

A comprehensive analysis of the La-motif protein superfamily

CÉCILE BOUSQUET-ANTONELLI and JEAN-MARC DERAGON

Université de Perpignan Via Domitia UPVD, CNRS UMR5096 LGDP, 66860 Perpignan, France

ABSTRACT

The extremely well-conserved La motif (LAM), in synergy with the immediately following RNA recognition motif (RRM), allows direct binding of the (genuine) La autoantigen to RNA polymerase III primary transcripts. This motif is not only found on La homologs, but also on La-related proteins (LARPs) of unrelated function. LARPs are widely found amongst eukaryotes and, although poorly characterized, appear to be RNA-binding proteins fulfilling crucial cellular functions. We searched the fully sequenced genomes of 83 eukaryotic species scattered along the tree of life for the presence of LAM-containing proteins. We observed that these proteins are absent from archaea and present in all eukaryotes (except protists from the *Plasmodium* genus), strongly suggesting that the LAM is an ancestral motif that emerged early after the archaea-eukarya radiation. A complete evolutionary and structural analysis of these proteins resulted in their classification into five families: the genuine La homologs and four LARP families. Unexpectedly, in each family a conserved domain representing either a classical RRM or an RRM-like motif immediately follows the LAM of most proteins. An evolutionary analysis of the LAM-RRM/RRM-L regions shows that these motifs co-evolved and should be used as a single entity to define the functional region of interaction of LARPs with their substrates. We also found two extremely well conserved motifs, named LSA and DM15, shared by LARP6 and LARP1 family members, respectively. We suggest that members of the same family are functional homologs and/or share a common molecular mode of action on different RNA baits.

Keywords: La motif (LAM); La autoantigen; LARP; RRM; DM15/LARP1 domain; RNA-binding proteins

INTRODUCTION

The La motif (LAM) (pfam05383/smart00715/InterPro006630) is an extremely well-conserved 90-amino acids-long domain found in several proteins from various eukaryotic species. It was first identified in the so-called La autoantigen (or genuine La protein), an abundant RNA-binding factor detected so far in all eukaryotic species studied (Wolin and Cedervall 2002). The genuine La protein mainly is localized to the nucleus, where it specifically recognizes and binds with a high-affinity terminal 3'UUU-OH motif (Stefano 1984; Wolin and Cedervall 2002). Since this motif is the hallmark of RNA polymerase III (pol III) transcription termination (Huang et al. 2006), the genuine La protein binds every RNA pol III primary transcript (including U6 snRNA, tRNA precursors, or pre-5S rRNA). In the yeast *Saccharomyces*

cerevisiae, the 3'-end maturation of RNA polymerase II (pol II) U3 snoRNA and spliceosomal snRNA primary transcripts generates a terminal 3'UUU-OH motif that also allows binding of the La protein (Kufel et al. 2000; Xue et al. 2000; Inada and Guthrie 2004). In most cases, the 3'-end binding of the La protein is only transient, the protein being displaced upon excision or modification of the 3'-UUU-OH motif. So far, the best-established role of the La protein is to provide protection against 3'-exonucleolytic degradation with various consequences depending on the nature of the transcript. For pre-tRNAs and pre-U3 snoRNA, La protein binding allows normal processing of the RNA (Yoo and Wolin 1997; Kufel et al. 2000). The La protein was also shown to exhibit an RNA chaperone activity (Belisova et al. 2005), promoting correct folding of certain pre-tRNAs (Chakshumathi et al. 2003) or facilitating the assembly of U snRNAs into functional ribonucleoprotein particles (Pannone et al. 1998; Xue et al. 2000). Also, the La protein most probably takes part in the quality-control mechanism of newly synthesized noncoding RNAs such as pre-tRNAs (Copela et al. 2006; Huang et al. 2006; Kadaba et al. 2006). But genuine La proteins are versatile factors that were

Reprint requests to: Jean-Marc Deragon, Université de Perpignan Via Domitia UPVD, CNRS UMR5096 LGDP, 52 avenue Paul Alduy, 66860 Perpignan, France; e-mail: jean-marc.deragon@univ-perp.fr; fax: 33-4-68668499.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.1478709>.

reported in yeasts, flies, and mammals to bind to certain mRNAs and to influence their translation efficiency (Holcik and Korneluk 2000; Kim et al. 2001; Cardinali et al. 2003; Trotta et al. 2003; Inada and Guthrie 2004; Vazquez-Pianzola et al. 2005), their stability, or subcellular localization (McLaren et al. 1997; Adilakshmi and Laine 2002; Brenet et al. 2005). In addition, several reports established that the HsLa (human genuine La) protein directly binds to several viral RNAs, often enhancing their translation and replication efficiency (Costa-Mattioli et al. 2004; Raha et al. 2004; Domitrovich et al. 2005; Xue et al. 2007; Bitko et al. 2008) and, at least for some viruses, shielding them from the host's defense systems (Bitko et al. 2008).

Genuine La proteins are modular factors that most of the time contain three structured domains: the LAM, a canonical RNA recognition motif (RRM1), and an atypical RNA recognition motif (RRM2) followed by a largely unstructured tail (Maraia and Intine 2001; Wolin and Cedervall 2002). The NH₂-terminal domain (NTD), composed of the LAM-RRM1 motifs, is extremely well conserved across eukaryotes and constitutes a structural hallmark of genuine La proteins. Functional as well as structural studies demonstrated that the genuine La NTD is necessary and sufficient to bind to the 3'-UUU-OH motif and hence is directly implicated in La nuclear functions (Maraia and Intine 2001; Wolin and Cedervall 2002; Curry and Conte 2006; Maraia and Bayfield 2006). The LAM was shown to adopt an elaborate winged helix-turn-helix fold (Alfano et al. 2004; Dong et al. 2004), whereas the RRM1 adopts the typical $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ fold (β : beta-sheet; α : alpha-helix) of canonical RRMs, which usually forms a β -sheet surface directly involved in RNA binding (Alfano et al. 2004; Maris et al. 2005; Teplova et al. 2006; Clery et al. 2008). Co-crystal structure analysis of the human La NTD showed that in the presence of an RNA ligand, while keeping their overall structure unchanged, LAM-RRM1 domains synergize and fold back together to form a tight RNA-binding pocket around the 3'-UUU-OH motif (Teplova et al. 2006; Kotik-Kogan et al. 2008). With the exception of genuine La homologs from a few species (Wolin and Cedervall 2002; Fleurdepine et al. 2007), the COOH-terminal domain (CTD) of genuine La proteins contains the atypical RRM (RRM2) followed by an unstructured terminal region that was shown in HsLa and Sla1p (*Schizosaccharomyces pombe* La) to contain several phosphorylation sites, a short basic motif (SBM), a nuclear localization signal (NLS), and a nuclear retention element (NRE) (Simons et al. 1996; Van Horn et al. 1997; Rosenblum et al. 1998; Maraia and Intine 2001; Intine et al. 2002). Based on crystal structure data, the HsLa RRM2 domain was shown to adopt a typical RRM fold with an additional long terminal α -helix that folds back on the β -sheet surface (Jacks et al. 2003). The RRM2 motif is conserved in the CTD of most La homologs, albeit not at the primary sequence level but in terms of secondary structure motifs

(Fleurdepine et al. 2007). At present, it seems that the RRM2 domain does not display the ability to bind RNA (Jacks et al. 2003), but several reports suggest that the C-terminal domain is involved in the regulation of the complex nuclear and cytoplasmic activities of genuine La proteins, at least in mammals (Park et al. 2007).

The LAM is also present on proteins otherwise unrelated to genuine La, referred to as La-related proteins (LARPs) (Yoo and Wolin 1994; Saget et al. 1998; Sobel and Wolin 1999; Remillieux-Leschelle et al. 2002; Aigner et al. 2003; Witkin and Collins 2004; Valavanis et al. 2007; He et al. 2008; Krueger et al. 2008; Markert et al. 2008; Nykamp et al. 2008). Contrary to the well characterized genuine La protein, functional and structural data on LARPs are very scarce, and to date only a few LARPs have been characterized. The yeast *S. cerevisiae* LARP proteins Sro9p and Slf1p are RNA-binding factors able to bind homopolymers in vitro. These two LARPs are mainly cytoplasmic proteins found in association with translating polysomes (Sobel and Wolin 1999). Although their mode of action is not clear, they have been proposed to act as modulators of translation elongation or termination and as regulators of the stability of a subset of mRNAs (Sobel and Wolin 1999; Tan et al. 2000). More recently, Sro9p was proposed to participate as a co-chaperone in the repression of Hap1p function, a heme-induced transcriptional activator (Lan et al. 2004).

In ciliates, the *Tetrahymena thermophila* p65 and *Euplotes aediculatus* p43 LARPs are active telomerase-specific subunits that directly and specifically bind to telomerase RNA (Aigner et al. 2000, 2003; Witkin and Collins 2004). They are key member proteins of the telomerase RNP and are required for the assembly, nuclear retention, activity, and processivity of the complex (Mollenbeck et al. 2003; Aigner and Cech 2004; Witkin and Collins 2004). P65 is one of the first factors binding to telomerase RNA. Upon binding, p65 induces structural rearrangements of the RNA providing a suitable binding platform for reverse transcriptase binding, an essential step in the formation of the active telomerase RNP (Stone et al. 2007; Teixeira and Gilson 2007).

