The EMBO Journal (2004) 23, 1000–1007 | © 2004 European Molecular Biology Organization | All Rights Reserved 0261-4189/04
www.embojournal.org

THE EMBO JOURNAL

Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch

Gang Dong[1], Ghadiyaram Chakshusmathi[1], Sandra L Wolin[1,2] and Karin M Reinisch[1,*]

[1]Department of Cell Biology, Yale University School of Medicine, New Haven, CT, USA and [2]Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT, USA

The La protein is a ubiquitous nuclear phosphoprotein that recognizes the 3′ uridylates found in all newly synthesized RNA polymerase III transcripts. La binding stabilizes these transcripts from exonucleases and may also assist their folding. Here we present the first structural insights into how the La protein specifically interacts with its RNA substrates. The most conserved region of the La protein is the La motif, a domain also found in several other RNA-binding proteins. We have determined the structure of the La motif from the *Trypanosoma brucei* La protein to 1.6 Å resolution (PDB code 1S29). The La motif adopts a winged helix–turn–helix architecture that has a highly conserved patch of mainly aromatic surface residues. Mutagenesis experiments support a critical role for this patch in RNA binding and show that it partly determines binding specificity for RNAs ending in 3′ hydroxyl, a defining characteristic of the La protein. These findings reveal that the La motif is essential for high-affinity binding and also contributes to specificity.

The EMBO Journal (2004) **23**, 1000–1007. doi:10.1038/sj.emboj.7600115; Published online 19 February 2004
Subject Categories: structural biology; RNA
Keywords: La autoantigen; RNA-binding protein; RNA–protein interactions; tRNA processing; winged helix domain

Introduction

The La protein was first described as an autoantigen in patients with the rheumatic diseases systemic lupus erythematosus and Sjogren's syndrome (Mattioli and Reichlin, 1974; Alspaugh and Tan, 1975). Although first characterized in humans, La homologs have since been found in all eucaryotes including trypanosomes, yeast, and plants (see Maraia and Intine, 2001; Wolin and Cedervall, 2002).

La proteins from humans, *Drosophila*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Trypanosoma brucei* have been characterized (Chambers *et al*, 1988; Bai *et al*, 1994; Yoo and Wolin, 1994; Van Horn *et al*, 1997; Marchetti *et al*, 2000; Wolin and Cedervall, 2002). La is localized primarily in the nucleus, where it binds nascent RNA polymerase III transcripts. RNAs bound by La include pre-tRNAs, 5S rRNAs, U6 RNA, and SRP RNA (see Wolin and Cedervall, 2002). In all of these RNAs, La specifically recognizes the 3′ UUU_{OH} as part of its binding site (Stefano, 1984; Wolin and Cedervall, 2002). In the yeast *S. cerevisiae*, La also binds nascent RNA polymerase II transcripts such as U1, U2, U3, U4, and U5 snRNA that have been partially processed to end in polyuridine (Kufel *et al*, 2000; Xue *et al*, 2000). One characteristic feature of La is its preference for a hydroxyl over a phosphate group at the 3′ terminus. This may allow La to distinguish intact nascent polymerase III transcripts from degraded RNA molecules with 3′ terminal phosphates and to discriminate between 3′ terminal and internal polyuridine stretches (Stefano, 1984; Long *et al*, 2001). So far, the best-established role for La is to protect newly synthesized RNAs from exonucleases, with various consequences for different transcripts. For pre-tRNAs, La is required for the normal endonucleolytic removal of the 3′ trailer (Yoo and Wolin, 1997) and may also be important for their 5′ maturation (Fan *et al*, 1998). For certain pre-tRNAs, binding by La facilitates correct folding (Chakshusmathi *et al*, 2003). For other transcripts, La binding facilitates their assembly into functional ribonucleoprotein particles (Pannone *et al*, 1998; Xue *et al*, 2000). Although La has also been proposed to be a transcription factor for RNA polymerase III (Gottlieb and Steitz, 1989; Maraia, 1996), it is not required for transcription in human, *Xenopus*, or yeast cell extracts (Yoo and Wolin, 1997; Lin-Marq and Clarkson, 1998; Chong *et al*, 2001).

Like many RNA-binding proteins, La is modular. An N-terminal 'La motif' (LM) is followed by either one or two RNA recognition motifs (RRMs, also referred to as RBDs or RNP domains) (see Varani and Nagai, 1998; Wolin and Cedervall, 2002). The size of La varies from ~32 kDa in yeasts, where La has only one RRM, to 50k Da in humans, where La has two RRMs. The most C-terminal region of La is highly charged and is known to be phosphorylated (Fan *et al*, 1997; Long *et al*, 2001). Whereas sequence identity breaks down in the RRM and C-terminal regions of La, the LM is conserved across species. Human La, for example, has 29% sequence identity with the *S. cerevisiae* and 40% identity with the *T. brucei* N-terminal domain. In all characterized La proteins, the LM is at the N-terminus (Wolin and Cedervall, 2002).

The LM is also found in a number of other proteins that are otherwise structurally unrelated to La. Two of these proteins, the closely related *S. cerevisiae* proteins Slf1p and Sro9p, are polyribosome associated and may function in translation (Sobel and Wolin, 1999). The other characterized LM protein, p43 from the ciliate *Euplotes aediculatus*, is a component of telomerase (Aigner *et al*, 2003). In contrast to the La protein, p43 lacks sequence specificity for the 3′ UUU_{OH} (Aigner *et al*, 2003).

Both the LM and the central RRM are necessary for polyuridine specific binding (Goodier *et al*, 1997; Jacks *et al*, 2003).

*Correspondence. Department of Cell Biology, Sterling Hall of Medicine, Yale University School of Medicine, 333 Cedar St, New Haven, CT 06520-8002, USA. Tel.: +1 203 785 6469; Fax: +1 203 785 7446; E-mail: karin.reinisch@yale.edu

Received: 21 November 2003; accepted: 16 January 2004; published online: 19 February 2004

The role of the second RRM that is present in some species is not understood. It does not appear to contribute to binding polyU (Goodier *et al*, 1997; Ohndorf *et al*, 2001; Jacks *et al*, 2003). Moreover, in the solution structure of this domain (Jacks *et al*, 2003), the canonical RNA-binding surface is acidic, suggesting that it may not interact with RNA at all. Possibly then, this domain is a module for interacting with other proteins.

Here we begin to address how the La protein specifically interacts with its RNA targets. We present the 1.6 Å crystal structure of the LM domain from the *T. brucei* La protein. The LM adopts a winged helix–turn–helix (wHTH) fold present in many transcription regulatory proteins (Gajiwala and Burley, 2000). An aromatic patch on the domain surface is strictly conserved in all LMs, and structure-based mutagenesis studies show that residues in this patch are critical for high-affinity RNA binding and 3′ end recognition.

