Interaction of the non coding RNA 7SK, a regulator of human transcription elongation, with the LaRP7 protein

Xiao Han

To cite this version:


HAL Id: tel-01581116
https://tel.archives-ouvertes.fr/tel-01581116
Submitted on 4 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
THÈSE DE DOCTORAT

de l’Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée dans le cadre d’une cotutelle entre
ECOLE NORMALE SUPÉRIEURE
et EAST CHINA NORMAL UNIVERSITY

Interaction of the non-coding RNA 7SK, a regulator of human transcription elongation, with the LaRP7 protein

Interaction de l’ARN non-codant 7SK, un régulateur de la transcription chez l’homme, avec la protéine LARP7

Ecole doctorale n°515
Complexité du vivant
Spécialité BIOLOGIE MOLECULAIRE

Soutenue par XIAO HAN
le 20 juillet 2016

Dirigée par
Anne-Catherine DOCK BREGEON
# INDEX

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDEX</td>
<td>1</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>2</td>
</tr>
<tr>
<td>CHAPTER I: INTRODUCTION</td>
<td>4</td>
</tr>
<tr>
<td>1. Non-coding RNAs</td>
<td>4</td>
</tr>
<tr>
<td>1.1 Non-coding RNAs</td>
<td>4</td>
</tr>
<tr>
<td>1.2 7SK RNA</td>
<td>4</td>
</tr>
<tr>
<td>2. 7SK RNA function in transcription control</td>
<td>5</td>
</tr>
<tr>
<td>2.1 P-TEFb releases the transcriptional pausing</td>
<td>6</td>
</tr>
<tr>
<td>2.2 P-TEFb phosphorylates RNA polymerase II</td>
<td>7</td>
</tr>
<tr>
<td>2.3 Recruitment of P-TEFb is essential for HIV transcription</td>
<td>9</td>
</tr>
<tr>
<td>3. The human 7SK RNA snRNP</td>
<td>10</td>
</tr>
<tr>
<td>3.1 Role of 7SK RNA in P-TEFb inhibition</td>
<td>10</td>
</tr>
<tr>
<td>3.2 Molecular description of 7SK RNA</td>
<td>14</td>
</tr>
<tr>
<td>3.3 Proteins binding the 7SK RNA</td>
<td>17</td>
</tr>
<tr>
<td>4. LARP7 and LARPs</td>
<td>23</td>
</tr>
<tr>
<td>4.1 La protein</td>
<td>23</td>
</tr>
<tr>
<td>4.2 La-related proteins (LARP) family</td>
<td>24</td>
</tr>
<tr>
<td>4.3 LARP7</td>
<td>25</td>
</tr>
<tr>
<td>5. Investigating the interactions between 7SK RNA and LARP7</td>
<td>28</td>
</tr>
<tr>
<td>CHAPTER II: Material and methods</td>
<td>29</td>
</tr>
<tr>
<td>1. Preparation of molecules</td>
<td>29</td>
</tr>
<tr>
<td>1.1 7SK RNA purification</td>
<td>29</td>
</tr>
<tr>
<td>1.2 LARP7 purification</td>
<td>32</td>
</tr>
<tr>
<td>2. Binding assays</td>
<td>36</td>
</tr>
<tr>
<td>2.1 Footprinting experiments</td>
<td>36</td>
</tr>
<tr>
<td>2.2 EMSA</td>
<td>43</td>
</tr>
<tr>
<td>Chapter III: Characterization of LARP7 binding to 7SK RNA by biochemical approaches: involvement of the 3'hairpin HP4. (Publication No.1)</td>
<td>45</td>
</tr>
<tr>
<td>Supplementary material</td>
<td>48</td>
</tr>
<tr>
<td>Discussion</td>
<td>60</td>
</tr>
<tr>
<td>Chapter IV: LARP7 binding to the middle-region of 7SK RNA</td>
<td>62</td>
</tr>
<tr>
<td>1. LARP7 still binds 7SK RNA after deleting HP4</td>
<td>62</td>
</tr>
<tr>
<td>2. LARP7 interacts with 7SK RNA sequence in domain 2</td>
<td>63</td>
</tr>
<tr>
<td>2.1 Mutation in domain 2 affects LARP7\’s binding</td>
<td>66</td>
</tr>
<tr>
<td>Discussion</td>
<td>68</td>
</tr>
<tr>
<td>Chapter V: LARP7 binding to HP1 and HP3</td>
<td>72</td>
</tr>
<tr>
<td>1. LARP7 and HEXIM\’s interactions with hairpins of 7SK RNA</td>
<td>72</td>
</tr>
<tr>
<td>2. Interactions with HP3</td>
<td>74</td>
</tr>
<tr>
<td>3. LARP7 interaction with HP1</td>
<td>76</td>
</tr>
<tr>
<td>4. Specificity of the C-terminal domain binding</td>
<td>83</td>
</tr>
<tr>
<td>Discussion</td>
<td>84</td>
</tr>
<tr>
<td>Chapter VI: LARP7 impact on HEXIM binding</td>
<td>86</td>
</tr>
<tr>
<td>1. HEXIM binds 7SK RNA using ARM sequences in both monomers</td>
<td>86</td>
</tr>
<tr>
<td>Supplementary Material</td>
<td>90</td>
</tr>
<tr>
<td>Chapter VII: General discussion and conclusion</td>
<td>96</td>
</tr>
<tr>
<td>1. LARP7 wraps around 7SK RNA</td>
<td>96</td>
</tr>
<tr>
<td>2. Assembly and dynamics of the 7SK RNA snRNP</td>
<td>99</td>
</tr>
<tr>
<td>3. Impact of LARP7 on HEXIM recognition</td>
<td>101</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>104</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I remember the first day when I came to ENS, a stranger in another side of the road said “bonjour” to me with sweet smile. It was a sunny day. Now, four years passed, I have almost finished my PhD study, full of grateful in my heart.

I am deeply grateful to my supervisor, Anne-Catherine Dock-Bregeon, not only for her endless passion in science inspired me, but also her honest pushed me to be better. I didn’t have a structure molecular biology background when I came to this lab. I have learned a lot from Anne-Catherine and I couldn’t finish my PhD work without her strong support.

I want to express my gratitude to Dr. Olivier Bensaude. He always tries his best when I need help. He taught me the method to conquer the problems I met. He also did many contributions to the cooperation between ENS and ECNU.

I appreciate all members of the Jury, Dr. Pascale Romby, Dr. Olivier Bensaude, Pro. Maria R Conte, Pro. Catherine Vénien-Bryan and Pro Yuan Chonggang. I would like to thank Thierry Foulon, Carine Tisne and Ye Shixin to be my committee members.

It is a pleasure to thank all the members in our lab. Lydia Kobbi, Florence Proux, Floriane Braye, Bo Gu, Nina Verstraete and Trung Nguyen Van. We work hard and always help each other sincerely. I will remember the time we spent together. Once, Florence told everybody “Xiao’s gel is the best in the world” when I was making a long gel. It really encouraged me. Thank you!
I want to thank Emiko Uchikawa and Denise Martinez Zapien, they are the PhD students before I joined in this lab. They have done many efforts to the projects that I benefit from.

I would like to thank the Genomics technology platform and all staffs and students who worked with me in IBENS. Especially Hervé Le Hir, Jérôme Saulière, Brigitte Arnaud, Abdul-Karim Konate, Zhen Wang, Isabelle Barbosa, Rahul Kumar Mishra, Emelie Marquenet, Veronique Tanty, Chloé Connan, and Francesco De Carli. I would also acknowledge Elodie Zhang and Emilie Moreau who did their internship with us.

I really appreciate for the four years financially support by Chinese Scholarship Council (CSC). I am heartily thankful to the cooperation between ENS and ECNU, which gives me the chance to study in France. Especially Professor Yuan Chonggang, who suggested me to join in this programme. I was also able to learn a lot of experimental skills at Yuan’s lab.

Thanks to Ye Shixin, who is not only my committee member, but also a very precious friend. Thanks to Xiang Feng, Tian Meilin, Elodie Zhang, Zhang Jing, Xu Lijun, Xu Qiongfei, Li Yuan and Wang Shichao. They are my best friends in Paris. Thanks to my parents, my husband and my family for their forever supports. Thanks to myself, still keep an optimistic attitude and a beautiful smiley face after almost 4 years very hard work. :)
CHAPTER I: INTRODUCTION

1. Non-coding RNAs

1.1 Non-coding RNAs

7SK RNA is a non-coding RNA. Unlike messenger RNA (mRNAs), non-coding RNAs (ncRNA) are not translated into proteins. The first non-coding RNA was found in 1965, a baker’s yeast alanine transfer RNA (tRNA) (Holley, 1965). tRNA and ribosomal RNA (rRNA) were wildly known as functional ncRNAs in 1970s. Eukaryotic non-coding RNAs are transcribed from all three RNA polymerases (RNA Pol) I, II, or III. Some of them appear to regulate RNA Pol II transcription (Goodrich and Kugel, 2006). More and more ncRNAs are discovered to be involved in regulation of gene expression. ncRNAs that regulate transcription are not as well understood as protein regulators. Nevertheless, ncRNAs can be divided into two groups according to their differences in regulation mechanism: Some control transcription by mediating changes in chromatin structure such as in X chromosome inactivation and others directly control the assembly or activity of transcription factor complexes (Kugel and Goodrich, 2012). 7SK RNA belongs to the latter category.

1.2 7SK RNA

7SK RNA was discovered in HeLa cell nuclei in 1977, it was named snK RNA initially (Zieve et al., 1977). In 1982, Ullu and coworkers renamed it as 7SK RNA (Ullu and Melli, 1982). 7SK RNA complete nucleotide sequence was obtained in 1984 (Reddy et al., 1984). It is an abundant nuclear RNA of medium size (331 nucleotides in mammals).
It is transcribed by RNA Pol III. 7SK RNA gene has an uridine rich region at its 3’ end used as a termination signal for RNA Pol III (Murphy et al., 1987). It is monomethylated on the γ-phosphate of the first nucleotide at its 5’end (Gupta et al., 1990) (Shumyatsky et al., 1990) (Shumyatsky et al., 1993). Methylation of the γ-phosphate of the first nucleotide in 5’ is shared with a few other RNA Pol III transcripts such as U6 and B2 RNAs. Methylation protects 7SK RNA from degradation.

7SK RNA is strongly conserved in higher vertebrates (Gursoy et al., 2000). Homologues are also found in annelid and mollusk (Gruber et al., 2008b). A related RNA Pol III transcript was found in Drosophila and insect genomes (Gruber et al., 2008a) and shown to display a function similar to that of human 7SK RNA (Nguyen et al., 2012). A 7SK RNA homolog might be present in C. elegans (Marz et al., 2009). However, there is biochemical and genetic evidence that the latter is involved in pre-rRNA processing (Hokii et al., 2010). 7SK RNA conservation during evolution suggests that it plays an important role in cell function which have been discussed in numerous reviews (Diribarne and Bensaude, 2009) (Rice, 2015) (Quaresma et al., 2016).

2. 7SK RNA function in transcription control

7SK RNA Function
The first function for 7SK RNA was proposed in 2001. In mammalian cells, 7SK RNA was found associated to an inactive form of the Positive Elongation Factor (P-TEFb) (Nguyen et al., 2001) (Yang et al., 2001). A similar association was later shown to exist in Drosophila (Nguyen et al., 2012). Other functions of 7SK RNA were discovered more recently and will be discussed further.
2.1 P-TEFb releases the transcriptional pausing

Transcription is the first step of gene expression, in which a particular gene is transcribed into RNA by an enzyme named RNA polymerase. Bacteria and Archaea have only one RNA polymerase. In eukaryotes, there are three nuclear RNA polymerases. RNA Pol I synthesizes ribosomal rRNAs. RNA Pol II is responsible for synthesizing messengers mRNA and some small nuclear snRNAs. RNA Pol III synthesizes non-coding RNAs less than 400nt, for example, tRNA, 7SK RNA, U6 snRNA. RNA Pol II plays an important role as it transcribes the genes coding for proteins.

Transcription includes three main steps: initiation, elongation and termination, which are all submitted to multiple regulation processes. In addition to tight regulation at the level of initiation, RNA Pol II is also regulated during elongation. After successfully synthesizing a short nascent RNA chain, RNA Pol II stops or pauses. This promoter-proximal pausing is caused by the DRB sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF). The pausing is mainly alleviated by phosphorylation of DSIF and NELF by the Positive transcription elongation factor b (P-TEFb) (Peterlin and Price, 2006) (Figure 1-1). Thus, P-TEFb is required to allow transcription of most genes by RNA Pol II.
2.2 P-TEFb phosphorylates RNA polymerase II

P-TEFb comprises a kinase subunit, the cyclin-dependent kinase 9 (CDK9), and its regulatory partner, a cyclin T subunit (Figure 1-3). CDK9 is inhibited by DRB. The structure of P-TEFb was first reported by Baumli (Baumli et al., 2008). Besides DSIF and NELF, P-TEFb also phosphorylates the CTD of the largest RNA Pol II. RNA Pol II is an enzyme with 12-15 subunits, depending on the organism. In contrast to RNA Pol I or RNA Pol III, RNA Pol II possesses an extended carboxy-terminal domain (CTD) at its largest subunit, Rpb1. The CTD contains heptapeptide repeats, YSPTSPS (up to 52 in mammals). Phosphorylation occurs in the CTD at serine 2 and 5 of the heptapeptide repeat (Ser2 and Ser5) (Buratowski, 2009). Serine 7 (Ser7) phosphorylation has been
discovered more recently (Boeing et al., 2010). Different modification states of the CTD are characteristic of different transcriptional stages (Egloff and Murphy, 2008) (Figure 1-2). Modification of RNA Pol II CTD varies during all the steps of transcription between hypophosphorylated (RNA Pol IIa) and hyperphosphorylated (RNA Pol IIo) forms. P-TEFb phosphorylates the CTD at the Ser2 position, yielding an RNA Pol IIo. This is required to ensure proper termination and co-transcriptional splicing (Ahn et al., 2004).

**Figure 1-2.** The phosphorylation state of the CTD changes during transcription. The phosphorylation positions at the heptad repeat of the CTD during the transcription cycle are indicated. (Cited from Denise Martinez Zapien’s PhD thesis).
Figure 1-3. Structural organization of P-TEFb bound to Tat (Tahirov et al., 2010). The 3D structure of CDK9 is colored in yellow. In the active site, a Mg ion (green sphere) and the ATP (pink stick) are pointed by a red arrow. The Thr186 at the T-loop is shown as “stick” model (pink). 3D structure of Cyclin T colored in blue. Tat is shown in green sticks.

2.3 Recruitment of P-TEFb is essential for HIV transcription

The importance of P-TEFb was first discovered in studies of the human immunodeficiency virus (HIV) infection. In productive HIV infection, full length transcription of the viral RNA depends on the recruitment of cellular P-TEFb to a region located at the 5’-end of the viral transcript (Peterlin and Price, 2006). This region forms an RNA structure named TAR (transactivation response) which is recognized by the viral protein Tat (transactivator of transcription) (Karn, 1991, 1999). Together they are able to recruit P-TEFb, thus acting as a switch to turn transcription “on”. In the latency state, Tat is not produced by cells having integrated the HIV genome and transcription ends just
after the TAR region. Poorly defined stimuli trigger a burst of transcription and production of Tat that then induces productive transcription of the whole genome. At the molecular level, the TAR/Tat interaction results in a combined interface able to bind human cyclin T1, the activation subunit of the binary P-TEFb complex (Figure 1-3). Mechanistically, this stabilizes the activation loop (regulation loop that is shown in Figure 1-3) of the CDK9 catalytic subunit in a conformation suitable for catalysis (Tahirov et al., 2010). This is the best-known case of a protein-RNA complex controlling P-TEFb activity.

3. The human 7SK RNA snRNP

3.1 Role of 7SK RNA in P-TEFb inhibition

During its functional cycle, 7SK RNA interacts with several protein partners. It is constantly embedded in a complex (7SK snRNP) where it is stabilized by proteins LARP7 (He et al., 2008; Krueger et al., 2008; Markert et al., 2008) and MePCE (Jeronimo et al., 2007), which will be described later. The stabilization of 7SK RNA appears to be necessary for its function, which is not regulated by cyclic processes of degradation and synthesis.

Inhibition of P-TEFb involves HEXIM proteins binding to 7SK RNA

The first proteins identified to bind 7SK RNA are the HEXIM proteins (Michels et al., 2003) (Yik et al., 2003). HEXIM knockdown in flies lead to strong phenotypes and organ failures (Nguyen et al., 2016). They found HEXIM affects the Hedgehog signaling pathway and this is the first time that the physiological function of HEXIM has been addressed in such details in vivo.
7SK RNA binds HEXIM proteins and turns them into P-TEFb inhibitors (Michels et al., 2004) (Yik et al., 2004). So far, two HEXIM proteins have been identified in mammals, HEXIM1 and HEXIM2 (Byers et al., 2005). In other species, only one HEXIM gene is found (Marz et al., 2009). The interaction characteristics of C.elegans protein homologues in a two-hybrid assay supported the existence of a functional Hexim homologue in nematodes (Verstraete et al., 2014). Since they appear to carry out the same function and show the same molecular organization, all are referred with the generic name HEXIM in this thesis.

As will be detailed later, 7SK RNA is constantly bound to its chaperone, LARP7 (He et al., 2008) and its capping enzyme, MePCE (Gupta et al., 1990). Thus, binding to P-TEFb results in a large 7SK RNA/HEXIM/ P-TEFb complex of approx, 700 kDa (Dulac et al., 2005). Stress (UV) or drugs (flavopiridol, DRB) that arrest transcription induce reversion of the inactive 7SK RNA/HEXIM/P-TEFb complex into free, active P-TEFb (Yang et al., 2001) (Nguyen et al., 2001). A fraction of 7SK RNA that is not involved in 7SK RNA/HEXIM/P-TEFb formation, interacts with RNA helicase A (RHA), heterogeneous nuclear ribonucleoprotein A1 (hnRNP), A2/B1, R and Q proteins (Barrandon et al., 2007; Van Herreweghe et al., 2007b). hnRNPs released from the nascent transcripts binds the released 7SK RNA and promote transcription-dependent dissociation of 7SK RNA/ HEXIM/ P-TEFb. Recent evidence suggests that AFF1 accompanies P-TEFb in 7SK RNA snRNP as well (Lu et al., 2014).

Recruitment of 7SK RNA on promoters

7SK RNA seems to function as a regulator of the storage and delivery of P-TEFb. There
are still major issues regarding the function of 7SK RNA. First is to understand how 7SK RNA is addressed to chromatin by the chromatin adaptor factors, then how P-TEFb is released from 7SK RNA snRNP. This may involve P-TEFb release factors, then how this allows the assembly of P-TEFb with transcription factor into super elongation complexes. Finally, the regulation of the sequestration of P-TEFb back into 7SK RNA snRNP upon transcription shutdown is still under investigation. Some data highlight this regulation in the case of HIV promoters (Barboric and Lenasi, 2010) (D'Orso and Frankel, 2010) (McNamara et al., 2013). 7SK RNA has been observed at the HIV promoter (D'Orso and Frankel, 2010). The HIV Tat protein binds 7SK RNA (Krueger et al., 2010). As this binding competes with HEXIM binding (Schulte et al., 2005), it might contribute to dissociate HEXIM from P-TEFb (Muniz et al., 2010). Recruitment of 7SK RNA snRNP to HIV-1 and specific human gene promoters could also be achieved via the non-histone chromatin protein HMGA1 (Eilebrecht et al., 2011). HMGA1 interacts with 7SK RNA both in vitro and in vivo. 7SK RNA snRNP anchored to chromatin recruits P-TEFb which would next be dissociated and stimulate RNA Pol II escape from pausing (McNamara et al., 2013). Indeed, KAP1 recruitment of the 7SK RNA snRNP complex to promoters was found to enable transcription elongation by RNA polymerase II (McNamara et al., 2016). The PPM1G phosphatase directly binds 7SK RNA and forms a heterodimer with HEXIM (henceforth referred to as “7SK RNA-PPM1G snRNP”) preventing P-TEFb to reassemble with HEXIM (Gudipaty and D'Orso, 2016). Thereby, 7SK RNA-PPM1G interaction promotes NF-KB mediated RNA Pol II transcription. Instead of contributing to gene activation, 7SK RNA inhibits pervasive enhancer transcription by modulating nucleosome position (Flynn et al., 2016).
Release of P-TEFb from 7SK RNA snRNP on chromatin

How is P-TEFb released from HEXIM:7SK? A possible conformational switch of the RNA was hypothesized (Krueger et al., 2010). RNA helicases might catalyze the conformational switch. RNA helicase RHA binds 7SK RNA (Van Herreweghe et al., 2007). Furthermore, the promoter-bound DEAD-box RNA helicase DDX21 widely associates with Pol I- and Pol II-transcribed genes and with diverse species of RNA, most prominently with non-coding RNAs involved in the formation of ribonucleoprotein complexes, including ribosomal RNA, small nuclear RNAs and 7SK RNA. Promoter-bound DDX21 facilitates the release of the positive transcription elongation factor b (P-TEFb) from the 7SK snRNP in a manner that is dependent on its helicase activity, thereby promoting transcription of its target genes (Calo et al., 2015).

An exchange, on the 7SK RNA scaffold, of the bound HEXIM for Tat, has been hypothesized (Muniz et al., 2010), but the mechanics of this exchange still needs to be clarified to explain how Tat displaces HEXIM from 7SK RNA. The splicing factor SRSF1 (SF2/ASF) and Tat recognize overlapping sequences within TAR and the 7SK RNA (Kamieniak et al., 2015). It is proposed that SRSF1 activates transcription in the early stages of HIV infection by recruiting TAR from the 7SK RNA, whereas in the later stages of viral infection Tat substitutes for SRSF1 to promote dissociation from P-TEFb and release of the stalled polymerase. The splicing factor SRSF2 collaborates with 7SK RNA and promoter-associated nascent RNA to release paused polymerase (Mo et al., 2013).
3.2 Molecular description of 7SK RNA

The first published secondary structure of 7SK RNA was deduced from probing experiments by Wassarman & Steitz (1991) (Wassarman and Steitz, 1991) (Figure 1-5 A). It is characterized by three hairpins (HP1, HP3, HP4) and a conserved 3-way junction domain (domain 2). From the view of evolution, 7SK RNA sequence and structure are conserved from vertebrates to insects (Gruber et al., 2008b) (Gruber et al., 2008a). Figure 1-4 shows 7SK RNA localized sequence conservation of small domains. Thus another structure was proposed, based on structure conservation and bioinformatics by Manja Marz et al. (Marz et al., 2009) (Figure 1-5 B). The major difference between the 2 models is a proposed interaction of the 5’-end of 7SK RNA with an internal sequence comprising nucleotides 291 to 297. The resulting sub-domain is called M1. This interaction results in a closed RNA. Both 7SK RNA models share similar domains which form hairpins (HP) HP1 (blue), HP3 (green), and HP4 (pink) shown in figure 1-5. There may be more than one 3D structures of 7SK RNA. Actually, it has been proposed that 7SK RNA may adopt several conformations (Krueger et al., 2010).
Figure 1-4. Secondary structure and sequence conservation of 7SK RNA. Model of the secondary structure of 7SK RNA according to Wassarman et al. (1991) (Wassarman and Steitz, 1991). The sequence is colored based on its conservation according to Marz et al. (Marz et al., 2009).

Figure 1-5. Comparison of two different 2D structures of human 7SK RNA. A) Wassarman and Steitz model (Wassarman and Steitz, 1991). B) Marz et al.’s model (Marz et al., 2009). Both structures share three hairpins, HP1 (in blue), HP3 (light green) and HP4 (pink) that have been conserved during evolution. Red circles show identified protein binding sites.
A 3D structure of HP4 has been determined by NMR (Durney and D'Souza, 2010) (Figure 1-6). Residues U321 of the internal bulge and G312 in the apical loop, points out to the solvent, while the second residue of the bulge, C320 stacks inwards G312 residue had been shown to be important for LARP7 recognition (Muniz et al., 2013), as will be discussed below. The structure was also solved in the presence of a derivative of arginine, which is sandwiched between U319 and C320. The authors mention a similarity with the situation in TAR, which also contains bulged pyrimidines (UCU). Similarly, in TAR, the residue located 5’-to the bulge, A22, and the pyrimidine U23 in the bulge sandwich an arginine (Figure 1-6). Thus, both structures show a common arginine sandwich motif. However, the functional similarity between TAR and 7SK RNA has since been shown to involve another hairpin of 7SK RNA, the 5’-hairpin (HP1) (Muniz et al., 2010).

Recently, the structure of the 5’-hairpin (HP1) has been determined by our team, from two types of data, both collected with a variant of the hairpin, modified in the apical loop. A solution structure from NMR data has been solved in collaboration with Isabelle Lebars (Bourbigot et al., submitted) and the crystal structure determined by X-ray crystallography (Martinez-Zapien, manuscript in preparation).
Figure 1-6. Solution structure of 7SK RNA-HP4 (SL4) (Durney and D'Souza, 2010). (a) Secondary structure of HIV-1 TAR and 7SK RNA-SL4 showing some similarity of bulged residues. (b) 3D structure obtained by NMR for 7SK RNA-HP4 with residue, U321 of the bulge pointing towards the solvent while C320 is stacked inwards. The loop residue C312 is indicated. 3’ CCC and 5’ GGG were added to ensure 7SK RNA-HP4 stability, which are colored in yellow.

3.3 Proteins binding the 7SK RNA

HEXIM

As mentioned before, 7SK RNA binds HEXIM proteins and turns them into P-TEFb inhibitors (Michels et al., 2004) (Yik et al., 2004). HEXIM proteins share a basic region (BR) (residues 149-179 in human HEXIM1) that is essential for binding to 7SK RNA (Figure 1-7). This motif is an arginine rich motif (ARM). A similar ARM was found in Tat (Yik et al., 2004). In contrast to amino-acid sequences C-terminal to this motif, sequences on its N-terminal side show no conservation through evolution. It has been hypothesized that the BR interacts with an adjacent acidic regions (AR) in the absence of RNA (Barboric et al., 2005). The removal of the positive or negative charges from these
regions in HEXIM1 leads to its sequestration into the large complex and inhibition of transcription independently of the BR. In addition, residues 200-220 have also been shown to be in contact with the 7SK RNA (Belanger et al., 2009).

HEXIM directly contacts the Cyclin T subunit of P-TEFb (Michels et al., 2003). A cyclin T binding domain has been localized in the C-terminal domain (residues 255-316) in human HEXIM1 (Schonichen et al., 2010). Furthermore, a conserved motif in the central part of HEXIM, the “PYNT” motif (residues 201-205 in human HEXIM1) was shown by mutational analysis to also contribute to P-TEFb binding (Byers et al., 2005; Michels et al., 2004) (Figure 1-7). This motif is essential for P-TEFb inhibition (Czudnochowski et al., 2012).

HEXIM proteins form dimers (Dulac et al., 2005) (Li et al., 2005). The dimerization interfaces comprise a bipartite coiled-coil at the C-terminus (residues 284-348 in human HEXIM1), where two helices (α2 and α3) wrap around the equivalent helicase in the dimer (Dames et al., 2007), as shown in Figure 1-7.
Figure 1-7. Domain organization of the HEXIM1 protein showing the regions of known functional interest: non-conserved N-terminal domain (grey), RNA-binding region (blue), t PYNT (green), acidic sequence (orange), and dimerization domain (pink). The sequence of the ARM peptide is given below. Above, the 3D structure of the C-terminal region from Dames et al., with one monomer dark pink and the other light pink (Dames et al., 2007). The boundaries are indicated for the Homo sequence (numbers on top).

