digestion. We report the 1.85 Å crystal structure of the N-terminal domain of human La, consisting of La and RRM1 motifs,bound to r(U1-G2-C3-U4-G5-U6-U7-U8-$U9_{OH}$).The U7-U8-$U9_{OH}$ 3′ end, in a splayed-apart orientation, is sequestered within a basic and aromatic amino acid-lined cleft between the La and RRM1 motifs. The specificity-determining U8 residue bridges both motifs, in part through unprecedented targeting of the β sheet edge, rather than the anticipated face, of the RRM1 motif. Our structural observations, supported by mutation studies of both La and RNA components, illustrate the principles behind RNA sequestration by a rheumatic disease autoantigen, whereby the UUU_{OH} 3′ ends of nascent RNA transcripts are protected during downstream processing and maturation events.

Introduction

Diverse aspects of RNA metabolism are dictated by the La autoantigen (reviewed in Kenan et al., 1991; Maraia and Intine, 2001; Wolin and Cedrervall, 2002; Kenan and Keene, 2004), an abundant RNA binding phosphoprotein found in the nucleus of all eukaryotes (sequence alignments are presented in Figure S1, in the Supplemental Data available with this article online) and originally identified as an autoantigen in patients with systemic lupus erythematosus (Mattioli and Reichlin, 1974) and Sjogren's syndrome (Alspaugh and Tan, 1975). La specifically targets UUU_{OH} elements of newly RNA polymerase III-transcribed RNA (Stefano, 1984), including pre-tRNAs, 5S rRNAs, and snRNAs, whereas it

*Correspondence: pateld@mskcc.org.

Supplemental Data

Supplemental Data include one table and three figures and can be found with this article online at http://www.molecule.org/cgi/content/full/21/1/75/DC1/.

Accession Numbers

Coordinates have been deposited in the Protein Data Bank under accession codes 1ZH5 (1.85 Å structure, native data set) and 1YTY (2.3 Å structure, SAD data set).

discriminates against 3′-phosphate-containing internal oligo U tracts and degraded RNA. La plays a key role in 5′ and 3′ end processing of pre-tRNA precursors (Fan et al., 1998; Yoo and Wolin, 1997) and exhibits RNA chaperone-like activity (Chakshumathi et al., 2003), thereby playing a key role in facilitating correct transcript folding, downstream processing and maturation, and ribonucleoprotein particle assembly (Pannone et al., 2001; Xue et al., 2000). In addition, La binds viral RNAs associated with hepatitis C virus (Ali et al., 2000; Pudi et al., 2004) and the X-linked inhibitor of apoptosis protein (Holcik and Korneluk, 2000) by site-specifically targeting their internal ribosome entry sites and stimulating translational initiation. Despite its diverse functional role in RNA folding, assembly, processing, and maturation, and as a rheumatic disease autoantigen, no details are currently available regarding the recognition principles by which La site-specifically discriminates amongst its diverse RNA targets.

The 408 residue human La protein is composed of N- and C-terminal domains, with the N-terminal domain (NTD) consisting of La and RRM1 motifs, whereas the C-terminal domain (CTD) consists of the RRM2 motif and an unstructured, long flexible element (Figure 1A). NMR studies on human La domains have defined a winged-helix fold for the La motif (Alfano et al., 2004), a classical RNA recognition module (RRM) for the RRM1 motif (Alfano et al., 2004), and an atypical RRM fold for the RRM2 motif (Jacks et al., 2003). Crystallographic studies have independently identified a winged-helix fold for the *T. brucei* La motif (Dong et al., 2004). NMR-based chemical shift mapping (Alfano et al., 2004) and site-specific mutagenesis (Dong et al., 2004) studies have identified highly conserved surface patches, containing both basic and aromatic residues, within both the La (Alfano et al., 2004; Dong et al., 2004) and RRM1 (Alfano et al., 2004) motifs that appear to be sufficient for specific targeting of 3′-terminal UUU_{OH} elements (Alfano et al., 2004).

We report the structure of the human La NTD protein-RNA complex (Figure 1C), which defines the molecular basis for binding affinity and specificity of a rheumatic disease autoantigen targeted to the 3′ end of nascent RNA polymerase III transcripts. Our structure, supplemented by mutational studies, has established how La and RRM1 motifs synergize to form a functional RNA binding scaffold for recognition and sequestration of UUU_{OH} 3′ ends of nascent Pol III transcripts. Strikingly, neither the β sheet face of the RRM1 motif nor the recognition helix of the winged-helix La motif is involved in UUU_{OH} recognition, making them available for additional RNA recognition-based activities. Instead, one edge of the RRM1 β sheet is sequence-specifically targeted by the central U of the UUU_{OH} segment, steric constraints discriminate against modifications (such as phosphorylation) of the 3′-terminal sugar, and the complex is stabilized by base and sugar stacking of individual U residues on invariant aromatic amino acids.

Our structural results are complemented by recent functional experiments, in which 3′ protection from the nuclear exosome and chaperoning of structurally impaired pre-tRNAs have been shown to represent distinguishable and modular La activities (Y. Huang, M.A. Bayfield, and R.J. Maraia, unpublished data). This study indicates that typical RRM-like interactions are not involved with 3′ end protection but play a critical role in La's chaperone function.

Results

Crystal Structure of the Complex

We have used NMR to screen a series of RNA oligomers terminating with UUU_{OH} 3′ ends for their ability to form complexes with La NTD in solution. The best NMR spectra were found for La NTD complexes with 3 nt and 9 nt RNA sequences ending in UUU_{OH}. The r(U1-G2-C3-U4-G5-U6-U7-U8-$U9_{OH}$) 9 nt sequence used in this study has been shown previously to bind La NTD with 5 nM affinity (Ohndorf et al., 2001).

We have successfully grown crystals of La NTD (1– 194) (sequence listed in Figure S1) bound to this 9 nt sequence (Figure 1B), which belonged to space group $P2_12_12$ and diffracted to 1.85 Å resolution. The structure of the complex was determined by multiwavelength anomalous dispersion (MAD) phasing on the crystal of the selenomethionine-labeled complex (see Experimental Procedures), with crystallographic statistics listed in Table 1. A simulated annealing omit map for the bound UUU_{OH} segment of the complex is shown in Figure 2A.

There are two copies of noncrystallographically related La NTD's bound to two copies of single-stranded 9 nt RNA in the asymmetric unit. The 9 nt RNA forms a self-complementary five-base-pair duplex involving two Watson-Crick and three mismatch base pairs and contains 5′-U1 and U7-U8-$U9_{OH}$-3′ overhangs (Figure 1B). The UUU_{OH} 3′ terminii, located toward either end of the five-base-pair duplex, are sequestered at the interface between the La and RRM1 motifs (Figure 1C), within pockets that include basic surface patches (Figure 1D). The buried surface area is 373 $Å^2$ and 319 $Å^2$ at the La motif-RNA and RRM1 motif-RNA interfaces, respectively, with minimal contact (164 $Å^2$) between the La and RRM1 motifs within a given La NTD in the complex. Molecular recognition is further supplemented through intermolecular contacts between the 5′-U1-G2 segment and the La NTD in the complex (Figures 1E and 1F).

The structures of the La (D6-R91) and RRM1 (D107-Y188) motifs in the crystal structure of the human NTD La-RNA complex can be compared with the corresponding structures of the individual human La and RRM1 motifs in the free state determined by NMR spectroscopy (Alfano et al., 2004). The rmsds (C_α atoms) are 1.52 Å for the La motif and 1.85 Å for the RRM1 motif (excluding loop segment L146-K151), consistent with these motifs retaining their conformations on complex formation. Because La and RRM1 motifs in isolation bind weakly to nascent RNA transcripts, the partial rigidification, through α helix formation, of a portion (T101-N106) of the linker segment on complex formation must contribute to the relative interfacial orientation of the two motifs in the complex (Figure 2B) necessary for synergistic UUU_{OH} 3′ end recognition.

The relative alignments of 5′-U1 and U7-U8-$U9_{OH}$-3′ overhangs at the protein-RNA interface in the complex are shown in Figure 1E. TheU1base stacks on the duplex but otherwise is unpaired in the complex, whereas extension of the phosphodiester backbone at the U6-U7 step prevents U7 from stacking on the duplex. The UUU_{OH} segment adopts a novel conformation, where U7 and U9 are partially stacked on each other and directed toward the La motif (in cyan), whereas U8 is looped out and directed toward the RRM1

motif (in green) (Figures 1E and 1F), which corresponds to a basic region of the interface (Figure 1D). Strikingly, the bases in each of the U6-U7, U7-U8, and U8-U9 steps are splayed apart, with their phosphodiester backbones adopting extended conformations (Figure 2C), facilitated by sugar puckers at U7 (O1′-*endo* and C2′ -*endo* in the two molecules in the asymmetric unit), U8 (C2′-*endo*), and U9 (01′-*endo*). Importantly, the recognition pocket for the UUU_{OH} 3′ end is lined by aromatic amino acids Y23, F35, and F55 (Figure 1E), which are highly conserved or invariant amongst La sequences (Figure S1).

Intermolecular Recognition of U7-U8-$U9_{OH}$ 3′ End by La NTD

The sequence specificity of La NTD for the UUU_{OH} 3′ end can be evaluated from the range of intermolecular contacts associated with complex formation (stereo Figure 2D and Figures 3A–3F). The O^2 atom of the U7 base is directed toward the side chain amide of partially conserved N56 (Figure 3A) but appears unlikely to form a hydrogen bond given the long heteroatom-heteroatom distances (3.54 Å and 5.50 Å for molecules 1 and 2, respectively) in the two molecules in the asymmetric unit.