In humans, a screen designed to identify new factors involved in the functional regulation of transcription elongation complex p-TEFb identified the HsLARP7/PiP7S protein (He et al. 2008). This LARP was shown to associate tightly to 7SK RNA and to participate in the stabilization, assembly, and repressive function of the 7SK RNP on p-TEFb (He et al. 2008; Krueger et al. 2008; Markert et al. 2008). HsLARP7 was also shown to have an antigrowth/antitumor function that correlates with its negative impact on p-TEFb. Similarly, the *Drosophila melanogaster* putative homolog of HsLARP7, the multi-sex comb (mxc) protein, was shown to be a tumor suppressor involved in the control of larvae blood cell proliferation and differentiation (Remillieux-Leschelle et al. 2002). The mode of action of mxc is unknown at present, but with the recent and

unexpected report of the existence of a 7SK RNA homolog in fly (Gruber et al. 2008a), it is tempting to speculate that mxc could bind fly 7SK and act in a similar fashion as described for HsLARP7. In addition to HsLARP7, five LARPs have been identified to date in the human genome (data from the HUGO Gene Nomenclature Committee at the European Bioinformatics Institute, <http://www.genenames.org>), but none of their functions have been readily addressed to date. The only available data at present concern HsLARP6/Acheron, a putative homolog of the Acheron protein of the moth *Manduca sexta*, the mRNA of which was identified in a screen for cDNAs overaccumulating in skeletal muscles upon their commitment to programmed cell death processes (Valavanis et al. 2007).

The *Caenorhabditis elegans* genome encodes two LARPs named CeLARP1 and CeLARP5 (Nykamp et al. 2008). The CeLARP1 and its putative fly homolog, the dlarp protein (Chauvet et al. 2000; Ichihara et al. 2007), appear to be required for normal metazoan development. Nykamp and colleagues demonstrated that CeLARP1 is an RNA-binding factor able to bind homopolymers in vitro and is required for normal oogenesis. CeLARP1 is enriched in P-bodies and is required for the accumulation at normal levels of Ras-MAPK-related mRNAs in oocytes. They hence proposed that CeLARP1 could play a role in oogenesis by controlling levels of specific mRNAs in relation to P-body-mediated degradation.

In summary, it appears that LARPs are involved in highly divergent putative functions and may not represent a homogenous group of proteins. In this work, we provide a more complete picture of the LAM protein superfamily. Phylogenetic analyses of LAM proteins from a large number of highly divergent eukaryotes, together with structural motif searches, allowed us to propose that besides the genuine La proteins family, LARPs can be divided into four clearly distinct families (LARP1, 4, 6, and 7). We postulate that members of each family share some functional properties.

RESULTS

Identification of LAM-containing proteins in eukaryotes

We performed an exhaustive search for LAM-containing proteins in the fully sequenced genomes of 46 archaea and 83 eukaryotes representing all four eukaryotic kingdoms: protists, plants, fungi, and animals (searches were made using databases available at the NCBI [<http://www.ncbi.nlm.nih.gov>] or the JGI [<http://genome.jgi-psf.org/>] websites) (see Supplemental Fig. 1; see Materials and Methods for a list of the genomes searched, a description of the search methods, and the number of LAM proteins coded by each genome). LAM-containing proteins were not found in any of the 46 archaeal genomes used. On the other hand, using the 83 eukaryotic genomes, we successfully retrieved

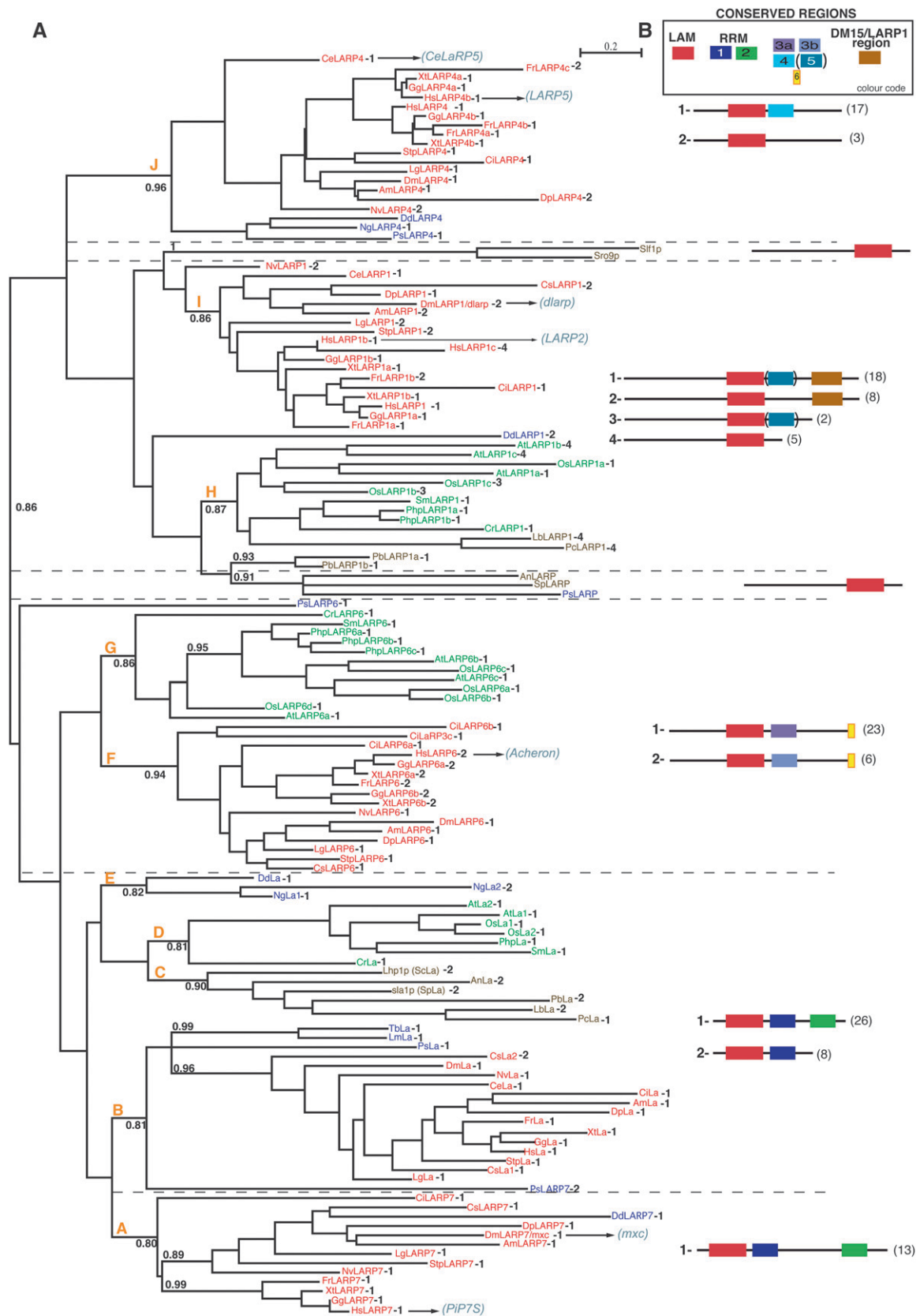
308 distinct LAM-containing proteins (Supplemental Fig. 1B), each protein having a single LAM. LAM-containing proteins were found in every plant, fungi, and animal genome we searched. LAM-containing proteins were also found in every protist genome, with the notable exception of species from the *Plasmodium* genus. The systematic use of several complementary search methods, the use as a query of several additional LAMs from closely related protist species, and the high level of conservation we observed for eukaryotic LAMs (see Supplemental Figs. 3–6) make it unlikely that LAM-containing proteins were missed out of the five *Plasmodium* genomes searched. We hence propose that LAM-containing proteins (genuine La and LARPs) are not coded in *Plasmodium* genomes. It is intriguing to observe that species classified in the same taxonomic group (the Apicomplexans) show either a presence or absence of LAM-containing proteins (see Supplemental Fig. 1A). Given that LAM-containing proteins are absent from archaea but are present in almost all other eukaryotes tested, we propose that LAM is an ancestral eukaryotic motif that emerged early after the archaea-eukarya radiation.

Phylogenetic relationship among eukaryotic LAM-containing proteins and identification of conserved structural features

To analyze phylogenetic relations among eukaryotic LAM proteins, we selected a subset of 134 proteins from 29 species scattered along the tree of life and divided as follows: five protists, five plants, six fungi, and 13 animals (Supplemental Figs. 1A, 2). We mostly retained a single organism per phylum, generally picking the organisms most frequently utilized by the scientific community.

We first explored the evolutionary relationships among these proteins by reconstructing a phylogenetic tree based on the alignment of the 134 LAMs (Fig. 1A). We also searched each protein for the presence of additional conserved domains by specialized BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the Conserved Domain Database (CDD) (Marchler-Bauer et al. 2007). Finally, we performed a multiple sequence alignment of the 134 full-length proteins to look for new conserved motifs not described in the CDD (Fig. 1B; data not shown).

The CDD search detected the presence of three referenced motifs in addition to the LAM. These motifs are the canonical and atypical RRM1 previously detected in the genuine La proteins, RRM1 (pfam00076) and RRM2 (pfam08777), respectively, and a short motif called DM15 (smart00684) previously detected in the *C. elegans* protein LARP1 (Nykamp et al. 2008). The RRM1 and RRM2 motifs are found only in proteins belonging to clusters A–E on the phylogenetic tree (Fig. 1). With the exception of groups A and E, each of these clusters contains previously recognized functional genuine La proteins (HsLa, DmLa, TbLa, Lhp1p [ScLa], Sla1p [SpLa],



and AtLa1) (Yoo and Wolin 1994; Van Horn et al. 1997; Wolin and Cedervall 2002; Foldynova-Trantirkova et al. 2005; Fleurdepine et al. 2007). The human, fly, and trypanosome proteins are found in group B, yeast proteins in group C, and plant proteins in group D. Proteins falling into these four groups are between 263 and 770 amino acids long with an average size of ~ 400 amino acids and adopt two types of domain arrangements: LAM-RRM1 or LAM-RRM1-RRM2. In all cases, LAM-RRM1 motifs are separated by a short region (18 amino acids on average) and are located at the protein's N terminus. Based on these structural and evolutionary relationships to known genuine La proteins, we propose that proteins in these four clusters are all genuine La homologs. Out of the 31 putative genuine La proteins reported here, only seven lack the terminal RRM2 domain. Most of these (five out of seven) are from fungi; nevertheless, this should not be considered as a general feature of this taxa since we found several fungi putative genuine La proteins displaying the second RRM2 motif (Fig. 1A, *Phanerochaete chrysosporium*; data not shown). In addition, shorter genuine La proteins are also found in animals and protists. With a few exceptions (*Arabidopsis thaliana*, *Oriza sativa*, and *Capitella* sp.), a single genuine La protein was found per organism. For *A. thaliana*, despite the presence of two proteins with characteristics of genuine La, only one (AtLa1) is able to bind to polymerase III transcripts and fulfill nuclear La functions (Fleurdepine et al. 2007). Therefore, it remains to be seen if two functional genuine La proteins can coexist in the same organism.