The hnRNP proteins

Transcription inhibition induces disassembly of the 7SK RNA/ HEXIM/ P-TEFb complex, the released 7SK RNA is then captured by a subclass of heterogeneous nuclear ribonucleo- proteins (hnRNPs) (Barrandon et al., 2007; Van Herreweghe et al., 2007b). hnRNPs are known for packaging newly synthesized pre-RNA and to be involved in alternative splicing, nucleo-cytoplasmic transport, stability and translation of mRNAs (Dreyfuss et al., 2002; Krecic and Swanson, 1999). Transcription-dependent dissociation of 7SK RNA/ HEXIM/ P-TEFb relies upon formation of 7SK RNA complexes with hnRNP A1, A2, Q1 and R (Barrandon et al., 2007; Van Herreweghe et al., 2007b). Interestingly, hnRNP Q2 differs from hnRNP Q1 in that it fails to associate with 7SK RNA. Both proteins are alternative splicing isoforms coded by the same gene. hnRNP Q2 just lacks a second RNA recognition motif (RRM) that is present in the N-terminal
section of hnRNP Q1. hnRNPs associate with hairpin HP3, a different region than the HEXIM-binding region (HP1). It is possible that apart from these partners, 7SK RNA snRNP carries or binds other proteins which participate in 7SK RNA remodelling, such as RNA helicases DDX21 (Calo et al., 2015) or RHA (Van Herreweghe et al., 2007). After recruiting the hnRNPs, 7SK RNA may undergo a conformational change that leads to release of HEXIM.

**MePCE**

Methylephosphate-capping enzyme (MePCE) was first investigated in Drosophila as a bicoid interacting protein (BIN3 or BCDIN3) (Zhu and Hanes, 2000). It was later shown to bind the 5’end of human 7SK RNA and to catalyze 5’ end capping of 7SK RNA (Jeronimo et al., 2007). MePCE has a methyltransferase domain, it monomethylates the γ-phosphate of the 5’-terminal guanosine-triphosphate of nascent 7SK RNA snRNA. MePCE interacts with a short 5'-terminal G1-U4/U106-G111 which in the representation of Wassarman & Steitz, would be present at the bottom of a long hairpin comprising HP1 (Muniz et al., 2013). MePCE caps 7SK RNA, probably co-transcriptionally and prior to its sequestration into the 7SK RNA snRNP. Despite its function in 7SK RNA biogenesis, MePCE remains stably associated with 7SK RNA. MePCE act cooperatively with LARP7 to stabilize 7SK RNA and maintain the integrity of 7SK RNA snRNP (Xue et al., 2010). Upon interaction with LARP7, MePCE loses its capping activity. LARP7 may occlude the catalytic center of MePCE. Thus MePCE displays a capping-independent function which is to promote the LARP7–7SK RNA interaction, which in turn stabilizes 7SK RNA. Indeed, siRNA-mediated silencing of MePCE reduced the cellular 7SK RNA level by about half (Jeronimo et al., 2007). However, among all subunits of 7SK RNA snRNP, MePCE remains the least studied protein.
LARP7

Nascent 7SK RNA transcripts bind the La autoantigen (La) protein (Reddy et al., 1984). The mature RNA remains associated with a La-Related protein initially named PIP7S (He et al., 2008). This protein has been renamed LARP7, as it belongs to the La-related (LARP) family of proteins. This will be developed in the next part of the Introduction.

Summary of the identified protein binding sites on the 7SK RNA

The multiple domains of 7SK RNA constitute binding sites for the numerous proteins which interact during its functional stages. This is summarized in Figure 1-8. Hairpin

![Figure 1-8](image-url)

Figure 1-8. Major protein binding sites identified on the 2D structure of 7SK RNA.

HP1 contains the HEXIM binding site (Egloff et al., 2006). It was further demonstrated that the repeated GAUC motif in the upper part of HP1, which constitutes the 7SK RNA-
signature motif, is essential for specific HEXIM1 recognition (Lebars et al., 2010) and Tat binding (Muniz et al., 2010). HMGA1 was shown to interact with a short motif in domain 2 both in vitro and in vivo (Eilebrecht et al., 2011). Hairpin HP3 has been shown to bind hnRNPs (Van Herreweghe et al., 2007). HP4 is required for P-TEFb recruitment (Egloff et al., 2006). On the whole, these 7SK RNA constitute a modular platform for interactions.

Pathologies linked to 7SK RNA and LARP7 disorder

7SK RNA is upregulated in the serum of multiple sclerosis patients (Santoro et al., 2016). It is enriched in the somatodendritic compartment (Briese et al., 2016). 7SK RNA over expression promotes apoptosis in cancer cells (Keramati et al., 2015). As pausing is the siege of multiple regulations leading to the maturation of messenger RNA, 7SK RNA function is directly related to several human diseases (for instance cancers such as leukemia) (Elagib et al., 2013). Importantly, cardiac hypertrophy and development of various types of human malignancies have been associated with increased P-TEFb activity, consequence of a disruption of this regulatory equilibrium. In addition, the HIV-1 Tat protein also releases P-TEFb from the 7SK/HEXIM complex during viral infection to promote viral transcription and replication (Muniz et al., 2010). Recently, it was found that LARP7 is a potential tumor suppressor gene in gastric cancer (Cheng et al., 2012). LARP7 suppresses P-TEFb activity to inhibit breast cancer progression and metastasis (Ji et al., 2014).
4. LARP7 and LARPs

4.1 La protein

La protein, or La autoantigen is an abundant nuclear phosphoprotein. It is conserved in eukaryotes. Following transcription by RNA Pol III, the La protein binds to the 3’ end poly U of the nascent transcript to assist in folding and maturation of the RNAs produced by this polymerase (Bayfield et al., 2010). The N-terminal domain of La protein comprises two domains, the LAM (La motif) and RRM (RNA Recognition Motif) which together form the La module. There is another RRM domain in the C-terminal part of the molecule (RRM2). The individual structures of the LAM and RRM1 domains of the La module of the La protein have been solved by NMR (Alfano et al., 2004; Kotik-Kogan et al., 2008), as well as the structure of RRM2 of La (Jacks et al., 2003). The La module binds to the poly-uridines sequence UUU-3’OH found at the 3’-end of the Pol III transcript. Crystal structures of the La module bound to oligonucleotides with a terminal UUU triplet showed how the 3’-end of the transcript is recognized (Teplova et al., 2006) (Kotik-Kogan et al., 2008). The RNA forms a hook, and binds in a cleft between the two domains, the LAM and the RRM1. The terminal uridines are bound to conserved residues from the LAM and RRM. The majority of the contacts are with the LAM. The terminal uridine (U-1) binds only LAM. Binding essentially involves H-bonds with the ribose 2’ and 3’ hydroxyl groups, and the stacking of the uridine base on a phenylalanine (F35). This suggests that other types of base can be accommodated, and it was shown indeed that mutating the uridine had only a small impact on binding affinity. The penultimate uridine (U-2) is tethered by the two domains. The O2 and O4 groups of the uridine base interact with Q20 (LAM) and I140 (RRM1), thus drawing the two domains together. The uridine U-3 is stacked on U-1, itself stacking on the phenylalanine F35 of LAM, and
binds additionally to a conserved asparagine (N56).

### 4.2 La-related proteins (LARP) family

La-related proteins (LARP) are defined as proteins containing a high degree of homology with La protein. However, they have taken on independent functions from those associated with La (Bayfield et al., 2010). The evolutionarily conserved LARP family currently comprises genuine La, LARP1, LARP1b, LARP4a, LARP4b, LARP6 and LARP7 (Bousquet-Antonelli and Deragon, 2009). At the molecular level, these proteins contain a characteristic LAM and a RRM or a RRM-L (RRM-like motif), generally located in their N-terminal region, which together form the La module. The La module of LaRP1 can bind poly U and poly G like protein La (Nykamp et al., 2008). LaRP4 proteins lack several conserved key side chains in the LAM, which are used by La proteins to bind UUU-3’OH (Bousquet-Antonelli and Deragon, 2009). Apart from La and LARP7, the RRM may diverge from standard, like RRM-L3, L4, L5 (Bousquet-Antonelli and Deragon, 2009). In the two La and LARP7 sub-classes, there is another RRM in the C-terminal region. Besides LAM and a RRM or a RRM-L, additional domains characterize some of the family members, like DM15 in LARP1 and SUZ-C domain (SUZC) in LARP6 (Figure 1-9). La, LARP1, LARP6, LARP7 and possibly LARP4a and 4b are dysregulated in cancer. Of these, LARP1 has an important role in translation regulation and it is the first to be demonstrated to drive oncogenesis (Stavraka and Blagden, 2015). LaRP4 is a cytoplasmic, polyribosome-associated protein that interacts with poly-A binding protein (PABP). LaRP4 is a positive factor promoting mRNA’s stability (Bousquet-Antonelli and Deragon, 2009). LaRP6 interacts with transcription factors, and functions upstream of the transcription factor MyoD, to control muscle
development (Valavanis et al., 2007). LARP7 can only specifically bind 7SK RNA in higher eukaryotes and will be presented in detail below. In ciliates, a protein of the LARP7 sub-family (P65 in Tetrahymena, P43 in Euplotes) is involved in telomerase complex assembly (Singh et al., 2012).

![Diagram of LARP protein family domains](image)

**Figure 1-9.** The principal domains common to members of the LARP protein family. Abbreviations: DM15: DM15-repeat containing region (“DM15 region”) also known as “LARP1 motif”; LAM: La Motif; NLS: Nuclear localisation signal; PAM2w: Atypical PAM2 domain; RCD: RNA chaperone domain; RRM: RNA Recognition Motif; RRM-L: RNA recognition-like motif; SUZ-C: SUZ-C domain. (Modified from Bayfield & Maraia, 2010 (Bayfield et al., 2010)).

### 4.3 LARP7

LARP7 is the closest paralogue to La-protein (Bousquet-Antonelli and Deragon, 2009) and like La, possesses an additional RRM in the C-terminal region. This second RRM (RRM2) located at the C-terminal minus bears some homology with the proteins involved in the telomerase assembly in ciliates, such as P65 in Tetrahymena (Bousquet-Antonelli
and Deragon, 2009). Human LARP7 consists of 582 amino acids (aa). The La module of LARP7, at the N-terminal region (aa 1-208, Figure 1-11) is separated from the RRM2 (aa 450-545, Figure 1-12) by a linker that is mainly unstructured. This was indicated by an analysis with a folding prediction program (Foldindex) as shown in Figure 1-10.

When LARP7 was discovered as associated to 7SK RNA, the La module, was hypothesized to bind the 3’-end uridines present in 7SK RNA, by analogy with the La protein. This was confirmed by two experiments (He et al., 2008). (1) Deletion of the three 3’-end uridines of human 7SK RNA reduced LARP7 binding. (2) Mutation of one essential conserved residue (Y127D) in RRM of the La module, known to be involved in RNA binding, reduced LARP7 binding to 7SK RNA in vivo. However, the replacing the La module of LARP7 by the La module of La did not provide an equivalent protein. The replaced chimera was able to bind 7SK RNA, but the specificity was impaired, as it was observed also bound to other RNAs such as U6 (He et al., 2008).

**Figure 1-10.** a. The prediction of structure formation (red for unfolded, green for folded) as computed with Foldindex. b. Ribbon representation of LARP7 (582 residues in human): LAM (28-112) yellow, RRM1 (120-185) orange and RRM2 (450-545) purple. Blue bars represent stretches of basic residues.
Figure 1-11. Sequence alignment of La protein and LARP7 N-terminal domain 1-208. 1ZH5 is PDB (protein data bank) number of a La protein structure. Blue area is additional region for LARP7. Purple is additional α helix for La protein. Red is additional β strand for La protein. Pink is the beginning of the linker region (from Emiko Uchikawa PhD thesis). Another, more detailed alignment can be found in the Chapter III and Uchikawa et al. 2015.

Figure 1-12. Sequence alignment of Tt (Tetrahymena telomerase)-p65, Ea-p43, Hs (homo sapiens)-La and Hs-LARP7 C-terminal domain (450-545) from (Singh et al., 2012).
5. Investigating the interactions between 7SK RNA and LARP7

We were interested in understanding the specificity interaction between LARP7 and 7SK RNA. Firstly, our aim was to discover which elements of 7SK RNA and LARP7 are responsible for the specificity and stability of the interaction. Secondly, we wanted to clarify what is the effect of LARP7 when binding to 7SK RNA. Is it responsible for stabilizing a special conformation? Does the binding of LARP7 affect the function of 7SK RNA, for example, by enhancing the specificity of HEXIM recognition? To address these questions, we used biochemical and structural approaches using in vitro reconstituted complexes from purified recombinant proteins and in vitro transcribed RNA.

Before my arrival, Emiko Uchikawa and AC Dock-Bregeon had started to investigate how 7SK RNA was recognized by LARP7. They focused on the recognition of the 3’-end of 7SK RNA by the La module. A truncated protein encompassing the N-terminal region (1-208) was produced in bacteria. This protein was crystallized in the presence of a 7SK RNA oligonucleotide (287-332) encompassing the HP4 and parts of the upstream sequence. Crystals were obtained, but they were small, difficult to handle, and diffracted poorly. After many attempts of crystal data collection, a data set could be obtained at the Swiss Light Synchrotron, and the structure solved with the help of Kundhavai Nachar and Bruno Klaholz at IGBMC, Illkirch. The resulting structure of the La module is described in the chapter III. My contribution was to use footprinting and EMSA to better understand the specificity for 7SK RNA binding. In the following chapters, I extend the use of these methods to investigate HEXIM1 and LARP7 binding to the other hairpins (HP1, domain 2 and HP3) of 7SK RNA.
CHAPTER II: Material and methods

1. Preparation of molecules

1.1 7SK RNA purification

7SK RNA and all RNAs in this work were prepared by in vitro transcription with polymerase T7 (home-made). The template for transcription was generally prepared by PCR from a plasmid, pHDV_7SK RNA derived from Walker (Walker et al., 2003) by D. Martinez-Zapien, where the 7SK RNA gene was inserted contiguously to the T7 promoter. The 7SK RNA gene was followed by the sequence of the HDV ribozyme, which folds and cleaves spontaneously when transcribed since the transcription buffer contains magnesium (the pHDV plasmid was described in (Walker et al., 2003)). A primer for general production was chosen 50 nucleotides upstream of the T7 promoter. For studies with LARP7, the 7SK RNA was transcribed without the HDV ribozyme because the ribozyme cleavage leaves a 3’-phosphate, which is detrimental to recognition (La module recognizes a 3’-OH). In that case, the primer was 7SK RNA-3U*, to end with U331.

PCR for template

<table>
<thead>
<tr>
<th>H$_2$O</th>
<th>Up to 50 µL</th>
<th>32.5 µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template plasmid</td>
<td>7SK RNA in pHDV</td>
<td>2.5 µL 10 ng/µL</td>
</tr>
<tr>
<td>Buffer</td>
<td>Tampon 5*Phusion</td>
<td>10 µL</td>
</tr>
<tr>
<td>Nucleotide</td>
<td>dNTP 2 mM</td>
<td>2.5 µL</td>
</tr>
<tr>
<td>Primer</td>
<td>Up-T7</td>
<td>1 µL</td>
</tr>
<tr>
<td>Primer</td>
<td>HDV primer or 7SK RNA-3U*</td>
<td>1 µL</td>
</tr>
<tr>
<td>Enzyme</td>
<td>Phusion</td>
<td>0.5 µL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Temp</th>
<th>Time</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98 °C</td>
<td>30 s</td>
<td>1 cycle</td>
</tr>
<tr>
<td>2</td>
<td>98 °C</td>
<td>30 s</td>
<td>50 cycle</td>
</tr>
<tr>
<td></td>
<td>55 °C</td>
<td>30-45 s</td>
<td></td>
</tr>
<tr>
<td></td>
<td>72 °C</td>
<td>6 s</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 2-1.** Checking template production. (1) DNA ladder (2) 2 µL of the PCR made with primers T7-50 and 7SK RNA_3U* and plasmid pHDV_7SK RNA were loaded into a 1.5% agarose gel. The product size is about 400 bp.

**Transcription**

<table>
<thead>
<tr>
<th></th>
<th>Stock</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O</td>
<td>Up to 1 ml</td>
<td></td>
</tr>
<tr>
<td>Template:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annealed oligos</td>
<td>100uM</td>
<td>100 µL</td>
</tr>
<tr>
<td>PCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasmid linear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcription Buffer</td>
<td>5X</td>
<td>200 µL</td>
</tr>
<tr>
<td>5X: Tris pH8 150mM;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spermidine 10mM;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MgCl₂ 30mM; Triton 0.05%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MgCl₂</td>
<td>1 M</td>
<td>10 µL</td>
</tr>
<tr>
<td>DTT</td>
<td>100 mM</td>
<td>50 µL</td>
</tr>
<tr>
<td>ATP</td>
<td>100 mM</td>
<td>40 µL</td>
</tr>
<tr>
<td>GTP</td>
<td>100 mM</td>
<td>40 µL</td>
</tr>
<tr>
<td>CTP</td>
<td>100 mM</td>
<td>40 µL</td>
</tr>
<tr>
<td>UTP</td>
<td>100 mM</td>
<td>40 µL</td>
</tr>
<tr>
<td>GMP</td>
<td>500 mM</td>
<td>40 µL</td>
</tr>
<tr>
<td>T7 Pol (Home made)</td>
<td>1 mg/ml</td>
<td>25 µL</td>
</tr>
</tbody>
</table>

---

30
Figure 2-2. Checking in vitro transcription. 5 µL of transcription mix was loaded into an analytical polyacrylamide gel. Band of 7SK RNA transcript is revealed by toluidine blue staining.

- Incubate 4 hours at 37°C.
- Extract with 500 µl Biophenol, vortex 1min/tube, centrifuge 2 mins 10 000 g. Collect supernatant. Add100 µl DEPC water to the phenol phase, vortex 1min/tube, and centrifuge 2 mins at 10000 rpm. Collect supernatant. *(Biophenol extraction repeats twice)*
- Precipitate with 2.5 volume of 100% ethanol. Incubate -20°C over night.
- Recover the pellets by centrifugation at 13 000 g for 20 mins. Dissolve the pellet in formamid blue buffer (< 500 µL for loading in one pocket of gel).
- Denature at 90°C for 2mins, load in 10% acrylamide RNA denaturing gel (30 cm x 20 cm x 0.15 cm).
- Migrate at 20 W.
- Lay the gel (without the glass plates) on a special plate for UV-shadowing. Analyse under UV light and draw the shape of the RNA bands. Cut the piece of gel and transfer into a tube.
- Add 500 µL elution buffer. Incubate at 4°C with gentle agitation over night. Precipitate with 2 vols ethanol 100%.
- Recover the pellets by centrifugation at 13 000 g for 20 mins followed by a wash in 80% ethanol. Dry 5 min. Dissolve in 50 µL RNA2 buffer.
- Measure OD to calculate RNA concentration. MW is 110KD for 7SK RNA.
1.2 LARP7 purification

The genes for human LARP7 (1-582) or domains (La module 1-208, RRM2 433-582) were inserted in a plasmid of the pnEA family for production in E. coli. This family, derived from the pET series, was designed by C. Romier (IBMC) to facilitate the change of tag, to allow tag cleavage by different proteases, and to vary tag position (Diebold et al., 2011). For the N-terminal region, the protein was well expressed with a N-terminal (His)6 tag. However, the tag at the N-terminal minus of the protein impaired binding to 7SK, thus tag cleavage was required. A plasmid comprising a sequence ensuring cleavage by TEV protease was thus chosen in general, and particularly for all constructs comprising the N-terminal domain. However, full length LARP7 was best expressed from the C-terminal (His) 6-tagged version.

Protein expression was realized at 18°C (for full length) or 26°C (N- and C-terminal domains). Temperature below 37°C favours proper proteins folding. No protein was sufficiently pure after Ni-affinity purification, and an additional chromatographic step was performed. This was done on a heparin-affinity column.

Expression

- Transform BL21 cells with pnEA-HV-Cter-LARP7 FL (ampicillin resistance).
- Pick up 5 colonies for a preculture in LB medium 37°C over night OD$_{600} = 3$ to 5.
- Inoculate 1 L of LB medium at OD = 0,01 / 0,02 and grow at 37°C (the doubling time is ~ 30 min).
- Induce with IPTG 1 mM when OD = 0.8 to .01 and grow overnight at 18°C (around 20 hours, OD$_{600}$ < 3).
- Pellet the bacteria at 4000 RPM for 20 min at 4°C.
- Freeze pellet at -20°C. Frozen pellet can be stored (1 pellet corresponds to 1 L culture).

**Figure 2-3.** Expression test. (1) Protein marker, size 100kd and 70kd are shown. (2) LARP7 is indicated in square.

**Lysis and Ni-affinity**

- Thaw and suspend the pellet (corresponding to 1 L culture) in 45 ml Lysis buffer.

| Lysis buffer | 500 mM NaCl, 50 mM Tris HCl pH 7.6, 5 mM MgCl$_2$, 1.4 mM mercaptoethanol; 1 pill Pierce™ Protease Inhibitor (Thermo Scientific) for 100 ml buffer. |

- Sonicate on ice at intensity 40% 5 sec on, 5 sec off for 3 min.
- Centrifuge 18000 g for 30 min at 4°C.
• Incubate supernatant with 2 ml Ni-NTA beads at 4°C for 1 hour.

*Beads preparation*:

| Imidazole stock | 1 M at pH = 8 (68 g/L - adjust with HCl) |

• Centrifuge 500 g for 5 mins at 4°C. Keep supernatant (FT).

• Load FT on the Ni-beads into a column (single use, empty). Wash beads with 8 ml lysis buffer.
• Wash beads with 12 ml Buffer with 30 mM Imidazole.
• Wash beads with 10 ml Buffer with 50 mM Imidazole.

| Wash Buffer | 500 mM NaCl, 50 mM Tris HCl pH 7.6, 5 mM MgCl2, 1,4 mM mercaptoethanol, + 30 mM/50 mM Imidazole |

• Elute with 500 µL Buffer with 300 mM Imidazole (Elution 1).
• Elute with 1 ml Buffer with 300 mM Imidazole (Elution 2) (line 7 in figure 2-4).
• Elute with 1 ml Buffer with 300 mM Imidazole (Elution 3) (line 8 in figure 2-4).
• Elute with 500 µL Buffer with 300 mM Imidazole (Elution 4).
• Elute with 500 µL Buffer with 300 mM Imidazole (Elution 5).

| Elute Buffer | 500 mM NaCl, 50 mM Tris HCl pH 7.6, 5 mM MgCl2, 1,4 mM mercaptoethanol, + 300mM Imidazole |

**Tag removal**

• Analyze on a denaturing gel
• Pool fractions enriched with LARP7 (generally 2 to 5)
• Determine concentration by Bradford and estimate yield (MW_LARP7 = 69 436 Da including tag).
**Bradford analysis**

- Prepare 10 ml Bradford reagent (diluting 2 ml 5X Bradford in 8 ml water).
- Incubate 1 ml 1X Bradford with different doses of samples, depending on the color change to blue. It lasts 5 mins for the reaction at room temperature.
- Calculate concentration

\[
\text{(OD}_{595}\times 17.5)/\text{volume in } \mu\text{L} = C \text{ in mg/ml}
\]

Example: (if 5 µL used multiply OD<sub>595</sub> by 3.5 (= 17.5/5)

- Add TEV (1 TEV molecule per 20 LARP7 molecules; the stock solution is (45 µM).
- Add the TEV and LARP7 mix into a dialysis tube and dialyse overnight at 4°C in 500 ml Dialysis Buffer.

<table>
<thead>
<tr>
<th>Dialysis Buffer</th>
<th>500 mM NaCl, 20 mM pH 7.2 NaHepes, 2 mM DTT, 1 mM EDTA</th>
</tr>
</thead>
</table>

**Figure 2-4.** Purification of LARP7. (1) Protein marker. (2) Total extract. (3)(4)(5) Wash solutions of Ni-chelating column. (6)(7)(8)(9)(10) Elutions of Ni-chelating column. We choose (7)(8) elutions for purification.

**Purification by affinity chromatography on Heparin**

- Centrifuge dialysate 10000rpm for 5 minutes at 4°C to eliminate aggregates that may block columns.
- Preequilibrature the Heparin-HiLoad column (1 ml) in 60% buffer A + 40% buffer B
• Load dialysate.
• Elution program at 0.5 ml/min, max pressure 0.5 MPa
  o equilibrate column 5 Cv (column volume) (5 ml) at 40% B buffer
  o inject and wash 10 Cv (10 ml) at 40% B buffer
  o gradient 20 Cv (20 ml) 40% to 100% buffer B with fractionation
• There are two peaks at the end of the salt gradient. LARP7 is in the second one (approx. 85% B).
• Check expression on acrylamide gel
• Dialyse in the storage buffer; add 10% glycerol, then flash-freeze aliquots in liquid nitrogen; store at -80°C

| Storage Buffer | 20 mM Hepes pH 7.2, 2 mM DTT, 0.35 M NaCl, add 10% glycerol after dialysis |

2. Binding assays

2.1 Footprinting experiments

Footprinting assays were performed to delineate the binding sites of proteins on RNA. In that type of experiment, the profile of a free RNA is compared with the profile of the same RNA in complex with protein(s). The profile results from the action of a probe (chemical reagent or RNase) which can cleave or not the RNA. With RNases, the RNA is cleaved at specific positions. The resulting fragments may be directly analyzed on a denaturing gel (direct footprinting). The action of chemical reagents, such as 1M7, DMS or CMCT, results in the formation of an adduct at specific positions of the RNA. Analysis of these positions requires a second step. Chemical processes may induce cleavage, producing fragments which can be analyzed directly. Another possibility requires reverse-transcription and is named indirect footprinting.
Each probe is specific. Thus, modification at a nucleotide by a probe will give insight into the local structural property of the RNA. In this project were used RNases and chemicals. There are several RNases used in this type of experiments: RNase V1, RNase T1, RNase ONE and RNase A. The probe used in this thesis was mostly RNase V1. Chemical probes: 1M7, CMCT and DMS are quite known (Shown in the table below). One thing of note for chemical probes is that the modification prevents the addition of the incoming nucleotide by the reverse-transcriptase at the modification site. Each of the bands generated is shifted by one base down on the gel.