The U8 residue, which is splayed apart relative to U7 and U9, is anchored in place through hydrogen bonding of its O^2 atom to the side chain amide of Q20 on the La motif and its O^4 atom to the peptide backbone of I140 on the RRM1 motif, as well as through stacking of its base with Y23 (Figure 3B). A nonbridging phosphodiester oxygen of the U8-U9 step is hydrogen bonded to the side chain hydroxyl of Y24 and the backbone amide of R57 (Figure 3C).

The U9 base and sugar ring stack on F35 (Figure 3D) and F55 (Figure 3A), respectively, but otherwise the U9 base does not form any base-specific intermolecular hydrogen bonds. The side chain carboxylate group of D33 is hydrogen-bonded to the 2′-OH and 3′-OH of U9 in the complex (Figure 3D).

Intermolecular Recognition of the 5′ End of Bound RNA by La NTD

There are also intermolecular contacts between the La NTD and the 5′ overhang U1 base and the G2-C3 step associated with the duplex segment of the RNA. The O^4 atom of U1 forms weak bifurcated hydrogen bonds to the guanidinium group of partially conserved R32 on the La motif and to the amide group of partially conserved N139 on the RRM1 motif (Figure 3E). Further, the guanidinium group of R32 is stacked over the purine ring of G2 of the terminal G2•U6 mispair (Figure 3E).

We also observe interactions between the sugar-phosphate backbone of the 5′ end segment and the RRM1 motif of the La NTD in the complex. The 2′-OH of G2 forms a hydrogen bond with the amide of partially conserved Q141 (Figure 3F), whereas the nonbridging phosphodiester oxygen of the G2-C3 step is hydrogen bonded to partially conserved K185 (Figure 3F).

Impact of La NTD Mutants on Binding Affinity

We have used gel electrophoretic mobility shift binding assays to monitor the binding of selective La NTD mutants to the r(U1-G2-C3-U4-G5-U6-U7-U8-$U9_{OH}$) 9 nt sequence, with

the emphasis on amino acids involved in intermolecular recognition of the UUU_{OH} 3′ end in the structure of the complex. The experimental data are shown in Figure 4A, with the binding curves (Figure 4B) quite distinct for the nonaromatic amino acid mutants N56A and Q20A, with the former exhibiting a K_D of 15 ± 3 nM, compared to 7 ± 1 nM for wild-type La NTD, whereas the latter exhibiting undetectable binding affinity. We observe a pronounced reduction in binding affinity for aromatic amino acid mutants, with a K_D > 500 nM for the Y24A mutant and no measurable binding affinities for the Y24F, Y23A, and F35A mutants (Figure 4B). We have generated three mutants of D33 that differ as a function of charge and size, with the binding affinities decreasing in the order D33E (K_D = 7 ± 1 nM), D33A (K_D = 21 ± 3 nM), and D33I (K_D = 129 ± 23 nM) (Figure 4C). These binding affinities are listed and compared in Table S1.

Impact of RNA UUU_{OH} 3′ End Substitutions on Binding Affinity

We have used gel electrophoretic mobility shift competition assays to monitor the binding of wild-type La NTD to the r(U1-G2-C3-U4-G5-U6-U7-U8-$U9_{OH}$) 9 nt sequence containing base substitutions at the U7, U8, and U9 positions (Figure S2). We observe a 2.4-fold and 1.4-fold loss in binding affinity for the U7C and U7A substitutions, respectively (Figure 5A). By contrast, a much larger reduction in binding affinity is observed for U8 substitutions, with a 14-fold and 9.0-fold loss in binding affinity for the U8C and U8A substitutions, respectively (Figure 5B). We observe a 3.3-fold and 1.4-fold loss in binding affinity for the U9C and U9A substitutions, respectively (Figure 5C). Replacement of U9 by dT resulted in a minimal 1.5-fold loss in binding affinity (Figure 5C). These binding affinities, based on competition assays, are listed and compared in Table S1.

Impact of Compensatory Mutations on Binding Affinity

We have used gel electrophoretic mobility shift competition assays to monitor the binding of wild-type La NTD and its D33A mutant to the r(U1-G2-C3-U4-G5-U6-U7-U8-$U9_{OH}$) 9 nt sequence (Figure S3A) and an analog containing $U9_m$ (m stands for 2′-OCH_3) for U9 substitution (Figure S3B). Replacement of U9 by $U9_m$ in the 9 nt RNA results in a pronounced 38-fold drop in binding affinity for wild-type La NTD (Figure 5D). The D33A mutant of La NTD, in which Ala replaces the larger, negatively charged Asp residue, results in a reduced loss (6-fold) in binding affinity, when U9 is replaced by $U9_m$ in the 9 nt RNA (Figure 5E).

Discussion

RNA UUU_{OH} 3′ End Segment Primarily Targeted by the La Motif

The majority of the hydrogen bonding and stacking contacts of the UUU_{OH} 3′ end segment are with the La motif (Figure 2B), consistent with this being the most conserved motif within the La protein; and one critical contact involving U8 is made with the RRM1 motif in the complex. The U1 5′ end and UUU_{OH} 3′ end contact segments of α1′, α1, α2, and α4 of the La motif and β5 and α6 of the RRM1 motif, which are brought into proximity to generate the binding site (Figures 2B and 6A).

The classical RRM fold is composed of a four-stranded antiparallel β sheet packed against a pair of α helices (reviewed in Varani and Nagai, 1998; Messias and Sattler, 2004). Structural studies of several RRM-RNA complexes, starting with the U1A-RNA complex (Oubridge et al., 1994; Allain et al., 1996), have established the key contributions to RNA recognition by conserved amino acids that project from the two central β strands of the four-stranded β sheet.

The RRM1 motif of La NTD adopts the classical RRM fold in both the free state (Alfano et al., 2004) and as part of the La NTD in complex with RNA (Figure 6A, this study). It was therefore surprising to find that the bound RNA is not positioned over the four-stranded β sheet in the La NTD-RNA complex (Figure 2B and the β sheet highlighted in red in Figure 6A), as it is in the *D. melanogaster* sex-lethal (SxL) protein bound to its transformer (*tra*) pre-mRNA site (Handa et al., 1999). The first RRM domain of SxL, with its β sheet highlighted in red, in complex with oligo-U tract RNA is shown in Figure 6B.

Rather, the U8 base forms a hydrogen bond with the backbone amide of I140 (Figure 3B) located on β5, with this interaction positioned in the plane of the β sheet rather than above it. This edge-wise use of the RRM for sequence-specific recognition still leaves the exposed face of the β sheet available for further recognition by either RNA or protein. Such recognition could be envisaged for an ssRNA that extends from the 5′ end of the 9 nt sequence used in the current study.

The classical winged-helix scaffold consists of three α helices, three β strands, and two loop segments or wings (reviewed in Gajiwala and Burley, 2000) and has been found in proteins that play critical roles in embryogenesis, development, and aging. Structural studies, starting with the hepatocyte nuclear factor-3 (HNF-3)-DNA complex, have established the key contributions to DNA recognition by an α helix within a helix-turn-helix fold, which targets the major groove, together with one wing, which targets the minor groove (Clark et al., 1993). Subsequently, a winged-helix domain was also observed for double-stranded RNA adenosine deaminase (ADAR1), and an α helix within it was shown to target the groove of left-handed Z-DNA (Schwartz et al., 1999). To date, the majority of winged helix motifs target DNA, with the only exception being SelB (Fourmy et al., 2002; Selmer and Su, 2002), which targets RNA in an as yet unknown manner.

The La motif of La NTD adopts the classical winged-helix fold (with three additional α helices) in both the free state (Alfano et al., 2004; Dong et al., 2004) and as part of the La NTD in complex with RNA (Figure 6A, this study). The E2F family of transcription factors, which control genes involved in growth and DNA replication, adopt winged-helix motifs. The structure of the EF4-DNA component of the EF4-DP2-DNA complex (Zheng et al., 1999) is shown in Figure 6C, with the recognition α helix (colored red) positioned in the major groove of the DNA duplex. It is readily apparent that the corresponding α helix in the La motif, α5 (colored red), is not involved in UUU_{OH} RNA recognition in the La NTD-RNA complex (Figure 6A). The majority of the surface of the La motif, including α5 (Figure 6A), is therefore available for further recognition by either nucleic acid or protein.

Uracil Alignments in the Complex

U7, the first U in the UUU_{OH} 3′ end segment, appears not to be anchored to the La NTD involving its base and sugar rings through either hydrogen bonding or stacking with aromatic amino acids in the structure of the La NTD-RNA complex (Figure 3A). The side chain of N56 is not involved in hydrogen bonding to U7, and it is therefore not surprising that there is only a 2-fold loss in binding affinity for the N56A mutant (Figure 4B). Similarly, only a 2.4-fold and a 1.4-fold loss in binding affinities are observed for the U7C and U7A substitutions, respectively (Figure 5A). Thus, our data suggest that the first U in the UUU_{OH} 3′ end segment is not critical for recognition and can be replaced by either C or A.

U8, the second U in the UUU_{OH} 3′-end segment, is anchored to the La NTD by both hydrogen bonding and stacking interactions (Figure 3B). Thus, the O^2 and O^4 atoms of the U8 base are hydrogen bonded in opposite directions to side chain (amino group of invariant Q20) and backbone (I140) atoms, respectively. Indeed, disruption of one such interaction in the Q20A mutant resulted in a dramatic loss in binding affinity (Figure 4B). Disruption of the stacking of U8 and highly conserved Y23 (Figure 3B),as in the case of the Y23A mutation, also resulted in the dramatic loss of binding affinity (Figure 4B). These conclusions are reinforced by the 14-fold loss in binding affinity for the U8C substitution and the 9-fold loss for the U8A substitution (Figure 5B). These specific intermolecular contacts imply that U8 cannot be replaced by either C or A and that the second U in the UUU_{OH} segment represents a key specificity determinant.