With the exception of DdLARP7 from the protist *Dictyostelium discoideum*, all cluster A proteins belong to animal species. This cluster can be further divided into two subgroups containing proteins from either invertebrate or vertebrate species. The vertebrate subgroup contains the human HsLARP7/PiP7S protein, while the other group contains the proposed *D. melanogaster* functional homolog, mxc/DmLARP7 (He et al. 2008). Like genuine La homologs, cluster A proteins display a LAM-RRM1-RRM2 structural organization with N-terminal LAM-RRM1 motifs separated by an average of 18 amino acids. However,

cluster A proteins are generally longer (average size of 550 amino acids), and the distance between the RRM1 and RRM2 domains is much larger. Indeed, genuine La protein RRM1 and RRM2 motifs are separated by an average of 68 amino acids, whereas for cluster A proteins this linker is on average 200 amino acids long. LAMs from proteins belonging to clusters A–E are evolutionarily very close, and it is difficult to discriminate between genuine La and LARP based solely on the LAM phylogenetic analysis presented in the tree shown in Figure 1. However, based on structural differences between genuine La and cluster A proteins, together with the fact that HsLARP7/Pip7S (Fig. 1A) does not fulfill authentic La functions (He et al. 2008; Krueger et al. 2008; Markert et al. 2008), we hypothesize that proteins falling into cluster A are LARPs. Cluster E contains three protist proteins displaying a LAM-RRM1-RRM2 or LAM-RRM1 organization. Given that no functional data are available to date on any of these factors, it is difficult to conclude whether they might fulfill genuine La or La-Related functions. Nevertheless, for DdLa and NgLa1, the short distance between the two RRMs (60 and 40 amino acids, respectively) is characteristic of genuine La proteins.

The remaining 88 proteins possess a LAM more distantly related to the LAM of genuine La proteins. Also, these proteins do not have the classical LAM-RRM1-RRM2 organization of genuine La proteins. Therefore, these proteins will also be referred to as LARPs. With a few exceptions, these LARPs are clustered into five subgroups (Fig. 1A, labeled F–J). Although statistical support for these subgroups is not always optimal, structural features obtained by the CDD search and the multiple alignment tend to validate this initial clustering.

We observed that a 90-amino acids-long region located immediately downstream from the LAM (at the place of the RRM1 motifs for genuine La homologs and cluster A LARPs) (see Supplemental Fig. 3) is highly conserved for almost all cluster F, G, and J LARPs belonging to distantly related species (see Supplemental Figs. 4, 5). Indeed, every protein belonging to clusters F and G possesses a conserved

FIGURE 1. Phylogenetic relationships among LAMs and structural organization of LAM-containing proteins. (A) Phylogenetic relationships among LAMs. The phylogenetic tree was obtained using LAM sequences from the 134 selected proteins as described in the Materials and Methods. Selected informative statistical supports (approximate likelihood-ratio test [aLRT] data) are indicated. (Blue) Proteins from protists, (green) proteins from plants, (brown) proteins from fungi, (red) proteins from animals. Protein names are given according to the nomenclature proposed in the present paper. For the corresponding accession numbers, see Supplemental Figure 2. Previously utilized names are reported in parentheses when available. Black numbers following protein names correspond to structural organizations presented in B. (B) Schematic representations of the different structural organizations found among the full-length LAM-containing proteins. The color code used is reported in the box at the top of the figure. (Numbers in parentheses) Number of proteins of a given structural organization, (dashed lines) putative significant groups of proteins based on phylogenetic and structural organization criteria. Species names abbreviations: (Am) *Apis mellifera*, (At) *Arabidopsis thaliana*, (An) *Aspergillus niger*, (Ce) *Caenorhabditis elegans*, (Cs) *Capitella* sp., (Cr) *Chlamydomonas reinhardtii*, (Ci) *Ciona intestinalis*, (Dp) *Daphnia pulex*, (Dd) *Dictyostelium discoideum*, (Dm) *Drosophila melanogaster*, (Fr) *Fugu rubripes*, (Gg) *Gallus gallus*, (Hs) *Homo sapiens*, (Lb) *Laccaria bicolor*, (Lm) *Leishmania major*, (Lg) *Lottia gigantea*, (Ng) *Naegleria gruberi*, (Nv) *Nematostella vectensis*, (Os) *Oriza sativa*, (Pc) *Phanerochaete chrysosporium*, (Pb) *Phycomyces blakesleanus*, (Php) *Physcomitrella patens*, (Ps) *Phytophthora sojae*, (Sc) *Saccharomyces cerevisiae*, (Sp) *Schizosaccharomyces pombe*, (Sm) *Selaginella moellendorffii*, (Stp) *Strongylocentrotus purpuratus*, (Tb) *Trypanosoma brucei*, (Xt) *Xenopus tropicalis*.

motif at this position (Fig. 1B, boxes 3a–3b; see Supplemental Fig. 4), while all but three proteins from cluster J share a different conserved region at this position (Fig. 1B, box 4; see Supplemental Fig. 5). We also observed two additional conserved motifs among LARPs. First, 27 out of 29 proteins from clusters F and G display a well conserved 20–30-amino acids-long motif (Fig. 1B, box 6; see below) located at the C-terminal end of the proteins (Figs. 1B, 4; see below). Second, most proteins from clusters I and H (26 out of 33) display one to four repetitions of a highly conserved 40-amino acids-long domain, referred to as DM15 in the CDD databank (Figs. 1B, 3, see below). Finally five proteins, namely the *S. cerevisiae* Slf1p and Sro9p proteins and the *Aspergillus niger* AnLARP, *S. pombe* SpLARP, and *Phytophthora sojae* PsLARP proteins do not belong to any of the above-mentioned clusters and do not appear to possess additional conserved boxes except for the LAM (Fig. 1).

Conserved regions following the LAM adopt an RRM-like fold

To get a better understanding of the meaning of conserved boxes 3a, 3b, and 4 found immediately downstream from the LAM, we analyzed the region encompassing the LAM plus the following 100 amino acids for all genuine La and LARPs, using the protein structure prediction program SAM-T08 (http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html; Karplus et al. 2005). This software was systematically successful in predicting the correct $\alpha 1\alpha 1'\alpha 2\beta 1\alpha 3\alpha 4\alpha 5\beta 2\beta 3$ topology (Alfano et al. 2004; Dong et al. 2004) of LAMs as well as predicting the typical α/β sandwich fold ($\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$) of canonical RRM1 domains for all genuine La homologs as well as for all cluster A LARPs (Supplemental Fig. 3; Maris et al. 2005). We found that conserved boxes 3a, 3b, and 4 are predicted to adopt an RRM-like structure composed of a $\beta 1\alpha 1\beta 2\beta 3\alpha 2$ (boxes 3a/3b) or a $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ (box 4) fold (Fig. 2A; Supplemental Figs. 4, 5; data not shown). Also, despite the fact that they do not share sequence homology in their region downstream from the LAM, most proteins from clusters H and I (18 out of 29) (see Fig. 1B), plus proteins PbLARP1a and PbLARP1b, are also predicted to adopt an RRM-like structure composed of a $\beta 1\alpha 1\beta 2\beta 3\alpha 2$ topology (Fig. 1B, box 5; Fig. 2A; Supplemental Fig. 6; data not shown). We hence decided to name these boxes RRM-like motifs (RRM-L) (Fig. 2A). We observed that, for a given RRM-L

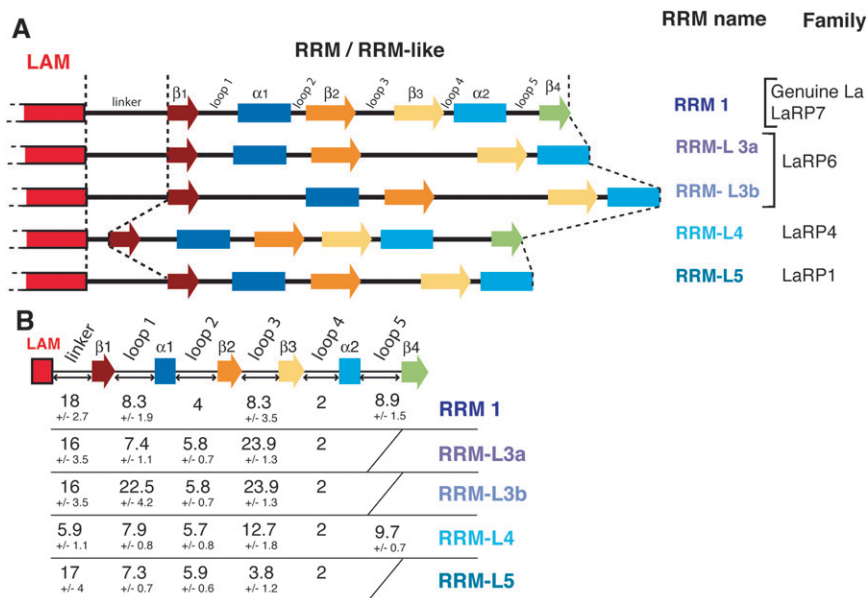


FIGURE 2. Structural organization of the LAM-RRM1 and LAM-RRM-L domains of genuine La homologs and LARPs. (A) Schematic representation of the predicted structural organizations of the 90 amino acids found immediately downstream from the LAM. (Plain red box) LAM, (colored arrows) β -sheets ($\beta 1$, $\beta 2$, $\beta 3$, and $\beta 4$), (colored boxes) α -helices ($\alpha 1$ and $\alpha 2$) of the RRM or RRM-L domains. The proposed names of these regions are reported on the right (see Fig. 5). (B) Size of the linker regions (in amino acids) connecting the LAM and RRM1 or RRM-L domains and of the loops connecting the structural features of each domain. Sizes of each of these connecting regions were determined for every protein, and the average sizes are reported together with standard deviations.