Results and discussion

Overview of the structure: the La motif is a variation of the winged helix–turn–helix fold

To obtain crystals of the La protein, we screened constructs based on the *T. brucei*, *S. pombe*, *S. cerevisiae*, and human La proteins. Crystals were obtained for the LM domain from *T. brucei*. The LM construct used consists of residues 1–89 of La and an additional Gly-Ser-His at the N-terminus. Although these three residues are not part of the *T. brucei* sequence, they are essential in forming crystal contacts. The structure with 100 water molecules has been refined to 1.6 Å with $R = 23.2\%/R_{free} = 25.0\%$ and excellent geometry (Table I).

Table I Summary of crystallographic data and refinement

Data collection	
Space group	$P2_12_12_1$
Cell constants (Å)	$a = 32.926$, $b = 53.833$, $c = 53.912$
Wavelength (Å)	0.9796
Resolution (Å)	1.6
Last shell (Å)	1.65–1.59
No. of unique reflections (last shell)	24 672 (2382)
Redundancy (last shell)	7.0 (4.1)
R_{sym} (last shell) (%)	3.2 (6.0)
I/sigma (last shell)	41 (21)
Completeness (last shell) (%)	99.5 (95.7)
Figure of merit (last shell)	0.45 (0.40)
Phasing power (last shell)	2.52 (1.78)
Refinement	
Resolution (Å)	20.0–1.6
R_{work}/R_{free} (%)	23.2 (25.0)
rmsd bond lengths (Å)	0.00526
rmsd bond angles (deg)	1.169
rmsd improper angles (deg)	0.76
rmsd dihedral angles (deg)	21.0
Residues in favored Ramachandran region	91
Residues in additionally allowed region	1
Residues in generously or disallowed region	0
B value rmsd (main/side)	1.0/2.2
No. of water molecules	100

$R_{sym} = \sum\sum_j |I_j - \langle I \rangle| / \sum I_j$.
Phasing power $= [\sum|F_+ - F_-|^2 / \sum|E|^2]^{1/2}$, where the denominator is the lack of closure error.
$R_{work,\ free} = \sum||F_{obs}| - |F_{calc}|| / |F_{obs}|$ for the reflections used in refinement (work) and 10% of reflections held aside (free).

The LM domain is mostly α-helical, consisting of six helices and a three-stranded antiparallel sheet (Figure 1A). One long α-helix (H1) and two shorter ones (H1′ and H2) precede the first β-strand (B1). Helices H3, H4, and H5, arranged to form a 'C' shape, reside between β-strands B1 and B2, and a six-residue loop links B2 and B3. Overall, the LM adopts a wHTH fold found in both eucaryotic and procaryotic transcription factors (Figure 1B) (Gajiwala and Burley, 2000). Two of the three N-terminal helices and H4 in the LM may be viewed as insertions into the wHTH topology. The wHTH transcription factors typically contain a three-helix bundle and a three-stranded antiparallel β-sheet. H1 precedes B1; H2 and H3 reside between B1 and B2; and a loop, the wing, links B2 and B3. In some wHTH proteins, a second wing follows B3. Winged HTH domains have different functions. In *Fok*I restriction endonuclease and replication protein A, for example, the wHTH is involved in protein–protein interactions (Wah *et al*, 1997; Mer *et al*, 2000). In many transcriptional regulators, it is involved in binding dsDNA. Most frequently, wHTHs interact with dsDNA by inserting the recognition helix H3 (H5 in the LM) into the DNA major groove, and often the wing also makes contacts with the DNA (Gajiwala and Burley, 2000).

To our knowledge, the elongation factor selB is the only previously known wHTH protein that binds RNA. The C-terminus of selB consists of four consecutive wHTH domains, where the two most C-terminal of these interact with an mRNA hairpin (Fourmy *et al*, 2002; Selmer and Su, 2002). The residues that bind RNA have not been identified. Sequence conservation and electrostatic potential calculations point to residues in the most C-terminal wHTH located on a surface between H3 (H5 in La), H2 (H3 in La), and the β-hairpin between B2 and B3 (Selmer and Su, 2002). This surface is similar to that used for DNA recognition in many of the wHTH transcription factors. In La, the residues on the corresponding surface are not well conserved. As discussed below, the LM may use an entirely different, highly conserved surface between helices H1, H1′, H2 and the loop between H3 and H4 to interact with RNA (Figures 1A and 2).

Conserved residues form an aromatic patch on the surface of La

LM-containing proteins share the highest sequence identity with the more N-terminal portions of the *T. brucei* LM

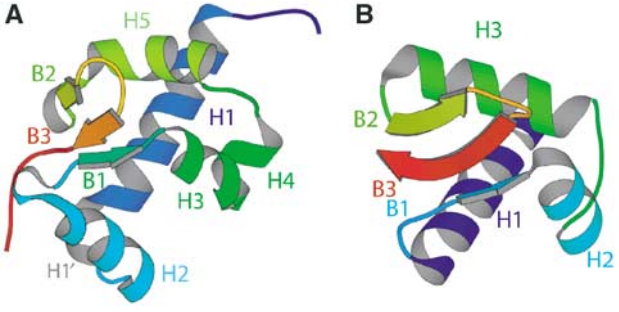

Figure 1 (**A**) Ribbon views of the LM of *T. brucei* La. (**B**) Ribbon view of a wHTH motif as observed in the Z-DNA-binding domain of the adenosine deaminase ADAR1 (PDB accession number 1QBJ). The topology of the LM is similar to that of the wHTH motif of some transcription factors. MOLSCRIPT (Kraulis, 1991) was used to prepare this figure.

sequence. In contrast, La proteins have sequence conservation throughout the domain (Figure 2D). The strictly conserved surface residues localize primarily to an extended surface patch (Figure 2). This patch is lined with the aromatic residues F17, Y18, F19, F29, and F50 as well as N23, D27, Q14, and R52. These residues are conserved in all characterized La proteins and are also conserved in LM-containing proteins. Thus, this patch may be involved in the only established La function, RNA binding. In the LM structure, this patch forms a crystal contact with a histidine (residue −1) from a symmetry-related molecule, which makes hydrogen bonds to N23 and D27 and interacts edge-on with F17 and Y18. The patch is long enough (~15–20 Å) to accommodate one or more nucleotide bases in place of this histidine. In a number of protein–ssRNA complexes, nucleotide bases are stacked with protein side chains, particularly from phenylalanines, tyrosines, and arginines (Allers and Shamoo, 2001). The stacking interactions and van der Waals contacts localize the bases to particular pockets, where hydrogen bonds to the peptide backbone or to other protein side chains determine specificity (Antson, 2000). It is possible that the conserved aromatic residues in the LM patch similarly serve to position one or more nucleotide bases. A second, smaller conserved patch centers on F41, which is surrounded by E77, E39, and R83, but mutagenesis experiments described below do not implicate this patch in RNA binding.