One nonbridging phosphate oxygen of the phosphate group linking U8 and U9 of the U8-$U9_{OH}$ 3′ end segment is hydrogen bonded to the side chain of invariant Y24 in the complex (Figure 3C). The intermolecular hydrogen bond to the hydroxyl of Y24 appears to be important, as binding is reduced from 7 ± 1 nM in the wild-type La NTD to >500 nM in the Y24A mutant, and no detectable binding affinity is observed for the Y24F mutant (Figure 4B).

U9, the last U in the UUU_{OH} 3′ end segment, forms a set of intermolecular contacts in the La NTD-RNA complex. There are no intermolecular hydrogen bonds between the U9 base and the amino acid backbone and side chain atoms (Figure 3D). This observation is consistent with only modest reductions in the binding affinity for the U9C (3.3-fold) and U9A (1.4-fold) substitutions (Figure 5C). By contrast, stacking of the U9 base on invariant F35 (Figure 3D) contributes significantly to stability, as no detectable binding affinity was observed for the F35A mutant (Figure 4B), emphasizing the importance of intermolecular stacking interactions of the last U to complex formation.

The U9 sugar stacks on invariant F55 (Figure 3A), whereas its 2′-OH and 3′-OH groups are hydrogen-bonded to both carboxylate oxygens of invariant D33 (Figure 3D). The binding affinity is unperturbed for the D33E mutant (Figure 4C), containing a longer side chain while retaining the carboxylate functionality, whereas it is reduced by a modest factor of three for the D33A mutant (Figure 4C). Thus, the pair of sugar hydroxyl-carboxylate hydrogen bonds (Figure 3D) appear not to be critical for high affinity recognition. By

contrast, the binding affinity drops by 18-fold for the D33I mutant (Figure 4C), suggesting that bulky substitutions are not tolerated opposite the sugar hydroxyls of U9 in the complex.

Substitution of U by its 2′-OCH_3 U_m counterpart at the U9 position resulted in a 38-fold drop in binding affinity to La NTD (Figure 5D). This methyl substitution appears to contribute a steric block to recognition by D33 of La NTD, as the D33A mutant, in which aspartic acid is replaced by the smaller alanine side chain, resulted in only a 6-fold loss in binding affinity when U is replaced by U_m at the U9 position (Figure 5E).

Overall, it appears that U9 can be replaced by A and to a somewhat lesser extent by C, with its alignment stabilized by base-aromatic and sugar-aromatic stacking interactions. Both sugar hydroxyls of U9 are hydrogen bonded to D33, but these hydrogen-bonding interactions appear to be less critical than steric compatibility, as bulky substitutions on either the sugar hydroxyl or at amino acid position 33 dramatically reduce binding affinity.

Comparison with *T. brucei* La Mutants Targeted to UUU_{OH} 3′ Ends

Our electrophoretic mobility shift assays on the binding of mutants of human La NTD to the 9 nt UUU_{OH}-containing RNA are in agreement with the corresponding published data on the binding of protein point mutants of intact *T. brucei* La targeted to UUU_{OH} 3′ ends of yeast pre-$tRNA^{Arg}$ (Dong et al., 2004). These studies report that mutating the *T. brucei* equivalents of Y23, Y24, F35, and F55 of human La, each in turn to alanine, resulted in a greater than 100-fold reduction in binding activity, independently emphasizing the importance of these aromatic amino acids as specificity determinants for recognition and stabilization. Competition assays established that the *T. brucei* La equivalent of D33 discriminates by a factor of up to 100-fold between UUU_{OH} and UUU_P 3′ ends (Dong et al., 2004), and this discrimination is reduced to a factor of 18 for the *T. brucei* La equivalent of the D33A mutant in human La. These results are consistent with the incompatibility of positioning bulky and negatively charged groups on the 3′ -OH group.

Earlier Predictions of UUU_{OH} 3′ End Recognition by La NTD

Several predictions have been made regarding La NTD recognition of UUU_{OH} 3′ ends based on the NMR structures of human La and RRM1 motifs in the free state (Alfano et al., 2004), the NMR chemical shift perturbations in these two motifs on complex formation with U_{10} (Alfano et al., 2004), the X-ray structure of the *T. brucei* La motif (Dong et al., 2004), and the effect of mutations in full-length *T. brucei* La on binding to pre-tRNA (Dong et al., 2004). Both solution (Alfano et al., 2004) and crystallographic (Dong et al., 2004) efforts identified highly conserved surface patches, lined by aromatic residues within the La motif, that exhibited NMR chemical shift perturbations upon complex formation with U_{10} (Alfano et al., 2004) and adversely impacted binding affinity when mutated (Dong et al., 2004). In addition, mutational studies on the *T. brucei* equivalent of D33 identified this residue as critical for discrimination between UUU_{OH} and UUU_P ends, implying its involvement in 3′-OH end recognition (Dong et al., 2004). These conclusions, though qualitative in nature, receive support from our crystal structure of the La NTD-RNA complex (Figures 1–3). By contrast, the prediction based on NMR chemical shift perturbations that the two central b

strands of the four-stranded β sheet of the RRM1 motif are involved in UUU_{OH} 3′-end recognition (Alfano et al., 2004), is not supported by our crystal structure of the complex.

Comparison with Other RNA 3′ Overhang Recognition Complexes

Our structure addresses a long-standing challenge associated with the molecular basis for La NTD recognition and protection of the 3′ ends of nascent RNA transcripts. The splaying apart of individual U-U steps in the UUU_{OH} segment of the 9-mer RNA in the La NTD-RNA complex was unanticipated (Figure 2C), as was the observation that 3′ end recognition did not involve the exposed b sheet face of the RNA binding domain of RRM1.

The present structure of the La NTD-RNA 9-mer complex, which forms a 5 bp duplex with 5′-U and UUU_{OH}-3′ overhangs, can be compared with a recently solved PAZ-RNA 9 nt complex, which forms a 7 bp duplex with 2 nt 3′ overhangs (Ma et al., 2004). The conformation of the UUU_{OH}-3′ overhangs in the La NTD-RNA complex are very different from the 2 nt overhang in the PAZ-RNA complex, because the three uracils in the former complex are splayed apart (Figure 1E), whereas the 2 nt overhangs in the latter complex stack on each other and are accommodated in the same highly conserved pocket. In addition, the central uracil U8, in the La NTD-RNA complex, forms base-specific intermolecular hydrogen bonds (Figure 3B), whereas the intermolecular contacts involving the 2 nt, 3′ overhang are solely to the sugar-phosphate backbone in the PAZ-RNA complex.

Role of Splayed-Apart Bases in Protein-RNA Recognition

The UUU_{OH} 3′ end of the r(UGCUGUUUU$_{OH}$) 9 nt sequence is most likely flexible in the free state but becomes ordered upon complex formation with the La NTD. The three uridines in the UUU_{OH} segment are splayed apart in the complex, with each base forming specific hydrogen bonding and stacking interactions. Principles associated with splaying apart of a trinucleotide segment, with individual bases targeting specific recognition pockets, was first reported for the anticodon bases in tRNA-tRNA synthetase complexes (Rould et al., 1989; reviewed in Arnez and Moras, 1997) and later was extended to exposed tetraloop bases in viral nucleocapsid protein-ψ-RNA complexes (De Guzman et al., 1998; D'Souza and Summers, 2004).

RNA 5′ End Segment Primarily Targeted by the RRM1 Motif

We have also observed intermolecular contacts between the 5′ overhang base U1 (Figure 3E) and the adjacent G2-C3 step (Figures 3E and 3F), which is part of a duplex segment in the La NTD-RNA complex. These intermolecular contacts involve base edge (U1), sugar (G2), and phosphate backbone (G2-C3 step) of the RNA and are primarily targeted by side chains of the RRM1 motif (Figures 3E and 3F).

These contacts could be of importance because autoantigens such as Ro are known to stably associate with Pol III transcripts that resemble Y-RNAs, which have UUU_{OH} termini that overhang stems comprised of base-paired 5′ and 3′ regions (Wolin and Steitz, 1984). Thus, it is conceivable that there is a partitioning of RNA recognition capabilities for the two domains associated with the La NTD, in that the La motif primarily contributes to UUU_{OH}

3′ end recognition, whereas the RRM1 motif primarily contributes to recognition of the 5′ end segment of adjacent stem segments.

Interdomain Linker

Previous structural studies on the folding topologies of the free La motif (Alfano et al., 2004; Dong et al., 2004) and the free RRM1 motif (Alfano et al., 2004) provide no insights into the linker segment that connects these two motifs. Such linkers are likely to be flexible in the free state and rigidify upon complex formation with RNA, as observed for linkers connecting zinc finger domains (Wuttke et al., 1997). We can monitor the linker segment, spanning S92 to N106, in our crystal structure of the La NTD-RNA complex. The segment S92 to V100, which contains three nonadjacent prolines, is unstructured, whereas the segment T101 to N106 is part of a6, which spans from T101 to R111 (Figure S1). There are no direct contacts between the linker and the RNA in the La NTD-RNA complex. Nevertheless, the linker plays a key topological role, as anticipated (Dong et al., 2004), in orienting the La and RRM1 motifs, resulting in the generation of a basic cleft that is site-specifically targeted by the UUU_{OH} 3′ end overhang of nascent RNA Pol III transcripts (Figures 2B and 6A). We observe a direct hydrogen bonding contact between the side chain of Y23 in the La motif and N139 on the RRM1 motif, which appears to be critical for the relative alignment of these two motifs in the complex (Figure 2B).

