

HEXIM1 targets a repeated GAUC motif in the riboregulator of transcription 7SK and promotes base pair rearrangements

Isabelle Lebars^{1,*}, D. Martinez-Zapien¹, A. Durand¹, J. Coutant², B. Kieffer¹ and Anne-Catherine Dock-Bregeon^{1,*}

¹IGBMC, BP10142, 1 rue Laurent Fries, 67404 Illkirch Cedex and ²Bruker BioSpin SA, 34 rue de l'Industrie BP 10002, 67166 Wissembourg Cedex, France

Received April 23, 2010; Revised July 12, 2010; Accepted July 13, 2010

ABSTRACT

7SK snRNA, an abundant RNA discovered in human nucleus, regulates transcription by RNA polymerase II (RNAPII). It sequesters and inhibits the transcription elongation factor P-TEFb which, by phosphorylation of RNAPII, switches transcription from initiation to processive elongation and relieves pauses of transcription. This regulation process depends on the association between 7SK and a HEXIM protein, neither isolated partner being able to inhibit P-TEFb alone. In this work, we used a combined NMR and biochemical approach to determine 7SK and HEXIM1 elements that define their binding properties. Our results demonstrate that a repeated GAUC motif located in the upper part of a hairpin on the 5'-end of 7SK is essential for specific HEXIM1 recognition. Binding of a peptide comprising the HEXIM Arginine Rich Motif (ARM) induces an opening of the GAUC motif and stabilization of an internal loop. A conserved proline-serine sequence in the middle of the ARM is shown to be essential for the binding specificity and the conformational change of the RNA. This work provides evidences for a recognition mechanism involving a first event of induced fit, suggesting that 7SK plasticity is involved in the transcription regulation.

INTRODUCTION

7SK is a highly abundant small nuclear RNA (snRNA) of 331 nt in human, transcribed by RNA polymerase III

(1–4). Since its discovery in the human nucleus in the middle 70s, secondary structure analysis in the 90s (5), and more recent elucidation of its function (6,7), 7SK has become an archetype of non-coding RNA involved in transcription regulation in higher eukaryotes. 7SK regulates RNA polymerase II (RNAPII) transcription at the level of elongation, by sequestering and inhibiting the Positive Transcription Elongation Factor (P-TEFb), a heterodimer formed by the kinase cyclin-dependent Cdk9 and the cyclin T1/T2 (8,9). P-TEFb activates transcription elongation by phosphorylating the C-terminal domain of RNAPII and antagonizes the negative regulation by NELF (Negative Elongation Factor) and DISF [DRB (5,6-dichloro-t-β-D-ribofuranosylbenzimidazole) sensitivity-inducing factor] (10–12). Sequestration of P-TEFb depends on 7SK snRNA binding to a protein HEXIM (HEXIM1 or the minor protein HEXIM2), and is reversible (13–16). Indeed, global inhibitors of transcription such as DRB or actinomycin D induce the release of P-TEFb from the 7SKsnRNP, enabling its recruitment by the bromodomain-containing protein Brd4 (17,18). 7SK snRNA is a stable RNA, capped by the methylphosphate capping enzyme MePCE (also known as BCDIN3) (19,20) and protected from cleavage by exonucleases by the La-related protein LaRP7 (PIP7S) (21–23). LaRP7 most probably binds 7SK by its 3'-UUU-OH sequence (Figure 1A). These proteins act cooperatively to ensure the stability of a core 7SKsnRNP and promote further assembly of the large complex containing P-TEFb and HEXIM (23,24). In the absence of P-TEFb, several hnRNP proteins have been shown to bind to the 7SKsnRNP (25–27). Sequestration and inhibition of P-TEFb in the 7SKsnRNP is a ubiquitous way to

*To whom correspondence should be addressed. Tel: +33 3 88 65 57 72; Fax: +33 3 88 65 32 76; Email: dock@igbmc.fr
Correspondence may also be addressed to Isabelle Lebars. Tel: +33 3 68 85 44 07; Fax: +33 3 68 85 47 18; Email: lebars@igbmc.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