motif (RRM-L3a, 3b, 4, or 5) and for the typical RRM1 motif, sizes of the loops connecting the structural features (Fig. 2B, loop 1: $\beta 1/\alpha 1$, loop 2: $\alpha 1/\beta 2$, loop 3: $\beta 2/\beta 3$, loop 4: $\beta 3/\alpha 2$, loop 5: $\alpha 2/\beta 4$), as well as the size of the linker between the LAM and the first structural feature of the RRM (or RRM-L), are conserved between proteins sharing identical boxes but vary from one RRM (or RRM-L) motif to another (Fig. 2). Important differences in size are observed for loops 1 and 3. RRM-L3b loop 1 is about three times longer than that of the other domains, and the size of loop 3 differs for each family. On the other hand, the size of loop 4 remains unchanged in every case. The LAM-RRM/RRM-L linker size remains roughly around 17 amino acids long, except in the case of the RRM-L4 domain where it is about three times shorter (Fig. 2B).

Analysis of the DM15/LARP1 regions

We found that most proteins from clusters I and H, plus proteins PbLARP1a and PbLARP1b (Fig. 1A), possess a highly conserved C-terminal region, named the DM15/LARP1 region in Figure 1B. Recently, Nykamp and collaborators found such a conserved region on LARPs from human, fly, mouse, *C. elegans*, *D. discoideum*, and *A. thaliana* and named it the LARP1 domain (Nykamp et al. 2008). Consistently with their results, we identified this

conserved domain from the same LARPs (Fig. 1A). The DM15/LARP1 domain was also found on numerous additional LARPs from a wide range of eukaryotes (Fig. 1A) and

consists of one to four tandem repeats of a motif referenced in the CDD as the DM15 box (smart00684) (Fig. 3). We made a multiple sequence alignment of the 55 DM15

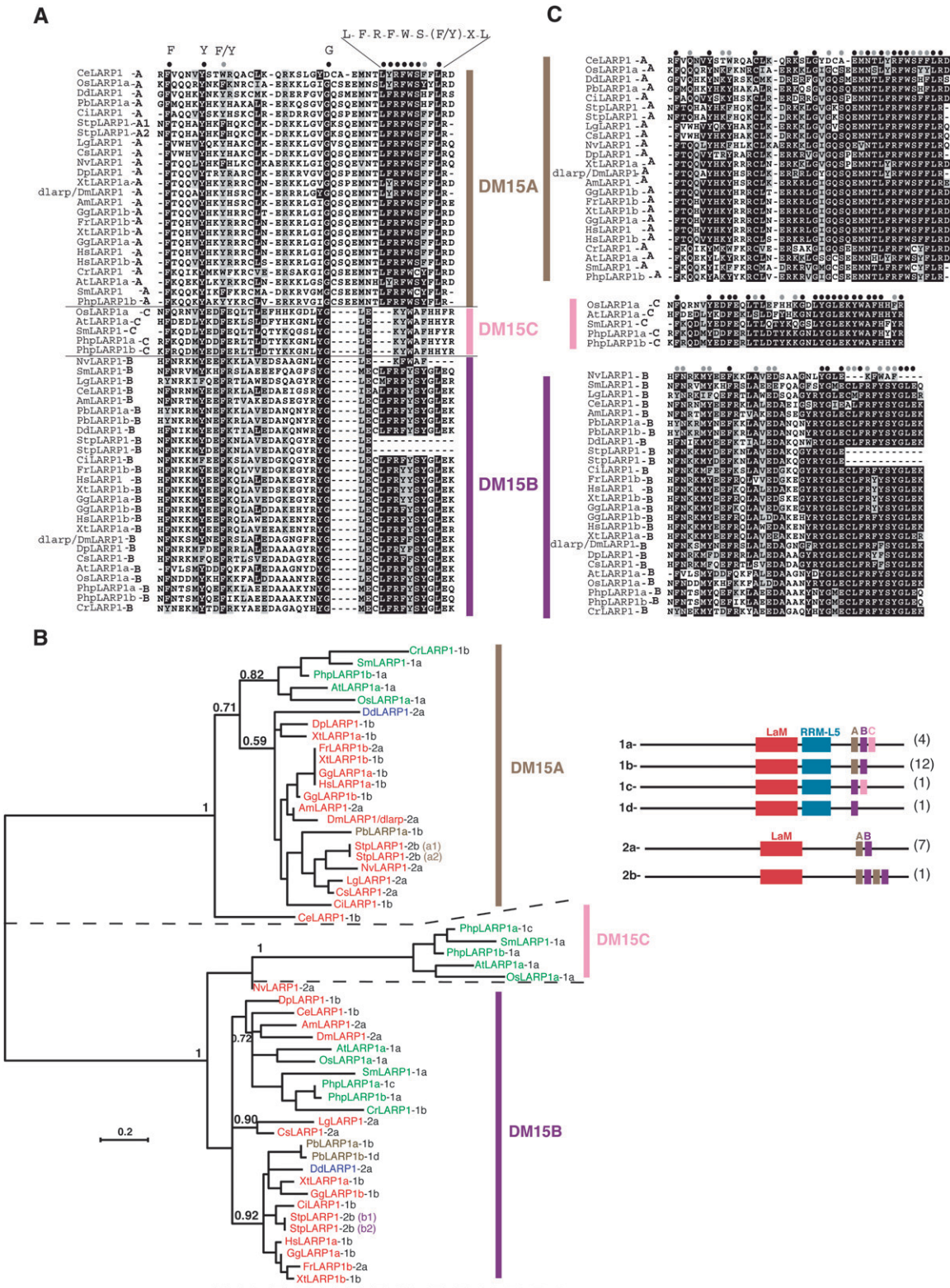


FIGURE 3. (Legend on next page)

boxes we retrieved from a total of 26 LARPs and found that these boxes are of three types we named DM15A, DM15B, and DM15C (Fig. 3A). The phylogenetic tree reconstructed out of this alignment corroborated this observation, as the DM15 boxes group into three well supported clusters containing the A, B, and C types, respectively (Fig. 3A,B). We hence found that the DM15/LARP1 domain is actually composed of one to four DM15 boxes of the A, B, and/or C type organized in different ways (Fig. 3B). Proteins containing DM15A and DM15B repeats are found in all four eukaryotic kingdoms, whereas proteins containing DM15C repeats are found only in plants. Therefore, it is likely that DM15A and DM15B repeats diverged early in evolution while DM15C repeats evolved from DM15B repeats at a later time in the plant lineage (Fig. 3B). Sequence conservation inside a subgroup of DM15 repeats is extremely high, as more than half of the residues are conserved in 90%–100% of the sequences (Fig. 3C). Comparison of the different DM15 subgroups (Fig. 3A) revealed that 12 amino acids are conserved in 90%–100% of the sequences and that the majority of these conserved amino acids are aromatic. The DM15/LARP1 region is always found in the C-terminal region of proteins containing a LAM or a LAM/RRM-L5 organization. We observed that in most cases, proteins have a DM15A-DM15B tandem and sometimes (only for plant proteins) a DM15A-DM15B-DM15C organization. With two exceptions, the DM15A motif is always present, and there is no case where DM15C is present alone.

A new conserved region found on LAM and S1-like containing proteins: The LSA motif

Multiple sequence alignment of full-length LAM proteins revealed the presence of an additional conserved motif (Fig. 1B, box 6). This 20–30 amino acids-long patch (Fig. 4) is exclusively shared by clusters F and G LARPs (Fig. 1) and localizes to the very C-terminal ends of the proteins. Plant proteins are slightly different, as they possess a 3-amino acids-long “PRM” insertion in the first half of the conserved consensus (Fig. 4). We found that factors annotated as cold-shock response protein 1 (CSP1) from various animals also possess, in addition to an S1-like nucleic acid binding domain referred to as the cold-shock domain

(CSD) (Bycroft et al. 1997; Sommerville 1999; Kloks et al. 2002), a similar conserved motif at their C-terminal ends (Supplemental Fig. 7). We propose to name this previously uncharacterized conserved region associated with LAM and S1-like proteins the LAM and S1 associated (LSA) motif. The evolutionary conservation of the LSA motif across eukaryotes and its association with two different nucleic acid binding motifs (LAM and S1) suggests that LSA could participate functionally in the selective binding of LARP6 and CSP1 proteins to their substrates.