Experimental Procedures

Protein and RNA Preparation

The PCR-amplified cDNA encoding human La NTD (1–194) was cloned into a modified pET28b vector that adds a Ulp1 protease-cleavable His_6-Smt3-tag at the N terminus. Point mutations in the La NTD construct were made with the QuikChange (Stratagene) system. The La NTD (residues 1–194 and an additional Ser at the N terminus) was expressed in *E. coli* strain BL21(DE3)pLysS (Novagen). Selenomethionine (SeMet)-substituted protein was expressed by growing cells in a M9 minimal media, using a standard protocol to saturate the biosynthetic pathway for methionine production (Doublié, 1997). Recombinant proteins were purified from the soluble fraction by Ni-chelating affinity column followed by His_6-Smt3 tag removal with Ulp1 protease and additional purification by sequential chromatography on monoQ, heparin, and Superdex 75 columns (Amersham). The wild-type, SeMet, and mutant La NTDs were prepared in a similar fashion.

RNA oligonucleotides were commercially synthesized (Dharma-con Research), deprotected, and purified by anion-exchange chromatography, followed by desalting. The 1:1 ratio of complex formation between La NTD and the 9 nt 5′-UGCUGUUUU-3′ was determined by analytical gel-filtration on a Superdex 75 column.

Crystallization and Data Collection

Crystals of SeMet-substituted La NTD-RNA 9 nt complex were grown by hanging drop vapor diffusion. Equal volumes of protein-RNA complex (1:1.2 ratio; protein concentration was 0.4 mM in 0.2 M NaCl, 25 mM Tris-HCl [pH 8.0], and 10 mM DTT) and reservoir solution (18% polyethylene glycol 3350, 0.2 M Li_2SO_4, 0.1 M KCl, and 0.1 M Na acetate

[pH 4.6]) were mixed and equilibrated at 4°C. Crystals typically grew as clusters of thin plates with dimensions of approximately 0.1 × 0.1 × ~0.01 mm after repeated seeding over a one-week period.

For data collection, crystals were flash frozen (100 K) in the above reservoir solution supplemented with 20% ethylene glycol. MAD (2.4 Å resolution) and native (1.85 Å resolution) data sets were collected on the SeMet La NTD-RNA 9-mer complex crystals on the X25 beamline at the National Synchrotron Light Source. The data were processed by HKL2000 (Otwinowski and Minor, 1997) by carefully adjusting the error modes to retain as much anomalous signal as possible. The crystal belonged to space group $P2_12_12$, with two protein-RNA complexes per asymmetric unit and an ~45% solvent content. Crystal and MAD data characteristics are summarized in Table 1.

Structure Determination and Refinement

The program SHELEXD (Schneider and Sheldrick, 2002) was used to locate eight out of ten selenium sites, and the program SHARP (De La Fortelle and Bricogne, 1997) was used to calculate MAD phases. The remaining two missing Se atoms are located within the structurally poorly defined N terminus. The phases were further improved by density modification and solvent flipping, assuming a solvent content of 45% with the SOLOMON program (Abrahams and Leslie, 1996). The electron density map was calculated in the CCP4 suite (CCP4, 1994). The RNA model and about 85% of the polypeptide backbone of one protein molecule were readily built manually by using the 5 base pair RNA duplex, the NMR structure of the human La RRM1 motif (Protein Database [PDB] ID: 1S79; Alfano et al., 2004), and the modified La motif coordinates from *T. brucei* (PDB ID: 1S29; Dong et al., 2004) as starting models. The second copy of the La NTD was next located by means of a molecular replacement. The MAD electron density map and the identified Se sites were used to validate the molecular replacement result. The initial experimental phases, based on MAD data to 2.8 Å resolution, were progressively replaced by model phases combined with data to a maximum resolution of 1.85 Å. The full model of the La NTD-RNA 9-mer complex was rebuilt with the program O (Jones et al., 1991) and refined by using the programs CNS (Brunger et al., 1998) and REFMAC (Murshudov et al., 1997) (CCP4, 1994). Refinement statistics are given in Table 1, and a portion of the electron density map corresponding to the UUU_{OH} segment of the complex is shown in Figure 2A. All residues are in allowed regions of the Ramachandran plot as evaluated by using PROCHECK from CCP4. The final model comprises two copies of La NTD, two molecules of 9 nt RNA, which pair to form a 5 base-pair duplex containing 5′-U and 3′-UUU_{OH} overhangs, three sulfate ions, and 455 water molecules.

Gel Electrophoretic Mobility Shift Binding Assays

Protein-RNA binding interactions were evaluated by electrophoretic mobility shift (gel shift) assay. 10 μl reactions containing 15 fmol of 5′-^{32}P-end-labeled RNA oligonucleotides and La NTD (or La NTD mutants) in 2.5-fold dilutions ranging from 2 μM to 0.52 nM concentrations were mixed in binding buffer (50 mM Tris-acetate, 50 mM K-acetate, 5 mM Mg-acetate, 10% glycerol, and 1 mM DTT; pH 8.0) and incubated for 10 min at room temperature. Reaction products were separated in 10% polyacrylamide gels carried out at a

constant voltage of 6 V cm^{-1} at 4°C in 50 mM Tris-acetate (pH 8.0) buffer containing 50 mM K-acetate and 5 mM Mg-acetate. RNA was visualized with a phosphorimager (Molecular Dynamics). Dissociation constants were determined graphically by plotting the fraction of bound RNA versus the concentration of protein (Irvine et al., 1991). The theoretical curves were fitted to the experimental data points by nonlinear least squares analysis with the equation:

$$\theta = \frac{(P_o + N_o + K_D) - \sqrt{\left((P_o + N_o + K_D)^2 - 4P_oN_o\right)}}{2N_o} \quad (1)$$

where θ is the fraction of bound RNA, N_o and P_o are total RNA and protein concentrations, respectively, and the K_D is the apparent dissociation constant. Gel shift assays were repeated for verification of dissociation constants.

Gel Electrophoretic Mobility Shift Competition Assays

Competition experiments were performed by a gel-shift assay with a constant concentration of La NTD (16 nM), 1.5 nM of 5′-^{32}P-end-labeled wild-type 9-nt RNA r(UGCUGUUUU), and a variable concentration of unlabeled 9-nt RNAs (0.5 nM to 20 μM). 20 μl reaction mixtures were incubated for 30 min at 4°C and resolved by electro-phoresis as described above. The apparent dissociation constants were determined graphically by plotting the fraction of bound RNA versus the concentration of competitor RNA. The theoretical curves were fitted to the experimental data points by nonlinear least squares analysis by using the equation (2) (Lin and Riggs, 1972) as follows:

$$\theta = \frac{(K_N + K_{rel}C_o + P_o + N_o) - \sqrt{\left((K_N + K_{rel}C_o + P_o + N_o)^2 - 4P_oN_o\right)}}{2N_o} \quad (2)$$

where θ is the fraction of bound radiolabeled RNA, K_{rel} is the ratio of the apparent dissociation constant of the labeled RNA (K_N) to the apparent dissociation constant of the competitor, and P_o, N_o, and C_o are the total concentrations of the protein, the radiolabeled RNA, and the competitor RNA, respectively.

The apparent dissociation constants are the average of at least two independent experiments with a standard error of approximately 20%–30%

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by funds from the Abby Rockefeller Mauze Trust and the Dewitt Wallace and Maloris Foundations to D.J.P. We thank Yuan Cheng for assistance with graphics. Drs. Elena Mossesova and Christopher Lima kindly provided the modified pET28b vector that adds an Ulp1 protease-cleavable His_6-Smt3-tag at the N terminus. We thank personnel at synchrotron beam lines X25 at the National Synchrotron Light Source at the Brookhaven National Laboratory and 19BM at the Advanced Photon Source (APS) at Argonne National Laboratory for their assistance. Use of the APS beamline was supported by the U.S. Department of Energy, Basic Energy

Sciences, Office of Science. D.J.P. is a member of the New York Structural Biology Center, supported in part by National Institutes of Health funds.

References

Abrahams JP, Leslie AGW. Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr. D Biol. Crystallogr. 1996; 52:30–42. [PubMed: 15299723]

Alfano C, Sanfelice D, Babon J, Keely G, Jacks A, Curry S, Conte MR. Structural analysis of cooperative binding by the La motif and central RRM domain of human La protein. Nat. Struct. Mol. Biol. 2004; 11:323–329. [PubMed: 15004549]

Ali N, Pruijn GJ, Kenan DJ, Keene JD, Siddiqui A. Human La antigen is required for the hepatitis C virus internal ribo-some entry site-mediated translation. J. Biol. Chem. 2000; 275:27531–27540. [PubMed: 10856291]

Allain FH, Gubser CC, Howe PW, Nagai K, Neuhaus D, Varani G. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. Nature. 1996; 380:646–650. [PubMed: 8602269]

Alspaugh MA, Tan EM. Antibodies to cellular antigens in Sjogren's syndrome. J. Clin. Invest. 1975; 55:1067–1073. [PubMed: 804494]

Arnez JG, Moras D. Structural and functional considerations of the aminoacylation reaction. Trends Biochem. Sci. 1997; 22:211–216. [PubMed: 9204708]

Brunger AT, Adams PD, Clore GM, De Lano WL, Gros P, Grosse-Kuntsleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 1998; 5:905–921.

CCP4 (Collaborative Computational Project, Number 4). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. 1994; 50:760–763.

Chakshumathi G, Kim SD, Rubinson DA, Wolin SL. La protein requirement for efficient pre-tRNA folding. EMBO J. 2003; 22:6562–6572. [PubMed: 14657028]

Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature. 1993; 364:412–420. [PubMed: 8332212]

D'Souza V, Summers MF. Structural basis for packaging the dimeric genome of Moloney murine leukemia virus. Nature. 2004; 431:586–590. [PubMed: 15457265]

De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. Structure of the HIV-1 nucleocapsid protein bound to the SL3 c-RNA recognition element. Science. 1998; 279:384–388. [PubMed: 9430589]

De La Fortelle E, Bricogne G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous and multiwavelength anomalous diffraction methods. Methods Enzymol. 1997; 276:472–494.