A basic region formed by residues K10, K13, R26, K33, K51, R52, and K72 borders the conserved aromatic patch (Figure 2B), consistent with a proposed role for the patch in RNA binding. These basic residues could make nonspecific contacts to the sugar–phosphate backbone of RNA. An acidic patch on the other side of the molecule contains E36, E39, E45, D59, E77, and D78, and is unlikely to interact with the negatively charged RNA backbone.

Biochemical characterization of T. brucei La

We used electrophoretic mobility shift assays (EMSA) and competition experiments to confirm that *T. brucei* La interacts with RNA like its human and yeast orthologs. First, we confirmed that *T. brucei* La interacts with pre-tRNA primarily through the 3′ UUU_{OH}. Gel shift assays revealed that the full-length *T. brucei* protein, as well as a truncated form containing only the LM and adjacent RRM (amino acids 1–192), binds to yeast pre-$tRNA^{Arg}$ (Figure 3A, left panel and data not

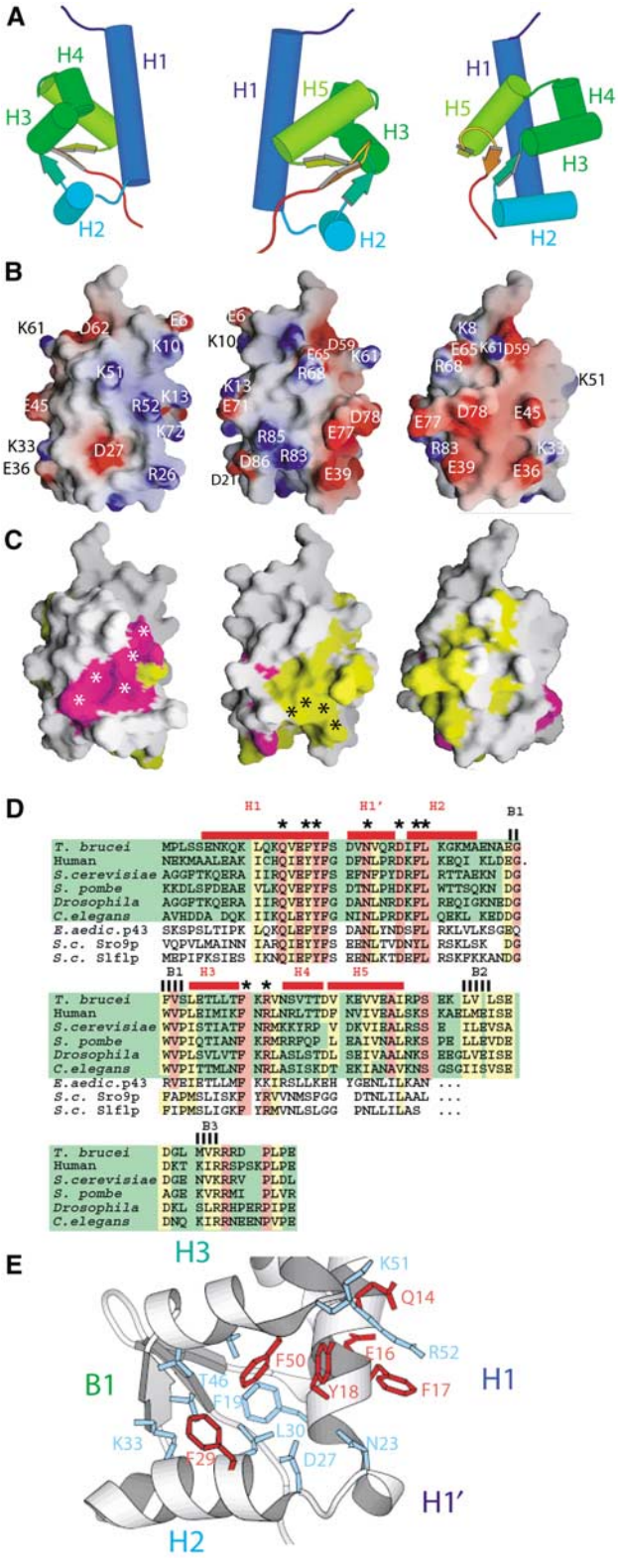

Figure 2 The orientations of the LM in (A–C) are the same. The three orientations are related by a rotation about an axis along the length of the page. (**A**) Ribbon representation of the LM with cylinders representing all helices but H1′. Helix H1′ is rendered as a squiggle. (**B**) Qualitative map of the electrostatic potential at the LM surface. The left-most view shows the positively charged surface adjoining the conserved aromatic patch (marked by D27). (**C**) Views of the LM are as in (B), but pink surfaces indicate residues that are invariant in La proteins. The conserved aromatic patch is visible in the left-most image. Yellow denotes less strictly conserved residues. Asterisks mark conserved surfaces that were mutated. (**D**) An alignment of LM sequences. Sequences from true La proteins, which bind nascent RNA polymerase III RNAs, are boxed in green. The *Caenorhabditis elegans* protein has not been biochemically characterized. Sequences for Slf1p, Sro9p, and p43, which contain an LM but are not La proteins, are also shown. As in (C), residues that are invariant in the true La proteins are boxed in pink and conserved residues are boxed in yellow. Red bars indicate α-helices, and striped black bars indicate β-strands. Stars indicate those residues that form an extended conserved patch on the surface of the LM. The patch is essential for RNA binding and recognition. (**E**) A rendering of the aromatic patch conserved in all LM proteins. Mutating the residues depicted in red eliminated La binding to pre-tRNA. Mutating residues D27, N23, and R52 to alanine had modest effects on binding affinity, and D27A had a reduced specificity for a hydroxyl over a phosphate group at the RNA 3′ terminus. MOLSCRIPT (Kraulis, 1991) was used to prepare (A, E) and Grasp (Nicholls *et al*, 1991) was used to prepare (B, C).