Phylogenetic relationships among the LA-RRM domains: Co-evolution of the LAM and its associated RRM (RRM-L) domain

The function of genuine La proteins necessitates that both the LAM and RRM1 act in concert to form a single active domain (Maraia and Intine 2001; Wolin and Cedervall 2002; Curry and Conte 2006; Maraia and Bayfield 2006). Consistently, their presence at the primary and secondary structural levels has been highly conserved during evolution. We found in this study that the LAM-RRM (or RRM-like) organization is much more conserved than originally thought since it concerns the vast majority of LARPs as well (113 out of the 134 LAM-containing proteins analyzed in this study). We therefore hypothesized that, in most cases, LAM should not be considered as an independent evolving unit but should be studied in association with its co-evolving partner, the RRM or the RRM-L. To test this, we performed a multiple sequence alignment of LAM-RRM or LAM-RRM-L regions for the 113 concerned proteins, as well as for the ciliate p43 and p65 LARPs, for which we have consistent functional data (Supplemental Figs. 3–6; Aigner et al. 2000, 2003; Witkin and Collins 2004) and used this alignment to reconstruct a phylogenetic tree (Fig. 5). Analysis of Figure 5 shows that clusters previously formed using the LAM alone are maintained (cf. Figs. 1 and 5), but they now form five statistically well supported groups (Fig. 5, labeled 1–5). Each group also contains proteins that share the same structural domain organization (Fig. 5).

With the exception of a protein from the protist *P. sojiae* (PsLARP7), group 1 proteins correspond to the putative

FIGURE 3. Analysis of the DM15/LARP1 region. (A) Complete sequence alignment of every detected DM15 box. The multiple sequence alignment was edited by Boxshade at the Mobyle portal (<http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=boxshade>). (Black) Identical residues, (gray) similar residues. Protein names are reported on the *left*. These names are according to the nomenclature used in Figure 1A. The bold letter added to the protein name (A, B, or C) corresponds to the type of DM15 box. Amino acids identical in 90%–100% of the sequences are reported on *top* of the alignment. (Black dots) Residues present in 100% of the sequences, (gray dots) amino acids present in 90%–100% of the sequences. (B) Phylogenetic relationships among the DM15 boxes and structural organization of the full-length proteins. A phylogenetic tree was reconstructed (see Materials and Methods) from the multiple sequence alignment presented in A. Protein names from which the DM15 box originates are reported at the tip of each branch. Selected informative statistical supports are reported. (Black numbers) Structural organization of the full-length proteins as pictured on the *right-hand* side. (Plain red box) The LAM, (blue box) the RRM-L5 motif, (brown boxes) type A DM15 repeats (pink boxes) type B DM15 repeats, (purple boxes) type D DM15 repeats. The number of proteins presenting a given type of organization is reported in parentheses. (C) Multiple sequence alignments performed between DM15 boxes of the same type: (brown bar) type A, (pink bar) type B, (purple bar) type C. Protein names reported on the *left* and alignment editing are as in A.



FIGURE 4. Identification of a novel conserved region, the LSA motif. Multiple sequence alignment of the C-terminal region of proteins with a LAM/RRM-L3a or LAM/RRM-L3b organization. (Black) Identical residues, (gray) similar residues. The names of the proteins are reported on the *left-hand* side as in Figure 1A. The consensus sequence is reported *below* the alignment. (X) Any amino acid. Plant protein consensus sequences contain an additional three amino acids (PRM) between positions six and seven and display two instead of four undefined amino acids between positions 13 and 14.

genuine La factor homologs previously belonging to clusters B–E, while group 2 proteins correspond to cluster A LARPs (Fig. 1A). The phylogenetic analysis of the LAM-RRM (RRM-L) region gives a sharper distinction between these two related sets of proteins and strongly supports this previous classification. We also found that the functionally characterized ciliate p43 (EaLARP7) and p65 (TthLARP7) proteins (Aigner et al. 2000, 2003; Aigner and Cech 2004; Witkin and Collins 2004; Stone et al. 2007; Teixeira and Gilson 2007) are group 2 factors and cluster with the other protist proteins of this group. As for genuine La proteins, not all group 2 factors possess an RRM2 motif (see PsLARP7 and p65 [TthLARP7]), rendering both structural and evolutionary analyses necessary for classification.

Group 3 contains every protein from clusters G and F in Figure 1. Group 3 proteins share structural features, although vertebrate proteins have an RRM-like motif slightly different (RRM-L3b) compared with other eukaryotes (see Fig. 2, RRM-L3a). This difference might reflect an adaptation to different substrates in the vertebrate lineage. Group 4 contains every protein of cluster J (except for those not having an RRM-L and not used in this study), while group 5 is composed of factors from clusters H and I, plus proteins PbLARP1a and b, which now clearly belong to this group. We can notice that, in contrast to other groups where 85%–100% of the proteins have an RRM or an RRM-L motif, only 60% of proteins from cluster H and I have such domain. This might be linked to the fact that some of these factors display the additional extremely well conserved DM15/LARP1 domain, which was proposed to be involved in direct RNA binding (Nykamp et al. 2008).

The overall comparison of the LAM- and LAM-RRM-based phylogenetic analysis shows similar results, although the latter tree gives more statistically well-supported clusters. This leads us to propose that the LAM and downstream RRM or RRM-L domains co-evolved, as suggested for genuine La proteins, and that the combination of both motifs should be used to define the functional region of interaction of LARPs with its substrate.

Definition of five families and systematic nomenclature for LAM-containing proteins

Data presented in Figure 5 clearly show that proteins sharing a common ancestral origin for their LAM-RRM/RRM-L region also share a common structural organization, supporting the classification of the LAM protein superfamily into five distinct families. To reflect this classification, we propose adopting a systematic naming of the La-motif-containing proteins (Fig. 6). Proteins belonging to cluster number 1 in Figure 5A constitute the group of genuine La proteins. With the exception of the *S. cerevisiae* and *S. pombe* La proteins, largely referenced as Lhp1p and Slal1p, we suggest systematically naming the genuine La proteins according to the nomenclature already employed for some organisms, such as humans, where the species abbreviation name is added before the “La” appellation, resulting in HsLa protein for *Homo sapiens* La protein.

Most of the human LARPs have previously been named and referenced by the HUGO Gene Nomenclature Committee (<http://www.genenames.org/genefamily/larp.php>), and two (HsLARP7/PiP7S and HsLARP6/Acheron) were reported in the literature (Valavanis et al. 2007; He et al. 2008; Krueger et al. 2008; Markert et al. 2008). Consequently, to comply as much as possible with the already existing names and to introduce as little change as possible, we chose to number the recognized families according to the numbering of the human LARP proteins present in each group. Group 5 in Figure 5A was named family 1, given that it contains the human LARP1 and the *C. elegans* LARP1 that was recently reported by Nykamp et al. (2008). Groups 4 and 3 contain HsLARP4 and HsLARP6/Acheron, respectively (Valavanis et al. 2007), and hence were labeled families 4 and 6. And, finally, since the HsLARP7/PiP7S protein falls into group 2, we labeled it family 7. To sum up, we propose to name the La-related proteins as such: XsLARPfz, with Xs being the abbreviation name of the species they belong to, f as the number of the family they belong to (i.e., 1, 4, 6, or 7), and z being a lower-case letter (i.e., a, b, c, ...) when there are several members of the same family in a given species (Fig. 6). According to this nomenclature, HsLARP2, HsLARP5, and CeLARP5 would need to be renamed HsLARP1b, HsLARP4b, and CeLARP4 to reflect their group of origin. Since no published data include these proteins, these changes should have little impact. For LARPs used in previously published work that do not comply with this nomenclature, mxc (Remillieux-Leschelle et al. 2002),