Dong G, Chakshusmathi G, Wolin SL, Reinisch KM. Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch. EMBO J. 2004; 23:1000–1007. [PubMed: 14976553]

Doublié S. Preparation of selenomethionyl proteins for phase determination. Methods Enzymol. 1997; 276:523–530. [PubMed: 9048379]

Fan H, Goodier JL, Engelke DR, Maraia RJ. 5′ processing of tRNA precursors can be modulated by the human La antigen phosphorylation. Mol. Cell. Biol. 1998; 18:3201–3211. [PubMed: 9584161]

Fourmy D, Guittet E, Yoshizawa S. Structure of prokaryotic SECIS mRNA hairpin and its interaction with elongation factor SelB. J. Mol. Biol. 2002; 324:137–150. [PubMed: 12421564]

Gajiwala KS, Burley SK. Winged helix proteins. Curr. Opin. Struct. Biol. 2000; 10:110–116. [PubMed: 10679470]

Handa N, Nureki O, Kurinoto K, Kim I, Sakamoto H, Shimura Y, Muto Y, Yokoyama S. Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. Nature. 1999; 398:579–585. [PubMed: 10217141]

Holcik M, Korneluk RG. Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosomal entry site element: role of La autoantigen in XIAP translation. Mol. Cell. Biol. 2000; 20:4648–4657. [PubMed: 10848591]

Irvine D, Tuerk C, Gold L. SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. J. Mol. Biol. 1991; 222:739–761. [PubMed: 1721092]

Jacks A, Babon J, Kelly G, Manolaridis I, Cary PD, Cury S, Conte MR. Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. Structure. 2003; 11:833–843. [PubMed: 12842046]

Jones TA, Zou JY, Cowan SW, Kjeldgaard GJ. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A. 1991; 47:110–119. [PubMed: 2025413]

Kenan DJ, Tsai DS, Keene JD. RNA recognition: towards identifying determinants of specificity. Trends Biochem. Sci. 1991; 16:214–220. [PubMed: 1716386]

Kenan DJ, Keene JD. La gets its wings. Nat. Struct. Mol. Biol. 2004; 11:303–305. [PubMed: 15048103]

Lin S-Y, Riggs AD. lac repressor binding to non-operator DNA: detailed studies and comparison of equilibrium and rate competition methods. J. Mol. Biol. 1972; 72:671–690. [PubMed: 4573844]

Ma JB, Ye K, Patel DJ. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature. 2004; 429:318–322. [PubMed: 15152257]

Maraia RJ, Intine RV. Recognition of nascent RNA by the human La antigen: conserved and divergent features of structure and function. Mol. Cell. Biol. 2001; 21:367–379. [PubMed: 11134326]

Mattioli M, Reichlin M. Heterogenity of RNA protein antigens reactive with sera of patients with systemic lupus erytematosus. Descriptions of a cytoplasmic nonribosomal antigen. Arthritis Rheum. 1974; 17:421–429. [PubMed: 4212051]

Messias AC, Sattler M. Structural basis of single-stranded RNA recognition. Acc. Chem. Res. 2004; 37:279–287. [PubMed: 15147168]

Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 1997; 53:240–255. [PubMed: 15299926]

Ohndorf UM, Steegborn C, Kniff R, Sondermann P. Contribution of the individual domains in human La protein to its RNA 3′-end binding activity. J. Biol. Chem. 2001; 276:27188–27196. [PubMed: 11342556]

Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature. 1994; 372:432–438. [PubMed: 7984237]

Pannone BK, Kim SD, Noe DA, Wolin SL. Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein. Genetics. 2001; 158:187–196. [PubMed: 11333229]

Pudi R, Srinivasan P, Das S. La binding protein at the GCAC site near the initiator AUG facilitates the ribosomal assembly on the hepatitis C virus RNA to influence internal ribosomal entry site-mediated translation. J. Biol. Chem. 2004; 279:29879–29888. [PubMed: 15138264]

Rould MA, Perona JJ, Soll D, Steitz TA. Structure of E. coli glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. Science. 1989; 246:1135–1142. [PubMed: 2479982]

Schneider TR, Sheldrick GM. Substructure solution with SHELEXD. Acta Crystallogr. D. Biol. Crystallogr. 2002; 58:1772–1779. [PubMed: 12351820]

Schwartz T, Rould MA, Lowenhaupt K, Herbert A, Rich A. Crystal structure of the Z-alpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. Science. 1999; 284:1841–1845. [PubMed: 10364558]

Selmer M, Su X-D. Crystal structure of an mRNA-binding fragment of Moorella thermoacetica elongation factor SelB. EMBO J. 2002; 21:4145–4155. [PubMed: 12145214]

Stefano JE. Purified lupus antigen La recognizes an oligouridylate stretch common to the 3′ termini of RNA polymerase III transcripts. Cell. 1984; 36:145–154. [PubMed: 6607117]

Varani G, Nagai K. RNA recognition by RNP proteins during RNA processing. Annu. Rev. Biophys. Biomol. Struct. 1998; 27:407–445. [PubMed: 9646873]

Wolin SL, Steitz JA. The Ro small cytoplasmic ribonu-cleoproteins: identification of the antigenic protein and its binding site on the Ro RNAs. Proc. Natl. Acad. Sci. USA. 1984; 81:1996–2000. [PubMed: 6201849]

Wolin SL, Cedrervall T. The La protein. Annu. Rev. Biochem. 2002; 71:375–403. [PubMed: 12045101]

Wuttke DS, Foster MP, Case DA, Gottesfeld JM, Wright PE. Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. J. Mol. Biol. 1997; 273:183–206. [PubMed: 9367756]

Xue D, Rubinson DA, Pannone BK, Yoo CJ, Wolin SL. U snRNP assembly in yeast involves the La protein. EMBO J. 2000; 19:1650–1660. [PubMed: 10747032]

Yoo CJ, Wolin SL. The yeast La protein is required for the 3′ endonucleolytic cleavage that matures tRNA precursors. Cell. 1997; 89:393–402. [PubMed: 9150139]

Zheng N, Fraenkel E, Pabo CO, Pavletich NP. Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F–DP. Genes Dev. 1999; 13:666–674. [PubMed: 10090723]

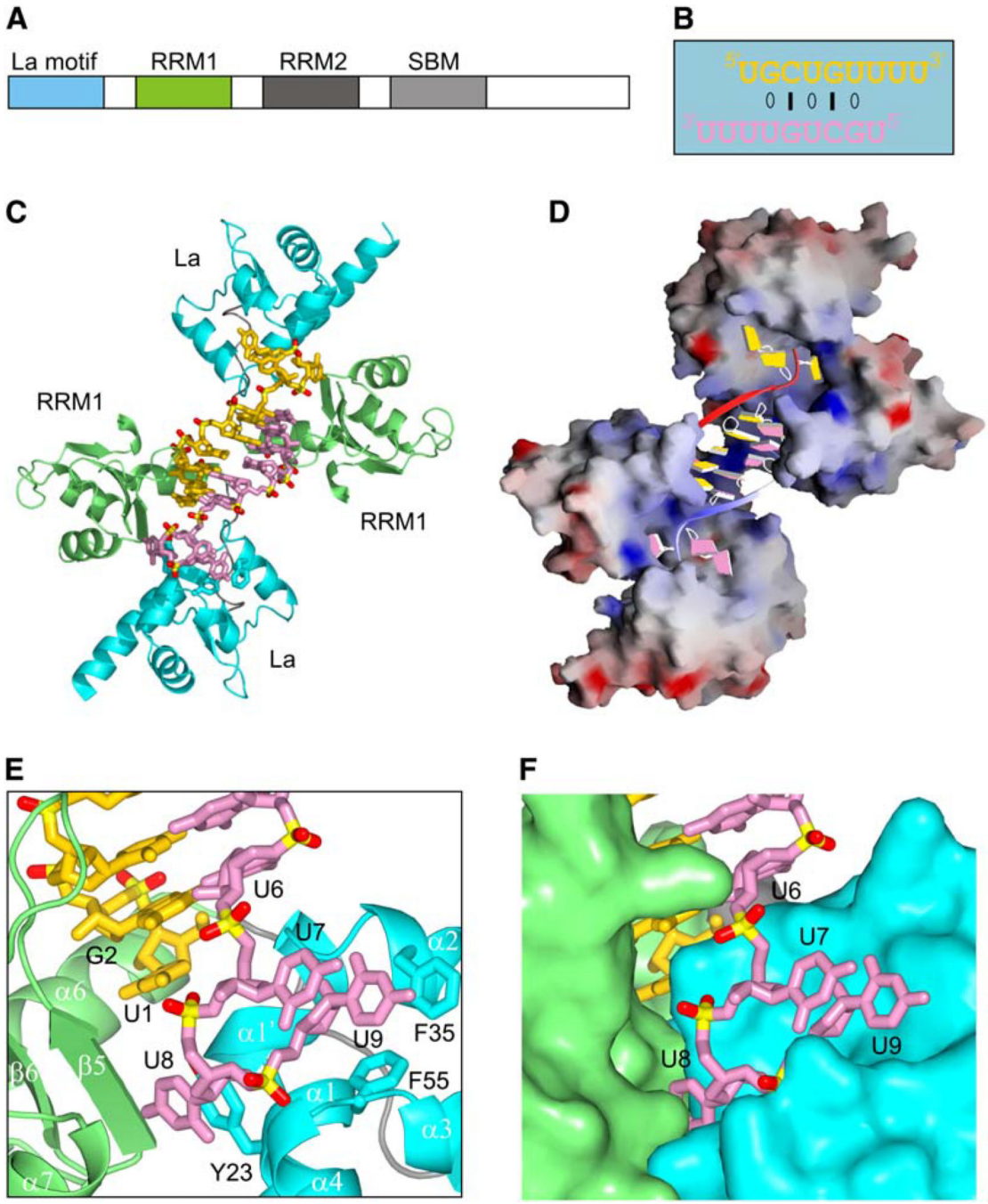

Figure 1. Structure of La NTD Bound to the r(U1-G2-C3-U4-G5-U6-U7-U8-U9) Sequence

(A) The domain architecture of the La protein.