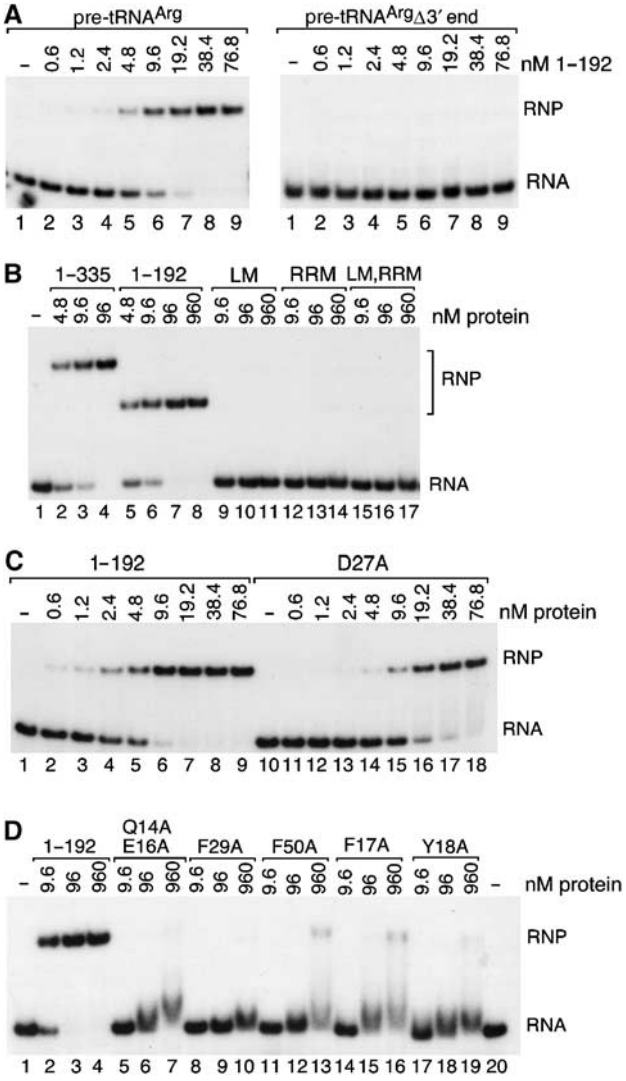

Figure 3 Binding of wild-type and mutant La proteins to RNA. (**A**) ^{32}P-labeled yeast pre-tRNA$^{Arg}_{CCG}$ (left panel) or pre-tRNA$^{Arg}_{CCG}$ lacking the 9 nt 3′ trailer (right panel) was incubated without protein (lane 1) or with the truncated *T. brucei* La protein (LM + RRM; amino acids 1–192) at the indicated protein concentrations. The pre-tRNA concentration in the reaction was 0.15 nM. Naked RNAs and RNPs were separated in native gels. (**B**) ^{32}P-labeled yeast pre-tRNA$^{Arg}_{CCG}$ (final concentration 0.15 nM) was incubated with either no protein (lane 1) or the indicated concentrations of full-length La (lanes 2–4), La 1–192 (lanes 5–8), the isolated LM (amino acids 1–89; lanes 9–11), the RRM (residues 94–192; lanes 12–14), or a 1:1 mixture of the LM and RRM (lanes 15–17). Naked RNAs and RNPs were separated by native gel electrophoresis. (**C**) Binding of La 1–192 (left panel) and the D27A mutant (right panel) to pre-tRNA$^{Arg}_{CCG}$ was performed as described in (A). Protein concentrations in the reactions are indicated above the lanes. In lanes 1 and 10, no protein was added. We quantitated the difference in binding affinity of La 1–192 and the D27A mutant by performing multiple experiments. These revealed that La 1–192 bound the RNA with a K_d of 4.1 ± 0.7 nM, while the D27A mutant bound the RNA with a K_d of 14.6 ± 2.9 nM, an approximately 3.6-fold decrease in affinity. (**D**) pre-tRNA$^{Arg}_{CCG}$ was subjected to gel shift analysis using either the wild-type La 1–192 (lanes 2–4) or the mutants Q14A/E16A (lanes 5–7), F29A (lanes 8–10), F50A (lanes 11–13), F17A (lanes 14–16), and Y18A (lanes 17–19). Protein concentrations in each lane are given in nM. Lanes 1 and 20, no protein.

shown). In the experiment shown, approximately 50% of the labeled pre-tRNA was shifted into RNA–protein complexes at a protein concentration between 4.8 and 9.6 nM. Deleting the UUU_{OH}-containing trailer or replacing it with polycytosine reduced La binding significantly (Figure 3A, right panel, and data not shown).

As with human La (Goodier *et al*, 1997; Ohndorf *et al*, 2001; Jacks *et al*, 2003), both the *T. brucei* LM and the adjacent RRM are necessary for high-affinity RNA binding. Neither the LM (residues 1–89) nor the RRM (residues 94–192) alone bind to RNA with high affinity (Figure 3B, lanes 9–11 and 12–14), but a construct consisting of both the LM and the adjacent RRM (residues 1–192) binds pre-tRNA with a similar affinity to the full-length protein (compare lanes 2–4 with lanes 5–7). Interestingly, a 1:1 mixture of the LM and the RRM also fails to bind pre-tRNA (Figure 3B, lanes 15–17). As the linker region between the two domains (residues ~89–96 in *T. brucei*) varies in length and is not well conserved in La proteins from different species, it seems unlikely that it is involved in important protein–RNA contacts. Most likely, the LM and the RRM do not have an extensive protein–protein interface and are flexibly tethered in the absence of RNA. Nucleic acid binding may orient the domains as is the case for Sxl, which binds ssRNA via two adjacent RRMs, or transcription factor TFIIIA, which binds dsDNA with three tandem zinc fingers (Wuttke *et al*, 1997; Crowder *et al*, 1999).

Residues in the conserved aromatic patch are essential for La interactions with pre-tRNA

To test which residues in the LM are important for RNA binding, we constructed a series of point mutations in the LM + RRM construct (residues 1–192). We mutated conserved surface residues in the LM to alanine and assayed the binding affinity for each construct by EMSA with pre-tRNA (Table II). Mutations of conserved surface residues not in the aromatic patch (the double mutations R83A/R84A and E39A/F41A) did not significantly affect RNA binding, as the amount of mutant protein required to shift half the labeled pre-tRNA varied less than two-fold from the wild type (data not shown). Mutation of three conserved residues in the patch (R52A, N23A, and D27A) all resulted in modest decreases in binding affinity compared to wild-type protein (shown for D27A in Figure 3C; Table II). In contrast, the double mutation Q14/E16A and mutations of aromatic residues in the conserved patch (F29A, F50A, F17A, Y18A, and the double mutation F17A/Y18A) diminish the RNA-binding affinity by more than 100-fold (Figure 3D). These results confirm that residues in the LM conserved patch are required for La interactions with RNA.