<table>
<thead>
<tr>
<th>RNase</th>
<th>Cleavage Selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNase T1</td>
<td>3’ of single-stranded G Residues</td>
</tr>
<tr>
<td>RNase V1</td>
<td>Base-paired Nucleotides</td>
</tr>
<tr>
<td>RNase ONE</td>
<td>3’ of single-stranded</td>
</tr>
<tr>
<td>RNase A</td>
<td>3’ of single-stranded C and U Residues</td>
</tr>
<tr>
<td>RNase H</td>
<td>RNA of RNA:DNA hybrids</td>
</tr>
<tr>
<td>Chemical probe</td>
<td>Modification</td>
</tr>
<tr>
<td>CMCT</td>
<td>Modify uridines and guanines</td>
</tr>
<tr>
<td>DMS</td>
<td>Modify adenines and cytosines</td>
</tr>
<tr>
<td>1M7</td>
<td>React with the 2'-hydroxyl group to form adducts on the 2'-hydroxyl of the RNA backbone</td>
</tr>
</tbody>
</table>

**Direct footprinting**

In direct footprinting, the RNA is previously end-labeled (for example at the 5’-end) and the fragments resulting from a cleavage are identified on a denaturing gel. Indexation results from the migration in parallel of sequencing lanes. Generally, a ladder obtained by alkaline hydrolysis and a G-index, obtained with RNase T1 in denaturing conditions allows to identify the positions of the cleavages (figure 2-5). Blue arrow’s line indicates
Figure 2-5. Direct footprinting by RNase T1: 7SK RNA alone (R) and complexes with LARP7 (F), N-terminal (N), C-terminal (C) and a mixture of N-terminal and C-terminal both (N+C) was incubated with RNase T1 (1/500) in a buffer with Hepes 20 mM, pH 7.2; MgCl2 5 mM, NaCl 125 mM (N-terminal and C-terminal) or 250 mM (LaRP7 full length) to allow complex formation, then migrated on a denaturing gel. Cleavage sites were indexed by sequencing, arrows show different cleavage sites in the RNA. O: control without probing. HA: ladder obtained by alkaline hydrolysis. G: guanine index obtained with RNase T1 with denatured RNA.

the accessibility to RNase T1 increases with LARP7 C-terminal and C-terminal+ N-terminal both. Green arrow’s line shows RNA footprinted by LARP7 C-terminal, C-terminal+ N-terminal both and LARP7 full length. However, the resolution of a gel is limited to fragments shorter than about 100 nucleotides.
**Indirect footprinting**

RNAs cleaved by RNase as well as RNAs modified by addition of a chemical may be analyzed by reverse-transcription. A reverse transcriptase uses the RNA fragments as templates (in the 3’ to 5’ direction) to produce cDNA. This requires priming. The primers used for reverse-transcription are labeled, either with a radioactive phosphate or fluorophores. Generally cDNAs produced with a radioactive primer are analyzed on denaturing gels. Variation of the primer position on the 7SK RNA allows to cover the entire molecule, with the exception of the priming sequence. In 7SK RNA, the priming sequences are shown in figure 2-6. Figure 2-7 shows the result of a gel analysis by indirect footprinting, probing 7SK RNA and complexes with LARP7, N-terminal and C-terminal by RNase V1, primer J. Lane R is the naked RNA, Green arrow shows the band is less digested by RNase V1 in the presence of LARP7 full length (lane F) which means a footprinting.

![Figure 2-6. Positions of the primers on the 7SK RNA.](image-url)
Figure 2-7. Probing naked 7SK RNA (R) and complexes with LARP7 (F), N-terminal (N) and C-terminal (C) by RNase V1 (1/10), primer J (302-324). The 4 lanes on the right (AUCG) correspond to Sanger sequencing. On the left of the gel shows the control without RNase, only incubated RNA with LARP7 full length (Of), N-terminal (On) and C-terminal (Oc).

To overcome the size limitation due to gel-migration, the cDNAs resulting from reverse transcription may be analyzed by sequencing on a sequencer. Primers with fluorophores were used for analyzes with a sequencing machine (an example is shown in Figure 2-8), for which we got access in the “Laboratoire de RMN et Cristallographie Biologique” thanks to Bruno Sargueil. We chose primer D, more efficient than others in the 3’ end of the RNA, but that lead to a large “blind region”, from about 250 to 330. Indexation is set-up by migrating in parallel (in another capillary) the products of reverse transcription of the same RNA with a mixture of dNTPs and ddXTP (where X= A, U, C, or G). After several experiments, one such sequencing lane is enough to index one probing reaction. Thus, a capillary will be injected with a mixture of one sequence (ddATP for instance), one control (without probe), the reference profile of the free RNA and the experiment with RNA and proteins. Each component of an experiment uses the same primer
sequence labeled with a color-coded fluorophore. In our experiments, we defined that blue curve stands for RNA and protein with probe. Green curve stands for RNA with probe. Black curve shows RNA and protein in the absence of probe, as control. Red curve is dideoxy-sequencing with ddATP reaction used to map reactivity to the RNA sequence. If the green curve is higher than the blue, either the protein protects this position (which is called footprinting), or the structure melts after protein binding (in the RNase V1 probing). If the blue curve is higher than the green; there is an enhancement of the structure or a change of conformation, the protein binds somewhere else and stabilize or reveals this position. A typical experiment is shown in figure 2-8. The observation was limited to 253-95. Missing residues (“blind region”) at the 3’ end was due to the position of the primer, 295-314. Unfortunately, primers located more 3’ in the 7SK RNA sequence did not allow an efficient transcription. Moreover, the sequencer technique induces the loss of information on short fragments (up to 50 nucleotides), because they are more abundant and saturate the fluorescent signal. Missing residues at the 5’ region was due to an unfortunate problem met with 7SK RNA. Reverse transcription stops at about position 95. Even when some extension is observed, this induces scaling problems. After sequencing, the chromatograms are analyzed with the program ShapeFinder (Vasa et al., 2008).
Figure 2-8. Footprinting experiment with V1 RNase. After RT sequencing, the chromatograms corresponding to read-outs of the free 7SK RNA (green curve) and LARP7 complex (blue curve) were superimposed. A control without probe (black curve) and a sequencing lane by incorporation of ddATP (red curve) were sequenced in the same capillary. Blue arrows show protection (footprint) from V1 cleavage by LARP7 (or melting of the structure). Orange arrows indicate enhancement of the V1 cleavage in the presence of LARP7 (strengthening of the structure or increase of accessibility of a structured motif).
2.2 EMSA

Electrophoretic mobility shift assays (EMSA) is a technique which permits the visualization of the complex(es) of RNA and protein. The RNAs are radioactively 5’-labeled with T4 oligonucleotide kinase and $\gamma P^{32}$-ATP. After labeling, RNA is separated from excess ATP by micro-gel filtration. Thermal treatment of the RNA is done if necessary. Labeled RNA is incubated with increasing concentration of protein(s) in a buffer allowing complex formation. Complex formation is analyzed by native gel, which separates the complexed RNA from free RNA. Normally, increasing amount of complex(es) is formed as the concentration of protein is increased. A typical gel is shown in figure 2-9.

![Figure 2-9. EMSA native gel. The figure shows the binding of 7SK RNA with LARP7. 7SK RNA alone and complexes are indicated. An increasingly concentration of LARP7 were added. It is migrated from negative charge (-) to positive charge (+).](image-url)
**Competition EMSA**

In order to compare different RNA mutants, we use competition EMSA experiment. The concentration of cold RNA (without labeling) is 5-20 times more than the hot RNA (labeled wild-type RNA). Protein was added after mixing cold and hot RNA firstly. The complexes are analyzed on a native gel, and compared with the incubation without cold RNA and only wild-type labeled RNA. If the cold RNA binds the protein, it competes with the hot RNA and no shifted band can be observed. This is the case of the positive control, the wild-type RNA. If the cold RNA doesn’t bind the protein at all, we can see the totally shift, this is the negative control. The principle of EMSA competition assay is schematized in figure 2-10.

![Competition EMSA Diagram](image)

*Figure 2-10. The principle of competition assay.*
Chapter III: Characterization of LARP7 binding to 7SK RNA by biochemical approaches: involvement of the 3’hairpin HP4. (Publication No.1)

Our aim is to clarify how 7SK RNA structural organization supports its specific relationship with its partners and the functional consequences of the interactions established between the RNA and the proteins. We use a multiple approach, combining biochemical and structural methods. The first investigations were aimed at highlighting what makes 7SK RNA specific to HEXIM (Denise Martinez-Zapien’s PhD project).

I will focus on LARP7, a major interactant of 7SK RNA. LARP7 was shown to bind 7SK RNA specifically, ensure its stability and binds its 3’end (He et al., 2008). LARP7 is a La related protein, very close to the La autoantigen (Bousquet-Antonelli and Deragon, 2009). An RNA Recognition Motif (RRM) from La –module recognizes RNAs transcribed by Pol III, with a stretch of uridines at the 3’end, as explained in the Introduction. LARP7 has a La –module but the binding to the 3’end uridines cannot account for the specificity of LARP7 binding to 7SK RNA, since all Pol III transcripts have 3’end uridines. We thus aimed to highlight which elements of 7SK RNA are responsible for the specificity and stability of the interaction. Crystallography is a technique of choice to highlight the details of molecular interactions. We thus started a project with the objective of a crystallographic study of a complex of LARP7 and 7SK RNA from human. Full length LARP7 was not easy to produce in the amount required for crystallization, so we quickly focused on the production of the two structured regions: the N-terminal region, comprising the La module, and the C-terminal region comprising the second RRM.
Crystals of the N-terminal region of LARP7 in complex with an oligonucleotide mimicking the 7SK RNA 3’end (UUUCUUUU) were obtained by Emiko Uchikawa during her PhD project, and the structure solved by Kundhavai Natchiar and Bruno Klaholz (IGBMC). The structure of the La module of LARP7 in our crystals is very close to the structure of the homologous domain in La protein. In particular, the LAM sub-domain superposes well with LAM from La protein. Importantly, the 3’end UUU was found at the same place, in the cleft between the La-motif (LAM) and RRM1 of the La module (Figure 3 in the paper). Most of the details of the RNA recognition are similar. This showed that the La module functions in the expected way. On another hand, the structural work also showed some differences. The RRM module is smaller than in the La protein, and some sequences are specific to LARP7. This is detailed in the first part of the following paper. However, these differences do not fully explain the specificity of LARP7 for 7SK RNA.

During our investigations, the group of Tamass Kiss showed the importance of nucleotides in the apical loop of the 3’-hairpin, HP4 for the recognition (Muniz et al., 2013). Interestingly, mutation of several residues of that loop destabilized the interaction with LARP7 in vivo. This prompted us to investigate which part of LARP7 is involved in the recognition of the HP4 apical loop, and encouraged us to continue investigating if that is the only part of 7SK RNA involved.

A good candidate for additional binding to 7SK RNA is the RRM2 in the C-terminal region. Cloning and purification of a protein domain comprising the RRM2 was performed at the time of my arrival at IBENS (Elodie Zhang’s Master project). At that
time, preparation of LARP7 full length was improved, too. These tools allowed me to further study the interaction of LARP7 with 7SK RNA. In the second part of the following paper, I used electrophoretic mobility shift assays (EMSA) with 7SK RNA or its sub-domains to show that while the La module of LARP7 binds the 3’-end of 7SK RNA, the RRM2 binds concomitantly the apical loop of the 3’-hairpin, HP4. The binding of RRM2 to HP4 was further investigated by footprinting. These experiments showed that RRM2 wraps around the apical region of HP4. This allowed to build a 3D model, using the published 3D structure of HP4 (Durney and D'Souza, 2010) and the RRM2 of P65, which has been proposed as homologous to the RRM2 of LARP7 (Singh et al., 2013).
Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7

Emiko Uchikawa¹,²,³,⁴,†, Kundhavai S. Natchiar¹,²,³,⁴,†, Xiao Han⁵,⁶,⁷,⁸,†, Florence Proux⁵,⁶,⁷, Pierre Roblin⁹,¹⁰, Elodie Zhang⁵,⁶,⁷, Alexandre Durand¹,²,³,⁴, Bruno P. Klaholz¹,²,³,⁴ and Anne-Catherine Dock-Bregeon⁵,⁶,⁷,*

¹Department of Integrated Structural Biology, Centre for Integrative Biology (CBI), IGBMC (Institute of Genetics and of Molecular and Cellular Biology, 67404 Illkirch, France, ²Centre National de la Recherche Scientifique (CNRS) UMR 7104, Illkirch, France, ³Institut National de la Santé et de la Recherche Médicale (INSERM) U964, Illkirch, France, ⁴Université de Strasbourg, 67000 Strasbourg, France, ⁵Department of functional genomics, Institut de Biologie de l’Ecole Normale Supérieure (IBENS), 75005 Paris, France, ⁶CNRS UMR 8197, 75005 Paris, France, ⁷INSERM U1024, 75005 Paris, France, ⁸Key Laboratory of Brain Functional Genomics, East China Normal University (ECNU), 200241 Shanghai, PR China, ⁹SOLEIL Synchrotron, 91192 Gif-sur-Yvette, France and ¹⁰INRA-URBIA, 44316 Nantes, France

Received October 03, 2014; Revised February 13, 2015; Accepted February 20, 2015

ABSTRACT

The non-coding RNA 7SK is the scaffold for a small nuclear ribonucleoprotein (7SKsnRNP) which regulates the function of the positive transcription elongation factor P-TEFb in the control of RNA polymerase II elongation in metazoans. The La-related protein LARP7 is a component of the 7SKsnRNP required for stability and function of the RNA. To address the function of LARP7 we determined the crystal structure of its La module, which binds a stretch of uridines at the 3′-end of 7SK. The structure shows that the penultimate uridine is tethered by the two domains, the La-motif and the RNA-recognition motif (RRM1), and reveals that the RRM1 is significantly smaller and more exposed than in the La protein. Sequence analysis suggests that this impacts interaction with 7SK. Binding assays, footprinting and small-angle scattering experiments show that a second RRM domain located at the C-terminus binds the apical loop of the 3′ hairpin of 7SK, while the N-terminal domains bind at its foot. Our results suggest that LARP7 uses both its N- and C-terminal domains to stabilize 7SK in a closed structure, which forms by joining conserved sequences at the 5′-end with the foot of the 3′ hairpin and has thus functional implications.

INTRODUCTION

The La-related proteins (LARPs) are involved in various important functions in RNA metabolism and are found in nearly all eukaryotes (1). Besides the essential role of the paradigmatic La protein in tRNA processing, and its involvement in transcription termination by binding to nascent transcripts generated by polymerase III (2), members of the LARP family are involved in the regulation of translation or demonstrate chaperoning activities (3). In addition to the characteristic domain containing the La motif (LAM), related to the winged-helix domain, they possess several RNA-binding domains akin to the RNA recognition motif (RRM) structural fold (4). LARPs share a conserved two-domain unit, called the La module, comprising the LaM and RRM1. High-resolution structures of the La module have been described in the case of HsLa, the human La protein (5,6), and very recently for human LARP6 (7). As such, LARPs are modular proteins, with intriguing possibilities for intricate RNA-binding combinations. LARP7 is the family member showing the highest sequence similarity to La, with the characteristic La module in the N-terminal third of the protein (3,8). However, while La binds all nascent transcripts synthesized by RNA polymerase III via their shared termination motif, UUUOH, LARP7 binds almost exclusively to the non-coding RNA 7SK (9–11). In Drosophila, LARP7 and the members of the 7SK snRNP have recently been identified (12,13). Other potential LARP7 homologs are found in ciliates, such as P65 in Tetrahymena thermophila, which has been found to assist in the correct folding of the telomerase RNA and hierarchical assembly of the RNP (14).
Although it was one of the first identified, 7SK still stands as an intriguing member among the fast-growing family of non-coding RNAs identified in humans (15,16). This abundant RNA found in the nucleus of higher eukaryotes functions as a regulator of P-TEFb, a transcription elongation factor required for the transition of promoter proximal paused polymerases into productive elongation (17–19), which is instrumental in regulating transcription in an appropriate temporal and spatial manner (20,21). 7SK sequesters and inactivates P-TEFb through the function of HEXIM proteins. Binding to 7SK enables HEXIM to interact with P-TEFb and inhibit its kinase activity (22–27). 7SK is a 331 nucleotide RNA transcribed by RNA polymerase III (Figure 1A). It has the usual stretch of uridines at the 3′-end that are required for efficient termination by RNA polymerase III. The 7SK-specific 5′ cap is mono-methylated at the gamma phosphate of the 5′ triphosphate by another component of the 7SKsnRNP, MePCE (also called BCDN3 in Drosophila) (10). Together, MePCE and LARP7 bind 7SK on both ends, thus forming a stable 7SKsnRNP core protecting the RNA from exonucleases (9,11,28–29). A model for the 7SK 2D structure (Figure 1A), based on experimental probing data, was proposed in the early 90’s (30), but while RNA domains involved in HEXIM-binding or P-TEFb regulation could be delineated according to it (31–35), it provided only poor information about how 7SK coordinates P-TEFb inhibition. Alternative 2D models for 7SK can be drawn with equivalent stabilities, suggesting that 7SK is intrinsically able to switch conformation. In fact, the original structural data can best be explained by the existence of at least two different conformations in the population of 7SK snRNPs in cells (36). An interesting model by Marz et al. (37) proposed the formation of a closed form of 7SK, based on the evolutionary conservation of sequences that would allow pairing of the first seven nucleotides of 7SK with a region just upstream of the terminal stem-loop (Figure 1A). This results in a lariat, a closed form with a dangling 3′-hairpin. In addition, this analysis highlighted the co-evolution of 7SK and LARP7, thus suggesting that LARP7 may have a chaperoning function for 7SK.

Several investigations carried out on human diseases highlighted the important role of LARP7 on the stability of the 7SK RNA, and consequently its function. Several frameshift mutations in LARP7 have been associated with gastric cancer (38). Mutations in the LARP7 gene were found associated with recessive cognitive disorders (39) and in primordial dwarfism associated with intellectual disability (40). These mutations seem to induce the loss of LARP7 protein through nonsense-mediated decay. Importantly, the loss of 7SK RNA as a consequence of the mutation was demonstrated in lymphoblasts from patients (40). This is in-line with previous experiments showing that the knock-down of LARP7 leads to decrease the level of nuclear 7SK in HeLa (11) and HEK293 cells (9).

As a consequence of the sequence similarity with La proteins, it was soon proposed that LARP7 uses its La module to bind the poly-uridine sequence at the 3′-end of 7SK (9,11). Indeed, almost half of the La module, the LAM region may be replaced with the LAM region of the genuine La protein, without dramatically reducing the binding to 7SK (11). In contrast, the specific recognition of 7SK involves two RRMs, the one adjacent to the LAM region (RRM1) and a C-terminal RRM (RRM2). This is demonstrated by the loss of binding specificity when the RRM1 is swapped with the RRM1 of La, or a point mutation introduced at a signature residue of RRMs (11).

To address the function of LARP7, we investigated the origin of LARP7 specificity for 7SK. The crystal structure of the La-homology domain of human LARP7, comprising LAM and RRM1, highlights specific features of the LARP7 RRM1 domain which suggest why it cannot be swapped for La RRM. The C-terminal RRM2 was shown by a combination of methods including binding assays, RNA footprinting and small-angle X-ray scattering (SAXS) to bind the apical loop of the 3′-hairpin. Taken together, our data support a model where both structural domains of LARP7 are combined to bind 7SK. LARP7 wrapping around the 3′ region includes the sequences closing the lariat form of 7SK. This constitutes a first evidence for the closed conformation of 7SK predicted by computational and phylogenetic analyses (37). LARP7 function would be to stabilize this closed conformation, thus bringing together the functional subdomains of 7SK.

MATERIALS AND METHODS

Preparation of RNAs and proteins

Several RNAs and protein constructs were designed, as detailed in the Supplementary Material section. All RNAs (apart from the 8-mer oligonucleotide UUUCUUUU, synthetic, from Dharmaco) were obtained by in vitro transcription. LARP7 full-length and the truncated versions were expressed in Escherichia coli. Mutagenesis was performed by the Quikchange approach (Stratagene).

Crystallization of complexes with RNA

Plate-shaped crystals were obtained with 30% PEG 3350 and 0.1-M succinic acid, pH 7.0 at 4°C with several RNAs (detailed in Supplementary material) but very few crystals diffracted well. All RNAs leading to crystal formation comprised the 7SK 3′-end oligonucleotide 325–332, but not systematically the HP4 hairpin. The structure was solved from a crystal obtained in a drop initially set up with RNA 300–332 comprising the HP4 hairpin but gel analysis of the drop showed that the RNA was degraded. After structural analysis, it appeared that the largest piece of RNA bound to the protein was a 5-mer corresponding to the 3′-end, which was probably protected from degradation by binding to the protein.

Crystal structure

The diffraction data were collected on beamline PX II on a Pilatus detector at SLS. The diffraction images were indexed and integrated using MOSFLM (41). The unmerged reflections were merged using the program SCALA (42) as a part of CCP4 suite of programs (43). The crystals belong to the space group C2 with the cell parameter \(a = 163.452\ \text{Å}, \ b = 33.50\ \text{Å}, \ c = 119.08\ \text{Å}, \ \alpha = 90.0^\circ, \ \beta = 128.99^\circ\) and
Figure 1. Domain organization of the molecules. (A) 2D model of 7SK with the 3′ U-triplet in orange and the additional residue 332 in gray. The two sequences that can form seven base-pairs are in red. The inset shows the closed 2D-structure thus formed. Arrows and numbers in purple indicate the 5′-boundary of the RNAs used in this work, for which the sequence is given. (B) Domain organization of LARP7 (582 amino acids in human) in linear representation with the color scheme adopted for the manuscript: LAM (28–111) yellow, RRM1 (120–188) orange and RRM2 (450–545) purple. The green horizontal arrow corresponds to the protein produced for the structural study. Blue bars represent stretches of basic residues. On top is the prediction of structure formation (red for unfolded, green for folded) as computed with Foldindex.

$\gamma = 90.0^\circ$. The structure solution was obtained by molecular replacement, using Phaser (44,45) with La protein (PDB 2VOO) as a search model (6). Initial rigid body and positional restraint refinement were carried out using CCP4 suite of programs (43). In the subsequent cycles, positional and B-factor refinements were performed using BUSTER and the simulated annealing refinements were carried out using CNS (46,47). Model building was carried out using COOT (48). Finally, the structure converged with R-factor and free R, 22.1 and 27.4%, respectively, with reasonable geometric parameters and B-factor (see Table 1 for statistics).

**Binding assessment with electrophoretic mobility shift assays**

The \( \gamma^3 \text{P}-5′ \)-labeled RNA (50 nM in all assays) was incubated 20 min at 4°C with increasing concentrations of proteins in a buffer containing 250-mM NaCl. Native gel analysis was performed as detailed in the Supplementary Material section.

**Footprinting**

The 5′-labeled 262-HP4 RNAs (50 nM) were mixed with protein in a similar buffer as in electrophoretic mobility shift assay (EMSA) and incubated at 4°C. The concentra-
of 7SK) linked with sequence AGA to the sequence 289–328 (comprising HP4 without the terminal uridine triplet). This AGA, adjacent to G8, was expected to favor a GAGA tetraloop, thus folding the M1 extension into an independent hairpin of seven base-pairs. This linear sequence was submitted to MC-fold (52), which proposed several 3D models of the M1–HP4. Those were indeed composed of two hairpins (M1 and HP4) whose orientations varied mostly at the level of the linker sequence AAAU (296–299). A series of 3D models, representing the most divergent families, were finally obtained by replacing the coordinates of the HP4 hairpin by those extracted from the solution structure (53) and removal of the helical extension introduced by the authors. Models for the RNA M1–HP4 in complex with the C-terminal domain were then manually assembled with P65, without attempting to change the coordinates of the protein (PDB id. 4ERD), as this structure has been suggested to be similar to the the structure of LARP7 C-terminal domain (54).

**SAXS experiments**

SAXS experiments were conducted on the SWING beamline at the SOLEIL synchrotron (λ = 1.033 Å). The Aviex charge-coupled device detector was positioned to collect data in the Q-range 0.008–0.33 Å⁻¹ (Q = 4πsinθ/λ, where 2θ is the scattering angle). All solutions were mixed in a fixed-temperature (15°C) quartz capillary with a diameter of 1.5 mm and a wall thickness of 10 μm, positioned within a vacuum chamber. Fifty microliter of a monodisperse sample of RNA–protein complex (70–130 μM) was injected onto a size-exclusion column (SEC-3 300, 150-Å Agilent), using an Agilent HPLC system, and eluted directly into the SAXS flow-through capillary cell at a flow rate of 0.2 ml min⁻¹. The elution buffer consisted of 20-mM Na HEPES, pH 7.2, NaCl 200 mM and 2-mM DTT. SAXS data were collected continuously, with a frame duration of 1.0 s and a dead time between frames of 0.5 s. Selected frames corresponding to the main elution peak were averaged using FOXROT, a dedicated home-made application. A large number of frames were collected during the first minutes of the elution, and these were averaged to account for buffer scattering, which was subsequently subtracted from the signal during elution of the protein. Data reduction to absolute units, frame averaging and subtraction were done using FOXROT. All subsequent data processing, analysis and modeling steps were carried out with PRIMUS and other programs of the ATSAS suite (http://www.embl-hamburg.de/biosaxs/atsas-online/).

Shapes of the M1–HP4 complex with the C-terminal domain were restored from the experimental data using the program GASBOR (55). These were averaged to determine common structural features and to select the most typical shapes using the programs DAMAVER suite.

The best model among those created for the RNA (M1-HP4) with P65 manually docked on the apical loop was sorted out by fitting with CRYSOL to the SAXS experimental data. The position of the P65 on HP4 was then further refined with program SASREF by rigid body molecular modeling against the shapes of the complex calculated from the SAXS data (56). In this last step of the modeling process, the

### Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>92.58–3.2 (3.37–3.2)</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
</tr>
<tr>
<td>α (Å)</td>
<td>163.45</td>
</tr>
<tr>
<td>β (Å)</td>
<td>33.50</td>
</tr>
<tr>
<td>γ (Å)</td>
<td>119.08</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>18.3</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>80.7</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Last resolution shell details are given in parentheses.

### Sequence alignments and figures

Two multiple sequence alignments were obtained independently. The first, for LARP7, resulted from a BLAST search starting with the human sequences Q4G0J3. Sequences were examined in the C-terminal region to distinguish LARP7 from other LARP sequences. The process led to about 50 sequences, treated with ClustalOmega, before visualization with Pymol (Figure 3B). The second alignment, of LARP7 and La, results from alignment with Muscle (49) of 15 sequences each, from species chosen to match as much as possible those presented in previous publications (1,50). The extracts presented in Supplementary Figures S2 and S7 were drawn with ESPRiPT (51). The figures were drawn using Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

### Molecular modeling

The starting model of the extended M1–HP4 RNA bound to the C-terminal domain was obtained as follows. First, a model for the RNA M1–HP4 was created from a linear sequence comprising the 5’-nucleotides (residues 1–8 of 7SK) linked with sequence AGA to the sequence 289–328 (comprising HP4 without the terminal uridine triplet). This AGA, adjacent to G8, was expected to favor a GAGA tetraloop, thus folding the M1 extension into an independent hairpin of seven base-pairs. This linear sequence was submitted to MC-fold (52), which proposed several 3D models of the M1–HP4. Those were indeed composed of two hairpins (M1 and HP4) whose orientations varied mostly at the level of the linker sequence AAAU (296–299). A series of 3D models, representing the most divergent families, were finally obtained by replacing the coordinates of the HP4 hairpin by those extracted from the solution structure (53) and removal of the helical extension introduced by the authors. Models for the RNA M1–HP4 in complex with the C-terminal domain were then manually assembled with P65, without attempting to change the coordinates of the protein (PDB id. 4ERD), as this structure has been suggested to be similar to the the structure of LARP7 C-terminal domain (54).
nucleotide G312 interaction with the C-terminal domain was considered as a supplementary distance constraint.