(B) The G2-C3-U4-G5-U6 segment of the 9 nt RNA forms a self-complementary duplex in the complex.

(C) Ribbon and stick representation of the complex containing two La NTD domains and two 9 nt RNAs related by noncrystallographic 2-fold symmetry axis. The La (E8-R91) and RRM1 (D107-Y188) motifs are colored cyan and green, respectively, whereas the connecting linker (R92-N106) is colored gray. The two RNA strands are colored gold and

magenta, with the backbone phosphorus atoms colored yellow and the nonbridging phosphate oxygens colored red. The U7-U8-U9 segment is sequestered at the interface between the La and RRM1 motifs.

(D) An electrostatic view of the complex generated by using the GRASP program. The RNA is shown in a slab representation, with blue and red patches corresponding to basic and acidic segments on the protein surface. The bases on the two symmetry-related RNA strands are colored gold and magenta.

(E) A stick and ribbon view of the intermolecular contacts between the U1-G2 and U6-U7-U8-U9 elements of the 9-mer RNA and La (in cyan) and RRM1 (in green) motifs of La NTD in the complex. The nucleotides of the RNA and the α helices and β strands of the protein are labeled. The U7 and U9 residues are directed toward the La motif, whereas the U8 residue is directed toward the RRM1 motif.

(F) The same view as in (E) except for the protein is in a surface representation, with the same color code as in (C).

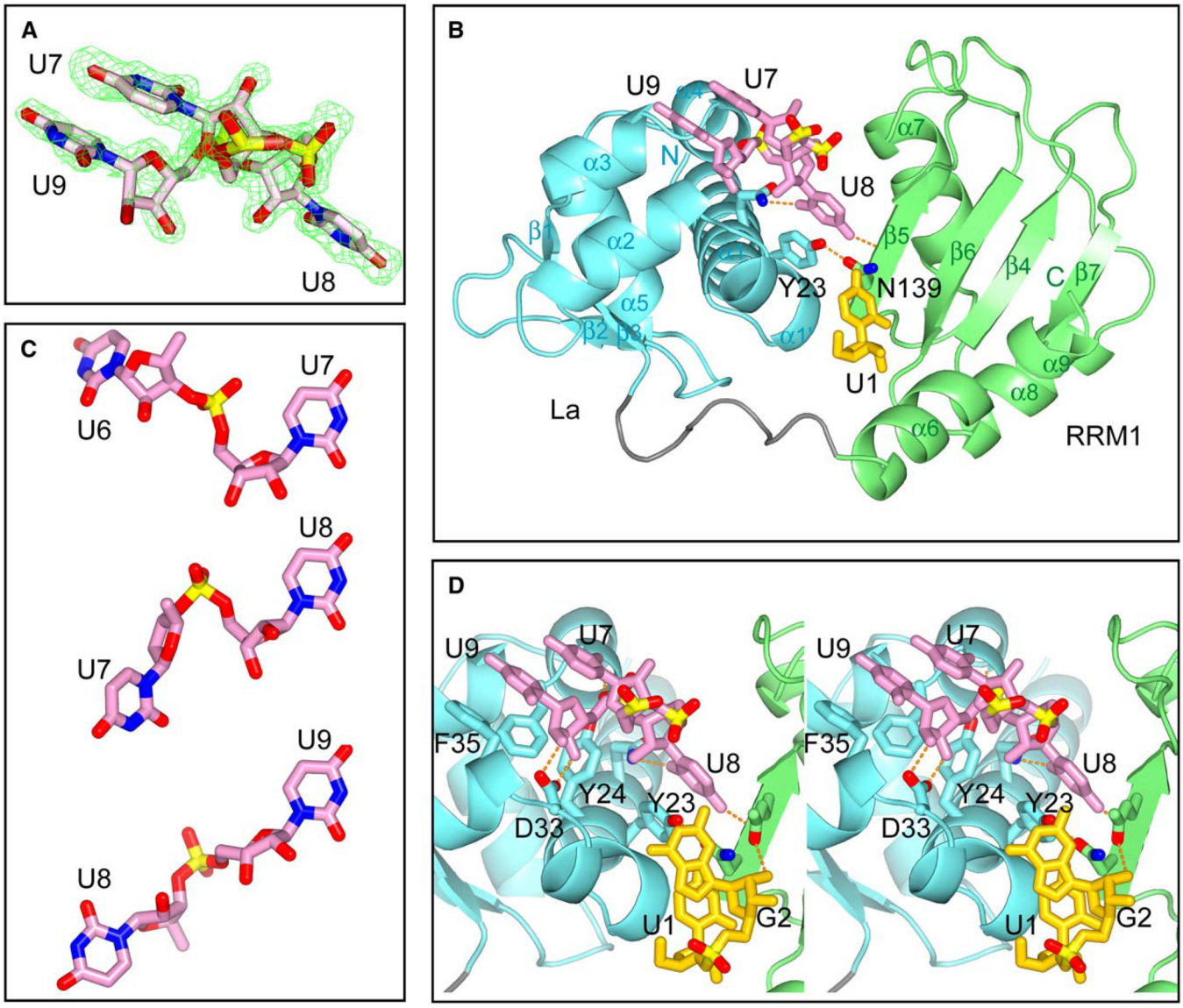

Figure 2. Overview of Protein-RNA Interactions in the Complex

(A) Simulated annealing omit map of the 3′-UUU_{OH} segment of the 9-mer in the complex. Electron density map (contoured at 2.5σ) calculated with a starting temperature of 5000 K and excluding UUU_{OH} from the model.

(B) A view of the alignment of the U1 and U7-U8-U9 RNA 9-mer segments relative to the La and RRM1 motifs of the La NTD in the complex. A single intermolecular hydrogen bond was detected between Y23 on the La motif (in cyan) and N139 on the RRM1 motif (in green) across the interface between these two motifs in the complex. The figure labels the a helices and b strands within the La NTD and also highlights the bridging intermolecular hydrogen bonding contacts from U8 to both the La and RRM1 motifs.

(C) A comparison between the extended conformations of the U6-U7, U7-U8, and U8-U9 steps in the complex.

(D) Stereo view of the protein-RNA interface highlighting intermolecular contacts between 5′-U1-G2 and U7-U8-$U9_{OH}$ 3′ ends and the La NTD. Stacking interactions between bases

and aromatic residues as well as hydrogen-bonding contacts involving U8 and U9 are highlighted.

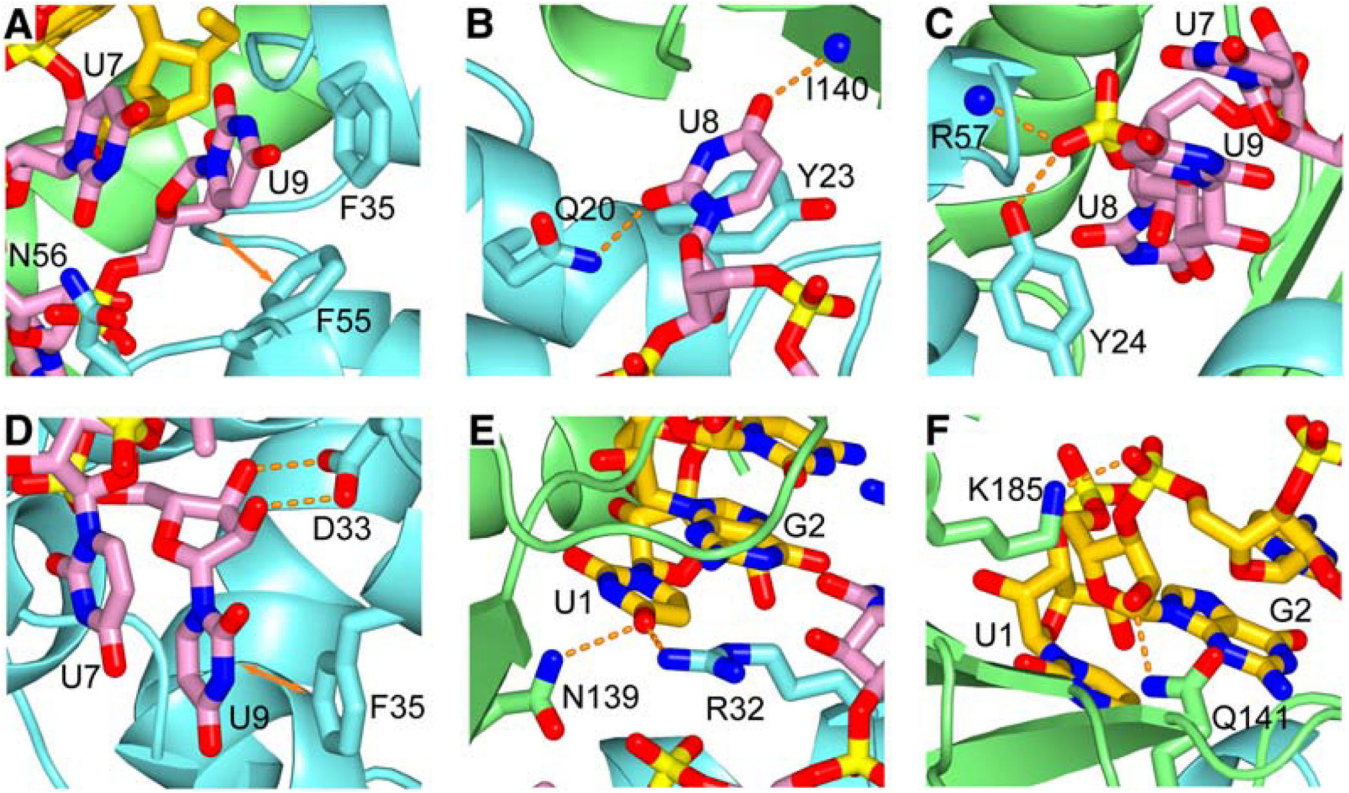

Figure 3. Details of RNA Architecture and Protein-RNA Contacts in the Complex

(A) Positioning of the O^2 of U7 with respect to the side chain amide of N56 of the La motif. Note stacking of the sugar ring of U9 on the aromatic ring of F55.