Table II Effect of mutations on RNA-binding affinity

Mutation	Location	Effect on RNA-binding affinity
Q14A/E16A	Conserved patch	>100× reduced
F17A	Conserved patch	>100× reduced
Y18A	Conserved patch	>100× reduced
F29A	Conserved patch	>100× reduced
F50A	Conserved patch	>100× reduced
N23A	Conserved patch	Slightly reduced (~2–4×)
D27A	Conserved patch	3.6× reduced
R52A	Conserved patch	Slightly reduced (~2–4×)
R83A/R84A	Not in patch	<2× reduced
E39A/F41A	Not in patch	<2× reduced

Mutating an aspartate in the conserved surface patch reduces La specificity for a 3′ hydroxyl versus phosphate group

One of the more interesting features of La recognition is its ability to distinguish RNAs ending in UUU_{OH} from those ending in UUU_p. We used competition experiments to examine those mutants that retained RNA-binding affinity (R52A, E39A/F41A, R83/R84, N23A, and D27A) for their ability to discriminate between RNAs differing only at the 3′ ends. In these experiments, we determined the amount of unlabeled competitor RNA necessary to dissociate the labeled pre-tRNA complex. Pre-tRNA containing a 3′ UUU_{OH} was a more effective competitor for *T. brucei* La binding than the same RNA terminating in UUU_p (Figure 4A), as previously described for the human and the *S. cerevisiae* La proteins (Stefano, 1984; Long *et al*, 2001). Mutations of residues outside the aromatic patch (E39A/F41A and R83A/R84A) did not affect the ability of *T. brucei* La to recognize 3′ OH, nor did the R52A mutation at the edge of the aromatic patch (data not shown). The N23A mutation in the patch also had little, if any, effect (Figure 4B). In contrast, the D27A mutation at the base of the patch has a diminished preference for the hydroxyl group, as less UUU_p-containing pre-tRNA was needed to dissociate the complex than required for the wild-type protein (compare Figures 4C and A, left panels). For the wild-type protein, the amount of 3′ UUU_{OH} competitor required to dissociate 50% of the labeled protein–RNA complex ($K_d^{1/2}$) was 9.9 ± 0.9 nM, while for the 3′ UUU_p competitor it was 988.6 ± 70.3 nM (Figure 4D). For the D27A mutant, these values were 19.9 ± 1.7 and 358.3 ± 62.3 nM, respectively (Figure 4D). For the native protein the relative binding affinity ($K_{rel} = K_d^{1/2}$ $UUU_p/K_d^{1/2}$ UUU_{OH}) was approximately 100, while for the D27A mutant the ratio was approximately 18, a five- to six-fold decrease in specificity. D27 then contributes at least a part of the specificity of the *T. brucei* La protein for the 3′ hydroxyl. In the native protein, D27 could hydrogen bond to the 3′ OH while sterically and electrostatically clashing with a negatively charged 3′ phosphate group. Replacing the aspartate with alanine would at least partially remove steric and electrostatic barriers that prevent a phosphate group from binding, thus resulting in diminished discrimination against it.

Models for RNA binding

The La protein makes specific interactions with the 3′ UUU_{OH}, optimally accommodating four uridine bases in its recognition site (Stefano, 1984), and it may also interact with other regions of its cognate RNAs (Wolin and Cedervall, 2002; Chakshusmathi *et al*, 2003). We have found that residues in the conserved aromatic patch of the LM are essential for high-affinity pre-tRNA binding. However, since RNA binding by La requires both the LM and the RRM, any model must take into account the role of the RRM. One possibility is that the LM functions indirectly to facilitate binding by the RRM. For example, the conserved aromatic patch may contribute to an LM:RRM interface that induces an RRM conformation able to interact with RNA. Our gel shift experiments show that the LM and RRM do not form a complex when separately expressed and then mixed together, either in the presence or absence of RNA (Figure 3B and data not shown), suggesting that the two domains do not share an extensive interface.

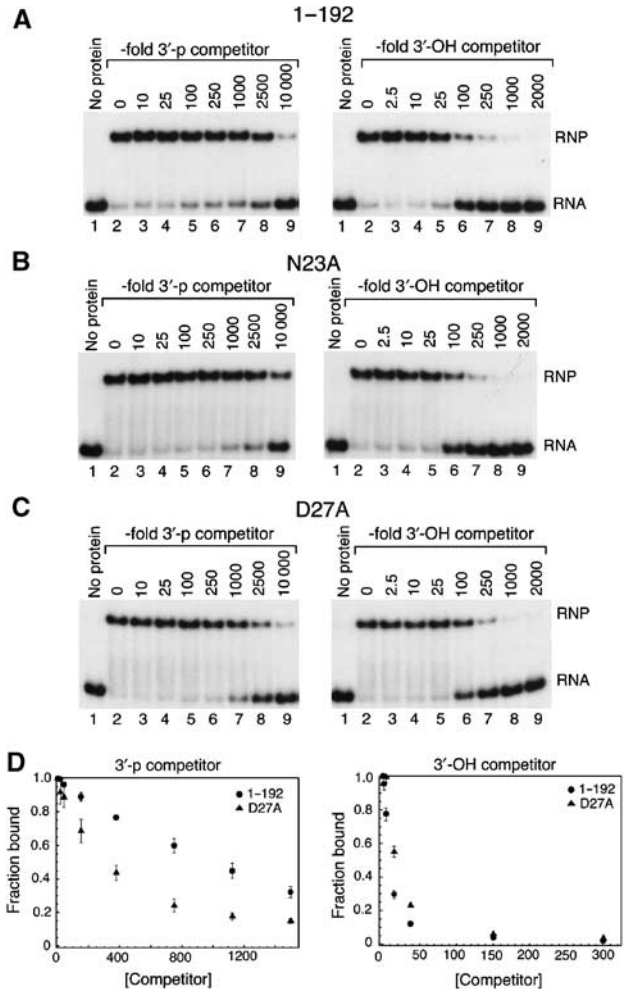

Figure 4 Mutating D27 reduces the specificity of La for RNAs ending in a 3′ hydroxyl. (**A**) La 1–192 (4.8 nM) was mixed with ^{32}P-labeled pre-$tRNA^{Arg}_{CCG}$ (0.15 nM) in the presence of a 0- to 10 000-fold molar excess of unlabeled pre-$tRNA^{Arg}_{CCG}$ terminating in either UUUp (left panel) or UUU_{OH} (right panel). Naked RNAs and RNPs were separated by native gel electrophoresis. The small differences in the fraction of RNA bound with 4.8 nM protein between this experiment and Figure 3C, lane 5, reflect experimental variation. Lane 1, no added protein. (**B, C**) The N23A mutant (B) or D27A mutant protein (C) was mixed with ^{32}P-labeled pre-$tRNA^{Arg}_{CCG}$ (0.15 nM) in the presence of the indicated molar excess of unlabeled pre-$tRNA^{Arg}_{CCG}$ as described in (A). The concentration of protein used in the reactions was 19.2 nM. Lane 1, no added protein. (**D**) Data from the competition titrations for La 1–192 and the D27A mutant are plotted as the fraction of labeled probe bound at each concentration of competitor RNA. Each data point represents the mean from at least three experiments. RNA concentrations are given in nM.