RESULTS

Defining the domains of LARP7 required for the study

The sequence of LARP7 (582 amino acids, in human) comprises three regions (Figure 1B). Following a short unfolded region containing positively charged amino acids (1–27), the La module comprises two structured domains, one containing the LAM (residues 28–111) and the second, an RNA-recognition motif (RRM1; residues 120–199) according to a global analysis of the LARP superfamily (1). At the C-terminus, a domain comprising RRM2 (residues 450–545) has been hypothesized to be similar to the xRRM domain found in P65, a protein involved in the telomerase complex in Tetrahymena (50). The xRRM fold differs from most RRMs because of the peculiar folding property of its C-terminal helix, which dramatically extends when binding to its RNA target (54). Between these folded N- and C-terminal domains, most of the central region of LARP7 is predicted to be unfolded, except a short region around residue 400 just before RRM2 (Figure 1B). In addition, the extreme N-terminus and the linker region contain stretches of basic residues.

In the prospect of elucidating how these modules are combined in LARP7 to bind specifically the 7SK RNA, we expressed in E. coli the full-length LARP7, the La module comprising the LAM and RRM1 regions with an N-terminal extension (1–208) and the C-terminal domain comprising the RRM2 (433–582). Several boundaries were tried for the La module, but only 1–208 was considered for a structural analysis (see details in the Supplementary Material section). Elucidation of the role of the C-terminal RRM2 was approached by biochemical experiments with the construct (433–582). Crystallization assays were focused on complexes of the La module with RNA.

Crystallization of a complex of the LARP7 La module with RNA

Binding to RNA was monitored in vitro by EMSA experiments (Figure 2). These showed that 7SK RNA and LARP7 interact without any additional partner and confirmed that 7SK truncation of the 3′ polyU reduces the binding of LARP7 about 1.6 times (Figure 2A), as anticipated from the homology with La (3). Further truncation of the entire 3′-hairpin in 7SK–ΔHP4 (1–295) led to further loss of binding (Figure 2A), reducing affinity about 2.2 times. Interestingly, this indicated that other parts of 7SK are involved in binding the full-length LARP7. The present study was mainly focused on the 3′-end domain of 7SK comprising the HP4 hairpin (300–331; Figure 1A), which is predicted in all 2D models of 7SK and is the only subdomain for which a 3D structure (PDB 2KX8) is available (53). With the La module (1–208) and the RNA corresponding to the 3′-domain, the truncation of the 3′-end uridines showed a drastic effect (Figure 2B), thus confirming that the 3′-uridines of 7SK are essential for the La module binding.

The crystallization assays included single-stranded oligonucleotides (325–332, 314–332) as well as RNAs comprising the HP4 hairpin (302–332, 300–332, 287–332). Thermoﬂuor experiments (57) showed a considerable increase of protein stability, with a Tm change from 26° to 43° upon RNA binding. Therefore, RNAs were mixed with puriﬁed protein prior to concentration and set-up of crystallization trials. Similar crystals were obtained in similar conditions with all RNAs, but very few of them diffracted well. The structure was solved from a crystal obtained in a drop initially set up with the hairpin HP4 (300–332). However, a check of the drop content after crystal mounting showed that the RNA was degraded. The formation of crystals with similar unit cells in drops initially containing RNAs which all comprised the 325–332 sequence immediately suggested that this short UUUCUUUU-3′ stretch of 7SK sequence was the longest oligonucleotide possibly present in the crystal. Indeed, difference Fourier showed densities for only three and ﬁve nucleotides in monomers A and B, respectively (Supplementary Figure S1). The electronic densities corresponded to pyrimidines and were interpreted as 5′-CUUUU-3′.

Global view of the structure of the La module of LARP7

The asymmetric unit of the monoclinic crystals contains two protein molecules. The best deﬁned monomer B will be described in the following text (Figure 3A). Although they share only 34% sequence identity, the La module of LARP7 showed a great degree of structural similarity with the La module of HsLa (6), as indicated by an RMSD of 2.0 Å for 188 residues. Figure 3A shows the two subdomains, LAM and RRM1 with the characteristic architectures observed previously in HsLa (5–6–58–59). Namely, the topology of the LAM subdomain is that of a winged helix-turn-helix, a fold often encountered in transcription factors involved in DNA binding, but with helices α2 and α3 inserted into the standard winged helix-turn-helix. It comprises six helices and two short β-strands. The N-terminal residues (1–28) are not visible in the map.

RRM1 shows a variant form of the RNA recognition fold found in many RNA-binding proteins (4), an ancient and abundant fold built around a central β-sheet, with two helices packing against one face (Figure 3A and B). Most RRMs use the central β-sheet surface to bind RNA. This surface is characterized by a cluster of aromatic residues, from hallmark sequences RNP1 and RNP2, located in strand 188, strands 5–6, 7 and 8, located on β strands, respectively (Figure 3B and Supplementary Figure S2). As expected from its early identiﬁcation as an RRM (60), the 3D structure of the La protein showed the RRM1 adjacent to LAM to be standard (5–6–58–59). Most LARP proteins contain RRM-like variants, as for example LARP6 (7). Interestingly, the present crystal structure of the human LARP7 shows RRM1 to be smaller than the standard fold, with a β-sheet of only three strands (Figure 3B and C). Although strand β4 is missing, the essential part of the β-sheet is maintained. The aromatic residues of the motif signatures, here Tyr128 from the RNP2 and Phe170 from the RNP1, are solvent-exposed (Figure 3B). Loop 3 connecting strands β2 and β3 is quite long (Figure 3A). It comprises two groups of two residues (158–159 and 164–165) facing each other, and forming two very short strands according to the secondary structure determination.
Figure 2. LARP7 N-terminal domain requires the 3′-end uridines for binding. (A) EMSA on a native agarose gel showing complex formation (arrow) after incubation of LARP7 full-length with 7SK: (7SK) wild-type (noU) RNA (1–328) deprived of the terminal U-triplet (ΔHP4) (1–295) deprived of the 3′-hairpin HP4. Complex quantification as a function of protein concentration is reported below. (B) Native gel after incubation with the La module (1–208) showing the free (lower arrow) and the complexed (top arrow) RNAs, which are schematized on top: (HP4) hairpin 300–331, (HP4noU) hairpin 300–326 deprived of the terminal U-triplet.

program STRIDE (61). Helices RRM-α1 (138–147) and RRM-α2 (176–184) pack against the other face of the β-sheet (Figure 3A). There is no additional C-terminal helix (α3 in HsLa), but helix α2 is one turn longer than in La, and extends to the last visible residue, Asn188. To rule out the possibility that our design of the protein was too short to include helix α3, we attempted to produce a larger protein construct encompassing residues 1–228, but unfortunately it was poorly soluble and could not be used in crystallization or binding studies. On the linker side, there is an N-terminal Helix RRM-α0 (121–125), as in La, but much shorter, and reduced to one helical turn. Together, these differences contrive to make the RRM1 domain of LARP7 singularly small. This is highlighted in Figure 3C, where the structures of the La modules of LARP7 and La were superimposed.

In LARP7, the linker connecting RRM1 to LAM follows a similar path as in HsLa, and has a similar size of nine residues (10 in HsLa) between the RR sequence from the ‘wing 2’ motif (Arg110-Arg111), which marks the C-terminal boundary of the LAM domain (7) and the first residue (Asp121) from the RRM-α0 helix. Interestingly, the connection between LAM and RRM has recently been hypothesized to impact the relative orientation of the LAM and RRM1 domains, and thus the RNA-binding property of the protein (7). In LARP7 the path of the linker is constrained by a salt bridge, between Arg118 and Glu122 and the relative orientation of the domains is maintained by a conserved salt bridge between Lys53 and Glu172 (Figure 3B). This results in an orientation similar as in HsLa. Together, the LARP7 La module composes a stable structure closing on the 3′-end of the RNA.

Specific features of the RRM1 of LARP7

The larger stretches of residues absent from the sequence and structure of LARP7 are highlighted in blue in the 3D structure of La (Figure 3C). The largest stretch corresponds to the missing fourth β-strand and the C-terminal helix α3 of the RRM1. Another missing stretch corresponds to the amputation of helix RRM-α0. Interestingly, these deletions are clustered in 3D and align continuously along the same surface, opposite to the uridine-binding cleft (Figure 3C). Together, these deletions suggest that RRM1 in LARP7
could have special properties. This is supported by the multiple sequence alignment shown in Supplementary Figure S2 where LARP7 and La sequences were compared. Several residues of the β-sheet are conserved only in LARP7, and not in La sequences. One of these is Glu130, adjacent to the Tyr128 of the RNP2, at the edge of the β-sheet opposite to the binding site of the 3′-uridine (which is described in the next section). Interestingly, Phe168, a hallmark residue of the RNPI in RRMs (4), present in LARP7, is a serine in La. On the other face, the two helices are packed closer to the β-sheet in the LARP7 structure. The contact involves several hydrophobic amino acids, as for example Trp140, Phe145 and Phe185. An interesting swapping of residues, conserved in evolution, is observed with Phe145 (Supplementary Figure S2). This phenylalanine comes from helix α1 in LARP7 but from helix α2 in La.

Taken together, these observations support the 3-stranded RRM as a genuine characteristic of LARP7 proteins. The La module comprising this shortened RRM1 binds the 3′ domain of 7SK (Figure 2B). The existence of a fourth β-strand and α3 helix of the RRM cannot however be totally ruled out in full-length LARP7, as a deep analysis of the alignment showed a weak sequence similarity of the α3 helix with a remote part of LARP7 (amino acids 375–390; Supplementary Figure S2B). If a 4-stranded RRM is formed in LARP7 proteins, it would thus involve a huge insertion of ~200 amino acids. This still makes the RRM1 in LARP7 very different from classical RRMs (62). Interestingly, LARP6 shows a very different situation, with additional helices blocking access to the RNA-binding face of the RRM (7). Here, the smaller RRM1 domain of LARP7 rather suggests an increased accessibility.

Recognition of the RNA 3′-terminal triplet

In the electronic map, three uridines in one monomer and five nucleotides in the second monomer are visible in the cleft formed between LAM and RRM domains (Figure 4A and Supplementary Figure S1). The three terminal uridines from the two monomers superpose well, and the following description depicts the monomer showing five residues. Most interactions involve H-bonds with residues of the LAM domain (Figure 4A and B). As was observed in La protein (6), the penultimate U-2 (numbering as in La) is anchored at the bottom of the crevice, and U-1 (the 3′-terminal uridine) at the surface of the LAM domain. The uridine U-3 stacks on U-1, leading to a characteristic fishhook shape of the backbone (Figure 4A).

The base of the terminal uridine U-1 stacks on Phe56 from the LAM-α3 helix. The stacking of the base of U-3 further restricts the binding pocket, which is limited at the bottom by Phe77, lying on the sugar ring. The terminal ribose binds Asn50 and Asp54 (Figure 4B). Both 2′ and 3′ hydroxyl groups from U-1 are bound simultaneously, thus ensuring that the ribose is 3′-terminal. All residues involved in U-1 binding are invariants in LARP7 and La (noted $ in the alignment; Supplementary Figure S2), including Asn50, which was not reported as binding the RNA in HsLa. The distance observed with the 3′-OH in LARP7 is quite long (3.9 Å), but still compatible with H-bond formation, and shorter than the distance in HsLa (4.4 Å). Interestingly,
this binding pocket, open to the solvent at the base edge, is not specific for uridine. In the course of our study, the binding of LARP7 N-terminal domain was measured with RNA variants of the terminal residue. The rationale was to test whether LARP7 could distinguish mature 7SK (331 nucleotides ending by CUUU) from transcript (332 residues, CUUUU) or maturation intermediates, such as a version with 332 nucleotides and a terminal adenine (CUUUA) that was mentioned in an earlier study (63). No binding differences were observed in EMSA experiments with these variants (Supplementary Figure S3). Similar absence of discrimination was observed with the full-length 7SK of 331 or 332 nucleotides. However, when the RNA was produced by self-cleavage from a transcript containing a 3′ ribozyme (which was attempted to produce homogeneous molecules with defined 3′-terminal ends to improve homogeneity for crystallization (64)), the binding efficiency was decreased. This was ascribed to the presence of a 2′-3′ cyclic phosphate resulting from the cleavage by the ribozyme. This highlights that the 2′ and 3′ hydroxyls of the terminal ribose must be free for LARP7 to bind, while the nature of the terminal base is of less importance. Similar results showing that the sequence of the 3′-end residue was not essential were obtained with HsLa (59).

Specific binding of the penultimate residue U-2 involves residues from both LAM and RRM1 domains (Figure 4C). The base ring is sandwiched between Phe44 from LAM and His138 from RRM1, in a pocket closed by Tyr153 from RRM1. Interestingly, in La, the bottom of the U-2 binding pocket is also closed by a tyrosine (Tyr23 in human La), which comes from the LAM side. Specific binding of the pyrimidine ring O2 atom results from H-bonds with Gln41 and Lys79. The O4 atom characteristic of uridine faces the RRM domain and points toward the β-sheet, between the β2 edge and the RRM-α1 helix (Figure 4C). This arrangement provides for one H-bond with the main chain nitrogen of Ile154. Residues His138 and Glu142, from the RRM-α1 helix are in correct orientation for H-bond formation; however, in the present crystal, the distances are slightly too long (above 4 Å). Most of the residues participating in the U-2 binding site belong to a group of residues specific of LARP7 proteins, as revealed in the sequence alignment (Supplementary Figure S2). Among those, His138 is 89% conserved. Ile154 is 50% conserved and can be replaced by a valine (50%). The others (Trp140, Phe145, Tyr153, Ser155, Pro157) show conservation above 94%. On the LAM side, Asn78 is 94% conserved and Lys79, 82% conserved, may be replaced by an arginine. Besides its stacking on U-1, U-3 is bound by Asn78 at its O2 atom, but is not further stabilized at the O4 edge, which faces the solvent. The same situation was observed in HsLa (6).

**Potential binding to other parts of the RNA**

The upstream nucleotides, U-4 and C-5, are only visible in one monomer of the asymmetric unit. They lie approximately in the same planes as U-3 and U-1, respectively, as shown in Supplementary Figure S4A. The ribose-phosphate chain is driven apart from helicity, with the bases U-4 and C-5 unstacked. This arrangement could be linked to the proximity of the long loop 3 connecting β2 and
β3, which was previously involved in RNA binding (4). In LARP7, loop 3 can be pictured as a guide pushing the RNA on the LAM surface into the binding cleft, as highlighted in Supplementary Figure S4A. In that context, Lys160, which here stacks on the U-4 base, could play a prominent role. Indeed, together with a Tyr159, this residue is quite conserved in LARP7 sequences (Supplementary Figure S2).

In the crystal, nucleotides U-4 and C-5 are involved in a packing contact involving the β-sheet of a neighboring molecule (Supplementary Figure S4A). The RNA is facing the aromatic residues from the signature sequences RNP1 (Phe168 and Phe170) and RNP2 (Tyr128). This suggests that the β-sheet in LARP7 RRM1 may bind RNA. Analysis of the packing contact does not reveal direct interactions with the RNA, apart from Tyr128 stacking on U-4. Interestingly, the mutation of Tyr128, hallmark of the RNP2, was reported in an earlier work to result in a loss of binding specificity to 7SK (65). Residue Glu130, one of the residues specific of LARP7s observed in the sequence alignment (Supplementary Figure S2), 94% conserved, is found in the vicinity of C-5. It is positioned near the nucleotidic base, suggesting that it could participate with Tyr128 to 7SK recognition. Phe168, a hallmark of RNP1, is stacking on Tyr128. To further clarify their role in RNA binding, residues Glu130 and Phe168 were mutated to alanines. We observed that RNA binding was not affected for the F168A mutant, but decreased for the E130A mutant (Supplementary Figure S4B). Checking by circular dichroism indicated that the E130A mutant protein showed the same global conformation as the wild-type. This suggests that the LARP7-specific residue Glu130 at the β-sheet edge of RRM1 is involved in 7SK binding.

Binding of the C-terminal RRM2 to 7SK

Most RRM-containing proteins have at least two RRMs, which often combine, leading to an expanding wealth of RNA- and protein-binding catalog. We therefore set out to investigate the function of the second, RRM2 domain at the C-terminus of LARP7.

A construct encompassing the C-terminal RRM2 (433–582) was assayed in binding experiments. It showed strong binding to several RNAs derived from 7SK, including those restricted to the 3′-hairpin, HP4 (Figure 5A and B). A recent breakthrough into specific recognition came from a mutational analysis establishing that position G312 in the apical loop of hairpin HP4 is essential for the 7SK to be correctly bound in vivo (32). In agreement with the in vivo experiment, EMSA assays in vitro with purified full-length LARP7 showed that the mutation G312C strongly reduced the binding with HP4. This mutation also abolished the binding with the RRM2-containing domain (Figure 5A and B), thus showing that this domain binds to the apical loop of HP4. Mutation G312C did not, however, decrease the binding of LARP7 N-terminal domain, indicating that it does not contact the G312 position. Considering that the La module binds the 3′-end uridines, this suggests that LARP7 folds back to position its C-terminal domain on the terminal hairpin of 7SK. Indeed, it was possible to bind simultaneously the N- and C-terminal domains of LARP7 on RNA constructs containing the 3′-hairpin of 7SK, and observe supershifted bands (Figure 5C and D), regardless of the order of addition of the two proteins.

Knowing the tendency of RRM domains to pack together, supported by packing contacts observed in the crystals of LARP7, we wondered whether an interaction could be established between RRM1 and RRM2. However, binding the La module to the 3′-end of a hairpin mutated at G312 did not recruit the RRM2 domain (Figure 6C). Simultaneous binding of the N- and C-terminal domains on the 3′-hairpin thus seems driven by RNA and not by interactions between the domains. More data will be required to investigate the situation when the domains are linked by the central region of LARP7.

Footprinting investigation of the positions of LARP7 N- and C-terminal domains on the 3′ hairpin

To position the N- and C-terminal domains on the 7SK 3′-hairpin, we used footprinting experiments (66) and compared the accessibility of nucleotides to RNases in free RNA and RNA complexed with the three protein constructs (Figure 6). The RNA was 262-HP4 (262–331; Figure 1A). We used RNase T1, which recognizes the guanosine base when it faces the solvent and cleaves the ribose-phosphate chain on the 3′ side of guanines, and RNase V1, which cleaves structured regions. In the absence of protein (Figure 6A, lanes R), RNase T1 strongly cleaves the hairpin loop at G312, while V1 cleaves in the helical regions of the hairpin on both sides of the loop. LARP7 C-terminal domain protects G312 from T1 cleavage, as well as the adjacent nucleotides 311 and 313 from in-line cleavage. The protection extends to the 5′-side of the loop, to the V1 cleavage at C306. The N-terminal domain protects a weak V1 cleavage at 325, in the stem just on top of the 3′-terminal single-stranded tail.

At the foot of the HP4 hairpin, the adenosines A296–A297 show strong V1 cleavages, indicative of base-pair formation. They might possibly pair with the terminal poly-U, but the V1 signal is unchanged upon complexation with the N-terminal domain, which captures the 3′-end. This suggests that A296–A297 connect elsewhere in 7SK. Binding of full-length LARP7 induces the disappearance of this V1 cleavage, which suggests a protection induced by the central linker region. An alternative explanation is that the structure probed by the V1 cleavage melts upon LARP7 binding, suggesting a conformational response of the RNA upon LARP7 binding, a situation reminiscent of the telomerase case (54). This hypothesis requires further investigation.

Before using the footprinting information to guide docking experiments, more information was required regarding the bulge at 320–321. The presence of a bulge at 320–321 in the upper part of hairpin HP4 was found to be essential for LARP7 binding in the in vivo study (32), but its sequence seemed to be free, as it could be changed without impacting LARP7 binding. Indeed, we checked that changing the bulge, even by a drastic reduction to one residue, did not compromise the binding to 7SK (Figure 7A). We took advantage of this freedom to change C320 for G320, and monitored the accessibility of this guanine with RNase T1. Footprinting of 262-HP4 with the mutation C320G (Figure 6B) showed profiles similar to the wild-type situation, but for
Figure 5. The C-terminal domain of LARP7 binds the apical loop of the 3′ hairpin of 7SK. (A) EMSA comparing binding of wild-type RNA HP4 (left panel) and mutant G312C (right panel) with LARP7 full-length (FL), La module (Ndom) or C-terminal domain (Cdom). (B) For each protein, the complex formation with wild-type or mutant HP4 reported as a function of protein concentration. (C) EMSAs of HP4 incubated with La module- (Ndom), C-terminal domain (Cdom) or both domains or mutant G312C (HP4mut) incubated with both domains. The supershift is labeled with a star (*). (D) Binding of the 5′-extended 262-HP4 RNA with the N- (Ndom), C- (Cdom) or both domains, analyzed with a native agarose gel. The supershift is labeled with (*).

LARP7 recognition at the foot of the 3′-hairpin supports a closed secondary structure of the 7SK RNA

Interestingly, the 5′-extension in the 262-HP4 RNA is not fully flexible, as indicated by the V1 cleavage at position 286, suggesting the formation of a structure. Indeed, Mfold (67) predicts this extension to form an additional hairpin, represented in Figure 6C. Moreover, we observed that assembly of N- and C-terminal domains was facilitated with 5′-extended constructs of HP4. For example, the RNA 262-HP4 (262–331) showed clear binding with each domain as well as clear supershifts (Figure 5D). Comparison of binding with shorter and longer constructs, as reported in Supplementary Figure S5, shows an increase of affinity for the 5′-extended RNA of ~25 times for the N-terminal domain, while it does not vary for the C-terminal domain. This suggests that the N-domain binds not only the 3′-end but also the 5′ foot of HP4.

The sequence upstream of HP4 (289–295) is well conserved (12). It has co-evolved together with seven nucleotides at the 5′-end of 7SK, with which it was hypoth-
Figure 6. Localization of the binding sites of the N- and C-terminal domains on the 3′ hairpin of 7SK. (A) Footprinting experiment of the 5′-labeled 262-HP4 RNA with RNases T1 and V1. Denaturing gel showing the cleavage products of the free (R) and complexed RNA with domains N- (N), C- (C), both (NC) or full-length LARP7 (F). Sequence was indexed with T1 in denaturing conditions (G) and ladder (AH); cleavages positions are indicated in green (V1) or blue (T1); control without treatment (C) shows in-line (purple) cleavages. (B) The same experiment but with the mutated 262-HP4 C320G. (C) Summary of the footprinting results. Arrows show cleavages with RNase T1 (blue), V1 (green) or in-line (pink). Circles represent protections with domains N- (orange) and C-terminal (purple) or full-length LARP7 (red).

SAXS study of the complex of LARP7 RRM2 with the 3′ region of 7SK

An RNA named M1-HP4 was designed by linking the 5′-sequence of 7SK (GGAUGUG) to the 3′ region at C299 by a GAAA sequence, and produced by in vitro transcription. Modeling with MC-fold (52) indicated that this RNA forms a structure with two hairpins, where the seven base-paired M1 extension closed by a GAAA tetraloop is appended to the hairpin HP4. M1-HP4 was bound efficiently by LARP7. The M1-HP4 sample was mixed with LARP7 C-terminal domain, and the complex submitted to SAXS analysis (Supplementary Figure S6). Starting models of the complex of M1-HP4 with RRM2 were generated by manually docking the structure of P65 on M1-HP4 RNA models obtained with MC-fold (52). This was based on the struc-
Figure 7. Model for the assembly of LARP7 N- and C-terminal domains on the 3′ hairpin of 7SK. (A) Impact of 7SK mutations on the binding with LARP7, estimated by competition experiments. R and X show the migration of the P32-labeled RNA when 7SK is free (R) or bound to full-length LARP7 (X). Increasing amounts (250, 500, 750 or 1000 nM) of mutant RNAs were incubated together with 7SK (50 nM) before adding LARP7. RNAs were 7SK full-length (WT), 1–295 (ΔHP4), 1–328 (noU), mutated at the HP4 bulge (Δ320, U321G) or apical loop (G312C) or at the sequence 289–295 (M1) or deleted of residues 1–8 (Δ9). (B) 2D model showing the M1 region, connecting the 5′-end of 7SK (orange) with the 290–295 sequence (red) and the binding positions of the LARP7 domains N- (orange) and C- (purple) with the linker in green. (C) Our working model, corresponding to the squared region in (B), showing the M1-HP4 RNA with the N-terminal domains (LAM, orange and RRM1, yellow) and the C-terminal domain, represented by the structure of P65 (PDB id. 4ERD, purple). The 5′-terminal phosphate of 7SK is indicated with a green sphere.
tural analysis of P65, which was suggested to have a similar structure as LARP7 C-terminal domain (50,54). It was immediately clear that the SAXS experimental curve was best fitted when the P65 was docked on the apical loop. Fitting with the SAXS experimental curve was then used to choose the best among the models provided by MC-Fold. The best model corresponded to a coaxial stacking of the HP4 and M1 hairpins, a favorite in RNA structures. In parallel, the shape of the complex shown in Supplementary Figure S6B was restored from the experimental data using the program GASBOR (55). Finally, starting with the best model for M1-HP4 with P65 grossly positioned at the apical loop, the position of the RRM2 (here P65) was refined by rigid body molecular modeling against SAXS data with SASREF (56). During this last stage of the process, a distance constraint was introduced, to maintain the interaction of nucleotide G312 with the C-domain. This process led to an excellent fit (chi² 2.1) as shown in Supplementary Figure S6A.

Model of LARP7 domains on the 3′ region of 7SK

The SAXS study thus confirmed that LARP7 C-terminal domain binds to the apical loop of HP4. Moreover, the atomic models generated with the SAXS study allowed a mutational analysis. Based on the alignment of P65 with LARP7 (Supplementary Figure S7A), we chose two residues close to the RNA in the model and conserved in LARP7, but different in P65 and not from the RNP sequence. Residues Tyr513 and Lys517 of P65 align with Lys535 and Asp539 of LARP7, respectively (Supplementary Figure S7A and B). In an EMSA experiment with M1-HP4, we observed that while D539A had no visible effect, the mutation K535A clearly decreased the binding to RNA (Supplementary Figure S7C).

The N-terminal domain was docked manually on the M1-HP4 model, by anchoring the U-4 nucleotide observed in the crystal on C328 from the model. The terminal U triplet was from the structure. This still leaves the N-terminal domain quite free to rotate around the connection. A more precise all-atoms modeling was not attempted, as it will require more data to orient the structural elements with confidence. Interestingly, the residue Glu130 which was suggested to be involved in 7SK-binding (Supplementary Figure S4) is positioned toward the M1 region of the RNA in the working model shown in Figure 7C.

DISCUSSION

The crystal structure of the LARP7 N-terminal domain, described here, is the first 3D structure of the La module of a member of the LARP family different from La showing the linked domains of the La module in a complex with RNA. Until the recent publication of the individual structures of the two domains of the La module from LARP6 (7), structures were available only for short fragments (68).

The triplet of uridines at the 3′-end of 7SK binds into the cleft between the LAM and RRM1 domains. It is constrained in a characteristic hooked conformation, allowing strict recognition of the penultimate uridine, with a contribution of the RRM1 revealing LARP7 specificity. The relative orientation of the two domains of the La module seems to be, as in HsLa essentially driven by the 3′-terminal uridines binding in the cleft between the domains, with U-2 located exactly at the same position. In the course of the La structural analysis, it was hypothesized that the two domains move freely in the absence of RNA (6), a hypothesis developed in the recent structural analysis of LARP6 that suggested a participation of the sequence at the exit of the LAM domain to the topological arrangements of the LAM and RRM1 domains (7). This movement may be restricted in LARP7, where a conserved salt bridge (Lys53 with Glu172) impacts the relative positions of the domains.