(B) Hydrogen bonding of the O^2 of U8 with the side chain amide of Q20 of the La motif and the O^4 of U8 with the backbone amide of I140 of the RRM1 motif. U8 also stacks with Y23 of the La motif.

(C) Hydrogen bonding between the nonbridging phosphate oxygen linking the U8-U9 step and the backbone amide of R57 and the side chain hydroxyl of Y24, both from the La motif.

(D) Hydrogen bonding between the 2′-OH and 3′-OH groups of U9 and the carboxylate oxygens of the side chain D33 of the La motif. U9 also stacks with F35 of the La motif.

(E) Hydrogen bonding of the O^4 of U1 with the side chains of R32 of the La motif and N139 of the RRM1 motif.

(F) Hydrogen bonding of the 2′-OH of G2 with the side chain of Q141 of the RRM1 motif and hydrogen bonding of the nonbridging phosphate oxygen linking the G2-C3 step and the side chain of K185 of the RRM1 motif.

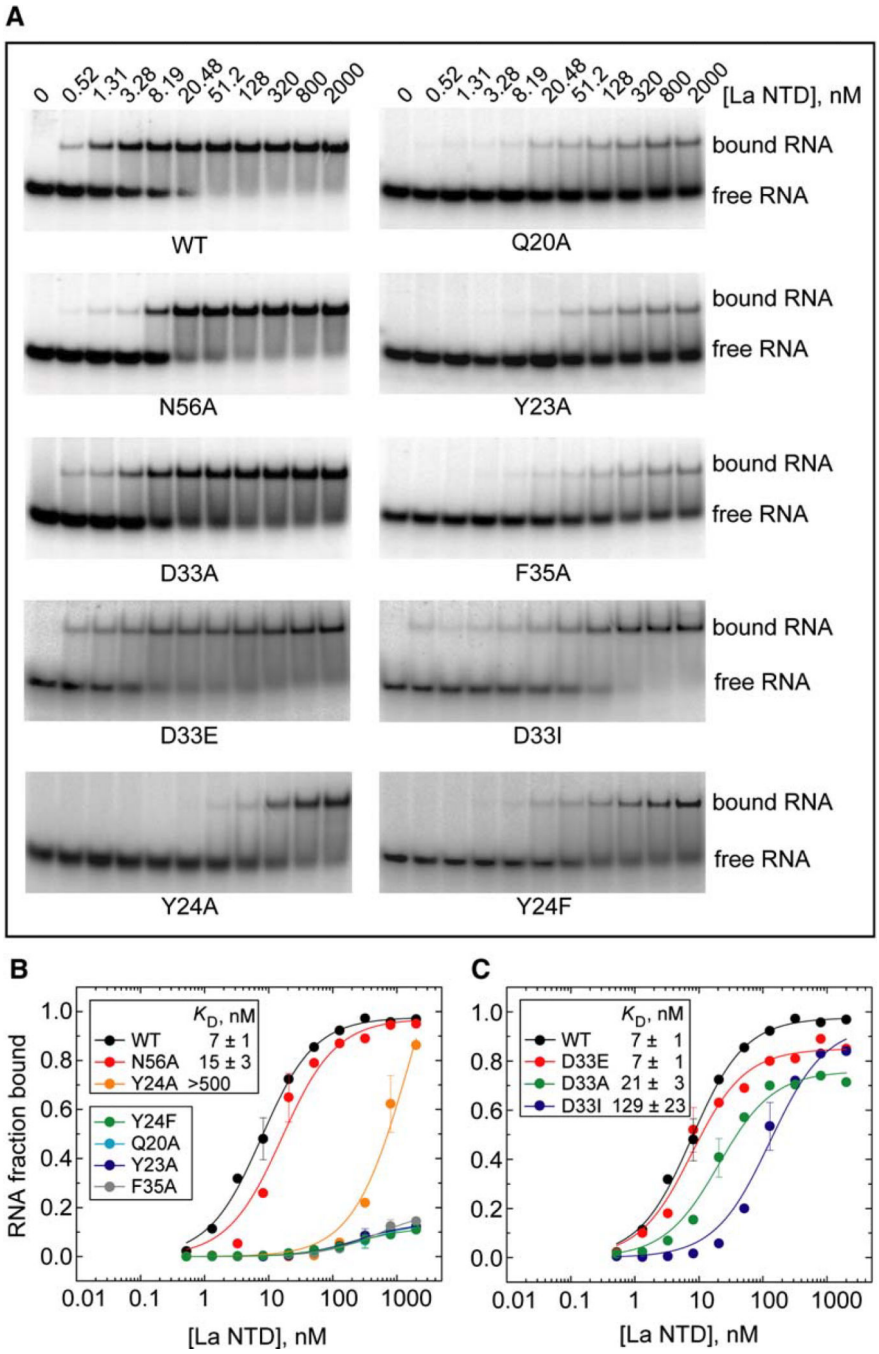

Figure 4. Gel Electrophoretic Mobility Shift Assays of Binding of Wild-Type and Mutant La NTD to the ^{32}P-labeled 9 nt 5′-UGCUG UUUU-3′ RNA

(A) Representative electrophoretic mobility gel shift assay data for binding of r(U-G-C-U-G-U-U-U-U) to La NTD point mutants. The specific La NTD point mutation under consideration is shown below each gel. Protein concentrations (nM) used are indicated above each lane and were identical for each protein construct.

(B) Plots of radiolabeled RNA fraction bound as monitored by gel electrophoretic mobility shift binding assays against La NTD concentration. The polar La NTD point mutants include N56A and Q20A, whereas the aromatic point mutants include Y24A, Y24F, Y23A, and

F35A. Solid lines indicate theoretical values of the best least squares analysis according to equation 1 in the Experimental Procedures section. Estimated dissociation constants for wild-type and mutant La NTD-rUGCUGUUUU 9-nt RNA complexes are listed in the boxes. Data represent mean ± standard deviation (SD) for at least two independent experiments.

(C) Corresponding plots for complex formation between either wild-type or D33 single point mutants (D33E, D33A, and D33I) of La NTD and the 5′-UGCUGUUUU-3′ sequence. Data represent mean ± SD for at least two independent experiments.