Nonetheless, we cannot exclude the possibility that the LM is only indirectly involved in RNA binding.

We favor a model in which the LM interacts directly with RNA. This is plausible in that the LM is the most conserved portion of La and is found in other RNA-binding proteins that lack recognizable RRMs (Sobel and Wolin, 1999; Wolin and Cedervall, 2002). Our result, that mutating residue D27 in the LM conserved aromatic patch to alanine reduces La's specificity for a 3′ OH over a 3′ phosphate, supports a role for the conserved patch in forming at least part of the recognition site for 3′ ends. With its many aromatic residues, the patch is well suited to interacting with one or more of the 3′ uridine

bases. Since RRMs are known to bind ssRNA and the La RRM is required for RNA binding, surfaces from both the LM and the RRM may contribute to the polyuridine-binding site. For both domains to bind to the relatively short RNA recognition sequence, they would have to form a V-shaped crevice lined on one side by the conserved aromatic patch of the LM and on the other by the RNA-binding surface of the RRM. The polyuridine would bind in this crevice, uridine bases interacting with each of the two domains. The length of the linker between the LM and the RRM varies in La proteins from different sources, and at eight amino acids the *T. brucei* linker is among the shortest. It is long enough, however, to allow juxtapositioning of the conserved aromatic La patch and the RNA-binding surface of the RRM. Since the LM and its conserved aromatic patch also occur in proteins that do not specifically bind UUU_{OH} (Sobel and Wolin, 1999; Aigner *et al*, 2003), it is possible that the specificity for polyU is a function of cooperative interactions by the LM and the RRM.

In summary, by combining structural information with mutational analysis, we have established that a conserved aromatic patch in the LM is required for high-affinity interactions of the La protein with RNA. Further, since D27 in the conserved aromatic patch contributes to the specificity for a hydroxyl versus a phosphate 3′ terminus, the patch plays a role in binding the 3′ terminus, possibly also contributing to the specificity of La for polyU. How the LM and the RRM cooperate in RNA binding and the precise nature of the protein–RNA interactions will best be resolved by a structure of the complex.

Materials and methods

Cloning

A full-length construct encoding *T. brucei* La was a generous gift of Dr Elisabetta Ullu. Wild-type full-length protein (residues 1–335) and three truncated versions (LM + RRM: residues 1–192; LM: residues 1–89; and central RRM: residues 94–192) of the *T. brucei* La protein were cloned into the plasmid pET-15b (Novagen) between the *Nde*I and *Bam*HI restriction sites. Each construct includes a thrombin-cleavable hexahistidine tag at the protein N-terminus. As the LM + RRM construct has the same affinity for pre-tRNA as the full-length protein, it was used as a template for mutagenesis. A total of 11 mutations (Table II) were introduced into the LM + RRM construct using the QuickChange site-directed mutagenesis kit (Stratagene). All plasmids were sequenced to confirm the correct sequence and the presence of the expected mutation.

Protein overexpression and purification

Escherichia coli BL21(DE3) cells harboring plasmid for the full-length La, La + RRM, the LM, or the central RRM were grown at 37°C in Luria broth supplemented with 50 μg/ml ampicillin. When the cells reached an OD_{600} between 0.6 and 0.8, they were shifted to 25°C (18°C for mutations) and protein production was induced with 1 mM IPTG. The cells were harvested 6–8 h after induction and stored at −80°C until further use. The protein purification was carried out entirely at 4°C. Frozen cells were thawed at room temperature and resuspended in 20 ml/l cell culture of cold lysis buffer (20 mM Tris–HCl (pH 8 at room temperature), 300 mM NaCl, 1 mM PMSF, and 0.5 mg/ml lysozyme). After a brief incubation, the cells were further lysed by pulse sonication. The lysate was cleared by centrifugation at 30 000 *g* for 30 min, followed by filtration through a 0.4 μm filter. The lysate was supplemented with 10 mM imidazole and loaded onto Ni-NTA resin (Qiagen) pre-equilibrated with buffer A (20 mM Tris–HCl (pH 8.0 at room temperature), 300 mM NaCl, and 10 mM imidazole). The resin-bound protein was washed with 10 volumes of buffer A. Thrombin (150 U/l of cell culture; Amersham) was added to release the protein from the resin. Protein, consisting of an N-terminal Gly-Ser-His sequence followed by the desired *T. brucei* La sequence, was washed from the column with 5–7 bed volumes of buffer A. The protein was concentrated to a small volume in the presence of 10% glycerol and loaded onto a Superdex200 gel filtration column (Amersham). Protein was collected from the gel filtration column in buffer B [10 mM Tris–HCl (pH 8.0 at room temperature), 100 mM NaCl, 5% v/v glycerol, and 5 mM $MgCl_2$]. LM + RRM mutants behaved like the native construct on the gel filtration column. Fractions corresponding to La were pooled and concentrated to 10–20 mg/ml, suitable for crystallization.

Selenomethionine-substituted LM was expressed using minimal media supplemented with all amino acids (2 mg/ml) except methionine. Just prior to induction, L-selenomethionine was added to 100 mg/l, and additional threonine, lysine, phenylalanine, leucine, isoleucine, and valine were added to inhibit the methionine biosynthetic pathway (Doublie, 1997). The selenomethionine protein was purified as described above, except for the addition of extra reductant. All steps preceding the gel filtration column were carried out in the presence of 20 mM β-me; after the Ni-NTA affinity purification, the protein was stabilized by the presence of 10 mM DTT.

Crystallization

Crystals of the LM were grown at 4°C by the hanging drop method against a well liquor containing 100 mM HEPES (pH 7.5), 5% (v/v) glycerol, 32–36% (w/v) PEG 3.5K, and, in the case of the selenomethionine-substituted protein, 10 mM DTT. Drops contained 1–6 μl of protein solution and an equivalent volume of well liquor. Crystals appeared within 1–2 days and grew to $\sim 0.2 \times 0.2 \times 0.8$ mm within a week. The crystals belong to space group $P2_12_12_1$ ($a = 32.926$ Å, $b = 53.833$ Å, and $c = 53.912$ Å) and contain one molecule per asymmetric unit. We were unable to crystallize the LM domain lacking the N-terminal Gly-Ser-His addition. The crystals were soaked in well liquor for about half an hour and flash-frozen in liquid nitrogen (Rodgers, 1997) prior to data collection. They diffracted to ~1.9 Å resolution in-house and to beyond 1.6 Å resolution at beamline X12C at Brookhaven National Laboratory and had a mosaicity of $\sim 0.5^{o}$. Data from selenomethionine-substituted crystals were not isomorphous with crystals of the unsubstituted protein although both crystal forms belonged to the same space group and had the same unit cell dimensions.