Comparing the present structure with HsLa shows that while the LAM domains are similar, the RRM1 are different. In LARP7, a smaller RRM1 is formed, with its β-sheet composed of three instead of four strands. The absence of the C-terminal α3 helix of RRM1 combines with the reduction of the length of the N-terminal helix of the RRM to increase the accessibility to the RNA-binding residues of the central β-sheet of RRM1. In RRM6s, the fourth β-strand often contributes H-bonds for the specific recognition of the RNA substrate. In LARP7, while it cannot be excluded that a fourth β-strand and α3 helix are recruited from downstream sequences, possibly via the RNA, the RRM1 shows several specific residues that could play a role in binding RNA. Residues including His138 and Ile154 were seen to participate in the penultimate uridine recognition, in a different way compared to La, thus showing a LARP7-specific response to a common task. Another residue, Glu130, also identified as LARP7-specific in the sequence alignment was shown by mutation to be involved in binding the 3′-terminal domain of 7SK. Glu130 is on the other side of the β-sheet, too far to be involved in the recognition of the terminal uridines. It may be participating in the specific function of LARP7, which is the recognition of 7SK. Further work will be necessary to identify the eventual binding site on 7SK. Interestingly, the RRM1 is very different in LARP6, where the RNA-binding surface is blocked by additional helices (7).

The 3′-hairpin of 7SK, HP4, was recently elegantly demonstrated to be the specific target of LARP7 in vivo (69). We now show that recognition of the 3′-end of 7SK occurs jointly through the N- and C-terminal regions of LARP7. Binding experiments and footprinting revealed that the C-terminal domain binds the apical loop of the HP4 hairpin. This domain is homologous to the telomerase protein P65, which forms an xRRM fold, an RRM with an extended C-terminal helix. In the telomerase, the RNA recognition depends on a two-nucleotide bulge, which is located in the middle of the hairpin. In 7SK, the HP4 hairpin also has a bulge, which was previously shown to be required for LARP7 binding in vivo (32), but without sequence specificity. The bulge in 7SK is not recognized by RRM2, but may be necessary to facilitate the packaging of 7SK into a functional conformation.

The N-terminal domain binds not only the terminal uridines but also the 5′ region at the foot of the 7SK 3′-hairpin. This sequence, which is highly conserved, has been previously hypothesized to form seven base pairs with the 5′-end of 7SK, thus forming a small stem named M1 (37). We show that mutations destabilizing this stem compro-
mise LARP7 binding, thus giving experimental evidence that 7SK is closed in the form of a lariat. Binding of the La module to the M1 stem, 5′ of the HP4 hairpin also explains why increasing the distance between HP4 and the 3′ uridine triplet compromises the binding of LARP7 in vivo (32).

The 7SK RNA is 5′-capped by the methyl-transferase MePCE (10), which has been shown to remain bound to 7SK after performing methylation (29). Interestingly, it was shown to bind LARP7 in that process (29). A closed 7SK, where the 5′ MePCE binding site is close to the 3′ LARP7 binding site, clearly facilitates this interaction. In our 3D model, accordingly, the 5′-end of 7SK (represented by a green sphere in Figure 7C) is free to bind MePCE.

We show that two domains of LARP7 bind 7SK, in a head-to-tail arrangement schematized in Figure 7B. The middle region of LARP7, which comprises stretches of basic residues, may also participate in the binding. This is suggested by the observation that full-length LARP7, but not the combined domains, induces protections from RNase cleavage at two positions in HP4, the bulged C320 and the A296–A297 at the foot of the hairpin. Additional binding to the 7SK core outside of HP4 is not excluded either. This is indicated by the observation that deletion of the entire 3′-hairpin still allows complex formation.

Several recent reports of major dysfunction in humans suggested that LARP7 and 7SK work as a pair to regulate the transcription factor P-TEFb. The present work showing how LARP7 is entwined with the 7SK RNA suggests functional correlations. Firstly, LARP7 binding could narrow the range of 7SK conformations, thus facilitating recognition by stabilizing a functional RNA structure. Secondly, LARP7 binding may help compact the RNA, by minimizing the phosphate–phosphate repulsions with its basic stretches of residues working as polyamines in the packaging of nucleic acids in viral capsids. Such a chaperoning mechanism could be aided by MePCE binding to the 5′-end. Thirdly, complex formation leads to reduction of the RNA surface accessible to other partners, such as HEXIM, PTEFb or hnRNPs. A heterologous RNA–protein surface, RNA surface accessible to other partners, such as HEXIM, accordingly, the 5′ green sphere in Figure 7B.

The 7SK RNA is 5′-capped by the methyl-transferase MePCE (10), which has been shown to remain bound to 7SK after performing methylation (29). Interestingly, it was shown to bind LARP7 in that process (29). A closed 7SK, where the 5′ MePCE binding site is close to the 3′ LARP7 binding site, clearly facilitates this interaction. In our 3D model, accordingly, the 5′-end of 7SK (represented by a green sphere in Figure 7C) is free to bind MePCE.

We show that two domains of LARP7 bind 7SK, in a head-to-tail arrangement schematized in Figure 7B. The middle region of LARP7, which comprises stretches of basic residues, may also participate in the binding. This is suggested by the observation that full-length LARP7, but not the combined domains, induces protections from RNase cleavage at two positions in HP4, the bulged C320 and the A296–A297 at the foot of the hairpin. Additional binding to the 7SK core outside of HP4 is not excluded either. This is indicated by the observation that deletion of the entire 3′-hairpin still allows complex formation.

Several recent reports of major dysfunction in humans suggested that LARP7 and 7SK work as a pair to regulate the transcription factor P-TEFb. The present work showing how LARP7 is entwined with the 7SK RNA suggests functional correlations. Firstly, LARP7 binding could narrow the range of 7SK conformations, thus facilitating recognition by stabilizing a functional RNA structure. Secondly, LARP7 binding may help compact the RNA, by minimizing the phosphate–phosphate repulsions with its basic stretches of residues working as polyamines in the packaging of nucleic acids in viral capsids. Such a chaperoning mechanism could be aided by MePCE binding to the 5′-end. Thirdly, complex formation leads to reduction of the RNA surface accessible to other partners, such as HEXIM, PTEFb or hnRNPs. A heterologous RNA–protein surface, RNA surface accessible to other partners, such as HEXIM, accordingly, the 5′ green sphere in Figure 7B.

The 7SK RNA is 5′-capped by the methyl-transferase MePCE (10), which has been shown to remain bound to 7SK after performing methylation (29). Interestingly, it was shown to bind LARP7 in that process (29). A closed 7SK, where the 5′ MePCE binding site is close to the 3′ LARP7 binding site, clearly facilitates this interaction. In our 3D model, accordingly, the 5′-end of 7SK (represented by a green sphere in Figure 7C) is free to bind MePCE.

ACKNOWLEDGEMENTS

We thank the PX II beamline staff of the Swiss Light source synchrotron and the SWING beamline staff at SOLEIL synchrotron for their support. We thank Olivier Bensaude for his constant support. We are particularly grateful to David Price, Dino Moras and Valerie Biou for their comments on the manuscript and help with writing. We thank Magali Blaud for invaluable help with the circular dichroism technique. We also thank the members of the Structural Biology and Genomics platform, IGBMC, Illkirch, and in particular Alastair McEwen and Pierre Poussin, who helped with crystals handling.

FUNDING

The French National Agency for Research (ANR) [Trscr-REGsnRNP ANR-06-BLAN-0072, DynamiC ANR-12-BSV5-0018]; Investissements d’Avenir, French Government, ANR [ANR-10-LABX-54 MEMO LIFE, ANR-11-IDEX-0001-02]; French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01]; European Strategy Forum on Research Infrastructures (ESFRI); CNRS to F.P., B.K., A.C.D.B.; JSPS ITP, Institutional Program for Young Researcher Overseas Visits and Toshiko Yuasa Memorial Scholarship Fund for Special Researchers to E.U.; Chinese Scholarship Council (CSC) to X.H.]. Funding for open access charge: CNRS. Conflict of interest statement. None declared.

REFERENCES


ACCESSION NUMBER
PDB: 4WKR.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

None declared.


Material and methods (supplementary details)

Design of RNAs

The choice of the RNA constructs was based on the sequence of the gene comprising an additional U at position 332. The RNAs assayed for crystallization of a complex with the La module of LARP7 were 302-332, 300-332 and 287-332 comprising the HP4 hairpin, and the oligonucleotides 314-332 and 325-332. While EMSA showed clear binding of the La module of LARP7 with RNAs comprising the HP4 hairpin (Figure 2), it did not with single-stranded oligonucleotides 314-332 or 325-332, suggesting the complex to be too unstable to withstand the electrophoretic process. However, 325-332 (UUUCUUUU) induced stabilization in a thermo fluor experiment, where Tm varied from 26°C for the free protein to 43°C for the bound protein. It was thus included in the crystallizations.

Design of proteins

Four proteins encompassing the La module of LARP7 were designed for the structural study, including or not the N-terminal low-complexity region: 1-228, 28-228, 1-208 and 28-208. Only 1-208 could be purified in quantity and quality compatible with crystallization. Similarly, several proteins encompassing the RRM2 of LARP7 were assayed for binding to HP4, among which 433-582 gave the best purification yield.
Preparation of RNAs and proteins.

All RNAs (see below), apart from the 8-mer oligonucleotide UUUCUUUU, synthetic, from Dharmaco, were obtained by \textit{in vitro} transcription of appropriate templates with T7 polymerase and purification of the RNAs on denaturing gels. Templates for transcription were obtained either by annealed oligos or by PCR, starting from plasmids of the type pHDV (Walker et al., 2003) into which each construct was cloned and primers chosen to produce a transcript ending with U331-3’OH or U332-3’OH. In the case of the RNA starting at 300, A301 was changed for a guanine to facilitate T7 transcription. Crystallization-grade RNAs were purified by anionic exchange on MonoQ followed by extensive dialysis against storage buffer (10 mM Na cacodylate, pH 6.5, 2 mM MgCl2, 0.25 mM EDTA). LARP7 full-length and the truncated versions were cloned in bacterial expression plasmids of the pnEA family (Diebold et al., 2011), derived from pET15. The plasmid producing an N-terminal His-tag (pnEA-NH) and TEV cleavage site was chosen for all proteins, except for full-length LARP7 which was produced with a C-terminal His-tag, and the crystallized N-terminal domain (amino acids 1-208) for which the P3C cleavage site was used. Expression of proteins, after transformation in \textit{E. coli} BL21-derived strain Rosetta (Novagen), was obtained by culture in auto-inducible medium at 28°C. Lysis was performed by sonication in the presence of 50 mM Tris, pH 7.6, 500 mM NaCl, 5 mM MgCl2, 1.4 mM \textbeta-mercapto-ethanol and protease inhibitors. After debris elimination, purification was performed in batch mode on Ni-beads. The tag was cleaved by overnight incubation with TEV (or P3C) during a dialysis into 20 mM Sodium HEPES, pH 7.2, 1 mM EDTA, 2 mM DTT and NaCl (300 mM for LARP full-length, 200 mM for domains). A cationic exchange chromatography was performed on HiLoad SP-Sepharose (GE Healthcare), followed by dialysis in storage buffer (same as dialysis buffer, above) and concentration by ultra-filtration. For biochemical usage, proteins at 10-50 \textmu M were kept in
aliquots at -80°C after addition of 10% glycerol and flash-freezing in liquid nitrogen. For crystallization, the final dialysis step was replaced by a size-exclusion chromatography on Superdex 75.

**Circular dichroism**

A potential unfolding of the E130A mutant version of the La module was checked by circular dichroïsm (CD). CD spectra were recorded using a Jobin-Yvon Mark VI circular dichrograph at a scan speed of 0.2 nm/s. Quartz spare cuvette with 0.1 cm path length was used. Blanks were run before each spectrum and subtracted from the raw data. Three spectra were averaged to increase the signal-to-noise ratio. The final proteins (wild-type and E130A) were in Hepes buffer (same buffer as in the final step of protein purification) or phosphate buffer (100 mM phosphate pH 7.5) and the assays were carried out at 20°C. The results are presented as normalized Δε values on the basis of the amino acid residue mass of 110 Da. Taking into account a sensitivity of δ(ΔA) = 10⁻⁶ for the apparatus, the protein concentration and the optical path-length of the cuvette, measurements were obtained at a precision of δ(Δε) = +/- 0.002 M⁻¹.cm⁻¹ per amino acid. The mutant E130A showed similar spectra as the wild-type protein in both buffers.

**Crystallization of complexes of La module with RNA**

Complexes were formed with 1.2:1 molar ratio of RNA to protein. The concentrations were calculated with MW and absorption coefficients from Expasy Website for the proteins, and OligoCalc (Kibbe, 2007) for the RNAs. Complexes were concentrated on Amicon ultrafiltration devices, controlling the concentration following absorption at λ=260 nm, up to about 5-10 mg/ml. Commercial crystallization kits were tried with a Cartesian robot, at
the Structural Biology Platform at IGBMC. Prior to flash-freezing the crystal in liquid nitrogen, the crystals were treated with the cryo protectant, 5% ethylene glycol dissolved in the crystallization solution.

**Experimental details for EMSAs and footprinting.**

The RNAs were labeled at the 5′-end with T4 polynucleotide kinase (Fermentas) and Y-\(^{32}\)P-ATP, following dephosphorylation with antarctic phosphatase (Biolabs). For 7SK, an additional purification on a denaturing gel was performed to eliminate fragments resulting from degradation. Thermal treatment (3 min at 90°C, followed by 5 min on ice) was necessary with 7SK to get one single band on a native gel. This treatment also helped to minimize the formation of duplexes, which formed spontaneously at RNA concentration higher than 0.5 µM, especially in the case of short RNAs such as 302-332. Interestingly, such duplexes elicited efficient binding (Supplementary Fig. S3).

Complexes were formed by incubation at 4°C of 50 nm RNA with proteins in the range of 0-2 µM for full-length and 0-10 µM for domains, in a buffer containing 25 mM Na HEPES, pH 7.2, 5 mM MgCl\(_2\), 250 mM NaCl, 2 mM DTT, 0.005% NP40, 10% glycerol, 0.05 mg/ml bovine serum albumin as well as 5 µM total tRNA to minimize non-specific binding. A small amount (1.5 µL) of dye mix containing 0.02% each of bromophenol blue and xylene cyanol in 60% glycerol was added to the 12 µL assay mix just before loading on a native gel in TBE (0.5X). For the small RNAs and hairpins, vertical gels (20 x 20 cm, 1.5 mm thickness) were prepared with 6% acrylamide (29:1), and run at 4 W for 75 minutes. For larger complexes with 7SK, gels (20 x 24, 5 mm thickness) were prepared with agarose (1.6%) and run horizontally in TBE 0.5X at 6W. In this case, RNAs were transferred on nitrocellulose membranes. Gels were revealed by phosphorimaging. Band intensities were
quantified with ImageJ. Curves [bound/ (bound+unbound)] were drawn with Excel.

For footprinting, the 262-HP4 RNA was gel-purified after labelling, and incubated at 4°C with or without proteins, in a volume of 10 μL, in a buffer containing 25 mM Na HEPES, pH 7.2, 5 mM MgCl₂, 250 mM NaCl, 2 mM DTT, 0.005% NP40, 0.05 mg/ml bovine serum albumin and 5 μM total tRNA. After 15 minutes, allowing for complex formation, 1 μL RNase T1 (Ambion; diluted 1/1000) or V1 (Ambion; diluted 1/10) was added. After 5 minutes cleavage at room temperature, gel-loading buffer was added with either 10 mM EDTA or 10 mM MnCl₂ for stopping V1 or T1 reactions, respectively. Reactions were immediately loaded on a sequencing gel (20x40 cm, 0.4 mm thick, 15% acrylamide in urea 8M and TBE), together with an alkaline ladder and RNase T1 in denaturing conditions for sequence indexation, and migrated at 20W.
**Figure S1.** Electron density maps at the RNA binding site

Stereoviews of electron density maps at the RNA binding site. (A) Final 2Fo-Fc map generated in the presence of the full model, including RNA, using BUSTER and contoured at 0.8 sigma. (B) Fo-Fc simulated annealing omit map (orange) generated in the absence of RNA using CNS, contoured at 2.5 sigma. (C) 2Fo-Fc simulated annealing omit map generated in the absence of RNA using CNS, contoured at 0.8 sigma.
Supplementary Fig. S2

Figure S2. Multiple sequence alignment of La and LARP7
Multiple sequence alignment of La and LARP7 showing the secondary structures depictions from the LARP7 structure (top) and La structure (2VOD, bottom). (A) La module with LAM in yellow, and RRM in orange. Blue arrows indicate the signatures sequences of RNP-1 and RNP-2. The residues binding the uridines triplet in LARP7 and HsLa are marked by a blue S. Specifically conserved residues are highlighted (blue for LARP7, green for La). Numbers on top correspond to residues number in the human LARP7 sequence, bottom numbers index the columns of the alignment. (B) Downstream sequences showing the 4th β-strand and C-terminal helix of the La-RRM1.
**Figure S3.** Tolerance towards variation at the terminal residue at the RNA 3’-end. EMSA gels showing the complex formation of full-length LARP7 (C) with hairpins with various 3’-ends. First panel: HP4 ending by 3 uridines (300-331); second panel: HP4 with a supplementary uridine (300-332); third panel: HP4 with a supplementary adenine. RNA dimers formation is indicated by (*) and complexes of the dimers by (**).
**Figure S4.** The RNA at a crystal packing interface 
(A) The 5-mer RNA (in grey sticks) with the 3’-end uridines bound between LAM (yellow) and RRM1 (orange) and nucleotides U-4 and C-5 in a crystal packing contact with the RRM1 of a neighboring RRM1 (in cyan, residues numbered with *). (B) EMSA with RNA M1-HP4 and wild-type (WT) or mutants E130A and F168A versions of the LARP7 La module.
Figure S5. Comparison of LARP7 binding with the 3’ hairpin of 7SK and its extended version. (A) Native gel analysis of complexes (arrows) of HP4 (left) and the 5’ extended 262-HP4 RNA (right) with LARP7 full-length (FL, green) or domains: La module (Ndom, orange) or C-terminal domain (Cdom, purple). (B) Complex formation as a function of protein concentration (full-length, green; La module, orange; C-terminal domain, purple).
Figure S6. SAXS analysis of the complex of RNA M1-HP4 with the C-terminal domain. (A) SAXS data analysis. Black line shows the experimental SAXS of a sample of M1-HP4 incubated with the C-terminal domain, and the red line shows the curve obtained from the atomic model of the complex. The logarithm of intensity is displayed as a function of the logarithm of scattering vector $q$ and a residual plot calculated from the relation $R(q) = (I_{\text{exp}}(q) - I_{\text{model}}(q)) / \sigma_{\text{exp}(q)}$ is also displayed with the corresponding $\chi^2$ value. (B) The low resolution molecular shape calculated with GASBOR (gray transparent spheres) and the atomic model performed with SASREF showed in two orientations.
Figure S7. Mutational analysis of the C-terminal domain binding to HP4
(A) Alignment of the P65 and P43 sequences with a set of 4 LARP7 sequences in the region of RRM2. Top numbers correspond to the tetrahymena P65 sequence, bottom green numbers to the human LARP7. Specific residues of LARP7 which were mutated are shaded green. (B) Our working model in the region of the apical loop of HP4, showing the P65 structure as a purple ribbon. The residues in stick are predicted to be involved in RNA binding, either because they belong to the RNP-2 (purple, Tyr407 in P65, Tyr483 in human LARP7), or correspond to LARP7-specific residues (Tyr513 and Lys517 in P65 align with Lys535 and Asp539 in LARP7, respectively). (C) EMSA of the C-terminal domain of LARP7 with M1-HP4, either wild-type (WT) or mutated at the LARP7-specific residues Lys535 and Asp539.
**Discussion**

To summarize, we built a 3D model, and showed that both the La module and the RRM2 can, together, wrap around at the 3’-end of 7SK RNA. This was supported by (1) footprinting data (2) measurement of the size and shape of the complex of the RRM2 and HP4 in solution, by SAXS and (3) analysis of the effect of mutations of residues conserved in the RRM2 of LARP7.

In the course of the study, it appeared that the La module of LARP7 binds also at the foot of hairpin HP4. Just upstream of HP4 lies the region where was indicated as M1 domain interacts with the 5’-end of 7SK RNA by the bioinformatic analysis of Manja Marz and colleagues (Marz et al., 2009). This interaction consists of a short stretch of 7 base pairs. We constructed an RNA comprising this region in addition to the HP4 (named M1-HP4). The binding of the La module to this RNA is more stable than with HP4 alone, suggesting that contacts are established with regions of the RNA outside of the 3’-end. Analysis of the structure suggests that these additional contacts could involve the RRM1, and more precisely a surface on the opposite side relative to the 3’-end binding site. As indicated in the manuscript, a mutation at one residue of this surface (E130) indeed weakened the interaction with the RNA M1-HP4. In the future, this mutation analysis should be extended to other amino acids of the RRM2, to best delineate the residues involved. Also, it is important to find out more precisely if there are specific position(s) of the RNA involved in the contact, or if the recognition is structural.

It is tempting to believe that these contacts help to pack 7SK RNA in a compact complex, and thus may help to stabilize the closed conformation of the RNA. This should be
further investigated. However, one important consequence of that finding for future studies is that the complex of M1-HP4 with the La module and the RRM2 makes a good target for crystallization and SAXS experiments. We aim to use this new construct in structural studies.
Chapter IV: LARP7 binding to the middle-region of 7SK RNA

1. LARP7 still binds 7SK RNA after deleting HP4

In the previous chapter we show that both LARP7 N- and C-terminal domains bind HP4, suggesting that the RRM2 in the C-terminal region participates to specific recognition of 7SK RNA. This may be only a part of the story. Indeed, Figure 4-1 shows that LARP7 still binds 7SK RNA even when HP4 is deleted (open red arrow, right side of figure 4-1). This suggests that other parts of 7SK RNA and LARP7 are involved in the interaction.

Figure 4-1. LARP7 binding extends further into 7SK RNA than the 3’end domain. EMSA on a native agarose gel showing complex formation (arrow) after incubation of LARP7 full length with 7SK RNA: (1) wild-type (2) RNA (1-328) deprived of the terminal U-triplet (3) (1-295) deprived of the 3’-hairpin HP4.
2. LaRP7 interacts with 7SK RNA sequence in domain 2

To investigate which part of 7SK RNA binds LARP7, we performed footprinting analysis of the full length RNA. V1, an RNase cleaving structured regions, was used to compare the folding of free 7SK RNA with the LARP7: 7SK RNA complex. The cleaved fragments were analyzed by sequencing with a reverse transcriptase (RT), starting from a primer hybridizing at 295-314. The primer was fluorescently labeled allowing to use a sequencer for analyzing the fluorescent cDNA products of the reverse transcriptase. The result is shown in Figure 4-2. Some peaks, corresponding to V1 cleavages in the free 7SK RNA (green curve), show decreased intensity when 7SK RNA was complexed with LARP7 (blue curve) before incubation with the RNase. Interestingly, a major effect is observed for 121-123 positions (green arrow). In this region, blue curve is lower than the green curve, which means V1 cleavages decrease in the presence of LARP7, indicating that V1 cannot see this region after LARP7 binding. One explanation is that 121-123
positions are bound by LARP7, which prevents RNase cleavage by steric hindrance. Another explanation is that the local structure is melted, and is no longer a target for RNase V1, which recognizes double-stranded regions.

In Figure 4-2, the comparison of curves reveals another change, at nucleotides 176-180. There, the effect is opposite: a V1 cleavage appears following complex formation. This could indicate a strengthening of the structure at 176-180 upon LARP7 binding, or an increase of accessibility to the probe. In both cases, this is a signature for a change of the 7SK RNA conformation.

After this exploration with RT sequencing, we investigated the accessibility to V1 by direct footprinting of 5’end labeled RNA. This allows to observe directly on the gel the 5’labeled oligonucleotides produced by the cleavage. This experiment is complementary to the observation of cDNAs reverse-transcribed from the 3’-end of the cleaved products. Assessing cleavage positions from the two parts of the product of cleavage (the 5’ and the 3’ oligonucleotides) is important. It allows to clarify whether the cleavage corresponds to a primary target of the probe (in the intact RNA), or results from a secondary cleavage (cleavage of an already cleaved fragment). Direct footprinting (with V1 probe) is represented in Figure 4-3. The experiment shows that position 122 is protected by LARP7 full length (VF) when compared to free RNA (VR). This result thus confirms that position 122 belongs to the LARP7 binding site. This is located in domain 2, the region which differs more between the published 2D models.
Figure 4-3. (a) Direct footprinting of 5’-labeled 7SK RNA with RNase V1. Denaturing gel showing the cleavage products of the free (VR) and complexed RNA with domains N-(VN), C-(VC), full length LARP7(VF), full length Hexim (VH) or both (VFH). Sequence was indexed with T1 in denaturing conditions (G) and a ladder (HA). On the left, are the controls without RNase. Cleavages position 122 is highlighted. (b) 2D structure of domain 2 in the 2D model of Wassarman & Steitz. (c) 2D structure of domain 2 in the 2D model of Marz. Nucleotide 122 is marked by black star.
This experiment was done also in parallel with the N-terminal and the C-terminal regions of LARP7, and HEXIM. We didn’t observe footprinting by N-terminal (VN) or C-terminal (VC) domains. This suggests that it is the linker of LARP7, the disordered region between RRM1 and RRM2 which is involved in the interaction with the domain 2 of 7SK RNA. This should be further investigated. HEXIM (VH) had no influence on the V1 RNase profile, and did not modify the profile obtained with LARP7. Since we checked that HEXIM was binding 7SK RNA in the conditions of the experiments, this indicates either that HEXIM binding site(s) is (are) located elsewhere in the RNA (HP1, which is known to comprise a major HEXIM-binding site, is not observed in this experiment), or do not correspond to V1 targets. Other probes should be investigated.

2.1 Mutation in domain 2 affects LARP7’s binding

Footprinting experiments indicated positions 121-123 in domain 2 to be involved in the LARP7 complex. To get further evidence that this region of the RNA interacts with the protein, we mutated these positions, and analyzed the binding capacity of LARP7 with the mutant RNA. The interaction of LARP7 was tested by EMSA in vitro (Figure 4-4). Binding was observed as slow-migrating bands corresponding to radioactively labeled RNA shifted by increasing concentration of LARP7. The mutant RNA (7SK RNA_C121G-C122G) was compared with the wild-type RNA (7SK RNA-WT). In figure 4-4(left), the shifted bands appear at a higher concentration of LARP7 for the mutant RNA when compared with 7SK RNA-WT. This indicates that LARP7 binds the mutant 7SK RNA, with a lower affinity than WT. Affinities are measured in the binding curves in Figure 4-4 (right). This result confirms that position 121-122 in domain 2 is important for LARP7 binding.
Figure 4-4. 7SK RNA-C121G-C122G mutation shows lower binding affinity than the wild type. The figure on the left shows the EMSA native gel, comparing binding of wild-type 7SK RNA (left panel) and mutant C121G-C122G (right panel) with LARP7. Figure on the right indicates the complex formation with wild-type and mutant reported as a function of protein concentration (in µM). The concentration of RNA in the assay was 50 nM.