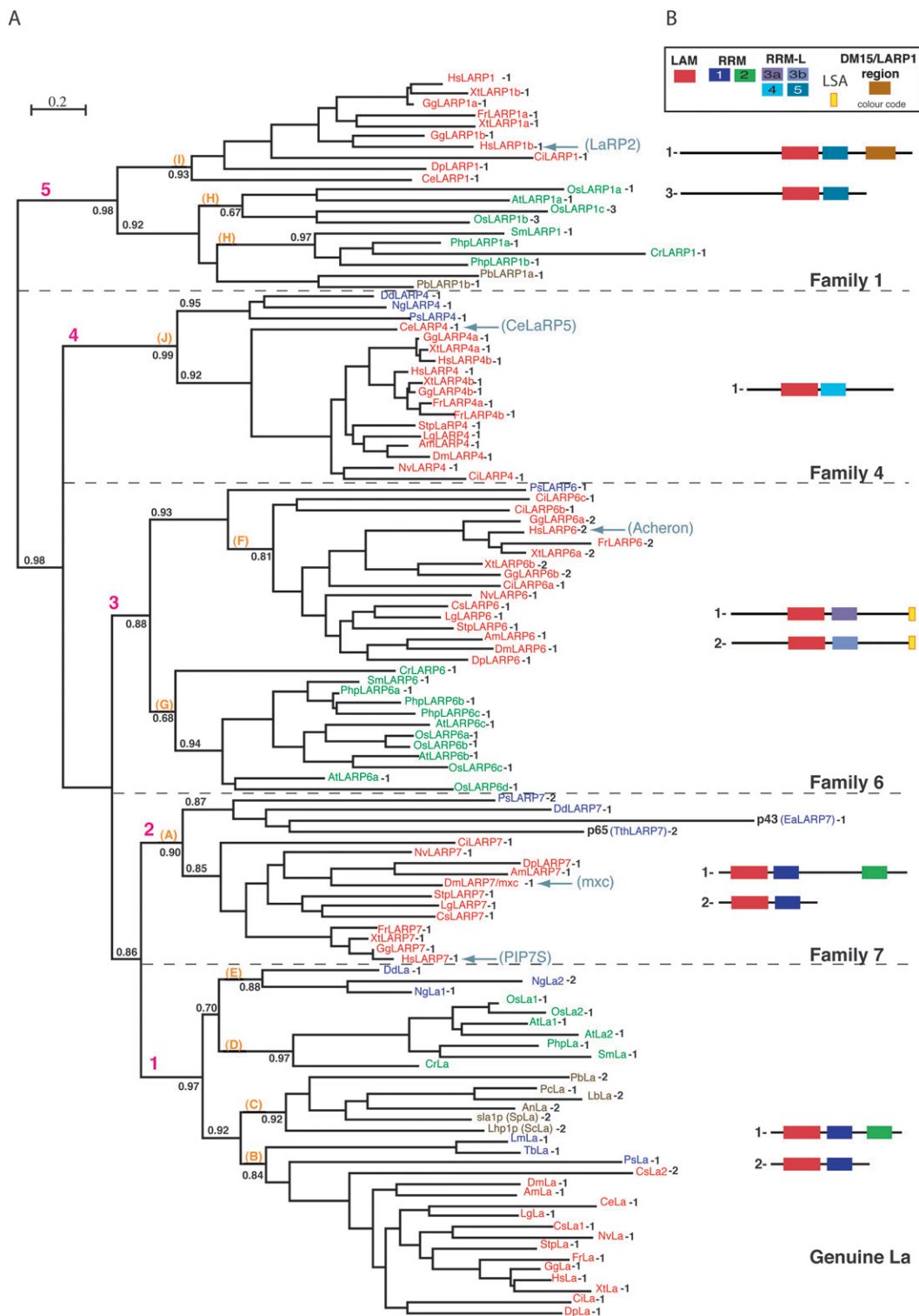


FIGURE 5. Phylogenetic relationships among the LAM-RRM1(RRM-L) domains. (A) Phylogenetic tree obtained using sequences of LAM-RRM1 or LAM-RRM-L regions (see Supplemental Figs. 3–6 for the sequences used) from the 113 LAM proteins and the ciliate p43 and p65 proteins. Selected statistical supports are reported and relevant clusters, defining the different LAM-containing protein families, are labeled with pink numbers (1–5). Phylogenetic groups of proteins previously identified in Figure 1A are also indicated (orange uppercase letters between parentheses). Protein names, species color codes, and species abbreviation names are as in Figure 1, plus: (Ea) *Euplotes aediculatus*, (Th) *Tetrahymena thermophila*. Black numbers following protein names correspond to the structural organization of the protein as represented in B. (B) Schematic representations of the different structural organizations found among the full-length LAM-RRM(RRM-L)-containing proteins. The color code used is reported in the box at the top of the figure. (Dashed lines) Boundaries of each significant cluster.

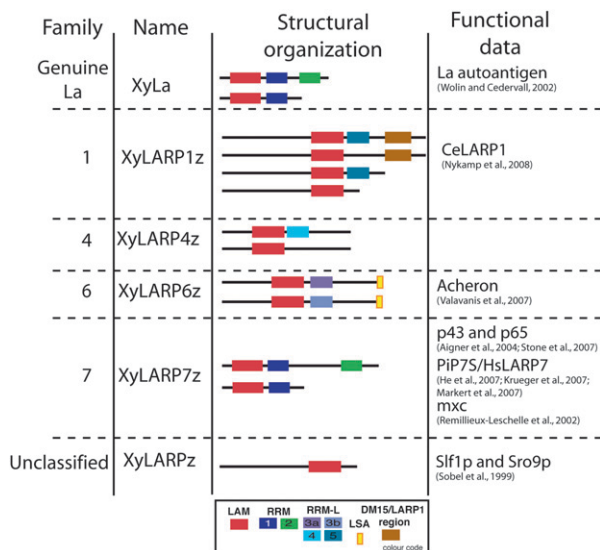


FIGURE 6. Nomenclature and definition of the five families of LAM-containing proteins, summarizing the proposed families definition of LAM-containing proteins. Names are defined according to these rules: Xy: species abbreviation name; La: genuine La or LARP: La-related Protein; 1, 4, 6, or 7; family number, z: lower case letter (a, b, c, ...) added in cases when there are several members of the same family in a given species. The structural organization of each family member is shown. Color code is identical to that of Figure 5B.

dlarp (Chauvet et al. 2000; Ichihara et al. 2007) p43, and p65, we suggest using both the published and presently proposed names (respectively, DmLARP7, DmLARP1, EaLARP7, and TthLARP7) (see Supplemental Fig. 2).

Several proteins of the LARP4 and LARP1 families do not share all structural features of their group and were classified only on the basis of the evolutionary analysis of their LAM (Fig. 1; Supplemental Fig. 2). We cannot exclude that some of these predicted proteins result from the in silico translation of pseudogenes, especially when a species displays other member(s) of the same family with appropriate structural features.

Finally, amongst the 134 factors we studied, five proteins (Fig. 1A, Slf1p, Sro9p, AnLARP, SpLARP, PsLARP) containing only a LAM cannot be clearly included in one of the five proposed families. With the exception of the *S. cerevisiae* proteins Slf1p and Sro9p, which were reported as such in several papers, we propose to name these factors according to the above-mentioned nomenclature but without a family number: XsLARPz (see Fig. 6).

DISCUSSION

Our analysis constitutes the most comprehensive and detailed identification and evolutionary as well as structural analysis of LAM-containing proteins available to date. It clearly demonstrates that LAM-containing factors belong to a superfamily of proteins present in most eukaryotic

organisms and that they can be classified into five families on the basis of their evolutionary as well as structural characteristics. We also identified previously unrecognized domains of probable functional significance.

Functional significance of the RRM-L domains

The RNA recognition motif is one of the most abundant protein domains in eukaryotes. Canonical RRM domains share two primary sequence homologies boxes: RNP1 [(K/R)-G-(F/Y)-(G/A)-(F/Y)-(I/L/V)-X-(F/Y)] and RNP2 [(I/L/V)-(F/Y)-(I/L/V)-X-N-L]; also, they present the characteristic $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ structural fold (Birney et al. 1993; Clery et al. 2008). Folding of this α/β sandwich allows the formation of a four-stranded anti-parallel β -sheet surface, on top of which five extremely conserved residues (underlined in the RNP1/RNP2 consensus sequences reported above) are exposed to the solvent and directly bind to the RNA bait (Maris et al. 2005). This is the prototypical mode of action of most RRM domains, but they are versatile domains that were shown to vary their fold and sequences to accommodate a wide range of substrates. There are numerous reported cases of noncanonical RRM domains that are still able to bind RNA and, conversely, of RRM domains with typical features, binding RNA through nonconventional surfaces or that are unable to bind RNA per se (for review, see Clery et al. 2008). Several reports showed that, in addition to the β -sheet surface, the connecting loops (mainly loops 1, 3, 5, and in one case loop 4), whose lengths vary from one RRM to another, can be crucial for nucleic acid binding (for review, see Clery et al. 2008). The q-RRM (quasi-RRM) domains of the human hnRNP-F protein, which adopt the canonical RRM fold but lack the RNP-1 or RNP-2 conserved residues, were shown to bind RNA solely through their loop 1, 3, and 5 regions (Dominguez and Allain 2006). The genuine La RRM1 domain constitutes another example of a motif that, although displaying the characteristic primary and secondary features of RRM domains, does not engage canonical surfaces or residues in the binding of the 3'-UUU-OH (Maraia and Bayfield 2006) or viral leader RNA (Bitko et al. 2008) and necessitates the concerted *cis* action of another conserved region of the protein, the LAM. In our case, none of the identified RRM-L regions possesses the RNP-1 or RNP-2 conserved residues (Supplemental Figs. 4–6). Nevertheless, they are predicted to adopt the typical RRM fold, and three of them display significant primary sequence conservation between members of the same family. We also found that their connecting loop sizes are strikingly conserved inside a given family even between proteins from extremely diverse origins, suggesting a functional role for these regions. We propose that the RRM-L domains we identified in the present study are new types of noncanonical RRM domains.

The phylogenetic analysis of the LAM-RRM/RRM-L and LAM results in trees that are similar overall, suggesting a

co-evolution of the LAM and RRM/RRM-L regions. We propose that these motifs co-evolved to accommodate different substrates and that, as is the case for genuine La proteins, they function in a concerted manner. Consistently, we noticed that the RRM-L domain systematically lies in the close vicinity of the LAM, spaced from it by a fixed size linker.

Functional significance of the DM15/LARP1 region

The function of the DM15 region remains largely unknown, but it has been proposed to be involved in nucleic acid binding. Nykamp and collaborators showed that the C-terminal half of the CeLARP1 protein retains the ability to bind RNA homopolymers *in vitro* (Nykamp et al. 2008). They found that this region binds to RNA as efficiently as the region containing the LAM-RRM-L5, but less well than the full-length protein (Nykamp et al. 2008). They suggested that the DM15 region has the ability to contact RNA and cooperates with the LAM-RRM-L5 in RNA-binding activity. These results are nonetheless only preliminary, especially since they used the complete C-terminal half of CeLARP1, which contains regions additional to the DM15A and DM15B boxes. Consistent with the idea of a cooperative action of the DM15 and LAM regions, we found DM15 repeats in association only with LAMs. In any case, the strong evolutionary conservation of the DM15 repeats (Fig. 3B) argues in favor of a significant role for this motif.