Structure determination

Data from a flash-frozen selenomethionine-substituted crystal were collected at the beamline X12C at Brookhaven National Laboratory. Full and highly redundant data sets collected at the anomalous peak wavelength sufficed to calculate high-quality electron density maps (Table I). Data were integrated and scaled using the program HKL2000 (Otwinowski and Minor, 1997). All subsequent calculations were performed using the software suite CNS (Brunger *et al*, 1998). Anomalous Patterson maps allowed us to identify three selenium sites. These were refined in CNS, and SAD phases together with solvent flipping produced highly interpretable experimental maps at 1.6 Å. The initial model was built in O (Kleywegt and Jones, 1997). Subsequent cycles of torsion angle dynamics, least squares minimization, individual B-factor refinement, and the addition of water molecules yielded $R_{work} = 23.2\%$ and $R_{free} = 25.0\%$ for all data between 20 and 1.6 Å. The final model contains all residues including Gly-Ser-His (−3, −2, −1) and 1–89 of the *T. brucei* La sequence and 100 water molecules. Coordinates and structure factors have been deposited in the Protein Data Bank (accession number 1S29).

Electrophoretic mobility shift assays

The clone containing pre-$tRNA^{Arg}_{CCG}$ under control of the T7 promoter has been described (Chakshusmathi *et al*, 2003). After linearization with *Dra*I, transcription with T7 RNA polymerase yielded a pre-tRNA containing three U's at the 3′ end. pre-$tRNA^{Arg}_{CCG}$ sequences lacking the 3′ trailer were amplified from the full-length clone using 5′-CGGCGAATTCTAATACGACTCACTATAGTTTATTATGCTCCTCTAG TGC-3′ and 5′-GCTCCTCCCGGGACTCGAACC-3′ and transcribed using T7 RNA polymerase. Pre-tRNAs were synthesized as described (Yisraeli and Melton, 1989), except that 50 μCi of [α-^{32}P]rCTP (400 Ci/mmol; Amersham) was used in place of rCTP. RNAs were gel purified and quantitated by scintillation counting. Prior to their use in binding assays, RNAs were resuspended in 50 mM Tris (pH 8.0) and 10 mM $MgCl_2$, refolded by heating to 60°C

for 1 min, and then slowly cooled to room temperature. For binding assays, 1.5 fmol of RNA was mixed with 6–768 fmol of protein and 0.2 μg/ml *E. coli* tRNA in binding buffer (20 mM Tris, pH 8.0, 3 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT, 100 mM NaCl, and 0.1% NP-40). The total reaction volume was 10 μl. Reactions were incubated at room temperature for 10 min and then on ice for 10 min, and loaded onto 5% polyacrylamide (38:2 acrylamide:bisacrylamide)/5% glycerol gels (prerun at 16 V/cm for 10 min at 4°C). Gels were run at 16 V/cm in 0.5× TBE (50 mM Tris, 45 mM boric acid, and 1.25 mM EDTA) at 4°C for 2 h and 30 min. The fraction of bound RNA (θ) was quantitated using a PhosphorImager (Molecular Dynamics). Using Imagequant, the data were fitted by nonlinear least squares analysis as a function of protein concentration using the equation

$$\theta = \frac{[\text{protein}]}{[\text{protein}]+K_d}$$

where θ is the fraction of bound RNA and K_d is the dissociation constant. Each K_d was calculated from at least three binding experiments.

Competition experiments

Competition experiments were largely as described (Long *et al*, 2001). For each mutant, we chose a protein concentration that resulted in 80–90% of the labeled RNA forming an RNP. For each reaction, 1.5 fmol of labeled pre-tRNA was mixed with unlabeled competitor RNA (3.75 fmol to 15 pmol) in the presence of truncated wild-type or mutant *T. brucei* La proteins. Reaction conditions and gel analyses were as described above. Competitor RNAs consisted of either the identical RNA (containing 3′ UUU_{OH}) or the RNA with a 3′ UUUp terminus. To prepare pre-$tRNA^{Arg}_{CCG}$ with a 3′ UUUp, a pre-tRNA containing four 3′ uridylates was first synthesized using T7 RNA polymerase. This was converted to the pre-RNA with a 3′ UUUp by incubating the RNA in 100 μl 1 M DL-lysine–Cl and 0.025 M $NaIO_4$ (pH 8.3) for 2 h at 45°C, followed by phenol extraction, ethanol precipitation, and gel purification. This Whitfield degradation reaction removes the last nucleotide, leaving a 3′ phosphate (Neu and Heppel, 1964). Data were quantitated using a PhosphorImager and analyzed as described (Long *et al*, 2001).

Acknowledgements

We thank Elisabetta Ullu for her gift of the *T. brucei* La cDNA. We are grateful to Anand Saxena and the staff at X12-C at Brookhaven National Laboratory for help with data collection. We thank David W Rodgers and Elisabetta Ullu for critically reading and discussing this manuscript and the reviewers for their comments. This work was supported by grants from the Pew Foundation and the Mathers Foundation to KMR and grant R01-GM48410 from the National Institute of Health to SLW. SLW is an Associate Investigator of the Howard Hughes Medical Institute.

References

Aigner S, Postberg J, Lipps HJ, Cech TR (2003) The Euplotes La motif protein p43 has properties of a telomerase-specific subunit. *Biochem* **42:** 5736–5747

Allers J, Shamoo Y (2001) Structure-based analysis of protein–RNA interactions using the program ENTANGLE. *J Mol Biol* **311:** 75–86

Alspaugh MA, Tan EM (1975) Antibodies to cellular antigens in Sjogren's syndrome. *J Clin Invest* **55:** 1067–1073

Antson A (2000) Single stranded RNA binding proteins. *Curr Opin Struct Biol* **10:** 87–94

Bai C, Li Z, Tolias PP (1994) Developmental characterization of a *Drosophila* RNA-binding protein homologous to the human systemic lupus erythematosus-associated La/SS-B autoantigen. *Mol Biol Cell* **14:** 5123–5129

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54:** 905–921

Chakshusmathi G, Kim SD, Rubinson DA, Wolin SL (2003) A La protein requirement for pre-tRNA folding. *EMBO J* **22:** 6562–6572

Chambers JC, Kenan DJ, Martin BJ, Keene JD (1988) Genomic structure and amino acid sequence domains of the human La autoantigen. *J Biol Chem* **263:** 18043–18051