Finally, to further assess whether this position in domain 2 is important for recognition, the mutant 7SK RNA_C121G-C122G was mixed with 7SK RNA-WT in a competition experiment. Figure 4-5 shows that the mutant does not compete with the wild type, even

Figure 4-5. EMSA competition assay. In this assay, 50nM radioactive labeled 7SK RNA was preincubated several concentrations (250/500/750/1000nM) of cold RNA mutants (CC121-122 mutated to GG, noU, deleted from HP4), or WT-7SK RNA, then 0.4µM LARP7 was added, and the samples were migrated on a native gel after further incubation allowing complex formation. Note that the concentrations of the cold RNA are at least 5 times more than radioactive labeled RNA. Lane R – shows migration of free 7SK RNA, X - 7SK RNA with LARP7.
when the concentration of the mutant was 25 times that of 7SK RNA-WT. In competition EMSA experiment, it is straightforward to compare the importance of mutations. Figure 4-5 indicates that this mutant is more competitive than 7SK RNA deleted from HP4, but quite similar as the 7SK RNA deprived only of the terminal uridines (noU). We concluded that LARP7 interacts with CC121-122 in domain 2 of 7SK RNA.

**Discussion**

The investigations described in this chapter were prompted by the observation that LARP7 still binds 7SK RNA even if HP4, the major binding site of LARP7, was deleted. A footprinting study indicated the CC121-122 positions of 7SK RNA as bound by LARP7. Furthermore, the mutant C121G-C122G binds LARP7 less efficiently than the wild type and was shown to be unable to compete with the wild-type RNA for LARP7 binding. We concluded that LaRP7 interacts with CC121-122 in the domain 2 of 7SK RNA. This is a new discovery. Interestingly, the sequence encompassing CC121-122 is conserved in evolution, will be discussed in the final discussion.

The investigation described in the present chapter suffered however from several limitations.

1. Clear results were obtained only with V1 probe. This RNase recognizes “structured” regions, which correspond to short stretches of stable helices. Investigation should have been completed by probing with an RNase specific for single-stranded regions, such as RNase ONE. Unfortunately, this RNase was not available at the time of the preliminary investigation with RT-footprinting. Moreover, when tried in direct footprinting, we
realized that it was quite difficult to set-up conditions favoring a majority of primary cleavages. RNase T1 was tried, but it suffers from a narrow specificity for guanines, and produced large peaks difficult to index in the preliminary experiments. Chemicals were also tried. Former experiments with the SHAPE technique (probe 1M7) on the HEXIM complex didn’t show clear protection signals. DMS was tried, but we met indexation difficulties. Probing with CMCT requires special preparation of proteins, since the probe is scavenged by the Tris.Cl and Na.Hepes buffers used in our purifications. It showed however promising preliminary information in direct footprinting experiments, and should be further explored.

2. Technical limitation preventing the analysis of the 5’ hairpin with the RT-footprinting technique was particularly detrimental to the analysis of the combination of LARP7 and HEXIM. This was unfortunate, since in the absence of any footprinting signal, we cannot ascertain that HEXIM is indeed bound to the RNA in the experiment (we did check, however, that in similar conditions HEXIM showed a shifted band in an EMSA experiment).

3. More mutations should have been tried, to see if it is the structure of the domain 2 that is recognized, or if the recognition depends on the sequence. Mutations C121U-C122U, or mutation on both sides of the strand could have been tried. It would be interesting perform G139C mutation to remain C122-G139 base pair and see if it compensates for the C121G-C122G’s deficiency.

4. The local secondary structure should have been analyzed to better understand the contribution of domain 2. Investigation by SHAPE (with the probe 1M7), performed by
Denise Martinez-Zapien, leads to the two experimental models presented in Figure 4.6. Interestingly, they are both compatible with a short hairpin encompassing residues 121-122. This hairpin is also present in the two published models of Wassarman & Steitz and Marz. Moreover, the region for which the V1 cleavages increase upon LARP7 binding in our experiment also forms a hairpin. This gives some weight to a hypothetical uncovering of this region, due to a conformational change of 7SK RNA in the presence of LARP7.
Figure 4-6. Two 7SK RNA secondary structure models constructed from SHAPE data by Denise Martinez Zapien. 7SK RNA secondary structure predictions were generated with RNA structure software (Reuter and Mathews, 2010) incorporating the normalized and averaged SHAPE reactivates into the energy function, and data from enzymatic probing. Images were generated using XRNA software (http://rna.ucsc.edu/rnacenter/xrna/xrna.html). A) Structure #1 and B) structure #2. Nucleotides are colored according to their reactivity to 1M7. Nucleotides in gray were not explored. HP1 and HP4 are indicated, as well as M1 and M6, and M6.
Chapter V: LARP7 binding to HP1 and HP3

1. LARP7 and HEXIM’s interactions with hairpins of 7SK RNA.

In the last chapter we showed that 7SK RNA domain 2 interacts with LARP7. This can explain why this protein still binds 7SK RNA after deletion of HP4. However, the binding may still be further extended to other regions. Since the protein LARP7 is large and contains a long unstructured region (210-430), it may bind 7SK RNA on many different sites. The same holds for HEXIM, which is also a large protein with disordered regions.

One initial major aim of the team was to understand where are the HEXIM binding sites on 7SK RNA and what happens when it binds: how is that interaction triggering the capture of P-TEFb? We showed (Denise Martinez-Zapien’s PhD project, in collaboration with Isabelle Lebars) (Lebars et al., 2010) by NMR mapping combined with mutagenesis and EMSA experiments that the interaction with HEXIM relies upon a repeated GAUC motif, which forms a short double-stranded helix within HP1. Our data suggested that HEXIM binding triggers a destabilization of the GAUC motif and stabilization of a sequence just below the motif (Figure 5-1)
In the course of the present PhD project, we wished to see if LARP7 binding to 7SK RNA impacts HEXIM recognition, or if the two molecules bind 7SK RNA in an independent fashion. We thus continued to investigate the possibility of interactions of other domains of 7SK RNA with LARP7 in parallel with further investigation of HEXIM interaction. For instance, the proven interaction of HP1 with HEXIM (Lebars et al., 2010) could be only a partial view of HEXIM interaction with 7SK RNA, and another region of 7SK RNA could be bound by HEXIM. This could be modulated by the presence of LARP7. The present chapter describes the results obtained with the two hairpin domains, HP3 and HP1 which besides HP4, constitute independent domains of 7SK RNA. This was shown by Denise Martinez-Zapien, in the course of the study of 7SK RNA 2D-structure by the SHAPE method, by comparing the modification profiles of the full length RNA and the hairpins. Hairpins HP1 (24-87) and HP3 (201-273) were thus produced and investigated with EMSA for binding of LARP7 and HEXIM.
2. Interactions with HP3

**HEXIM and LARP7 show a tendency to bind HP3**

An EMSA experiment with HP3 is represented in Figure 5-2. It shows that HEXIM and LARP7 both bind HP3. The binding of HEXIM was only observed at high protein concentrations, larger than the concentration of the tRNA included in the assay to ensure specificity (5 µM). Thus, the binding is weak and may be not specific. However, preliminary experiments with SHAPE done before my arrival in the team did indicate some protections at the HP3 hairpin by HEXIM. This should be further investigated, to see whether this weak binding of the isolated hairpin might reflect something about the situation in the full length 7SK RNA. There could be a contact due to proximity, since HEXIM is a large protein, which once bound to HP1 may be close to HP3. The present experiment may also reveal a specific contact of the second monomer of the HEXIM dimer binding only after one monomer has bound to HP1. This type of binding was nick-named fly-casting (Mackereth and Sattler, 2012). It may require the full length 7SK RNA to find properly the second site. This hypothesis will be developed in the next chapter, discussing the dimerization property of HEXIM.

Experiment of Figure 5-2 shows that binding with LARP7 happens in a lower concentration range, as compared with HEXIM, since at 1.8 µM there is almost no free RNA left. We then further investigated which part of LARP7 may be involved. The experiment with both HEXIM and LARP7 will be discussed in the next chapter.
**Figure 5-2.** LARP7 and HEXIM bind the hairpin HP3. EMSA experiment with LARP7 (left panel), HEXIM (middle panel) and both LARP7 and HEXIM (right panel). Increasing concentrations (0.6/1.2/1.8/2.4/3/4/6 µM) of proteins were incubated with radioactively labeled HP3 (50 nM) in the presence of 5 µM tRNA.

The middle-region of LARP7 is involved in binding HP3.

No direct binding with HP3 could be evidenced for the La module (Figure 5-3) or the C-terminal region (shown in figure 5-9). As explained previously, different mutations of

**Figure 5-3.** K2 (266-289 deleted) deletion affects binding with HP3. The figure on the left shows the EMSA experiment with HP3 and wild type LARP7 (left panel), K2 deletion (middle panel) and N-domain (right panel). Increasing concentrations of proteins were incubated with radioactively labeled HP3 (50 nM) in the presence of 5 µM tRNA. Figure on the right quantifies the complex formation with LARP7 and K2 deletion reported as a function of protein concentration in µM.
LARP7 were done to investigate the role of the linker. The linker comprises several sequences encompassing stretches of basic residues, as indicated in the Introduction (see Figure1-9 where they are represented by blue bars). Three of these regions were deleted, including the N-terminal extension. The protein mutants KN (28-582), K1 (291-231 deleted) and K2 (266-289 deleted) were produced. The binding capacity with HP3 of these mutants was investigated by EMSA (Figure 5-3). We observed that the deletion mutant K2 (266-289 deleted) impacted slightly the binding to HP3 (the shifted band appears at a higher concentration of protein than with the wild-type LARP7). The control with La module in a similar range of concentration shows no shifted band.

The possible binding of LARP7 to HP3 in addition to HP4 should be further investigated, as it may help delineate the full grip of LARP7 on the 7SK RNA.

3. **LARP7 interaction with HP1**

Binding of full length LARP7 and full length HEXIM to HP1 was compared in a single experiment, shown in Figure 5-4. HP1 is binding with HEXIM as described (Lebars et al., 2010), however it also binds LARP7. With HP1, a stronger binding affinity is observed with HEXIM, which displaces the RNA at 0.6 µM compared to LARP7, which requires 1.5 µM. No supershift is observed when HP1 is incubated with both proteins suggesting that they do not form HEXIM1/LARP7/7SK RNA complex.

We then further investigated which part of LARP7 may be involved in HP1 binding.
Figure 5-4. LARP7 and HEXIM bind the 5' hairpin HP1 (24-87). EMSA experiment with full length LARP7 (left panel), HEXIM (middle panel) and both LARP7 and HEXIM (right panel). Increasing concentration (0.3/0.6/0.9/1.2/1.5/2/3 µM) of proteins were incubated with radioactively labeled HP1 (50 nM) in the presence of 5 µM tRNA.

HP1 binds LARP7 C-terminal domain.

Hairpin HP1 binding with LARP7-full length, C-domain and N-domain were investigated by EMSA experiments. The experiment shown in Figure 5-5, indicates that LARP7 C-domain, binds HP1 at a concentration of 1.25µM. However, N-domain doesn’t bind the HP1 hairpin at 4µM.

Figure 5-5. LARP7 C-terminal domain binds HP1. EMSA experiment: binding of LARP7 (left panel), LARP7-Cdom (middle panel) and LARP7-Ndom (right panel) with HP1 (24-87). Increasing concentration (LARP7: 0/0.16/0.31/0.63/1.25/2.5/5µM C-terminal domain: 0/0.13/0.25/0.5/1/2/4µM N-dom: 4uM) of proteins were incubated with radioactively labeled HP1 (50 nM) in the presence of 5 µM tRNA.
LARP7 binds the apical loop of HP1

In order to further investigate which part of HP1 is involved in binding with LARP7 C-terminal domain, direct footprinting was used with the isolated hairpin HP1 (24-87) and the full length 7SK RNA labeled at the 5’end, since the gel electrophoresis allows to analyze fragments of about a hundred nucleotides. This encompasses the full HP1. The probes RNase ONE and V1 were chosen because they are complementary, as they cleave at single stranded (ONE) or structured regions (V1).

The footprinting experiment with the isolated hairpin is shown in Figure 5-6. In free RNA (lanes R) the apical loop of HP1 is cleaved by RNase ONE at positions 49-55. Nucleotides 50-52 in the apical loop are protected by LARP7 C-terminal domain, as indicated by green arrows. Accordingly, the protection is also observed with the mixture of C- and N-terminal domain (tried in case of synergy between the two LARP7 domains) and with LARP7 full length. Interestingly, positions corresponding to two bulges, C71-U72 and C75-U76 showed increased cleavages by RNase ONE when the RNA was incubated with the C-terminal domain (indicated by blue arrows). This tells that these residues are accessible to the solvent in the presence of the protein, and may reflect a change of conformation of the hairpin. In the recent structure of HP1 (Martinez-Zapien et al., in preparation) the uridines U72 and U76 are indeed protruding out of the helical stack. It is possible that the apical loop binding to the LARP7 C-terminal domain stabilizes that conformation. RNase V1 cleaves free RNA at nucleotide 44, in the GAUC HEXIM1 binding motif. The V1 cleavage is protected by the C-domain.
Figure 5-6. (a) Direct footprinting experiment with 5’-labeled HP1 and probes RNase ONE and V1. Denaturing gel showing the cleavage products of the free (R) and complexed RNA with domains N-terminal (N), C-terminal (C), both (N+C) and full length LARP7 (F). Sequence was indexed with T1 in denaturing conditions (G) and a ladder (HA). In the left part, are the controls without RNase. (b) 2D structure of HP1. The cleaved regions are indicated.

A similar footprinting was obtained with the full length 7SK RNA (Figure 5-7). The apical loop of HP1 (residues 50-52) is again protected from RNase ONE cleavage in the presence of the C-terminal domain (indicated by green arrows). However, this protection is no longer observed with the full length LARP7. This suggests that, when the 3’-end of
7SK RNA is present, LARP7 uses its C-terminal domain to bind preferentially the HP4 domain of 7SK RNA. There might be a preferential recruitment of C-terminal domain to HP4 as a consequence of the La module binding to the 3’-end. The apical loop of HP1 is still protected by C-terminal domain (C-lane), probably because the concentration of the isolated C-terminal domain is large enough to saturate the site on HP4 and bind at HP1. The concentration of full length protein in the F-lane should also be large enough. Thus, the absence of footprint at HP1 suggests that the C-terminal domain is not free, and may be involved in an interaction with another part of the LARP7. Interestingly, an increase of accessibility to RNase ONE was observed at the bulge C71-U72 (blue arrows in figure 5-7), similarly as in the isolated HP1. Here, this is protected in the presence of full length protein.

In the same experiment with full length 7SK RNA, HEXIM binding was also investigated. HEXIM induced no visible protections at the apical loop. The region where HEXIM was expected to bind, the UU(GAUC)_{2}U sequence, was not cleaved by RNase ONE. However, a weak protection from RNase V1 cleavage is observed at nucleotide 44, which is included in HEXIM1 binding motif (GAU_{44}C). The cleavage at nucleotide 49 is enhanced. These protections are similar to those observed in HP1 with the NMR analysis (Lebars et al., 2010). However full confidence in the result demands to reproduce the footprinting experiment again (the footprinting/enhacement was observed only twice out of three experiments). The residues of nucleotides 32-39 in figure 5-7 predicted to form a central stem in HP1 stem. It was thus quite surprising to find them cleaved by RNase ONE, which recognizes single-stranded residues. The recent crystal structure of HP1 (Martinez-Zapien et al., manuscript in preparation) shows indeed a central stem encompassing
Figure 5-7. Direct footprinting experiment with 5’-labeled 7SK RNA and probes RNase ONE and V1. Denaturing gel showing the cleavage products of the free (R) and complexed RNA with domains N, C, full length LARP7 (F), full length HEXIM (H) and both (FH). Sequence was indexed with T1 in denaturing conditions (G) and a ladder (HA). In the left part, are the controls without RNase.

residues 35 to 39. It is interesting, however, to note that in a solution structure based on NMR data (Isabelle Lebars, personal communication), this region seems to be more open than in the crystal structure (the helix is unwound and base pairs show less stacking surface). The footprinting data showing cleavage of the middle region of HP1 both RNases (ONE, specific for single-stranded RNA and V1, specific for structured region)
would thus give support to a conformational flexibility of the central region of HP1. HEXIM protects these cleavages against RNase V1 and ONE (figure 5-7 compare lane H, or FH with the others). A recent mutational analysis of the HP1 domain (Martinez-Zapien et al., manuscript in preparation) proved the importance of a Watson-Crick base pair at 39-68, thus suggesting HEXIM binding site to extent to the stem below the recognition sequence UU(GAUC)\textsubscript{2}U. The binding surface revealed here by footprinting is thus in accordance with the extended binding site for HEXIM revealed by the structural analysis.

The footprinting experiments suggested that the apical loop of the HP1 5’-hairpin of 7SK RNA is bound by the C-terminal domain of LARP7. This was an unexpected result, and to further prove that interaction, we investigated the effect of a mutation of the apical loop. Residues 50-58 of the apical loop of 11 nucleotides where replaced by a short tetra loop of sequence UUCG or GAAA. We compared the binding of those mutants with that of HP1 wild type. The binding affinity of the mutants for LARP7 full length and LARP7 C-terminal domain decreased strongly (Figures 5-8).

Both LARP7 and HEXIM bind HP1 on close but different residues. It is LARP7 C-terminal domain footprints the residues in the apical loop of HP1.
Figure 5-8. Effect of mutations of the HP1 apical loop on the binding of the C-terminal domain of LARP7 (C-domain) and LARP7 full length. EMSA experiments: HP1-wild type (left panel), HP1-UUCG (middle panel) and HP1-GAAA (right panel) with increasing concentrations (0/0.08/0.16/0.31/0.63/1.25/2.5/5 µM) of protein. were incubated with radioactively labeled RNA (50 nM) in the presence of 5 µM tRNA. LARP7 C-domain

4. **Specificity of the C-terminal domain binding**

The observation that the C-terminal domain of LARP7 binds HP1 is quite puzzling with regard to the fact that this domain was shown to bind to HP4, the 3’-domain of the RNA. This raises the issue of specificity. Indeed, RNA-binding proteins may possess characteristics, like positively charged surfaces, that are adapted to bind any positively charged RNA with a certain affinity in an un-specific way. In our case, does the C-terminal domain recognize any apical loop of RNA or it is specific for both HP1 and HP4? (In that case, which are the structural determinants in common to these loops?) To answer that question, HP3 was used as a control, as it also contains an apical loop. LARP7 C-terminal domain doesn’t bind HP3 (Figure 5-9). This experiment reinforces our confidence in our observation: the C-terminal domain of LARP7 binds HP1 with specificity.
**Discussion**

This series of experiments lead to the conclusion that the C-terminal domain of LARP7 can bind the apical loops of both the 3’ (HP4) and the 5’- (HP1) hairpins of 7SK RNA. This raises several issues, as binding to HP1 seems to happen only when the HP4 domain is not present. This should be investigated further, however. This discussion provides a list of queries that should be investigated.

The difference of binding affinities (C-terminal domain binds strongly HP4 and more weakly HP1) should be quantified. A method of choice is ITC. This would allow to estimate the competition between HP1 and HP4 for C-terminal binding.
It is essential to ascertain whether the C-terminal domain may bind the two hairpins at the same time or not. The footprinting experiment does not totally exclude that the weak binding site is not functional. EMSAs with both hairpins should be tried.

Simultaneous binding, if it exists, means that different surfaces of the protein are involved. Mutations of the C-terminal domain should be performed and binding analyzed for both hairpins, choosing residues located on opposite faces (a model structure may be done from the homology with P65).

If it is proven that the C-terminal domain cannot bind HP1 and HP4 at the same time, this would suggest two alternative modes of binding of LARP7 to 7SK RNA. This is interesting. In view of the recent finding that the 7SK RNA snRNP may exist in the absence of HEXIM and P-TEFb. It would be involved as a transcription factor for the transcription (by Pol II) of the snRNA genes, the small nuclear RNAs involved in the splicing machinery (Sylvain Egloff’s presentation at SifrARN, Toulouse 2016). The situation with the C-terminal domain could reflect alternative conformations of the 7SK snRNP corresponding to different functions.

On HP1, the HEXIM binding site and the LARP7 C-terminal domain binding site do not seem to overlap. Mutation of the apical loop into the tetra loop affects HEXIM binding only slightly (Martinez-Zapien, manuscript in preparation). However, the sites are close, perhaps contiguous. It is thus possible that the C-terminal domain of LARP7 “senses” the presence or absence of HEXIM. The weak binding site revealed here would thus have a functional role.
Chapter VI: LARP7 impact on HEXIM binding

1. HEXIM binds 7SK RNA using ARM sequences in both monomers.

One major axis of investigation of the team was to understand how HEXIM recognizes 7SK RNA, and what results from this binding (PhD Denise Martinez-Zapien). The RNA-binding sequence of HEXIM is an Arginine Rich Motifs (ARM) located in the middle of the molecule (aa. 149-179, see Figure I-7). It was shown by NMR chemical shift mapping that one ARM peptide binds to (GAUC)_2 motif in HP1 (Lebars et al., 2010). Knowing that HEXIM is a dimer, the question of how and where the second ARM bind 7SK RNA was raised. Monomeric forms of HEXIM were produced by truncating the C-terminal domain responsible for dimer stability. Binding experiments (EMSA and Native Mass Spectrometry) demonstrated that two monomers can bind one HP1. Their binding is sequential, with a second binding event showing less affinity. This observation raised several hypotheses, described in a manuscript published in Biochimie and summarized here.

Does the second binding event depend on the first? The first binding event was observed in the NMR experiment and shows one ARM peptide binding to the (GAUC)_2 motif in HP1. Perturbations of the RNA structure were observed, such as loosening of the central base-pairs of the motif and stabilization of the A39-U68 pair in the lower stem of HP1. Binding of one ARM could induce the presentation of a second binding site. However, although several nucleotides were mutated in the course of the study, such secondary binding site was not evidenced. Our mutational analysis could never pinpoint any nucleotide which could be ascribed to a second binding site.
The second binding site seems thus to be rather diffuse and might be opportunistic. The ARM sequence is highly positively charged, and can form electrostatic interaction with any RNA. Once the first ARM is bound, the second could bind the ribose-phosphate chain in the vicinity without being much affected by the sequence at the base level. In this case, the second binding site may be considered as non-specific, or of low specificity. However, it is not an « opportunistic » binding of a second monomer recruited in the vicinity of the 7SK RNA motif. That was ruled out by the experiments described in the Biochimie manuscript. Basically, the two monomers were separated by deletion of the dimerization domain of HEXIM. After checking that the construct (HEXIM 136-273) was indeed monomeric. Binding was assessed by Native Mass Spectrometry (experiment described in Figure 2 of the following manuscript, and confirmed by EMSA (Figure 4A). Native mass spectrometry shows the formation of a 1:1 complex when equivalent amounts of RNA and protein are mixed. At higher ratio of protein/RNA, there is formation of a complex corresponding to 1 RNA and 2 proteins. The EMSA experiment shows two complexes, the second appearing as the protein concentration increases. The observation of two binding events means that the two protein fragments find their way to the RNA. Thus, it seems that there are two binding sites on the HP1 domain of the 7SK RNA. Interestingly, contrasted results were obtained when considering the monomeric HEXIM (136-273) or only the ARM (149-179 linked to GST), as shown in Figure 4A of the manuscript (compare the central and right panels of the gel). The second binding event occurred only with the 136-273 construct, and thus seems to be involving sequences outside of ARM. Interestingly, it has been shown that U30 crosslinks HEXIM1 at residues 210 and 220 (Belanger et al., 2009).
In the following manuscript, the experimental investigation was limited to the domain HP1 because, unfortunately, the native mass spectrometry approach could not be used with the full length 7SK RNA (110 kDa). Thus, clearly, the picture of HEXIM binding to the full length RNA remains quite blurred. It is still possible that another part of the RNA, outside HP1 may be bound as the second site (HP3, for example). Preliminary footprinting experiments were performed (Nicolas Djordjevic, Master1 project) showing indeed some protections on the 5’-side of hairpin HP3. These merit further investigations.

However, it is also important to consider LARP7 as a participant to HEXIM recognition. Is LARP7 binding to full length 7SK RNA impacting HEXIM recognition? My interest in LARP7 prompted to question whether this protein could play a role in the HEXIM recognition process. I realized HEXIM binding assays in the presence of LARP7. LARP7 or HEXIM alone shifts full length 7SK RNA (figure 6-1). A supershift is observed when added together. In contrast, when using 7SK RNA HP1, no supershift observed (figure 6-2).

![Figure 6-1](image)

**Figure 6-1** LARP7 and HEXIM bind 7SK RNA together. EMSA experiment showing the binding of LARP7 full length (left panel), HEXIM (middle panel) and both (right panel, where 7SK RNA full length was pre-incubated with 0.25 µM LARP7. Increasing concentrations (0/ 0.008/ 0.016/ 0.03/ 0.06/ 0.12/ 0.25/ 0.5 µM for LARP7, 0/ 0.016/ 0.03/ 0.06/ 0.12/ 0.25/ 0.5/ 1 µM for HEXIM) of proteins were incubated with radioactively lab RNA (50 nM) in the presence of 5 µM tRNA.
Figure 6-2. LARP7 full length and HEXIM bind the 5’ hairpin HP1. EMSA experiment with LARP7 (left panel), HEXIM (middle panel) and both LARP7 and HEXIM (right panel). Increasing concentration (0.3/0.6/0.9/1.2/1.5/2/3 µM) of proteins were incubated with radioactively labeled HP1 (50 nM) in the presence of 5 µM tRNA.

These experiments showed that both HEXIM and LARP7 bind together the full length 7SK (a supershifted band is observed in Figure 6-1). However, with HP1, there is no concomitant binding of LARP7 and HEXIM, as shown in Figure 6-2 (same experiment as in Figure 5-4, shown here again for convenient reading). Another experiment, with the monomeric HEXIM (136-273), shown in Figure 4-B of the Biochimie publication (Martinez-Zapien et al., 2015) suggests that LARP7 presence changes the nature of the binding to HP1. Now, one unique binding event is observed (one band on the EMSA gel). This band seems to correspond to a complex with the two monomers (as it has the same migration). Interestingly, this complex migrates faster than the complex formed by LARP7 and HP1 only.
Research paper

Intermolecular recognition of the non-coding RNA 7SK and HEXIM protein in perspective

Denise Martinez-Zapien a, b, c, d, e, Jean-Michel Saliou f, g, h, Xiao Han i, j, k, l, Cedric Atmanene f, g, Florence Proux i, j, k, Sarah Cianférani f, g, Anne-Catherine Dock-Bregeon i, j, k, *

a Department of Integrated Structural Biology, Centre for Integrative Biology (CBI), Institute of Genetics and of Molecular and Cellular Biology (IGBMC) 67404 Illkirch, France
b Centre National de la Recherche Scientifique (CNRS) UMR 7104, Illkirch, France
c Institut National de la Santé et de la Recherche Médicale (INSERM) U964, Illkirch, France
d Université de Strasbourg, Strasbourg, France
e Biotechnologie et Signalisation Cellulaire UMR 7242, Ecole Supérieure de Biotechnologie de Strasbourg, Boulevard Sébastien Brant, BP 10413, F-67412 Illkirch, France
f BioOrganic Mass Spectrometry Laboratory (LSMBO), Université de Strasbourg, IPHC, 25 rue Becquerel, 67087 Strasbourg, France
g CNRS, UMR7178, 67087 Strasbourg, France
h Center for Infection and Immunity of Lille (CIIL), Institut Pasteur de Lille, CNRS UMR 8204, INSERM U1019, Université Lille Nord de France, 1 rue du Pr. Calmette, 59000 Lille, France
i Department of Functional Genomics, Institut de Biologie de l’Ecole Normale Supérieure (IBENS), 75005 Paris, France
j INSERM U1024, 75005 Paris, France
k Key Laboratory of Brain Functional Genomics, East China Normal University (ECNU), Shanghai, PR China

Article info

Article history:
Received 23 January 2015
Accepted 25 March 2015
Available online 8 April 2015

Keywords:
non-coding RNA 7SK
HEXIM
Molecular interaction
Native mass spectrometry
Transcription regulation

Abstract

A 7SKsnRNP complex, comprising the non-coding RNA 7SK and proteins MePCE and LARP7, participates in the regulation of the transcription elongation by RNA-polymerase II in higher eukaryotes. Binding of a HEXIM protein triggers the inhibition of the kinase complex P-TEFb, a key actor of the switch from paused transcription to elongation. The present paper reviews what is known about the specific recognition of the 7SK RNA by the HEXIM protein. HEXIM uses an arginine-rich motif (ARM) peptide to bind one specific site in the 5’-hairpin of the 7SK RNA. Since HEXIM forms a dimer, what happens with the second ARM impacts the assembly symmetry. In order to help sort through possible models, a combination of native mass spectrometry and electrophoretic mobility shift assays was used. It provides evidence that only one ARM of the HEXIM dimer is directly binding to the RNA hairpin and that another sequence downstream of the ARM participates in a second binding event allowing the other monomer of HEXIM to bind the RNA.