The LARP7 family

LARP7 family members are the most studied LARP proteins to date. The HsLARP7 protein was shown to bind to the pol III-encoded 7SK snRNA and to be required for its stability and role in pol II functional regulation (He et al. 2008; Krueger et al. 2008; Markert et al. 2008). This noncoding RNA was long believed to be restricted to vertebrates, but recent data reported the existence of 7SK homologs from several invertebrates (Gruber et al. 2008a,b). This includes basal deuterostomes such as *Ciona intestinalis*, mollusks such as *Lottia gigantea*, or worms like *Capitella capitella*. A 7SK snRNA homolog was also found from most arthropods, including the bee *Apis mellifera* and fly *D. melanogaster* (Gruber et al. 2008a). Consistently, we found that these species as well as every tested animal species (with the exception of *C. elegans*) have a potential homolog to the HsLARP7 protein. Moreover, it has been demonstrated that most of the 7SK RNA cellular pool is associated with HsLARP7 and, conversely, that there is very little protein not bound to the 7SK RNA (Krueger et al. 2008). This suggests that the main target of LARP7 proteins in animal species will be the 7SK snRNA. Moreover, since most genomes from

invertebrate species also encode other homologs of the 7SK RNP (Gruber et al. 2008a), it also suggests that the negative regulatory function of the 7SK particle in transcription elongation was maintained during animal evolution. The role for LARP7 proteins in protist species is likely to be different, as no 7SK snRNA has been detected in this group of species and since the p43 (EaLARP7) and p65 (TthLARP7) proteins from *Euplotes aediculatus* and *Tetrahymena thermophila*, respectively, have been shown to be involved in the biology and function of the pol III-encoded telomerase RNA.

More generally, we propose that LARP7 proteins, although structurally very close to genuine La factors, have lost their ability to bind transiently to most pol III transcripts and evolved the capacity to stably associate with a single pol III transcript (7SK snRNA for animals, or the telomerase RNA for certain protists). At the molecular level, LARP7 is likely the first factor to remain stably associated with its RNA partner (7SK RNA, the telomerase RNA, or other noncoding RNA yet to be identified) and to form an initial assembly platform for the RNP (Teixeira and Gilson 2007; He et al. 2008; Krueger et al. 2008).

In conclusion, our data suggest that LAM-containing proteins belonging to the same family not only share structural and evolutionary features, but also might display common functional characteristics, whether by functioning in the same biological process and binding to homologous baits (RNA, DNA, or even proteins) or by adopting the same molecular mode of action on different baits.

MATERIALS AND METHODS

Identification of LAM-containing proteins

Fully sequenced genomes were searched for La-motif-containing proteins using the JGI genome portal (<http://genome.jgi-psf.org/>) or, when not available through JGI, at the NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) genomic databases. Searches through JGI were performed for each complete genome as follows. The genome was first searched using the LAM InterProHit ID number (0066300). Retrieved LAM-containing proteins were curated and redundancy was eliminated. Next, LAMs retrieved in this first step were used as a query to search the protein database using blastp and the translated nucleotide database using tblastn. When possible, i.e., for fully sequenced genomes also available at the NCBI databank, the genome was also searched by iterative psi-Blast. Fully sequenced genomes available only through the NCBI database were searched as follows. An initial protein blast (blastp) was performed using the LAM of the human genuine La protein. Subsequently, the expressed genome was searched as for the JGI site (i.e., the retrieved LAMs were used as the query using blastp, tblastn, and iterative psiblast). For Archaeal and *Plasmodium* genomes, where no hit was obtained by the initial search, each LAM from protist proteins was used as a query to search these genomes by tblastn and psiblast.

Multiple sequences alignments, phylogenetic reconstruction, and secondary structure predictions

Sequences were aligned using the multiple sequence comparison by log-expectation (MUSCLE v3.7) software (Edgar 2004) and the tree was reconstructed using the fast maximum likelihood tree estimation program PHYML (Guindon and Gascuel 2003) using the LG amino acids replacement matrix (Le and Gascuel 2008). Statistical support for the major clusters was obtained using the approximate likelihood-ratio test (aLRT) (Anisimova and Gascuel 2006). Full-length sequences of each of the 134 LAM proteins were analyzed with the hidden-Markov-model (HMM)-based protein structure prediction program SAM-T08 (http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html; Karplus et al. 2005) to identify putative structural segments.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at <http://www.rnajournal.org>.

ACKNOWLEDGMENTS

This work was supported by the CNRS and the Université de Perpignan. We thank Carole Santi and Julio Saez-Vasquez for critical reading of the manuscript.

Received November 24, 2008; accepted January 22, 2009.

REFERENCES

- Adilakshmi, T. and Laine, R.O. 2002. Ribosomal protein S25 mRNA partners with MTF-1 and La to provide a p53-mediated mechanism for survival or death. *J. Biol. Chem.* **277**: 4147–4151.
- Aigner, S. and Cech, T.R. 2004. The Euplotes telomerase subunit p43 stimulates enzymatic activity and processivity in vitro. *RNA* **10**: 1108–1118.
- Aigner, S., Lingner, J., Goodrich, K.J., Grosshans, C.A., Shevchenko, A., Mann, M., and Cech, T.R. 2000. Euplotes telomerase contains an La motif protein produced by apparent translational frameshifting. *EMBO J.* **19**: 6230–6239.
- Aigner, S., Postberg, J., Lipps, H.J., and Cech, T.R. 2003. The Euplotes La motif protein p43 has properties of a telomerase-specific subunit. *Biochemistry* **42**: 5736–5747.
- Alfano, C., Sanfelice, D., Babon, J., Kelly, G., Jacks, A., Curry, S., and Conte, M.R. 2004. Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. *Nat. Struct. Mol. Biol.* **11**: 323–329.
- Anisimova, M. and Gascuel, O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**: 539–552.
- Belisova, A., Semrad, K., Mayer, O., Kocian, G., Waigmann, E., Schroeder, R., and Steiner, G. 2005. RNA chaperone activity of protein components of human Ro RNPs. *RNA* **11**: 1084–1094.
- Birney, E., Kumar, S., and Krainer, A.R. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**: 5803–5816.
- Bitko, V., Musiyenko, A., Bayfield, M.A., Maraia, R.J., and Barik, S. 2008. Cellular La protein shields nonsegmented negative-strand RNA viral leader RNA from RIG-I and enhances virus growth by diverse mechanisms. *J. Virol.* **82**: 7977–7987.
- Brenet, F., Dussault, N., Borch, J., Ferracci, G., Delfino, C., Roepstorff, P., Miquelis, R., and Ouafik, L. 2005. Mammalian peptidylglycine α -amidating monooxygenase mRNA expression can be modulated by the La autoantigen. *Mol. Cell. Biol.* **25**: 7505–7521.
- Bycroft, M., Hubbard, T.J., Proctor, M., Freund, S.M., and Murzin, A.G. 1997. The solution structure of the S1 RNA binding domain: A member of an ancient nucleic acid-binding fold. *Cell* **88**: 235–242.
- Cardinali, B., Carissimi, C., Gravina, P., and Pierandrei-Amaldi, P. 2003. La protein is associated with terminal oligopyrimidine mRNAs in actively translating polysomes. *J. Biol. Chem.* **278**: 35145–35151.
- Chakshumathi, G., Kim, S.D., Rubinson, D.A., and Wolin, S.L. 2003. A La protein requirement for efficient pre-tRNA folding. *EMBO J.* **22**: 6562–6572.
- Chauvet, S., Maurel-Zaffran, C., Miassod, R., Jullien, N., Pradel, J., and Aragnol, D. 2000. dlarp, a new candidate Hox target in *Drosophila* whose orthologue in mouse is expressed at sites of epithelium/mesenchymal interactions. *Dev. Dyn.* **218**: 401–413.
- Clery, A., Blatter, M., and Allain, F.H. 2008. RNA recognition motifs: Boring? Not quite. *Curr. Opin. Struct. Biol.* **18**: 290–298.
- Copela, L.A., Chakshumathi, G., Sherrer, R.L., and Wolin, S.L. 2006. The La protein functions redundantly with tRNA modification enzymes to ensure tRNA structural stability. *RNA* **12**: 644–654.
- Costa-Mattioli, M., Svitkin, Y., and Sonenberg, N. 2004. La autoantigen is necessary for optimal function of the poliovirus and hepatitis C virus internal ribosome entry site in vivo and in vitro. *Mol. Cell. Biol.* **24**: 6861–6870.
- Curry, S. and Conte, M.R. 2006. A terminal affair: 3'-end recognition by the human La protein. *Trends Biochem. Sci.* **31**: 303–305.
- Dominguez, C. and Allain, F.H. 2006. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: A novel mode of RNA recognition. *Nucleic Acids Res.* **34**: 3634–3645.
- Domitrovich, A.M., Diebel, K.W., Ali, N., Sarker, S., and Siddiqui, A. 2005. Role of La autoantigen and polypyrimidine tract-binding protein in HCV replication. *Virology* **335**: 72–86.
- Dong, G., Chakshumathi, G., Wolin, S.L., and Reinisch, K.M. 2004. Structure of the La motif: A winged helix domain mediates RNA binding via a conserved aromatic patch. *EMBO J.* **23**: 1000–1007.
- Edgar, R.C. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113. doi: 10.1186/1471-2105-5-113.
- Fleurdepine, S., Deragon, J.M., Devic, M., Guillemot, J., and Bousquet-Antonelli, C. 2007. A bona fide La protein is required for embryogenesis in *Arabidopsis thaliana*. *Nucleic Acids Res.* **35**: 3306–3321.
- Foldynova-Trantirkova, S., Paris, Z., Sturm, N.R., Campbell, D.A., and Lukes, J. 2005. The *Trypanosoma brucei* La protein is a candidate poly(U) shield that impacts spliced leader RNA maturation and tRNA intron removal. *Int. J. Parasitol.* **35**: 359–366.
- Gruber, A.R., Kilgus, C., Mosig, A., Hofacker, I.L., Hennig, W., and Stadler, P.F. 2008a. Arthropod 7SK RNA. *Mol. Biol. Evol.* **25**: 1923–1930.
- Gruber, A.R., Koper-Emde, D., Marz, M., Tafer, H., Bernhart, S., Obernosterer, G., Mosig, A., Hofacker, I.L., Stadler, P.F., and Benecke, B.J. 2008b. Invertebrate 7SK snRNAs. *J. Mol. Evol.* **66**: 107–115.
- Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**: 696–704.
- He, N., Jahchan, N.S., Hong, E., Li, Q., Bayfield, M.A., Maraia, R.J., Luo, K., and Zhou, Q. 2008. A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. *Mol. Cell.* **29**: 588–599.
- Holcik, M. and Korneluk, R.G. 2000. Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: Role of La autoantigen in XIAP translation. *Mol. Cell. Biol.* **20**: 4648–4657.