Chong SS, Hu P, Hernandez N (2001) Reconstitution of transcription from the human U6 small nuclear RNA promoter with eight recombinant polypeptides and a partially purified RNA polymerase III complex. *J Biol Chem* **276:** 20727–20734

Crowder SM, Kanaar R, Rio DC, Alber T (1999) Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. *Proc Natl Acad Sci USA* **96:** 4892–4897

Doublie S (1997) Preparation of selenomethionyl proteins for phase determination. In *Macromolecular Crystallography*, CW Carter, RM Sweet (eds) Vol. 276, pp 523–529. New York: Academic Press

Fan H, Sakulich AL, Goodier JL, Zhang X, Qin J, Maraia RJ (1997) Phosphorylation of the human La protein on serine 366 can regulate recycling of RNA polymerase III transcription complexes. *Cell* **88:** 707–715

Fan H, Goodier JL, Chamberlain JR, Engelke DR, Maraia RJ (1998) 5′ processing of tRNA precursors can be modulated by the human La antigen phosphoprotein. *Mol Cell Biol* **18:** 3201–3211

Fourmy D, Guittet E, Yoshizawa S (2002) Structure of prokaryotic SECIS mRNA hairpin and its interaction with elongation factor SelB. *J Mol Biol* **324:** 137–150

Gajiwala KS, Burley SK (2000) Winged helix proteins. *Curr Opin Struct Biol* **10:** 110–116

Goodier JL, Fan H, Maraia RJ (1997) A carboxy-terminal basic region controls RNA polymerase III transcription factor activity of human La protein. *Mol Cell Biol* **17:** 5823–5832

Gottlieb E, Steitz JA (1989) The RNA binding protein La influences both the accuracy and the efficiency of the RNA polymerase III transcription *in vitro*. *EMBO J* **8:** 841–850

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

Kleywegt G, Jones AT (1997) Electron density map interpretation. In *Methods in Crystallography*, CW Carter, RM Sweet (eds) Vol. 277, pp 173–208. New York: Academic Press

Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* **24:** 946–950

Kufel J, Allmang C, Chanfreau G, Petfalski E, Lafontaine DL, Tollervey D (2000) Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. *Mol Cell Biol* **20:** 5415–5424

Lin-Marq N, Clarkson SG (1998) Efficient synthesis, termination, and release of RNA polymerase III transcripts in *Xenopus* extracts depleted of La protein. *EMBO J* **17:** 2033–2041

Long SK, Cedervall T, Walch-Solimena C, Noe DA, Huddleston MJ, Annan RS, Wolin SL (2001) Phosphorylation of the *Saccharomyces cerevisiae* La protein does not appear to be required for its function in tRNA maturation and nascent RNA stabilization. *RNA* **7:** 1589–1602

Maraia RJ (1996) Transcription termination factor La is also an initiation factor for RNA polymerase III. *Proc Natl Acad Sci USA* **93:** 3383–3387

Maraia RJ, Intine RV (2001) Recognition of nascent RNA by the human La antigen: conserved and divergent features of structure and function. *Mol Cell Biol* **21:** 367–379

Marchetti MA, Tschudi C, Kwon H, Wolin SL, Ullu E (2000) Import of proteins into the trypanosome nucleus and their distribution at karyogenesis. *J Cell Sci* **113:** 899–906

Mattioli M, Reichlin M (1974) Heterogeneity of RNA protein antigens reactive with sera of patients with systemic lupus erythematosus. Descriptions of a cytoplasmic nonribosomal antigen. *Arthritis Rheum* **17:** 421–429

Mer G, Bochkarev A, Gupta R, Bochkareva E, Frappier L, Ingles CJ, Edwards AM, Chazin WJ (2000) Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell* **103:** 449–456

Neu HC, Heppel LA (1964) Nucleotide sequence analysis of polyribonucleotides by means of periodate oxidation followed by cleavage with an amine. *J Biol Chem* **237:** 2927–2934

Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11:** 281–296

Ohndorf U-M, Steegborn C, Knijff R, Sondermann P (2001) Contribution of the individual domains in human La protein and its RNA 3′ end-binding activity. *J Biol Chem* **276:** 27188–27196

Otwinowski Z, Minor W (1997) Processing X-ray diffraction data collected in oscillation mode. In *Macromolecular Crystallography*, CW Carter, RM Sweet (eds) Vol. 276, pp –326. New York: Academic Press

Pannone BK, Xue D, Wolin SL (1998) A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J* **17:** 7442–7453

Rodgers D (1997) Practical cryocrystallography. In *Methods in Enzymology*, CW Carter and RM Sweet (eds) Vol. 276, pp 183–202. New York: Academic Press

Selmer M, Su X-D (2002) Crystal structure of an mRNA-binding fragment of *Moorella thermoacetica* elongation factor SelB. *EMBO J* **21:** 4145–4155

Sobel SG, Wolin SL (1999) Two yeast La motif-containing proteins are RNA-binding proteins that associate with polyribosomes. *Mol Biol Cell* **10:** 3849–3862

Stefano JE (1984) Purified lupus antigen La recognizes an oligouridylate stretch common to the 3′ termini of RNA polymerase III transcripts. *Cell* **36:** 145–154

Van Horn DJ, Yoo CJ, Xue D, Shi H, Wolin SL (1997) The La protein in *Schizosaccharomyces pombe*: a conserved yet dispensible phosphoprotein that functions in tRNA maturation. *RNA* **3:** 1434–1443

Varani G, Nagai K (1998) RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct* **27:** 407–445

Wah DA, Hirsch JA, Dorner LF, Schildkraut I, Aggarwal AK (1997) Structure of the multimodular endonuclease *Fok*I bound to DNA. *Nature* **388:** 97–100

Wolin SL, Cedervall T (2002) The La protein. *Annu Rev Biochem* **71:** 375–403

Wuttke DS, Foster MP, Case DA, Gottesfeld JM, Wright PE (1997) 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* **273:** 183–206

Xue D, Rubinson DA, Pannone BK, Yoo CJ, Wolin SL (2000) U snRNP assembly in yeast involves the La protein. *EMBO J* **19:** 1650–1660

Yisraeli JK, Melton DA (1989) Synthesis of long, capped transcripts *in vitro* by SP6 and T7 RNA polymerases. *Methods Enzymol* **180:** 42–50

Yoo CJ, Wolin SL (1994) La proteins from *Drosophila melanogaster* and *Saccharomyces cerevisiae*: a yeast homolog of the La autoantigen is dispensible for growth. *Mol Cell Biol* **14:** 5412–5424

Yoo CJ, Wolin SL (1997) The yeast La protein is required for the 3′ endonucleolytic cleavage that matures tRNA precursors. *Cell* **89:** 393–402