© 2015 Elsevier B.V. and Société Française de Biochimie et Biologie Moléculaire (SFBBM). All rights reserved.

1. Introduction

1.1. The 7SKsnRNP function in transcription regulation

The non-coding RNA 7SK constitutes the scaffold of the 7SKsnRNP complex regulating the Positive Transcription Elongation Factor P-TEFb, thus impacting transcription by RNA polymerase II in higher eukaryotes [1]. P-TEFb is a kinase complex required for the transition of promoter proximal paused polymerases into productive elongation by phosphorylation of the DSIF and NELF
pausing factors [2–4]. Pauses of polymerase II constitute a transcription regulation step, which affects the temporal and spatial coordination of gene expression [5]. P-TEFB is also a key actor in the viral infection by HIV, as it is required for efficient synthesis of the viral RNA [6]. Interaction of P-TEFB with the viral TAR-Tat complex formed at the early steps of the HIV infection cycle leads to kinase activation [7,8]. In the eukaryotic cell, P-TEFB is sequestered and inactivated in a large complex containing 7SK. The 7SK RNA is embedded in a 7SKsnRNP core, where it is protected from degradation by the proteins MePCE and LARP7 [9–12]. The methylase MePCE is involved in 7SK capping [9] and binds the 5′-end of 7SK [13], while LARP7 binds to the 3′-terminal region of the RNA [14,15]. The recruitment of a HEXIM protein by the core 7SKsnRNP triggers the sequestration and subsequent inactivation of P-TEFB [16–20]. Following binding to the 7SK RNA, HEXIM binds to the Cyclin T subunit of the P-TEFB. This leads, by a mechanism still poorly understood, to inhibition of the catalytic activity supported by the CDK9 subunit [21–23]. Interestingly, the viral protein Tat has been shown to also bind 7SK [24]. Even if this suggests a viral molecular mimicry of TAR and Tat to 7SK and HEXIM, respectively, how such interactions with P-TEFB induce opposite outcomes is still an open question.

Transcription regulation results from the complex assembly of numerous factors [5,25]. Besides the dynamics impacting building and release of the interactions in equilibrium, such mechanisms are essentially dependent on the accuracy with which molecular recognition is established between partners. Specific recognition stands most often on an intricate set of interactions, such as H-bonds or hydrophobic interactions, all occurring in a three-dimensional context specifically adapted to these interactions. Structural analysis of many complexes of proteins and RNAs showed a diversity of means allowing specific interactions covering a wide span from direct “lock and key” mechanism to adaptive recognition [26]. With respect to specificity, the 7SK and HEXIM molecules are quite intriguing. They both exhibit sequence features shown to be required for their specific interaction, but there are still open issues regarding how specific binding may occur as the protein contributes to binding with unstructured peptides. The following text reviews, from a structural perspective, what is known about the recognition of 7SK by HEXIM.

1.2. The 7SK RNA

The human 7SK RNA (331 nucleotides) has a modular organization of four domains indicated in the 2D model of Fig. 1A, as was shown by Wassarman and Steitz [27]. This non-coding RNA has been identified in higher eukaryotes only, including insects such as Drosophila [28–30]. The hairpins HP1 (nucleotides 24–87) and HP4 (300–331) are the most conserved domains [29,30]. HP1 contains a HEXIM-binding site [24,31,32], which comprises a short palindromic GAUC sequence at the apical stem of the hairpin, bordered by two bulges composed of uridines [33]. HP1 is also bound by the Tat protein [24]. HP4 is involved in LARP7-binding [14,15,34]. Interestingly, HP4 seems to be directly involved in P-TEFB regulation, possibly by binding to P-TEFB [14,31]. Hairpin HP3 (200–273) is also conserved, but in vertebrates only, and has been implicated in the turnover of the system [29,35,36]. Several 2D models with equal stability can be drawn for 7SK. They differ mostly at the interconnections between the three conserved hairpins [27,37]. The model shown in Fig. 1B shows a closed conformation supported by an analysis of 7SK sequence conservations [37]. It suggests a more compact structure than the original model from Wassarman and Steitz [27], shown in Fig. 1A. This model proposes alternative connections. The existence of several models of similar stability may have functional relevance to the mechanism by which the HEXIM-binding site recognition triggers P-TEFB binding. A working hypothesis is that HEXIM binding induces conformational changes of 7SK [38]. This would explain why neither HEXIM nor 7SK inhibits P-TEFB on its own. On the whole, 7SK appears as a modular, semi-flexible molecule, with folded domains connected by flexible links. The role of the core proteins of the 7SKsnRNP should also be taken into account. Particularly, LARP7 could help restrict the range of 7SK conformations.

1.3. The HEXIM proteins

So far, two HEXIM proteins have been identified, HEXIM1 and HEXIM2 [39]. Since they appear to carry out the same function and show the same molecular organization, they will be referred with the generic name HEXIM in the following text. The HEXIM molecule, 359 residues-long in the human HEXIM1, comprises a large unstructured region of about two third of its sequence (1–273) [21]. The region identified to bind the 7SK consists of a conserved Arginine Rich Motif (ARM) [40] located in the middle of the molecule (residues 149–179), as shown in Fig. 1C. It is very similar to the RNA-binding motif in Tat [41], but for the HEXIM-ARM being twice the length of the Tat-ARM. ARMs are found mostly in viral proteins [40]. The RNA recognition mode of ARMs is quite peculiar since they are essentially unstructured in solution, but fold on the RNA [42]. Interestingly, the same peptide can adopt different folds when adapting to different RNA structures [43]. Moreover, the RNA structure is also changing upon ARM binding [44]. ARMs bind most often into the narrow major groove of the RNA type A-helix. This requires enlarging the groove to facilitate the peptide positioning deep into the groove [45]. The resulting interactions combine salt bridges between positive charges of the amino acids and the phosphates, and more specific recognition modes, such as H-bond or stacking with specifically positioned residues. On the whole, ARM-RNA recognition illustrates the induced fit concept. Further downstream, HEXIM proteins feature stretches of acidic residues (211–217 and 234–249 in human HEXIM1, in red in Fig. 1C). These were hypothesized to bind to the basic residues of the ARM, thus generating a self-inhibitory conformation of HEXIM [20]. RNA binding is supposed to liberate the acidic residues, which could then play a more direct role in cyclin binding or kinase inhibition [20]. Interestingly, part of the acidic region is included in a peptide comprising residues 196–220, which was shown by UV-crosslinking experiments to be involved in 7SK-binding [32]. A prominent property of HEXIM proteins is that they form dimers in the cells [46]. The dimerization interfaces comprise a long bipartite coiled-coil at the C-terminus (residues 284–348, Fig. 1C), where long helices turn around their symmetrical correspondents in the dimer. The coiled coil is interrupted by a stretch of small residues, which confers flexibility to this domain [47]. The C-terminal long helix (z3, residues 319–348) forms a stable dimerization interface, showing strong interconnections between residues from the two monomers [48]. The more upstream helix z2 (residues 284–313) while still participating to the dimer interface, as it coils around the other monomer, shows weaker intermolecular interconnections. It has been shown to bind the cyclin T1, together with residues from a very short upstream helix z1 (residues 276–281, Fig. 1C). The dimeric character of HEXIM raises an issue regarding the composition of the inhibitory complex, which has functional impact [46,47,49]. Considering that cyclin T binds to a surface shared by two HEXIM monomers, it is still not clearly established how many cyclin molecules bind to HEXIM.

Several teams suggested, by complex experiments in vivo, that the HEXIM dimer binds one single 7SK RNA [46,49,50]. This does not, however, mean that the complex thus formed is symmetric, as
the in vivo experiments did not tell if both RNA-binding peptides, one from each monomer, are involved directly, or if the second monomer is recruited to the RNA because it is linked by the dimerization interface. Moreover, it is intriguing how HEXIM, which exhibits an apparent low structural complexity with no structurally defined RNA recognition domains, is able to single out 7SK in the nucleus, where RNA is abundant, in the form of nascent transcripts and splicing factors. Indeed, HEXIM has been described as a promiscuous RNA-binding protein [51], suggesting that additional factors could be found contributing to the specificity of recognition. To clarify our point, the following text describes models that could stand for this peculiar complex. To further unravel how the dimeric state of HEXIM could impact the assembly of the complex with 7SK RNA, we present new experimental data obtained from a combination of native mass spectrometry and electrophoretic mobility shift assays. These were intended to clarify how many HEXIM monomers are directly binding one RNA molecule. They show that the ARM peptide is not the sole contributor to RNA-binding, and suggest asymmetry in the complex formation.

2. Material and methods

2.1. HEXIM expression and purification

The full-length human HEXIM1 was cloned in a pET28 vector containing a C-terminal His-tag, the HEXIM_136–273 in a plasmid of the pnEA serie [52], with a N-terminal His-tag, and the ARM peptide (149–179) in another plasmid of the same series, but with a N-terminal His-GST-tag. Overexpression in BL21 DE3 was induced overnight at 25 °C. The cells were lysed by sonication in 50 mM Tris pH 8, 5 mM MgCl₂, 500 mM NaCl and 1.4 mM beta-mercaptoethanol. After debris removal, the cellular extract was purified by nickel-affinity in the same buffer, followed by chromatography on HiLoad 16/60 Superdex S200, in a buffer containing 50 mM Tris pH 7.6, 5 mM MgCl₂, 500 mM NaCl and 2 mM DTT. The proteins were flash-frozen and stored at −80 °C after addition of 10% glycerol.

2.2. 7SK hairpins production and purification

7SK and hairpins were cloned in a pHDV vector encoding for a hammerhead ribozyme in 3′ of the target RNA. RNAs were produced by T7 in vitro transcription at 37 °C for 4 h, either from the linearized plasmid, or a PCR template obtained with adequate primers. When the ribozyme was co-transcribed, RNAs were cleaved from the 3′ ribozyme by incubating at 65 °C for 10 min followed by an incubation at 37 °C for 20 min, in the presence of 40 mM MgCl₂. All RNAs were gel-purified.

2.3. Electrophoretic mobility shift assays (EMSA)

The RNAs were labelled at the 5′-end with T4 polynucleotide kinase (Fermentas) and Y-32P-ATP, following dephosphorylation with antarctic phosphatase (Biolabs). The 5′-32P-RNA (50 nM in all assays) was incubated 20 min at 4 °C with increasing concentrations of proteins in the range of 0–0.8 μM for full-length and 0–2.5 μM for 136–273 or GST-ARM, in a buffer containing 25 mM Tris pH 7.5, 5 mM MgCl₂, 500 mM NaCl and 1 mM DTT. The proteins were flash-frozen and stored at −80 °C after addition of 10% glycerol.

Fig. 1. Schematic representation of 7SK and HEXIM molecules. (A) Secondary structure of 7SK showing the domain organization; the 108 boundary is indicated; the HEXIM-recognition sequence UUGAUC/UGAUC is highlighted in red. (B) Another representation of 7SK in a closed conformation. The yellow and pink ovals stand for the positions of the bound N-terminal and C-terminal domains of LARP7 (from Uchikawa et al., submitted). (C) Domain organization of the HEXIM1 protein showing the regions of known functional interest: N-terminal domain (grey), RNA-binding region (blue), acidic sequences (red), cyclin-binding domain (light green) and dimerization helix (dark green). The light blue panel indicates the region found by Belanger et al. [32] to crosslink with 7SK. The boundaries are indicated for the Homo sequence (numbers on top). The sequence of the ARM region is given below. The structure of the C-terminal region is from Dames et al. (PDB id 2GD7), with one monomer green and the other grey. Helices α1, α2 and α3 are indicated.
Na HEPES, pH 7.2, 5 mM MgCl₂, 250 mM NaCl, 2 mM DTT, 0.005% NP40, 10% glycerol, 0.05 mg/ml bovine serum albumin as well as 5 μM total tRNA to minimize non-specific binding. A small amount (1.5 μL) of dye mix containing 0.02% each of bromophenol blue and xylene cyanol in 60% glycerol was added to the 12 μL assay mix just before loading on a native 6% acrylamide gel (20 × 20 cm, 1.5 mm thickness) and run in TBE (0.5X) at 4 W for 75 min. Gels were revealed by phosphorimaging.

2.4. Native mass spectrometry

Mass spectrometry (MS) experiments were performed on an electrospray time-of-flight mass spectrometer (LCT, Waters, Manchester, UK) equipped with an automated chip-based nano-electrospray source (Triversa Nanomate, Advion Biosciences, Ithaca, NY, USA) operating in the positive ion mode. External calibration was performed with the multiply charged ions produced by 2 μm horse heart myoglobin diluted in 1:1 (v/v) water/acetoni tromite acidified with 1% (v/v) formic acid. Prior to native MS analysis, protein buffer was exchanged against 250 mM ammonium acetate buffer, pH 7.5, using microcentrifuge gel-filtration columns (Zeba 0.5 ml, Thermo Scientific, Rockford, IL, USA). For native MS analysis, the Accelerating voltage (Vc) was set to 180–200 V, while pressure in the interface region (Pi) was 6–7 mbar in order not to disrupt weak noncovalent interaction. Mass measurements under denaturing conditions were carried out by diluting samples to 2 μM in water/acetoni tromite/formic acid (50:50:1). The Accelerating voltage (Vc) was set to 40 V, while pressure in the interface region (Pi) was 1.1 mbar. Data analysis was performed with MassLynx 3.5 (Waters, Manchester, UK).

3. Hypotheses, results and discussion

3.1. One HEXIM dimer binds one HP1 hairpin of 7SK

Our previous contribution to understand HEXIM binding to the hairpin HP1 of 7SK included an NMR approach [33] which delineated the recognition site in the RNA to the palindromic sequence (UU)GAUC/(U)GAUC in the HP1 hairpin of 7SK. It showed that specific effects on the NMR spectra of the RNA were observed upon addition of peptides containing the ARM sequence up to a ratio of 1.3 peptide for one molecule of HP1. The same effects were observed for two peptides, 149–165 and 149–179. In each case, further addition of peptide lead to resonance broadening, suggesting that one ARM peptide of HEXIM was sufficient to recognize the specific (UU)GAUC/(U)GAUC sequence [33]. Electromobility shift assays (EMSA) experiments with full-length HEXIM showed that this RNA sequence was determinant for HEXIM binding, but did not indicate how many molecules of RNA were bound by HEXIM. Since one ARM was binding one HP1, one HEXIM dimer could bind two RNA molecules. We thus turned to native mass spectrometry (MS), which allows precise binding stoichiometry assessment from accurate mass measurements. Table S1 summarizes all the mass assessments of the present work. Fig. 2A–D shows the mass spectra obtained with HP1 and HEXIM either as standalone partners or in complex. When injected alone, HP1 shows the expected mass of 20 860 ± 1 Da (Fig. 2A). The HEXIM sample contains mainly a species of mass 83 389 ± 13 Da, corresponding to a dimeric HEXIM, and a small amount of monomers with a mass of 41 558 ± 3 Da (Fig. 2B). After incubation of both partners, native MS reveals the formation of a 104.5 kDa complex (104 267 ± 13 Da), corresponding to one HP1 bound to one HEXIM dimer (Fig. 2C). No higher order stoichiometry species were detected after increase of the HEXIM to RNA (Fig. 2D). In particular, complexes with one HEXIM dimer and two HP1 were never observed.

The presence of HEXIM monomers offered an interesting opportunity to compare the affinities of the monomer and the dimer for HP1. A small population with a mass of 62 676 ± 8 Da, consistent with a complex of one HP1 for one monomer HEXIM is also observed on native mass spectra (Fig. 2C and D). The fact that HEXIM dimer is always observed saturated with HP1 while a monomeric HEXIM is still detected in the free state indicates that the RNA affinity is higher for dimeric HEXIM compared to monomeric protein.

Formation of a complex containing one HEXIM dimer and one RNA may be exploited in several ways (Fig. 3). Fig. 3A–C shows various situations where both ARMs are involved. The first, in Fig. 3A shows a symmetrical solution, which is the simplest way to bind one RNA with a homodimer. Fig. 3B shows an asymmetric solution. In that pattern, HEXIM monomers bind different parts of the HP1 hairpin. The second binding site on the RNA (labelled with a star in the Figure) could be revealed consequently to the first binding event, or could require the local increase of concentration of monomer, due to the first binding, to be observable. In the same line of idea, a second binding site could be revealed in another part of 7SK, as pictured by Fig. 3C. More sophisticated models (Fig. 3D–F) picture cases where the ARM of the second monomer is not directly involved in binding the RNA, but participates indirectly, by establishing interactions with a protein region. In Fig. 3D, the second ARM is represented binding to the acidic sequence, as in the auto-inhibitory model postulated by Barbovic et al. [20]. Fig. 3E conveys the idea that a relationship established between monomers could increase the stability of the assembly, or could form a higher structural organization better adapted to bind RNA. In Fig. 3F, such higher structural organization involves LARP7, bound to HP4 through its N- and C-terminal domains and bound to HP1 thanks to a lysine-rich sequence present in its linker region. In all these cases (Fig. 3D–F), we suggest that the interaction of the ARM peptide with another protein region (materialized by large dotted lines in the figure) may induce a three-dimensional organization of residues that would be conducive to specific recognition.

3.2. Two HEXIM monomers, but only one ARM, bind one HP1 hairpin of 7SK

To get insight into that, we designed a monomeric HEXIM by deletion of the dimerization interface. The construct 136–273, also deprived of the N-terminal region, which is not involved in RNA-binding [19], was analysed by native MS, and demonstrated to be monomeric, with a mass of 18 712 ± 1 Da (Fig. 2F). Titration experiments of HP1 with HEXIM 136-273 showed that one HP1 was bound by one monomer, forming a complex of 39 573 ± 7 Da, even when the ratio was 2 proteins/RNA in the sample (Fig. 2G). At higher protein concentrations (excess of 4 proteins/RNA), a second population corresponding to a mass of 58 325 ± 5 Da suggested the formation of complexes comprising two HEXIM 136–273 and one HP1 (Fig. 2H).

In order to rule out possible artefacts due to the experimental conditions of native MS experiments (incubations in ammonium acetate without additives), EMSAs experiments were performed in parallel. The buffer conditions of EMSA allow addition of magnesium and reducing agent, and most importantly, inclusion of tRNA as a decoy to capture nonspecific complexes that HEXIM could form with any RNA. Fig. 4A shows complex formation following incubation of HP1 with three constructs of HEXIM: the full-length dimer, the 136–273 monomer, or a construct named GST-ARM, where the ARM peptide (149–179) was grafted on a GST protein as a carrier. This carrier was required to visualize complex formation, since the ARM
peptide is very small. As expected, one band is observed for the complex of HP1 with the HEXIM dimer (noted H in Fig. 4A). Titration with HEXIM 136–273 shows two complexes. A first complex is formed at protein concentrations below 1 μM. At higher protein concentrations, the conversion of that complex into a larger one is observed as a slower band. This agrees well with our observation in native MS analysis. The presence of 5 μM tRNA in the incubation medium excludes the possibility that this second complex is nonspecific. Two independent experiments thus show that two HEXIM monomers bind the RNA, but not with the same affinity. This suggests that the complex is not symmetric. Interestingly, titration with GST-ARM showed only one complex. The same result was obtained with Maltose-binding protein as a carrier. The observation of only one complex with the ARM peptide is in accordance with our previous NMR footprinting data. This clearly excludes binding patterns with two ARMs together on HP1, as indicated in Supplementary Fig. S1. The second complex may reconcile these observations. Model 3D suggests that binding of ARM to HP1 reveals another binding site, in HP1, for binding the downstream sequence. Reciprocally, binding of ARM would be a prerequisite for the downstream peptide to bind RNA. Model 3E involves cross-interactions between HEXIM monomers. Interestingly, early studies indicated that 7SK binding was stabilizing the dimerization of HEXIM [46]. The cartoon in Fig. 3E pictures a possible role for the acidic region, which could be involved in recruiting the other monomer, thus participating in the global stabilization. The linkage essentially established by the coiled-coil interface may be completed by interactions involving the acidic regions (as in Fig. 3E). Linking the monomers could help increase the local concentration of the second monomer, thus facilitating the second binding to a weaker binding site.

Recent evidence for a second HEXIM-binding site comprised in the apical loop of HP1 [15] contrasts with the early experiments of Egloff et al. [31], and the fact that this loop is not conserved [29]. Another site was proposed at the basal stem of a 5′-hairpin comprising nucleotides 1–108 [24]. This hairpin (7SK 1–108) comprises HP1, but includes sequences for which the available models of 7SK secondary structure differ considerably, as indicated in Fig. 1A and B [27,37]. In our hands, the 5′-hairpin 1–108 version bound the HEXIM dimer, the monomer HEXIM 136–273 and GST-ARM by forming one, two or one complex, respectively, in the same way as HP1 (24–87). Native MS analysis with the 5′-hairpin 1–108 showed an RNA mass of 34 798 ± 1 Da (Supplementary Fig. S1, panel A). Titration with HEXIM 136–273 showed formation of two species of 53 557 ± 5 Da and 72 300 ± 5 Da, corresponding to complexes of one hairpin 1–108 with one or two HEXIM monomers, respectively (Supplementary Fig. S1, panels C and D). Moreover, when both HP1 and the 5′-hairpin 1–108 were mixed and incubated with HEXIM full-length, there was no preference for the larger RNA, as species with HP1 and with the 5′-hairpin 1–108 were

---

**Fig. 2.** Analyses with MS of HEXIM binding to the HP1 hairpin of 7SK. (Panels A–D: HEXIM full-length) Native MS of HP1 (A), HEXIM (B) and titration of HP1 (5 μM) with 10 μM HEXIM (C) or 20 μM (D). (Panels E–H: HEXIM 136–273) Titration of HP1 (E) with HEXIM 136–273 (F): 2.5 μM HP1 and 5 μM HEXIM 136–273 (G) or 10 μM HEXIM 136–273.
simultaneously detected (Supplementary Fig. S1, panel G). Taken together, these experiments indicate that the two binding sites are comprised in the short version of HP1 (24–87). This is compatible with the involvement of the apical loop, recently suggested [15]. The observed effect of mutations at the basal part of the long (1–108) 5′-hairpin may be reconsidered as affecting the conformation or the dynamics of 7SK.

### 3.3. A possible contact of HEXIM with the middle hairpin of 7SK

HEXIM has been described as a promiscuous RNA-binding protein [51]. The difficulty to clarify the RNA-binding sites of HEXIM is not surprising with respect to the low complexity nature of the ARM since this positively charged peptide is expected to bind readily any negatively charged molecule. In the native MS experiments, the molecules were incubated at high concentration of salt to ensure that specific binding was distinguished from nonspecific binding. EMSA were performed in the presence of tRNA as a decoy for minimizing nonspecific binding, but it was still difficult to evidence single nucleotide position for which HEXIM binding was clearly impaired, besides those of the (UU)GAUC/(U)GAUC sequence. For example, changing both internal loops downstream of the determinant sequence, i.e. 71–72 and 75–77, for a stem without bulges did impair HEXIM binding, but it had only a limited effect [33]. This precluded straightforward interpretation of this lower bulge as another binding sequence rather than a feature assisting HEXIM binding by favouring RNA conformational dynamics. On another hand, we did observe binding of HEXIM to another hairpin of 7SK, HP3 (Fig. 5). This was observed both with EMSA experiments (Fig. 5A) and native MS (Fig. 5B, detection of a 1:1 HP3:HEXIM complex at 107 123 ± 10 Da). However, when HP1 and HP3 were mixed, native MS showed that binding was clearly in
favour of HP1, with masses of 104 177 ± 8 Da for the complex with the HEXIM dimer and 62 502 ± 16 Da for that with the monomer (Fig. 5B, panel d). This experiment thus indicates that the HP1 binding site is favoured, but that HEXIM can make additional contacts with HP3. This is not surprising when taking into account the large size of the HEXIM protein.

3.4. Is LARP7 contributing to a higher specificity of HEXIM:7SK recognition?

7SK is not free in the nucleus, but packed into a 7SK snRNP core. We showed recently that LARP7 specifically recognize the 3’ hairpin of 7SK, using two folded domains located at its N- and C-terminal ends [34]. Surprisingly, EMSA experiments showed that LARP7 also binds HP1 (Fig. 4B). When HP1 was incubated with LARP7 prior to incubation with HEXIM, the band corresponding to the LARP7:HP1 complex (noted L in Fig. 4B) disappears during titration with HEXIM, concomitantly with the appearance of the HEXIM:HP1 complex (noted H in the figure). Surprisingly, the bands corresponding to the HP1:HEXIM complex migrate at the same position in the experiments made in the absence or in the presence of LARP7 (compare Fig. 4A and B). Experiments migrated on the same gel never showed a new (supershifted) band in addition to those corresponding to the complexes with HEXIM and LARP7, thus indicating that the two proteins do not bind the hairpin together. A possible interpretation is that HEXIM captures the RNA from the LARP7:RNA complex. The domains of LARP7 binding HP4 are linked by a long, mostly unfolded sequence (residues 210–450 in human LARP7), which contains several stretches of basic residues. It is conceivable that the basic sequences of the linker are tethered at several places of the 7SK, including HP1. Interestingly titration with monomeric HEXIM 136–273 shows only one complex when HP1 is pre-incubated with LARP7 (Fig. 4B). Comparison with Fig. 4A suggests it to be the complex with two HEXIM monomers (it migrates at a similar distance as the complex with GST-ARM). Not observing the complex with one monomer suggests that formation of the complex with two monomers occurs at lower protein concentrations when LARP7 is present, as indicated by the fast disappearance of free RNA (F) in Fig. 4B. This further suggests that LARP7 helps HEXIM to bind. The sketch F in Fig. 3 pictures an involvement of the linker of LARP7 (with basic sequences materialized in cyan), which could be involved in a local higher structure of the low-complexity regions to define a new RNA-binding domain. This potential intermolecular HEXIM–LARP7 interplay, possibly involving HEXIM acidic sequences, merits further investigations. On another hand, nonspecific, but direct binding of LARP7 to the RNA is not excluded. On the whole, the LARP7 linker may function as an insulator, by screening charges of the polyphosphate chain of the RNA or the acidic region of HEXIM.