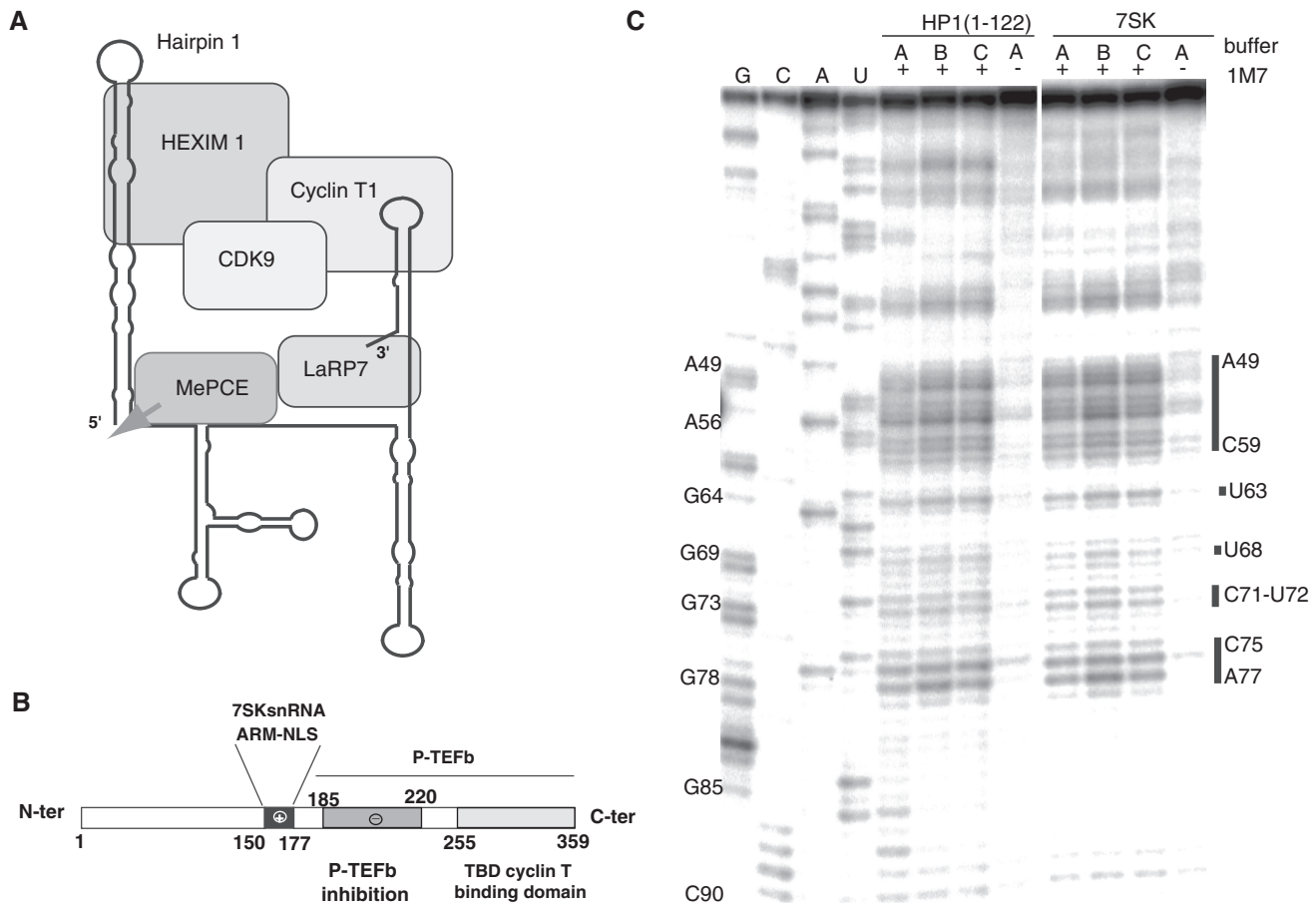


Figure 1. 7SK, HP1 and HEXIM1. (A) Schematic representation of the inhibitory 7SKsnRNP complex. (B) Domain organization of HEXIM1 protein. The central NLS overlaps with the 7SK RNA-binding domain (ARM). (C) SHAPE analysis of the 5' region of 7SK in full length context (7SK lanes) or isolated (HP1 lanes). Primer extension analysis of 1M7 reactions (+) in three different buffers (A: 50 mM Na HEPES pH 7.6, 6 mM MgCl₂, 0.25 mM EDTA; B: 50 mM Na HEPES pH 7.6, 6 mM MgCl₂, 0.25 mM EDTA, 200 mM KCl; C: 100 mM Na HEPES pH 8.0, 6 mM MgCl₂, 100 mM NaCl) or controls without reactant (-) were migrated in parallel with sequencing reactions (lanes G, C, A, U, identification on the left) to map the single-stranded, 1M7-modified nucleotides, which are identified on the right.

control elongation kinetics and transcription pauses. It influences the maturation and alternative splicing of the nascent messenger RNA in higher eukaryotes and seems essential for vertebrate development (24). P-TEFb plays also a crucial role in the life cycle of HIV through its recruitment by the viral protein Tat to the trans-activating responsive (TAR) RNA element located at the 5' end of the nascent viral transcript (28–33). Although similarly involving RNA and protein binding, the effect of 7SK:HEXIM1 mediation is opposite to TAR:Tat recruitment, where P-TEFb activity is hijacked to maximize HIV transcription.

The sequence of 7SK is highly conserved in vertebrates (34,35). A first approach of its structure came from chemical and enzymatic probing experiments performed on 7SK extracted from human cells (5), and led to a secondary structure composed of four regions (Figure 1A). A recent work based on computational survey, confirmed experimentally, identified 7SK in many more species than mammals (36–38). This increase in sequence information led to the proposition of an alternative fold for 7SK, which still contains most hairpins, but suggests

some compaction. The functionally relevant 7SK partners, HEXIM proteins, are conserved in the same species, and co-evolved with 7SK (38). In several species, including lower eukaryotes, there is only one protein HEXIM. HEXIM proteins can be divided in three functional domains centered on the RNA-binding region, an arginine-rich motif (ARM) that overlaps with a bipartite nuclear localization signal (NLS) (39) (Figure 1B). The ARM is identical for both HEXIM1 or HEXIM2 in human, and is highly conserved between species. The N-terminal proline-rich region, of variable length, is not conserved. Sequence analysis predict that both the N-terminal and central regions sustain very poor, if any, structured feature (40). The C-terminal domain, is composed of an acidic region and a coiled-coil domain, that has been shown to support the P-TEFb binding/inhibiting function and to mediate the dimerization of HEXIM (13,41–43). The acidic region was proposed to interact with the ARM in the absence of 7SK RNA. Release of the acidic region upon 7SK binding is hypothesized to induce the recruitment of the cyclin T on the coiled-coil (42). In this model, 7SK would be the

key to an HEXIM effector. However, the possibility that 7SK and HEXIM participate together in P-TEFb binding, as seen in the case of the TAR-Tat-cyclinT1 (44) cannot be ruled out.