- Huang, Y., Bayfield, M.A., Intine, R.V., and Maraia, R.J. 2006. Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. *Nat. Struct. Mol. Biol.* **13**: 611–618.
- Ichihara, K., Shimizu, H., Taguchi, O., Yamaguchi, M., and Inoue, Y.H. 2007. A *Drosophila* orthologue of larp protein family is required for multiple processes in male meiosis. *Cell Struct. Funct.* **32**: 89–100.
- Inada, M. and Guthrie, C. 2004. Identification of Lhp1p-associated RNAs by microarray analysis in *Saccharomyces cerevisiae* reveals association with coding and noncoding RNAs. *Proc. Natl. Acad. Sci.* **101**: 434–439.
- Intine, R.V., Dundr, M., Misteli, T., and Maraia, R.J. 2002. Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs. *Mol. Cell* **9**: 1113–1123.
- Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P.D., Curry, S., and Conte, M.R. 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.
- Kadaba, S., Wang, X., and Anderson, J.T. 2006. Nuclear RNA surveillance in *Saccharomyces cerevisiae*: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA* **12**: 508–521.
- Karplus, K., Katzman, S., Shackelford, G., Koeva, M., Draper, J., Barnes, B., Soriano, M., and Hughey, R. 2005. SAM-T04: What is new in protein-structure prediction for CASP6. *Proteins (Suppl. 7)* **61**: 135–142.
- Kim, Y.K., Back, S.H., Rho, J., Lee, S.H., and Jang, S.K. 2001. La autoantigen enhances translation of BiP mRNA. *Nucleic Acids Res.* **29**: 5009–5016.
- Kloks, C.P., Spronk, C.A., Lasonder, E., Hoffmann, A., Vuister, G.W., Grzesiek, S., and Hilbers, C.W. 2002. The solution structure and DNA-binding properties of the cold-shock domain of the human Y-box protein YB-1. *J. Mol. Biol.* **316**: 317–326.
- Kotik-Kogan, O., Valentine, E.R., Sanfelice, D., Conte, M.R., and Curry, S. 2008. Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein. *Structure* **16**: 852–862.
- Krueger, B.J., Jeronimo, C., Roy, B.B., Bouchard, A., Barrandon, C., Byers, S.A., Searcey, C.E., Cooper, J.J., Bensaude, O., Cohen, E.A., et al. 2008. LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. *Nucleic Acids Res.* **36**: 2219–2229.
- Kufel, J., Allmang, C., Chanfreau, G., Petfalski, E., Lafontaine, D., and 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.
- Lan, C., Lee, H.C., Tang, S., and Zhang, L. 2004. A novel mode of chaperone action: Heme activation of Hap1 by enhanced association of Hsp90 with the repressed Hsp70-Hap1 complex. *J. Biol. Chem.* **279**: 27607–27612.
- Le, S.Q. and Gascuel, O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**: 1307–1320.
- Maraia, R.J. and Bayfield, M.A. 2006. The La protein–RNA complex surfaces. *Mol. Cell* **21**: 149–152.
- Maraia, R.J. and Intine, R.V. 2001. Recognition of nascent RNA by the human La antigen: Conserved and diverged features of structure and function. *Mol. Cell. Biol.* **21**: 367–379.
- Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.L., Jackson, J.D., et al. 2007. CDD: A conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* **35**: D237–D240.
- Maris, C., Dominguez, C., and Allain, F.H. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**: 2118–2131.
- Markert, A., Grimm, M., Martinez, J., Wiesner, J., Meyerhans, A., Meyuhas, O., Sickmann, A., and Fischer, U. 2008. The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. *EMBO Rep.* **9**: 569–575.
- McLaren, R.S., Caruccio, N., and Ross, J. 1997. Human La protein: A stabilizer of histone mRNA. *Mol. Cell. Biol.* **17**: 3028–3036.
- Mollenbeck, M., Postberg, J., Paeschke, K., Rossbach, M., Jonsson, F., and Lipps, H.J. 2003. The telomerase-associated protein p43 is involved in anchoring telomerase in the nucleus. *J. Cell Sci.* **116**: 1757–1761.
- Nykamp, K., Lee, M.H., and Kimble, J. 2008. *C. elegans* La-related protein, LARP-1, localizes to germline P bodies and attenuates Ras-MAPK signaling during oogenesis. *RNA* **14**: 1378–1389.
- Pannone, B.K., Xue, D., and Wolin, S.L. 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.
- Park, J.M., Intine, R.V., and Maraia, R.J. 2007. Mouse and human La proteins differ in kinase substrate activity and activation mechanism for tRNA processing. *Gene Expr.* **14**: 71–81.
- Raha, T., Pudi, R., Das, S., and Shaila, M.S. 2004. Leader RNA of Rinderpest virus binds specifically with cellular La protein: A possible role in virus replication. *Virus Res.* **104**: 101–109.
- Remillieux-Leschelle, N., Santamaria, P., and Randsholt, N.B. 2002. Regulation of larval hematopoiesis in *Drosophila melanogaster*: A role for the multisex combs gene. *Genetics* **162**: 1259–1274.
- Rosenblum, J.S., Pemberton, L.F., Bonifaci, N., and Blobel, G. 1998. Nuclear import and the evolution of a multifunctional RNA-binding protein. *J. Cell Biol.* **143**: 887–899.
- Saget, O., Forquignon, F., Santamaria, P., and Randsholt, N.B. 1998. Needs and targets for the multi sex combs gene product in *Drosophila melanogaster*. *Genetics* **149**: 1823–1838.
- Simons, F.H., Broers, F.J., Van Venrooij, W.J., and Pruijn, G.J. 1996. Characterization of cis-acting signals for nuclear import and retention of the La (SS-B) autoantigen. *Exp. Cell Res.* **224**: 224–236.
- Sobel, S.G. and Wolin, S.L. 1999. Two yeast La motif-containing proteins are RNA-binding proteins that associate with polyribosomes. *Mol. Biol. Cell* **10**: 3849–3862.
- Sommerville, J. 1999. Activities of cold-shock domain proteins in translation control. *Bioessays* **21**: 319–325.
- Stefano, J.E. 1984. Purified lupus antigen La recognizes an oligouridylylate stretch common to the 3' termini of RNA polymerase III transcripts. *Cell* **36**: 145–154.
- Stone, M.D., Mihalusova, M., O'Connor, C.M., Prathapam, R., Collins, K., and Zhuang, X. 2007. Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446**: 458–461.
- Tan, Q., Li, X., Sadhale, P.P., Miyao, T., and Woychik, N.A. 2000. Multiple mechanisms of suppression circumvent transcription defects in an RNA polymerase mutant. *Mol. Cell. Biol.* **20**: 8124–8133.
- Teixeira, M.T. and Gilson, E. 2007. La sets the tone for telomerase assembly. *Nat. Struct. Mol. Biol.* **14**: 261–262.
- Teplova, M., Yuan, Y.R., Phan, A.T., Malinina, L., Ilin, S., Teplov, A., and Patel, D.J. 2006. Structural basis for recognition and sequestration of UUU(OH) 3' termini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol. Cell* **21**: 75–85.
- Trotta, R., Vignudelli, T., Candini, O., Intine, R.V., Pecorari, L., Guerzoni, C., Santilli, G., Byrom, M.W., Goldoni, S., Ford, L.P., et al. 2003. BCR/ABL activates mdm2 mRNA translation via the La antigen. *Cancer Cell* **3**: 145–160.
- Valavanis, C., Wang, Z., Sun, D., Vaine, M., and Schwartz, L.M. 2007. Acheron, a novel member of the Lupus Antigen family, is induced during the programmed cell death of skeletal muscles in the moth *Manduca sexta*. *Gene* **393**: 101–109.
- Van Horn, D.J., Yoo, C.J., Xue, D., Shi, H., and Wolin, S.L. 1997. The La protein in *Schizosaccharomyces pombe*: A conserved yet dispensable phosphoprotein that functions in tRNA maturation. *RNA* **3**: 1434–1443.

- Vazquez-Pianzola, P., Urlaub, H., and Rivera-Pomar, R. 2005. Proteomic analysis of reaper 5' untranslated region-interacting factors isolated by tobramycin affinity-selection reveals a role for La antigen in reaper mRNA translation. *Proteomics* **5**: 1645–1655.
- Witkin, K.L. and Collins, K. 2004. Holoenzyme proteins required for the physiological assembly and activity of telomerase. *Genes & Dev.* **18**: 1107–1118.
- Wolin, S.L. and Cedervall, T. 2002. The La protein. *Annu. Rev. Biochem.* **71**: 375–403.
- Xue, D., Rubinson, D.A., Pannone, B.K., Yoo, C.J., and Wolin, S.L. 2000. U snRNP assembly in yeast involves the La protein. *EMBO J.* **19**: 1650–1660.
- Xue, Q., Ding, H., Liu, M., Zhao, P., Gao, J., Ren, H., Liu, Y., and Qi, Z.T. 2007. Inhibition of hepatitis C virus replication and expression by small interfering RNA targeting host cellular genes. *Arch. Virol.* **152**: 955–962.
- Yoo, C.J. and Wolin, S.L. 1994. La proteins from *Drosophila melanogaster* and *Saccharomyces cerevisiae*: A yeast homolog of the La autoantigen is dispensable for growth. *Mol. Cell. Biol.* **14**: 5412–5424.
- Yoo, C.J. and Wolin, S.L. 1997. The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell* **89**: 393–402.