4. Conclusive remarks

During our investigations of HEXIM recognition of its target 7SK and the mechanism leading to P-TEFb inhibition, we were puzzled
by the small size of the discriminant sequence of the RNA and the low complexity of the RNA-binding sequence of HEXIM, an ARM peptide. The palindromic sequence (UU)GAUC/(U)GAUC discriminating 7SK among RNAs is symmetric. When modelled in 3D, as may be done with MC-Fold [53], it represents only half a helical turn (Supplementary Fig. S2A). It is thus difficult to imagine how it could bind two ARMs from the HEXIM dimer. This difficulty can be grasped by looking for example at proteins of the bZIP family. Like HEXIM, they form dimers and bind with basic sequences to a nucleic acid, DNA in the case of bZIP. The DNA recognition site corresponds to two half-sites allowing the two monomers to bind simultaneously [54,55], as can be appreciated in Supplementary Fig. S2B. This mode of interaction seems not possible in the 7SK:HEXIM case, because the HEXIM-binding sequence is too small, and doesn’t offer the correct symmetry in 3D. ARM peptides have been shown to fold in various ways dictated by the RNA architecture. They may form helices or strands, or a β-hairpin in the case of Tat on TAR [43,44], as shown in Supplementary Fig. S2C and D. Only the latter case (Supplementary Fig. S2D) could sustain a binding mode with two peptides joining together in the same groove. However, this model is ruled out by our EMSA experiments, showing that HP1 binds only one ARM. Our results indicate that a second binding event follows the binding of a first HEXIM monomer in HP1, at higher protein concentration. This makes sense with a second HEXIM binding site of lower affinity or with successive binding events. Recent results from another group indicated that the apical loop of HP1 is involved in HEXIM binding [15]. The existence of a second site in the same hairpin agrees well with our native MS analysis and EMSA experiments.

A second monomer binding was observed only when the peptide was extended to downstream sequences. Thus, it appears that the downstream sequence favours assembly of the dimer on the RNA. This is in line with an early observation on this system, that dimerization of HEXIM could be favoured by the presence of 7SK [50]. The extension contains an acidic region, which is tempting to propose as assembling in the recruitment of the second HEXIM monomer. However, our observations indicate that an intramolecular interaction with LARP7 is also plausible. Interestingly, the downstream sequence 178–220, which does not bind RNA directly, was shown to participate in the RNA binding selectivity, as a HEXIM containing this sequence is more sensitive to mutations in the RNA as compared to the ARM [15]. This is in accordance with a model where this sequence could participate in building a higher order structure conferring better specificity to the low-complexity regions of the proteins involved (Fig. 3E).

Interestingly, HEXIM shows some similarity with Tat, another protein of low complexity, but Tat is a monomer, and turns P-TEFb into activity. Could the oligomerization state bear consequence on the effect on P-TEFb activity? A cyclin T binding site has been located at the C-terminus of HEXIM, close to the dimerization interface [21,47,48]. Two modes for cyclin interaction can thus be hypothesized. Either one cyclin straddles the dimer, or two cyclins and consequently two P-TEFbs bind the HEXIM dimer [49]. The latter hypothesis is interesting in view of the difference of impact on the activity between the dimeric HEXIM, which inactivates, and monomeric Tat, which activates P-TEFb. Inactivation might result from mutual inhibition of two P-TEFb molecules. In that line of hypothesis, the role of 7SK can be further questioned in view of the induced stabilization of the HEXIM dimer. A common hypothesis is that binding to 7SK uncovers the HEXIM binding site to cyclin T [20]. Thus, P-TEFb would depend on a conformational change of HEXIM. Another proposition is that 7SK binding impacts cross-interactions between the monomers and consequently stabilizes the assembly of an inactive complex comprising two P-TEFbs.

A reasonable model of the 7SKsnRNP regulation is that HEXIM binding to 7SK generates a mixed RNA-protein surface specifically adapted to P-TEFb inhibition. This could be positioned near the CDK9 moiety of P-TEFb, as an additional contribution to the direct HEXIM-CyclinT interface [15,23]. However, the model could be further complicated, to account for the observation that at least two mutations of HEXIM residues (Phe208 or Tyr271) impair P-TEFb inactivation, but not RNA binding [15,49]. Moreover, another important partner in this assembly is LARP7, the 7SK chaperone. We report here experimental data on a possible role of LARP7 in helping the system to gain efficiency. LARP7 could facilitate RNA recognition by stabilizing a functional 7SK conformation. LARP7 binding could also reduce the RNA surface accessible to HEXIM or other molecules. Since HEXIM is a promiscuous RNA-binding protein [51], this would increase the specificity of recognition by occluding spurious binding sites. On the whole, specificity in the 7KsnRNP system could result from a complex combination of this shielding effect with direct recognition of the small HEXIM-binding site, molecular adaptation to allow accommodation of the ARM peptide in the groove, and conformational changes of the protein with an interesting cross-talk between monomers. Taken together, these events contribute to build up a functional structural units from low complexity molecules.

Conflict of interest

There is no conflict of interest.

Acknowledgements

This work was supported by the Centre National de la Recherche Scientifique, the University of Strasbourg (UDS), the “Fondation pour la Recherche Médicale”, the Alsace Region, and the GIS IbisA. It was started with funding from French National Agency for Research (grant TrsCrREGnRNP ANR-06-BLAN-0072) and continued with Dynamic [ANR-12-BSV5-0018]. It has received support under the program « Investissements d’Avenir » launched by the French Government and implemented by the ANR [ANR-10-LABX-54 MEMO LIFE] and [ANR-11-IDEX-0001-02]. DMZ benefited from doctoral fellowships from CONACYT (Mexico) and Association pour la Recherche contre le Cancer. XVH was supported by PhD fellowship from Chinese Scholarship Council (CSC).

We thank Olivier Bensaude for constant support and comments on the manuscript and Isabelle Lebars for discussions and useful suggestions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biochi.2015.03.020.

References

## Supplementary Material

### Supplementary Table

<table>
<thead>
<tr>
<th>RNA name</th>
<th>Protein name</th>
<th>Stoechiometry</th>
<th>Sequence information</th>
<th>Theoretical masses (Da)</th>
<th>Measured masses in denaturing condition (Da)</th>
<th>Measured masses in native condition (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP1</td>
<td>Mono</td>
<td>24-87 5'triP - 3'Pcyc</td>
<td></td>
<td>20 861.2</td>
<td>20 860.2 ± 0.1</td>
<td>20 860 ± 1</td>
</tr>
<tr>
<td>7SK 1-108</td>
<td>Mono</td>
<td>1-108 5'triP - 3’OH</td>
<td></td>
<td>34 797.6</td>
<td>n.a.</td>
<td>34 798 ± 1</td>
</tr>
<tr>
<td>HP3</td>
<td>Mono</td>
<td>200-273 5’triP - 3’Pcyc</td>
<td></td>
<td>23 709.2</td>
<td>23 709.1 ± 0.2</td>
<td>23 709 ± 1</td>
</tr>
<tr>
<td>HEXIM</td>
<td>Mono</td>
<td>1-359-LEH₆</td>
<td></td>
<td>41 557.3</td>
<td>41 557.1 ± 0.8</td>
<td>41 558 ± 3</td>
</tr>
<tr>
<td>HEXIM</td>
<td>Dimer</td>
<td>(1-359-LEH₆)²</td>
<td></td>
<td>83 114.6</td>
<td>/</td>
<td>83389 ± 13Da</td>
</tr>
<tr>
<td>HEXIM</td>
<td>Mono</td>
<td>MGSSHHHHHHHSSTGSGSGSPGSHM 136-273-GS</td>
<td></td>
<td>18 711.3</td>
<td>18 712 ± 1</td>
<td>18 712 ± 1</td>
</tr>
<tr>
<td>HP1</td>
<td>HEXIM</td>
<td>1/1</td>
<td></td>
<td>62 418.5</td>
<td>/</td>
<td>62 516 ± 13</td>
</tr>
<tr>
<td>HP1</td>
<td>HEXIM</td>
<td>1/2</td>
<td></td>
<td>103 975.8</td>
<td>/</td>
<td>104 267 ± 13</td>
</tr>
<tr>
<td>HP1</td>
<td>HEXIM</td>
<td>1/1</td>
<td></td>
<td>39 572.5</td>
<td>/</td>
<td>39 573 ± 7</td>
</tr>
<tr>
<td>HP1</td>
<td>HEXIM</td>
<td>1/2</td>
<td></td>
<td>58 283.8</td>
<td>/</td>
<td>58 325 ± 5</td>
</tr>
<tr>
<td>7SK 1-108</td>
<td>HEXIM</td>
<td>1/1</td>
<td></td>
<td>53 508.9</td>
<td>/</td>
<td>53 557 ± 5</td>
</tr>
</tbody>
</table>
**Figure S1.** Analyses with MS of HEXIM binding to the 5’ hairpin 1-108 of 7SK (Panels A-D: HEXIM 136-273) Native MS of 7SK 1-108 (A), HEXIM 136-273 (B) and titration of 7SK 1-108 (2.5 µM) with 5 µM (C) or 10 µM HEXIM 136-273 (D). (Panels E-G: HEXIM full-length) Titration of 5 µM HP1 with 5 µM HEXIM (E) 5 µM 7SK 1-108 with 5 µM HEXIM (F) 5 µM HP1, 5 µM 7SK1-108 and 15 µM HEXIM (G).
Figure S2. Examples of ARM-peptide binding sites in nucleic acids.

(A) A model for the 7SK hairpin HP1 showing the deep and narrow major groove of the RNA helix. The HEXIM binding site (UU)GAUC/(U)GAUC is in cyan and the A39-U68 base pair which was shown to open upon ARM-peptide binding (1) is in green. (B) The complex of the heterodimer c-Fos: c-Jun binding to DNA; coordinates from PDB 1FOS (2). (C) The complex of an HIV-1 rev peptide with RNA; coordinates from PDB 1ETG (3). (D) The Tat:TAR complex from bovine immunodeficiency virus; coordinates from PDB 1MNB (4).
Supplementary References


Discussion and Conclusion

In the manuscript presented in this chapter were proposed several models to explain how two monomers of HEXIM could bind the HP1 domain of 7SK RNA in a sequential way. In one set of models, one ARM binds to the (GAUC)₂-motif and the second ARM binds elsewhere, at a site yet unknown. We considered the presence of LARP7, which in vivo is associated with the 7SK RNA. An EMSA experiment with HP1 and the monomeric HEXIM (136-273) realized in the presence of LARP7, indicates that unique binding event occurs (one band on the EMSA gel). Thus, the presence of LARP7 seems to diminish the amount of complexes formed with one HEXIM monomer to favor the binding of two monomers on the hairpin. However, it cannot be excluded that the band observed is not unique, but represents a mixture of two complexes, one of HP1 with two monomers (as observed in the absence of LARP7) and another one, comprised of LARP7, one monomer of HEXIM, and HP1. In both cases, the observation can be interpreted as a «chaperoning» role of LARP7. LARP7 might stabilize HP1 in a competent conformation for binding readily the two monomers of HEXIM. LARP7 could alternatively play a chaperoning role by protecting HP1 from diffuse binding to the backbone, thus orienting HEXIM binding towards its specific binding site.

To fully clarify the models, new experiments should be performed in the context of the full length 7SK RNA. Since a weak binding of HEXIM to HP3 was revealed by EMSA (Chapter V), we think now that footprinting experiments with the full length molecules should be realized. Preliminary footprinting experiments were performed, but no clear result can be presented for different reasons.
The prominent problem is the lack of clear footprinting data at the 7SK RNA-motif. Due to the stalling of the reverse transcriptase at sequence 95 (described in Chapter II: Material and Methods), we could not ascertain the binding of HEXIM to its principal site while exploring the rest of the molecule. Moreover, due to priming, the HP4 region and the 3’-side of HP3 were also out of reach. In direct footprinting, the gel doesn’t show more than about 100 nucleotides, which excludes also HP3.

The choice of probes was technically limited. 1M7 (SHAPE reactant) did not show very clear protections in the presence of proteins, we then turned to RNases. Since RNase T1 didn’t show clear data, V1 was the only probe to give informations. Experiments were started with DMS (no clear results at HP3). CMCT required modifications of the storage buffer of the proteins. This was done, but time restrictions did not allow for the experiments to be finalized.
Chapter VII: General discussion and conclusion

The 7SK RNA, involved as the scaffold of a small snRNP in higher eukaryotes requires the protein LARP7 for its stability (He et al., 2008). One of the functions of the 7SK snRNP is to capture and inhibit the transcription elongation factor, P-TEFb, and thus to participate to the regulation of the elongation pausing by polymerase II (Nguyen et al., 2001). This function depends on 7SK RNA recognition by the HEXIM protein, which in turn binds P-TEFb (Michels et al., 2004). The present PhD project was aiming at understanding how LARP7 binds 7SK RNA and to investigate if that binding impacts HEXIM binding.

1. LARP7 wraps around 7SK RNA

In the first part of the project, I contributed to unveiling the origin of the specific recognition of 7SK RNA by LARP7. My experiments confirmed that binding requires the uridines at the 3’-end, and showed that the La module in the N-terminal domain of LARP7 is not only responsible for that, but also contributes to binding a 3’sequence at the foot of HP4. This sequence has been implicated in a closed conformation of 7SK RNA (Marz et al., 2009). Interestingly, this means that in the 7SK RNA snRNP, the La module is close to the 5’-end of 7SK RNA, which is capped by MePCE. It has been shown that once MePCE has methylated the cap, it stays bound to the RNA and contacts LARP7 (Xue et al., 2010). A closed form of 7SK RNA would be compatible with and favor an interaction between MePCE and LARP7. This may have functional significance, since both molecules are involved in 7SK RNA stabilization.

I showed that the C-terminal domain of LARP7 contributes to specific binding to 7SK
RNA by binding the apical loop of its 3'-hairpin, HP4. Nucleotide G312 in the apical loop had been recovered \textit{in vivo} to be important for specific recognition of 7SK RNA (Muniz et al., 2013), it was directly recognized by LARP7 C-terminal domain. My results contributed to a model where the C-terminal region of LARP7 comprising the RRM2 binds to the apical loop while the La module binds at the foot of HP4 (Uchikawa et al., 2015).

Further exploration of the interaction indicated that full length LARP7 also binds a sequence in the domain 2, in the middle-region of the 7SK RNA. LARP7 N or C-terminal domains does not bind this sequence, the linker region between the 2 domains of LARP7 might bind it. No clear 2D structure of 7SK RNA domain 2 is available. Published models differ in this region (Wassarman and Steitz, 1991) (Marz et al., 2009). Preliminary V1 RNase footprinting data of the team suggests it adopts a hairpin structure.

Interestingly, the sequence in domain 2 is conserved. The sequence conservation of 7SK RNA was reported in the two published 2D models (Wassarman and Steitz, 1991) (Marz et al., 2009) (Figure 7-1). The highest conservations are related to function. The 7SK RNA-motif in the 5'-hairpin was identified to contain the major HEXIM-binding site (Egloff et al., 2006, Lebars et al., 2010). The 3'-hairpin (HP4) was involved in P-TEFb regulation (Egloff et al., 2006; Muniz et al., 2013). Interestingly, it corresponds also to the LARP7 binding site. The sequence at the foot of the hairpin which was the support of the closed model of 7SK RNA, is also conserved. We found that it is also involved in LARP7 binding. This is the case also of the domain 2 region, highly conserved, now appearing as to be involved in LARP7 binding. On the whole, the LARP7 binding sites
identified correspond to highly conserved sequences of 7SK RNA. It has been established that LARP7 is responsible for 7SK RNA stability (He et al., 2008). The fact that it binds along the RNA at the conserved regions strongly suggests that the two molecules co-evolved. It may also indicate that the essential functions of 7SK RNA are shared with LARP7.

Our investigations indicated that the central hairpin, HP3 might also contribute to LARP7 binding. A sequence comprising several lysines (266-289) in the central linker of LARP7 may be involved. This result needs to be confirmed and further investigated. In particular, it would be interesting to know which part of HP3 is involved. There are some conserved nucleotides in HP3 which might be involved. The hairpin HP3 is present only in sequences from vertebrates (Gruber et al., 2008b). This could reflect special functions of the 7SK snRNP in vertebrates. Binding of the middle region of LARP7 to the middle-region of the RNA in addition to the two terminal domains binding to the 3’ hairpin suggests that LARP7 is wrapping around the RNA. On the whole, our present model of

Figure 7-1. Secondary structure and sequence conservation of 7SK RNA. Model of the secondary structure of 7SK RNA according to Wassarman et al. in the left (Wassarman and Steitz, 1991) and Marz et al. (Marz et al., 2009). The sequence is colored based on its conservation according to Marz et al. (Marz et al., 2009).
7SK RNA/LARP7/HEXIM binding is shown in Figure 7-2.

![Figure 7-2](image)

**Figure 7-2.** Model of 7SK RNA/LARP7/HEXIM binding on the Marz 2D 7SK RNA structure. HP1, domain 2 and HP3 of 7SK RNA are shown in black as 2D structure. HP4 is shown in 3D structure. HEXIM-binding region is shown in light blue dotted circle. Linker (purple dotted line) of LARP7 and possible binding region of the RNA are indicated with purple circles. N-terminal (La module 3D structure) and C-terminal (RRM2 3D structure) of LARP7 are in yellow and navy, respectively. 3’, 5’ end and G312 are pointed with arrows.

### 2. Assembly and dynamics of the 7SK RNA snRNP

In chapter IV, we observed that LARP7 also binds HP1. The hairpin comprising the major (GAUC)$_2$ HEXIM binding motif. This signature sequence for 7SK RNA is also essential for HEXIM-binding (Lebars et al., 2010). LARP7 binding to HP1 was unexpected, and was investigated by EMSA, footprinting, and confirmed by analysis of a
mutant RNA (Chapter IV). It involves the structured C-terminal domain. We showed that this domain recognizes the apical loop of the isolated domain HP1 by footprinting and by mutating the loop, which results in a loss of binding. Interestingly, these sequences are adjacent to the major HEXIM binding motif UU(GAUC)_2U. However, the situation is less clear when the full length RNA is investigated. The isolated C-terminal domain still binds. This shows that binding to HP1 competes with the primary target, HP4, and may be revealed in this context, provided enough molecules are present. More puzzling is the fact that the binding to HP1 is no longer observed with the full length LARP7. One explanation is that the C-terminal domain binding to HP4 (not visible in our experimental set-up) is favored in the full length situation. This could be due to the presence of the La module, which binds the foot of the hairpin. Binding of the La module might restrict the access of the C-terminal domain to HP1, or favor its interaction with HP4. An interaction between the La module and the C-terminal domain of LARP7, possibly mediated by the HP4 RNA, is not excluded.

Our observation that the C-terminal domain of LARP7 has the capacity to bind HP1, the 5’-hairpin raises interesting issues. In particular, with respect to the 7SK snRNP assembly. One can imagine a sequential binding of LARP7. The C-terminal domain might bind HP1 first, thus “sensing” the presence of this domain which comprises the 7SK RNA-motif, thus ascertain the nature of the RNA, while the 3’-end is bound by the La-domain. In an intermediate step, the assembly could be further stabilized by the La-domain contact with MePCE. The final assembly requires the linker to bind to domain 2 and perhaps HP3, as well as binding of the C-terminal domain to HP4. We have no insight, for the moment into the order of successive binding events. Obviously, many other explanations can be
given to explain how the C-terminal domain of LARP7 binds two domains of 7SK RNA, including alternative conformations of the 7SK RNA snRNP. The difference of affinities for the two hairpins should also be taken into account. We observed affinities ratio of 1:6 in favor of HP4. This may vary according to the presence of other proteins (preference for HP4 may be modulated by the presence of MePCE or HEXIM, or hnRNPs). Clearly, to gain further insight into that process, MePCE should be included in our experiments. On the whole, the role of LARP7 seems thus more complicated than just stabilizing one functional conformation of 7SK RNA.

Future investigations should first clarify if binding to HP1 and HP4 involves the same or different surfaces of the C-domain. A first control could be made with our available mutants, but the best way to compare the two situations is NMR chemical shift mapping. In that purpose, the spectra of the N^{15} labeled C-terminal domain is compared with the spectra obtained in the presence of either HP1 or HP4. If the two sites differ (i.e. perturbations observed for different residues), then a simultaneous addition of the two RNAs would give interesting information.

3. **Impact of LARP7 on HEXIM recognition**

When I started this project, HEXIM-binding was known to occur in the region of the 7SK RNA motif, in HP1. HEXIM depends on an ARM peptide to bind the RNA. NMR-chemical shift mapping of the RNA showed perturbation of the central base-pairs of the (GAUC)_{2} repeat, and stabilization of the A39-U68 base-pair below (Lebars et al., 2010). Contacts involved the upper part of the HP1 stem, just below the apical loop. The central
region appeared to be stabilized by peptide binding. Additionally, EMSA experiments with mutated RNAs showed that the bulged uridines were essential for recognition (Figure 5-1). Further insight into the structure of HP1 was gained recently with the crystal structure (Martinez-Zapien, manuscript in preparation) and the solution structure (Bourbigot et al., submitted) of the same molecule, a variant of HP1 (24-87) with the apical loop replaced by the tetraloop UUCG. Interestingly, the structures differ. This indicates a conformational flexibility of HP1, which was further confirmed by a molecular dynamics study (Martinez-Zapien, manuscript in preparation). Such flexibility may be exploited by HEXIM and LARP7 to bind alternatively to the hairpin.

While LARP7 and HEXIM bind together the full length 7SK RNA (Figure 6-1), they do not bind the isolated domain HP1 simultaneously (Chapter V). In Figure 5-4, we saw that HEXIM is able to extract and bind the RNA from a pre-formed complex of LARP7 and HP1. Simultaneous binding of the two proteins is not sterically impossible. Figure 7-3 shows the crystal structure of HP1\textsubscript{UUCG} with the nucleotides colored according to HEXIM recognition (the most recent data are reported), and arrows indicating the regions were contacts with LARP7 have been observed: at the apical loop, (not included in the structure) and U72 and U76 as reported in Chapter V. The hypothesis of HEXIM and LARP7 binding to different conformations is still open.

In the course of these investigations we observed that in the presence of LARP7, the binding of HEXIM to HP1 seemed somewhat different (Martinez-Zapien et al., 2015). In the absence of LARP7, two different complexes (corresponding to two bands migrating at different levels on a native EMSA gel) are formed with the monomeric HEXIM construct
(136-273). These were interpreted as complexes of HP1 with one or two monomers of HEXIM. In the presence of LARP7, only one complex is formed, which seems to correspond to the upper band, the complex with two monomers. This suggests that one monomer of HEXIM may bind to the LARP7-bound RNA in a first step, but that the incoming second monomer eliminates the binding with LARP7. This suggests a competition for the secondary binding site. Interestingly, this opens the possibility that LARP7 could be involved in specific recognition of 7SK RNA by HEXIM. At some time of the process, LARP7 binding to HP1 could either stabilize a competent conformation of the RNA to make it suitable for recognition by HEXIM or, alternatively, bind the RNA to restrict the primary access to HEXIM to the specific determinants of recognition. This could explain how such a low complexity molecule as HEXIM could perform its specific function.

Figure 7-3. 3D-structure of the hairpin HP1_UUCG indicating HEXIM and LARP7 binding sites. The nucleotides which after mutation where shown to impact HEXIM recognition, (Martinez-Zapien et al. manuscript in preparation) are highlighted; the strongest effect were observed for U40, U41 (red), the central base-pairs of the GAUC2 (yellow) and A39 (green). The spheres represent imino protons which were observed by NMR chemical shift mapping (Lebars et al., 2010). They are coloured according to the observed effect upon titration with the BR peptide: unaffected (blue), involved in binding (green), perturbated (pink) or not observed (grey). LARP7 binding is materialized in purple.
REFERENCES


inhibit breast cancer progression and metastasis. eLife 3, e02907.


Lu, H., Li, Z., Xue, Y., Schulze-Gahmen, U., Johnson, J.R., Krogan, N.J., Alber, T., and Zhou, Q. (2014). AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-


Mc, R.P., Bacon, C.W., and D’Orso, I. (2016). Transcription Elongation Control by the 7SK snRNP Complex: Releasing the Pause. Cell Cycle, 0.


increases the stability of 7SK, B2 and U6 small RNAs in Xenopus oocytes. Nucleic acids research 21, 4756-4761.


Résumé

L’ARN non-codant 7SK forme la charpente d’un complexe, 7SKsnRNP, qui régule l’activité du facteur d’élongation de la transcription P-TEFb, intervenant dans la levée des pauses transcriptionnelles chez les métazoaires. Le 7SKsnRNP comprend les protéines LARP7, essentielle pour la stabilité de l’ARN 7SK et MePCE, participant à sa coiffe. Dans le cadre d’une investigation du rapport entre structure et fonction de l’ARN 7SK, le projet était de comprendre comment la protéine LARP7 reconnaît et assemble l’ARN dans le 7SKsnRNP. La protéine LARP7, membre d’une famille reliée à la protéine La, est spécifique de 7SK. Les éléments responsables de l’interaction ont été analysés par des méthodes biochimiques dans des complexes reconstitués à partir d’ARN synthétique et de protéines recombinantes. Le module La, dans la région N-terminale, reconnaît et lie les trois uridines à l’extrémité 3’ de l’ARN et, additionnellement, une séquence conservée au pied de la tige-boucle en 3’, induisant une conformation fermée de l’ARN. L’autre extrémité de la protéine comprend un domaine RRM de reconnaissance de l’ARN, qui se lie à la boucle apicale de la tige-boucle 3’. La protéine LARP7 reconnaît également une région conservée au centre de l’ARN. Dans l’ensemble, LARP7 utiliserait ses domaines terminaux et central pour envelopper l’ARN et le stabiliser. Au cours de ces travaux, une interaction directe du domaine C-terminal avec la tige-boucle 5’ a également été mise en évidence. Celle-ci comprend le site de liaison à la HEXIM, la protéine qui déclenche l’interaction avec P-TEFb et un rôle fonctionnel de LARP7 est envisagé.

Mots Clés

ARN non-codant, 7SK, LaRP, Interaction RNA-protéine, Transcription régulation, Structure

Abstract

The non-coding RNA 7SK is the scaffold for the 7SKsnRNP complex that regulates P-TEFb, the positive transcription elongation factor, which relieves transcription pauses in metazoans. The 7SKsnRNP comprises the proteins LARP7, essential for 7SK stability and MePCE, involved in capping. In the frame of an investigation of how the structure of the 7SK RNA sustains its function, the project was to understand how is the RNA recognized and assembled in the 7SKsnRNP by the associated protein LARP7. LARP7, a La-related protein is specific for 7SK. The elements responsible for the interaction were investigated by biochemical approaches in vitro with complexes reconstituted from purified recombinant proteins and transcribed RNA. The La-module of LARP7 recognizes and binds the triplet of uridines at the 3’-end of the 7SK RNA and additionally binds to a conserved region at the foot of the 3’-hairpin. This may stabilize a closed conformation of the 7SK. On the other end of the LARP7 molecule, the C-terminal domain comprising a RRM (RNA Recognition Motif) binds to the apical loop of the 3’hairpin. Further investigations showed that a conserved region in the core of the RNA is also involved. On the whole, this strongly suggests that LARP7 wraps around 7SK using its N-terminal, C-terminal and linker domains to ensure the RNA stabilization into a functional core. In the course of the investigation, was revealed a direct interaction of the C-terminal domain of LARP7 with the 5’-hairpin of the RNA, which is responsible for 7SK function as it contains the binding site of HEXIM, the protein which bridges 7SK and P-TEFb. A possible functional role of LARP7 is envisioned.

Keywords

non-coding RNA, 7SK, LaRP, Molecular interaction, Transcription regulation, Structure