In a first attempt to decipher 7SK identity, a functional analysis of 7SK *in vivo*, using sequence alterations and truncations, demonstrated that two RNA substructures were essential for P-TEFb and HEXIM1 binding (35). The 5' hairpin (HP1) binds to HEXIM, and is required for the subsequent recruitment of P-TEFb. The 3'-hairpin is also required for P-TEFb inhibition, but does not bind HEXIM. Further analysis reduced the minimal HEXIM-binding site to the tip of the 5' terminal hairpin of 7SK, the 24–87 (HP1) region, shown to fold as a hairpin comprising internal loops and bulges (Figure 2A). Sequence alterations indicated that the nucleotide composition of the apical helix of HP1 and the 5' bulge were involved in HEXIM1 binding (35). Recently, U30 in the lower stem was also shown to contact HEXIM using photocrosslinking (45). In spite of these advances, the molecular mechanism of recognition of 7SK by HEXIM1 and the events that lead to P-TEFb inhibition still escape our full

understanding. In the present work, we focused on the recognition of the 7SK 24–87 region by the RNA-binding region of HEXIM. We used NMR to map the interaction at the atomic level, and with combined mutagenesis and biochemical experiments, we showed how the ARM of HEXIM1 recognizes specifically a conserved motif located in the upper part of the hairpin.

MATERIALS AND METHODS

Plasmids and RNA preparation

Plasmids pHDV_7SK and pRZ_HP1 for production were obtained by cloning the 7SK or HP1 sequence into pHDV or pRZ encoding for a hammerhead ribozyme in 3' of the target RNA (46). Mutations were introduced by conventional PCR-based techniques, in both plasmids when possible. RNAs were prepared by *in vitro* transcription with T7 RNA polymerase, after linearization of plasmids, or, in the case of HP1-long (1–108), from a PCR product obtained with primers corresponding to the T7 promoter, and the 90–108 sequence, as template.

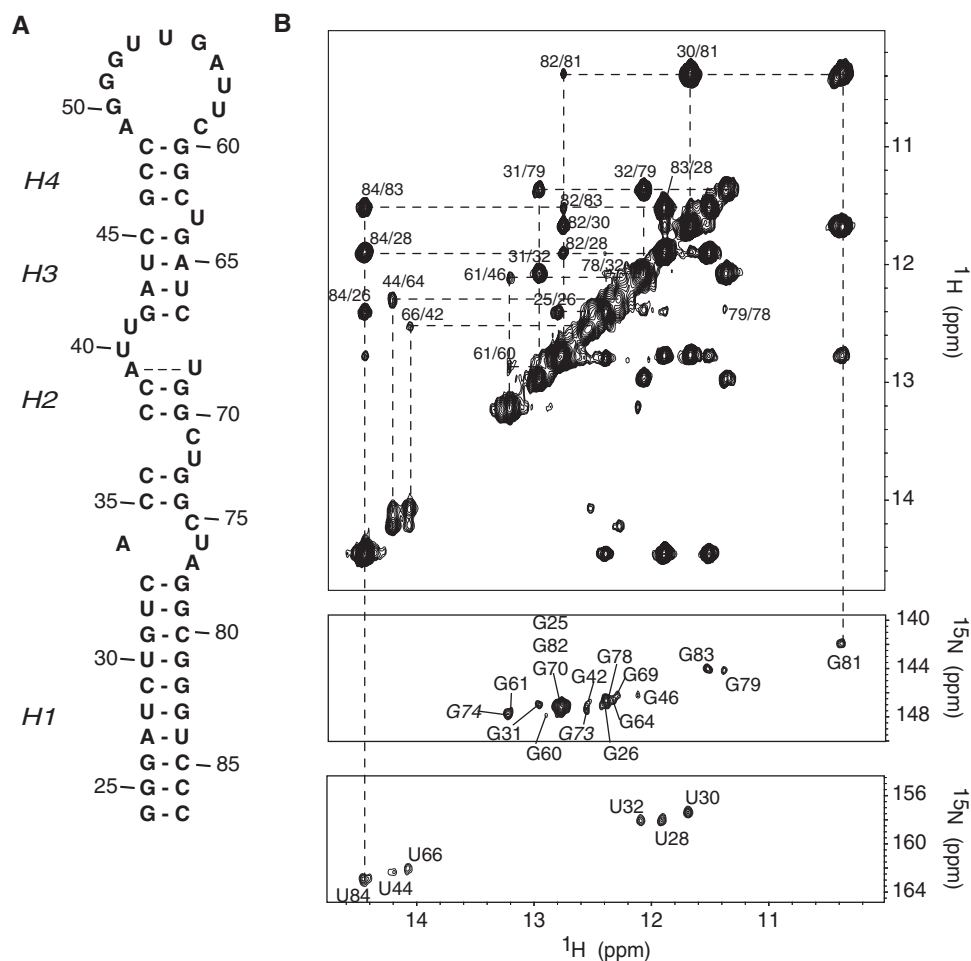


Figure 2. Determination of the secondary structure of the HP1 (24–87) hairpin. (A) Sequence of HP1 used for NMR studies. (B) Imino/imino protons region of a NOESY spectrum recorded at 10°C in 90/10 H₂O/D₂O with a mixing time of 300 ms. Dashed lines represent NH/NH sequential assignment (top). Regions of SOFAST-HMQC, recorded at natural abundance at 10°C, corresponding to the 15N guanosine imino protons (middle) and to 15N uridine imino groups (bottom).

When necessary, RNAs were cleaved from the 3' ribozyme by incubation with 40 mM magnesium. All RNAs were gel-purified.

Preparation of RNA samples for NMR

Milligram quantities of HP1 RNA and its mutants were prepared unlabeled by *in vitro* transcription. HP1 RNA was also prepared selectively ^{13}C - ^{15}N labeled at A and U residues with labeled UTP and ATP purchased from CortecNet (France). RNAs were purified as described by Wyatt *et al.* (47). After electroelution and ethanol precipitation, resuspended RNAs were dialyzed for 24 h against the buffer used for NMR experiments. Samples were concentrated by lyophilization and resuspended in 90/10 $\text{H}_2\text{O}/\text{D}_2\text{O}$ for experiments involving exchangeable protons. Each sample was refolded by heating at 95°C (2 min) and snap-cooled at 4°C. The concentrations of RNA samples were determined on a Nanodrop Spectrometer using molar extinction coefficients calculated with the Applied Biosystems calculator (www.ambion.com/techlib/misc/oligocalculator.html).