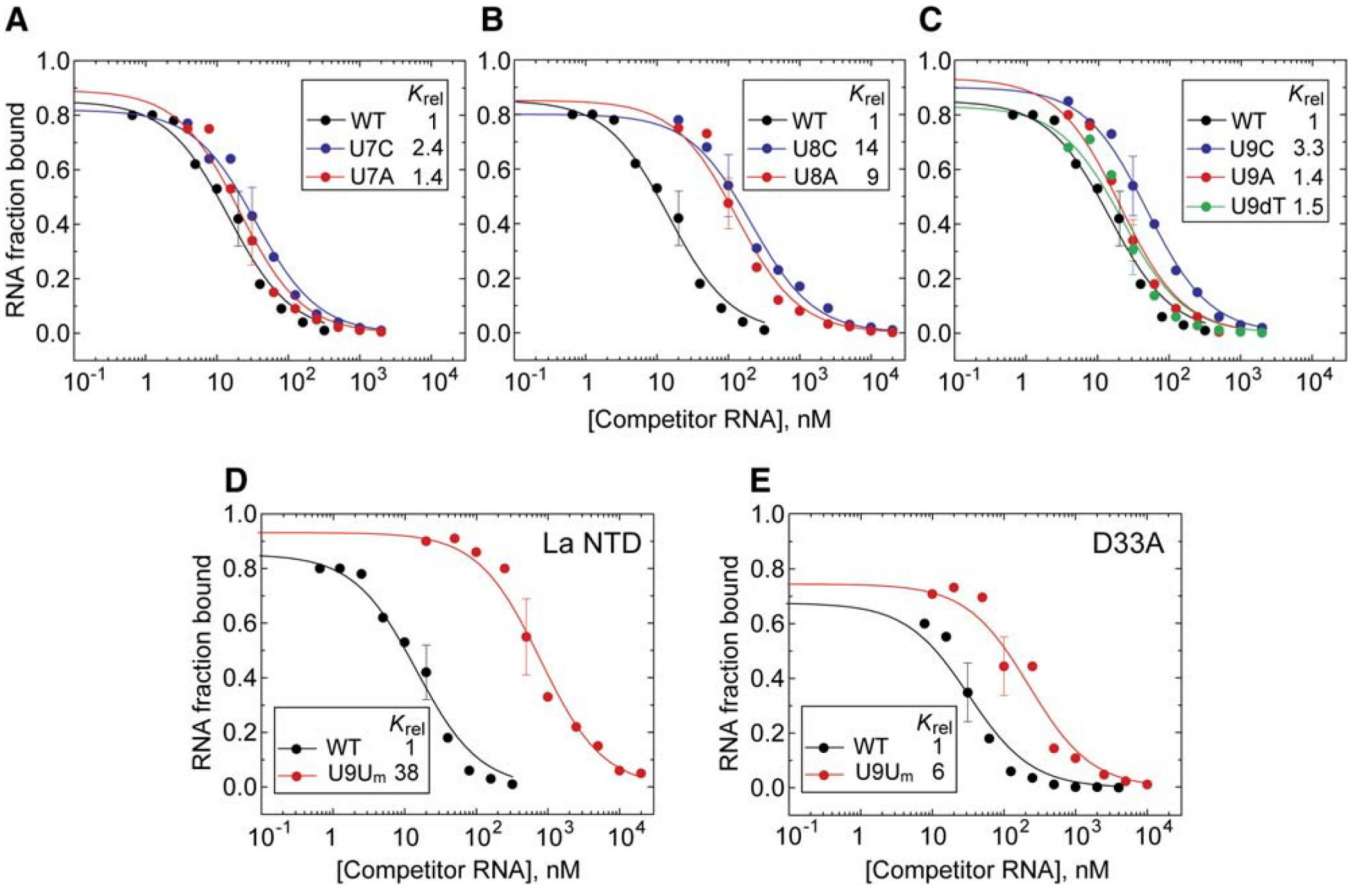

Figure 5. Plots of Gel Electrophoretic Mobility Shift Competition Assays Monitoring Addition of Unlabeled 9 nt RNAs to a Complex of Either Wild-Type La NTD or D33A Mutant and ^{32}P-Labeled 9 nt 5′-UGCUGUUUU-3′ RNA

(A) Plots of competition assays with unlabeled competitor 9 nt RNAs, which include wild-type sequence as a control and U7C and U7A substitutions.

(B) Plots of competition assays with unlabeled competitor 9 nt RNAs containing U8C and U8A substitutions.

(C) Plots of competition assays with unlabeled competitor 9 nt RNAs containing U9C, U9A, and U9dT substitutions.

(D) Complex contains wild-type La NTD. Plots of competition assays with unlabeled competitor 9-nt RNAs, which include the wild-type RNA sequence as a control and its U9U$_m$ substitution. U$_m$ stands for 2′-OmeU.

(E) Complex contains D33A La NTD mutant. Plots of competition assays for the wild-type RNA sequence and its U9U$_m$ substitution.

The plots monitor a fraction of radiolabeled 9 nt wt-RNA bound to La LTD or D33A mutant against concentrations (nM) of unlabeled competitor 9 nt RNA containing either wild-type sequence, single point mutations in the 3′-UUU$_{OH}$ segment, or U9U$_m$ substitution. Solid lines plot theoretical values of the best least squares analysis according to equation (2) (Lin & Riggs, 1972). K_{rel} indicates the relative binding strength and is expressed as a ratio of the apparent dissociation constant (K_D) measured for the La NTD mutant 9 nt RNA complex to the K_D measured for the La NTD-wild-type RNA. Estimated K_{rel} values are listed in the

box. The gel shift data are plotted in Figures S2 and S3. Data represent the mean ± SD for at least two independent experiments.

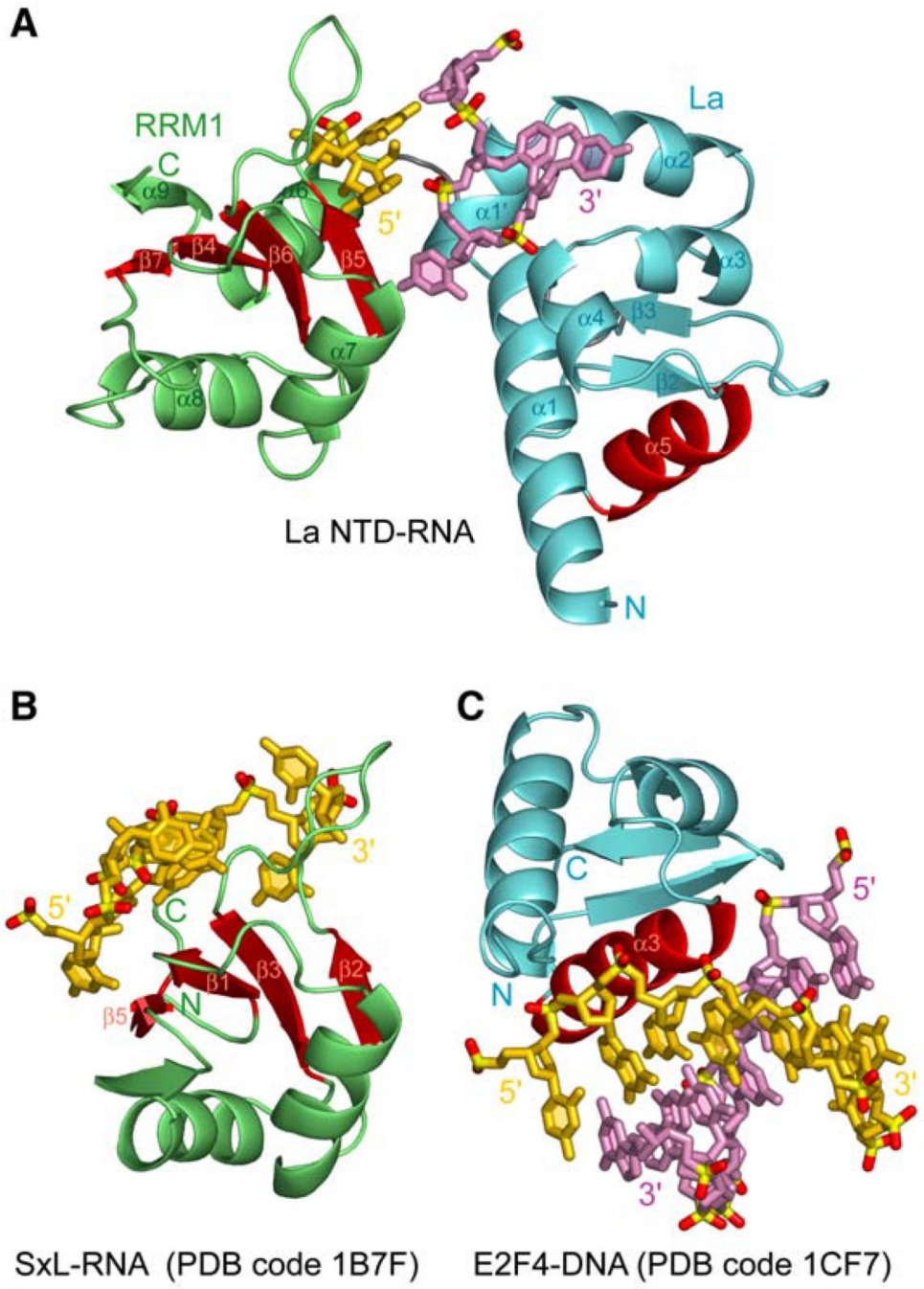

Figure 6. Comparison of Protein-RNA Interactions in the La NTD-RNA Complex with Those Reported for Classical RRM Motif-ssRNA and Winged-Helix Motif-DNA Duplex Complexes

(A) Protein-RNA recognition in the La NTD-9-mer RNA complex. Helix α5 within the La motif and β sheet composed of β7, β4, β6, and β5 in the RRM1 motif are colored red.

(B) Protein-ssRNA recognition in the Sex-lethal protein-*tra* mRNA precursor complex (PDB code: 1B7F; Handa et al., 1999). The four-stranded β sheet of the first RRM of the Sex-lethal (SxL) protein, which interacts with a segment of ssRNA, is colored red.

(C) Protein-DNA recognition in the heterodimeric cell cycle transcription factor E2F–DP bound to duplex DNA (PDB code: 1CF7; chains A, C, and D; Zheng et al., 1999). The α helix of the winged-helix motif of E2F domain, which interacts with a segment of duplex DNA, is colored red.

Table 1

Crystallographic Data Collection and Analysis

Data Collection	**MAD**			**Native**
Space group	**$P2_12_12$**			**$P2_12_12$**
Unit cell dimensions	a = 153.10 Å			a = 156.21 Å
	b = 53.90 Å			b = 55.03 Å
	c = 56.75 Å			c = 57.78 Å
	α = β = γ = 90°			α= β = γ = 90°
Wavelength (Å)	0.9798	0.9795	0.9611	1.1000
Resolution range (Å) (last shell)	20–2.58 (2.67–2.58)	20–2.40 (2.49–2.40)	20–2.70 (2.80–2.70)	40–1.85 (1.92–1.85)
Redundancy	7.6	7.6	6.8	5.3
Unique observations	26,887	35,083	23,001	42,919
Completeness (%) (last shell)	100.0 (100.0)	100.0 (100.0)	99.9 (100.0)	99.4 (99.4)
R_{sym} (%)[a] (last shell)	14.2 (61.2)	15.3 (58.5)	15.0 (60.4)	10.3(49.6)
Phasing power	0.660	1.108	0.863	
Overall figure of merit (acentric/centric)		0.44 (0.28)		
Refinement				
Resolution range (Å)				20~1.85
$R_{factor}(R_{free})$(%)[b,c]				20.0 (24.5)
Rms bond length deviations (Å)				0.011
Rms angles deviations (°)				1.662
Average temperature factor ($Å^2$)				24.3
Percent core (allowed) in Ramachandran plot				93.5 (5.9)

[a] R_{sym} is the unweighted R value on I between symmetry mates.

[b] $R_{factor} = \sum hkl||F_{obs}(hkl)| - |F_{calc}(hkl)||\sum hkl|F_{obs}(hkl)|$.

[c] R_{free} is the crossvalidation R factor for 5% of reflections against which the model was not refined.