Peptides synthesis, purification and analysis

Peptides were synthesized on a Applied biosystem 433A peptide synthesizer using standard Fmoc chemistry and resins. Peptides were purified by HPLC using a preparative scale C18 column (Waters: PrepPak cartridge, 21 × 250 mm, 300 Å, 5 μM) with an acetonitrile gradient in 0.1% trifluoroacetic acid. The molecular weight and purity of peptides were confirmed by mass spectroscopy. The sequence of ARM is GKKKHRRRPSKKKRHWK and the sequence of NLS-ARM peptide is GKKKHRRRPSKKKRHWKPYKLEWEEKKKFD. Each peptide was dissolved in the buffer used for NMR experiment. The concentrations of peptides samples were determined on a Nanodrop Spectrometer using molar extinction coefficients calculated with the ExpASY Proteomics Server (<http://www.expasy.ch/tools/protparam.html>).

Preparation of HEXIM1

HEXIM1 from *Homo sapiens*, was cloned in a pET28 plasmid, to express the protein with a C-terminal His-tag. Expression in *Escherichia coli* was induced overnight at 25°C. The cell pellet was sonicated after suspension in 50 mM Tris-Cl pH 8, 5 mM MgCl_2 , 500 mM NaCl and 1.4 mM beta-mercapto ethanol. After debris removal, the cellular extract was applied to a HiLoad-chelating chromatography column (BioHealthcare) derived with nickel and eluted with imidazole. We found important for reproducibility of the electrophoretic mobility shift assays (EMSA) experiments, to get rid of all spuriously bound RNAs coming from the *E. coli* cytoplasm and to remove aggregates. For that, lysis, Ni-chelation and gel-filtration steps were performed at high ionic strength (500 mM NaCl). The Ni-bound HEXIM was washed thoroughly (at least 3 × 10 ml for 750 μl resin) with this buffer before elution with imidazole. Finally, a further step of gel-filtration chromatography on Superdex S200 (in the same buffer with 2.5 mM DTE) allowed to separate a peak eluting earlier than the HEXIM peak, with OD_{260}

larger than OD_{280} , and considered as RNA-bound HEXIM aggregates. The protein was used as fresh as possible, without further treatment. Its concentration was measured optically with a molar extinction coefficient of $34\,170\text{ M}^{-1}\text{ cm}^{-1}$.

Proton and heteronuclear NMR

NMR experiments were recorded at 600 MHz (DRX and Avance III Bruker) or 700 MHz (Avance III Bruker) on spectrometers equipped with z-gradient cryoprobes. NMR data were processed using TopSpin (Bruker) and analyzed with Sparky software packages (Goddard, T.D. & Kneller, D.G., SPARKY 3, University of California, San Francisco). NMR experiments were performed in 50 mM sodium phosphate buffer (pH 6.2). The concentration of RNAs samples ranged from 0.2 to 0.8 mM. Samples volumes were 280 μl in Shigemi NMR tubes. Complexes were formed by addition of peptide to RNA monitoring the imino region of 1D spectra. ^1H assignments were obtained using standard homonuclear experiments. NMR data were acquired at 4, 10, 15, 20 and 30°C. Solvent suppression for samples in 90/10 $\text{H}_2\text{O}/\text{D}_2\text{O}$ was achieved using the 'Jump and Return' sequence combined to WATERGATE (48–50). Two-dimensional NOESY spectra in 90/10 $\text{H}_2\text{O}/\text{D}_2\text{O}$ were acquired with mixing time of 400, 150 and 50 ms. Base pairing was established via sequential nuclear Overhauser effects (NOEs) observed in 2D NOESY spectra at different mixing times. ^1H - ^{15}N SOFAST-HMQC experiment optimized for imino protons were acquired at 4, 10 and 15°C in 90/10 $\text{H}_2\text{O}/\text{D}_2\text{O}$ (51). ^1H - ^{13}C HSQC experiments were measured at 15 and 20°C in 90/10 $\text{H}_2\text{O}/\text{D}_2\text{O}$. ^{15}N or ^{13}C decoupling during acquisition time was achieved using globally optimized alternating phase rectangular pulse (GARP).

EMSA

Radioactive RNAs were prepared by *in vitro* transcription, in the presence of P^{32} -CTP, and gel-purified. Incubations with 0–1.2 μM recombinant purified HEXIM1 were performed in NaCl 350 mM, Tris 50 mM pH 7.6, MgCl_2 6 mM, DTT 2.5 mM, EDTA 0.5 mM, Glycerol 15%, NP40 0.01% in the presence of 0.1 mg/ml BSA and 0.3 mg/ml tRNA (12 μM). After 15 min incubation at 20°C, reactions were loaded into a 4% polyacrylamide gel and migrated in TBE 0.5× buffer at 4 W for 1 h 20 min at 4°C. Bands were visualized and quantified by phosphorimaging.

Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) experiments

Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) experiments were performed with 1M7, as described in (52,53). The RNAs [full-length 7SK or HP1L(1–122) transcripts] were incubated for 30 min in low salt buffer A (50 mM Na HEPES pH 7.6, 6 mM MgCl_2 , 0.25 mM EDTA), or higher salt B (50 mM Na HEPES pH 7.6, 6 mM MgCl_2 , 0.25 mM EDTA, 200 mM KCl) or C (100 mM Na HEPES pH 8.0, 6 mM MgCl_2 , 100 mM NaCl) at 20°C. RNAs (1 μM) were

treated with 8 mM 1M7 for 2 min at 37°C. Control reactions contained DMSO instead of 1M7. All reactions were stopped with 100 µl of stop solution (NaCl 200 mM, Glycogen 0.2 mg/ml, EDTA 2 mM) and precipitated with ethanol. Pellets were dissolved in 10 µl water for reverse transcription. The 5'-[P³²]-radiolabeled DNA primer (corresponding to residues 103–122) was added to modified RNAs and annealed by heating at 90°C for 10 min, then 20°C for 10 min. Reverse transcription was performed for 50 min at 42°C with 2 U Superscript II reverse transcriptase in the buffer provided with the enzyme and in the presence of 5 mM DTT and 0.125 mM dNTPs. Sequencing reactions of unmodified RNAs, with 0.5 mM of one of each ddNTPs, were performed in parallel. All reactions were stopped with 1 µl of NaOH 4 M and heated at 90°C for 10 min to get rid of the RNA. After precipitation, reactions were suspended in 10 µl of loading buffer (6 M Urea, Xylene cyanol 0.005%, Bromophenol blue 0.005%, 138 mM unbuffered acid Tris), and the products separated in 10% polyacrylamide, 8 M urea denaturing gels. Bands were visualized by phosphorimaging.

RESULTS AND DISCUSSION

Characterization of the secondary structure of 7SK HP1 (24–87)

The NMR structural studies were performed on the 64-nt hairpin HP1 corresponding to the 24–87 region of 7SK RNA, with a sequence modification (base pair A₂₅-U₈₆ to G₂₅-C₈₆) and a base pair inversion (C₂₆-G₈₅ to G₂₆-C₈₅) designed to obtain a better transcription yield (Figure 2A). We verified that this region folds similarly whether isolated or in the context of the full-length 7SK, by a probing experiment performed on the full-length 7SK RNA and an isolated construct (1–122) encompassing HP1 (HP1-long). This was achieved with the SHAPE technique (52,53), which allows to probe 2' hydroxyls of riboses, independently of the sequence. 2'-O-adducts are formed when local flexibility of the ribose-phosphate chain permits, and analysis by primer extension allows to locate the flexible regions in the sequence. In the region 20–90 encompassing HP1, loops and single-stands were the same when analyzed in full-length 7SK or isolated HP1-long (Figure 1C).

NMR spectra display imino protons resonances provided that they are protected from exchange with protons from the solvent, allowing assessing their involvement in H-bonds at atomic level. All base-paired imino protons were assigned via sequential NOEs observed in 2D NOESY (Figure 2B, top). The G:U wobble base pairs, located in the first helix (H1), were easily identified from the very strong NOE between the H1 guanine imino proton and the H3 uracil imino protons in NOESY experiments at short mixing time (50 ms). The A:U Watson–Crick base pair were next discriminated from G:C and G:U base pair by the strong correlation between the uracil H3 imino proton and the H2 proton of adenine. In a G:C Watson–Crick base pair, two strong NOEs occur between the guanine H1 proton and the cytosine

amino protons. The starting point for helix assignment was based upon identification of the A27:U84 pair that is the only A:U exhibiting a connectivity with a G:U pair. This U28:G83 pair allowed the assignment of the imino protons in H1. The remaining A:U imino protons were assigned to U44 and U66, further allowing the identification of G42 and G64. The last resonances, corresponding to imino protons of G:C pairs were assigned to helices H2 and H4 and the two remaining to C43-G74 and C36-G75. Assignments of A:U, G:U and G:C were finally confirmed by the analysis of ¹H-¹⁵N SOFAST-HMQC experiment recorded at natural abundance of nitrogen 15 (Figure 2B, middle and bottom) (51). Analysis of NOESY experiments revealed supplementary correlations involving U66 and G42 imino protons and three resonances at 10.44, 10.74 and 11.7 p.p.m. These last three imino protons could correspond to the bulged U40, U41 and U63 residues that appear to be involved in a hydrogen bonds network with helix H3 (Supplementary Figure S5).

In summary, four sequential nOes connectivities observed for (G₂₅-C₈₆/C₃₃-G₇₈), (C₃₇-G₇₀/A₃₉-U₆₈), (G₄₂-C₄₅/G₆₄-C₆₇) and (G₄₆-C₄₈/G₆₀-C₆₂) support the existence of the four helical segments suggested earlier (35). This corresponds globally to the secondary structures proposed by Wassarman and Steitz (5) or by Marz *et al.* (38), but with two modifications (Figure 2A). First, the central region between H1 and H2, comprises two C-G pairs, as indicated by the resonances of G73 and G74, in contrast to the internal loop proposed by folding programs. Second, the base pair A39–U68 is not formed (imino of U68 was not observed). This agrees with the SHAPE experiment (Figure 1C), where U68 is indeed flexible and accessible to the probe.

The binding of HEXIM ARM domain to a repeated GAUC motif leads to base pair rearrangements

Attempts to investigate full-length HEXIM1 binding to HP1 following imino proton shift by NMR did not give interpretable data. We observed uniform resonance broadening which can be partly due to the size of the complex and partly ascribed to conformational exchange resulting from the unfolded state of HEXIM1, a characteristic of this protein clearly supported by another NMR experiment with the protein alone (not shown). Previous reports showed that the ARM motif within the central domain of HEXIM (the 18 residues 149–165, in human) was fully functional for 7SK binding *in vivo* and *in vitro* (39). Therefore, we chose to use peptides encompassing this region to map the interaction on the RNA (Figure 5A).

The chemical shifts of imino protons were monitored as a function of the peptide concentration, and were used to map the interaction in the RNA. ¹H-¹⁵N HSQC spectra were also recorded for different concentrations of peptide (Supplementary Figure S1). Successive addition of ARM peptide on HP1 resulted in few changes in the imino proton region. In particular, two non-averaged signals from the free and bound RNA forms were observed for U66 during titration (Figure 3A, middle and Supplementary Figure S1), indicating a slow exchange

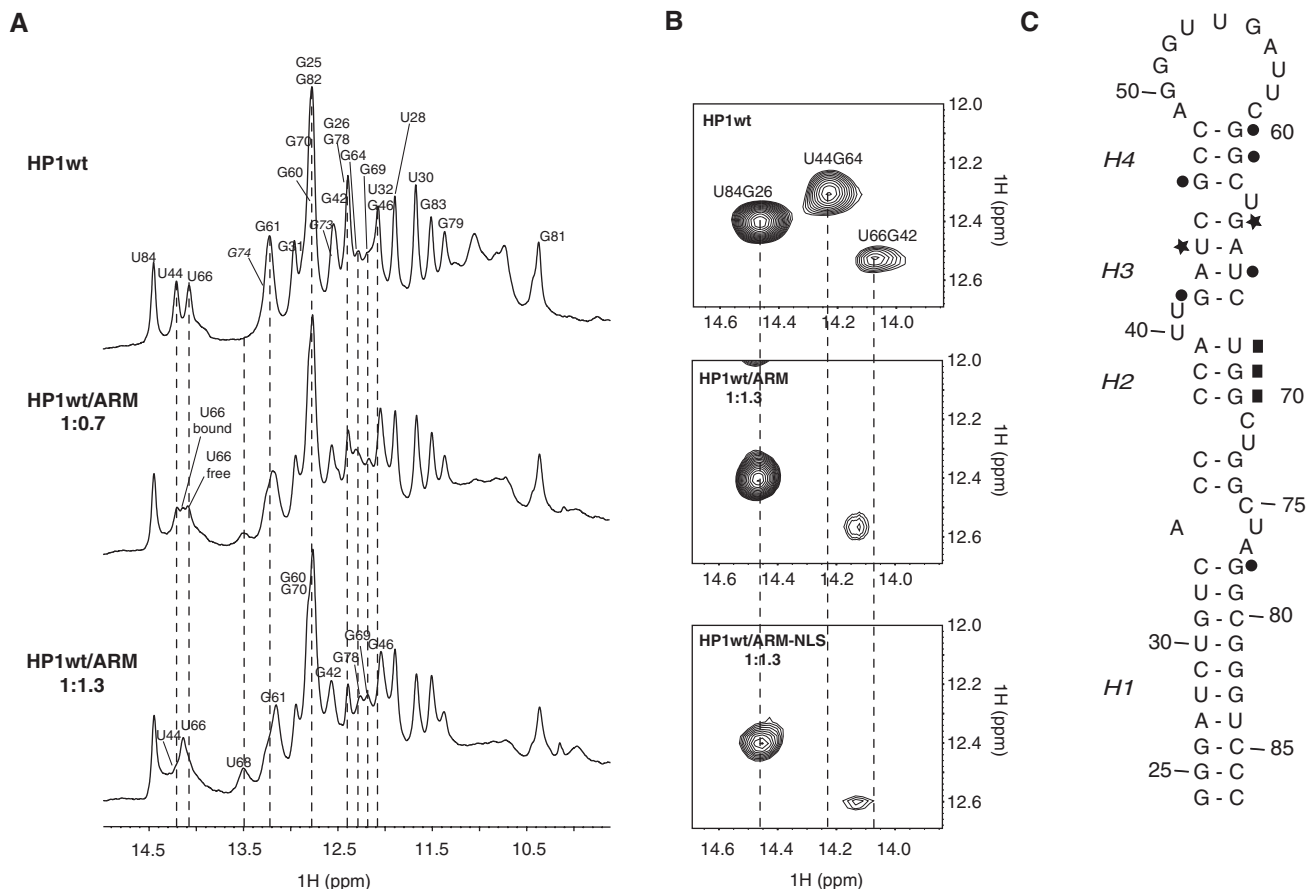


Figure 3. Mapping of the interaction between HP1 and RNA-binding domain of HEXIM1 revealed by NMR. **(A)** Titration of HP1 RNA by ARM peptide. Spectra for free RNA (top) and for HP1:ARM complexes at ratios of 1:0.7 and 1:1.3 are shown. The assignment of imino proton is based on analysis of NOESY experiments. Assignments of G73 and G74 resonances is indicated by italic characters as they can be inverted. Broken lines indicate imino proton that undergo chemical shift changes. **(B)** Imino/imino protons regions of NOESY spectra recorded at 10°C in 90/10 H₂O/D₂O with a mixing time of 300 ms. Correlations U84G26, U44G64 and U66G42 are observable in the free HP1 (top). U44G64 disappears in the HP1:ARM complex while U66G42 is shifted (middle). The same effect is observed in the complex HP1:ARM-NLS (bottom). **(C)** Summary of the effects observed on HP1 upon interaction. The opening of base pairs is indicated with stars. The stabilization of base pairs is indicated with squares. Residues for which imino protons undergo chemical shifts changes upon addition of peptide are highlighted with circles.

regime between the two states. This is characteristic of a tight binding (54). Saturation of RNA was observed at an ARM/RNA ratio of 1.3:1. Further addition of peptide to the RNA led to uniform broadening of all resonances indicating a non-specific binding.

Most of spectral changes involve residues of the upper half of HP1, as summarized in Figure 3C (the loop imino protons being not observable in the experience). Five imino protons, G42, G46, G60, G61 and U66, clustering in the H3–H4 stems undergo frequency shifts upon titration, pointing out the importance of this region for specific binding. Apart from G78, no changes are observed in the lower half of HP1.

Beside these shifts, changes in peak intensities at specific locations are observed, revealing modifications of solvent accessibility. The appearance of a new resonance at 13.50 p.p.m. (Figure 3A), in the region characteristic of imino protons involved in Watson–Crick base pairs, indicates the formation of an additional base pair. This resonance was assigned to the U68 imino proton by analysis of NOESY spectra and confirmed with ¹H-¹⁵N HSQC (Figure 3A and Supplementary Figure S1). Further

analysis also revealed that the intensity of the correlation between G69 and G70 is stronger in the complex than in the free RNA (data not shown). These observations indicate that the ARM peptide stabilizes the 37–39/68–70 (H2) region.

In contrast, the U44 resonance broadened as illustrated in Figure 3A and Supplementary Figure S1. The titration induced also a loss of the G64 resonance, which was not observable at the end of the titration. Furthermore, the correlation between U44 and G64 imino protons completely disappeared in NOESY spectra (Figure 3B). The intensity of the correlation between U66 and G42 also significantly decreased. As the exchange rate of imino proton with the solvent is related to the dissociation constant of base pairs (55), these observations may be interpreted as a melting of the H3 region upon addition of ARM.

Possible conformational changes of HP1 RNA upon ARM peptide binding was further investigated using non-exchangeable protons. For this purpose, HP1 was selectively labeled at adenosines and uridines using ¹³C-¹⁵N enriched nucleotides. C2H2 chemical shift changes upon peptide binding were monitored using ¹H-¹³C HSQC

spectra as shown in Figure 4. The free RNA spectrum displays five strong correlation peaks together with two weak peaks. Three of these correlations could be assigned non-ambiguously using NOE between H3 of

uridine and H2 of adenine (Figure 4A). Upon ARM peptide addition, a new strong correlation peak shows up upfield at 6.7 p.p.m., which could be unambiguously assigned to A39 (Figure 4E). New peaks are also

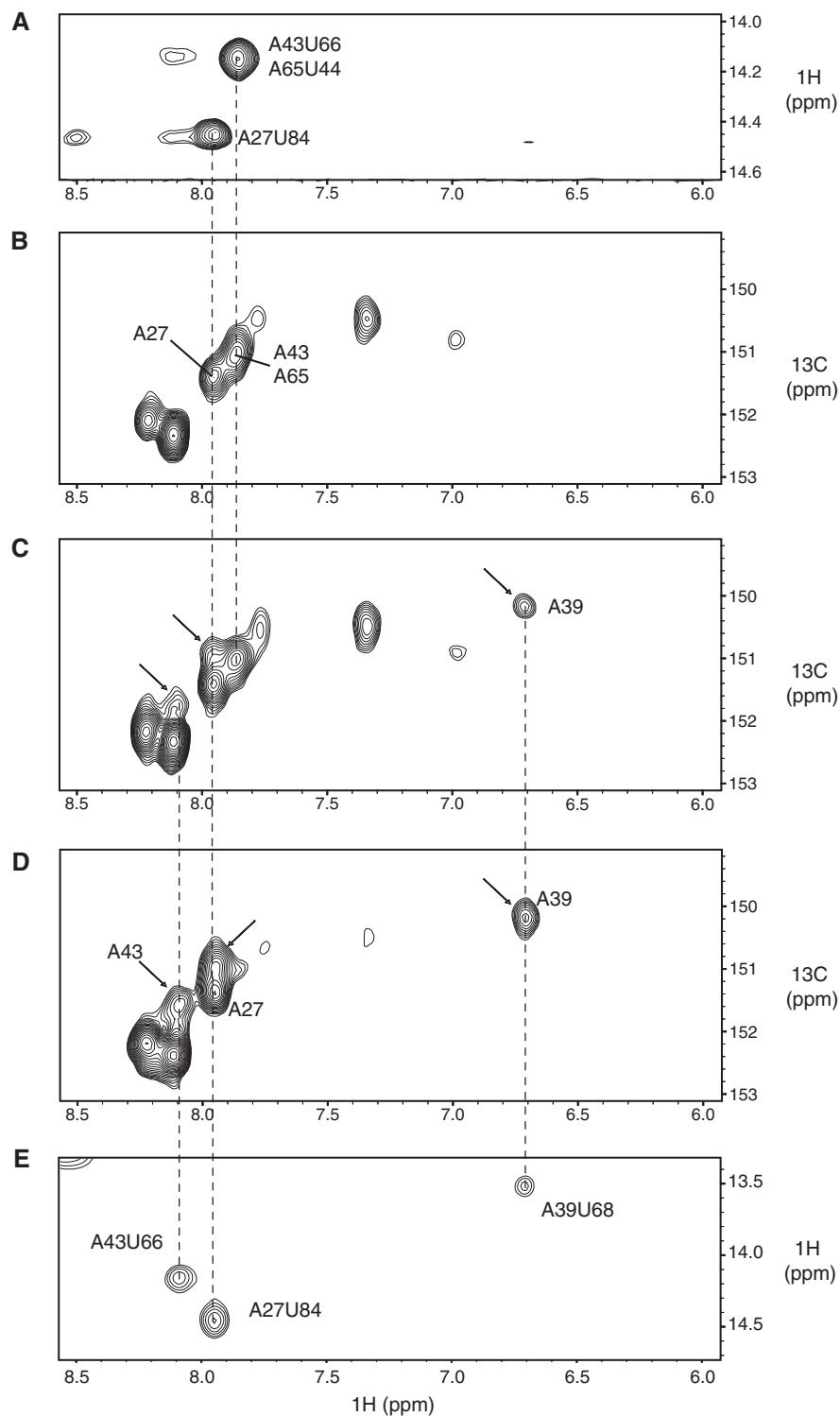


Figure 4. Effect of binding on non-exchangeable protons. (A) ^1H NOESY spectrum showing imino/H2 correlations of the free HP1 at 20°C, (B) ^1H - ^{13}C HSQC spectrum showing the aromatic H2-C2 correlations of the free HP1 at 20°C, (C) ^1H - ^{13}C HSQC spectrum showing the aromatic H2-C2 correlations of HP1 at 20°C with a ratio 0.5:1 peptide/RNA, (D) ^1H - ^{13}C HSQC spectrum showing the aromatic H2-C2 correlations of HP1 at 20°C with a ratio 1.3:1 peptide/RNA and (E) ^1H NOESY spectrum showing imino/H2 correlations of HP1 at 20°C with a ratio 1.3/1 peptide/RNA. Correlation peaks that appear upon titration are indicated by arrows.

observed in the downfield region, one of which resulting from the shift of A43 (Figure 4D). The correlation assigned to A65 in the free form is shifted upon binding to a downfield position which can no longer be assigned. Thus, the A65 and A39 residues shift in opposite directions, which can be interpreted as RNA conformational changes. The same behavior (shift downfield) is observed with A43, adjacent to the A65U44 base pair, further indicating the melting of GAUC motif. This study strengthens our interpretation of an opening of the U44-A65 base pair and concomitant closing of the A39-U68 base pairs in the H3/H2 region.

Finally, connectivities observed in the free RNA, that probably correspond to the three U bulges, also disappeared upon interaction with the peptide, indicating that the hydrogen bonds network involving U40, U41 and U63 would also open upon binding (Supplementary Figure S5).

In HEXIM proteins, the ARM RNA-binding region (150–165) and the bipartite NLS (150–177) are entwined (39). In order to avoid missing recognition events that could occur with the neighboring sequence, we extended our mapping with a longer peptide, spanning the ARM-NLS (149–179) region. We observed similar chemical shifts changes of HP1 imino protons upon titration, and a similar saturation at a ratio of 1.3:1 (Figure 2B, bottom). These two independent experiments allow to conclude that 7SK RNA recognition elements are located in the ARM region of HEXIM1. The contacts involve mainly the upper part of HP1 and the last C33:G78 base pair of the lower stem. The shift of U66 and broadening of the U44 resonances, concomitant with the appearance of a resonance corresponding to U68 at 13.5 p.p.m., are signatures of a specific recognition process where ARM-NLS, as ARM, opens the GAUC region of HP1 and stabilizes the H2 internal loop (Figure 3C).

P157 and S158 of the HEXIM1 ARM motif play an essential role in the specific binding to HP1

The ARM motif of HEXIM proteins (amino acid 149–165 in human HEXIM1) comprises two segments of seven and five basic amino acids linked by a proline–serine sequence conserved in most species (exceptions are insects, urchin and ciona). The first segment (KKKHRRRP) is highly homologous to the TAR-binding sequence of HIV-1 Tat protein (KRKHRRRP) (39). The sequence KHRR (amino acid 152–155) has been shown to be essential for 7SK binding in human cells (56). *In vivo* experiments showed that the Tat sequence is not sufficient to substitute the full HEXIM ARM (150–165). However, replacing the second segment of HEXIM1 ARM with Tat maintained HEXIM binding *in vivo*, but interestingly, a loss of specificity was observed, and other RNAs were bound, in addition to 7SK (39). ARM are found in other RNA-binding protein (57,58), but they do not comprise the proline (P) and serine (S) insertion observed in the middle of HEXIM ARM motif. Aiming at probing the role of these residues, several mutants were designed and

their binding properties were studied using NMR (Figure 5A).

In contrast with the titration with wild-type peptide, where the perturbations of peak intensity were observed at specific locations, the experiments performed with mutant peptides all led to a uniform broadening of all imino protons resonances. This indicates a loss of binding specificity. Following the imino protons located in the H3 region, saturation was obtained with a higher ratio of 2:1 compared to wild-type. The intensity of the resonance corresponding to U44 does not significantly decrease (Figure 5), showing that U44 is still engaged into a base pair. With the three ARM mutants, the nature of the complex is different: a less specific binding occurs, and the GAUC–GAUC motif does not open.

The resonance at 13.5 p.p.m., assigned to U68, appeared very weakly at the end of the titration with ARM-S158C or ARM-P157G (Figure 5), showing that the A39–U68 base pair is not stabilized as in the wild-type case. With the ARM-P157K variant, the U68 imino proton resonance was observed. It was narrower compared to wild-type, which can be ascribed to a stabilizing effect of a non-specific interaction provided by the addition of a positive charge.

In conclusion, NMR mapping with variant ARM shows that modified peptides contact all the surface of the RNA in a non-specific manner. The conserved proline and serine of the ARM motif target specifically, and promote the opening of the GAUC–GAUC motif.

Recognition elements for HEXIM encompass the U-bulges and GAUC–GAUC motif

The observation of base pairs opening upon peptide interaction suggests that the stability of helix H3 containing the GAUC motif and bordered by the U-bulges is directly involved in the binding mechanism. Bulges not only display recognition signals to a protein partner, they also favor the opening of grooves allowing insertion of protein helices or hairpin (59,60). To further investigate the mechanism of recognition, mutants were designed to modify the stability of H3 helix and their binding properties were probed by NMR (Figure 6A–C) and biochemical assays (Figure 7).

Three mutants were designed with either the deletion of U40U41 or U63 bulges (HP1 Δ U4041 and HP1 Δ U63) or the mutation of GAUC–GAUC to GGCC–GGCC motif (HP1dm) (Figure 4A–C). The secondary structure of HP1 Δ U4041, as determined by NMR, is similar to the HP1 fold, with the exception of the A39–U68 pair, not observed in the wild-type, but observed here. Suppression of the U40U41 bulge induces a more rigid local structure consisting of stacked base pairs in a helical stem, as expected. The deletion of the one-nucleotide bulge in the HP1 Δ U63 mutant produces also the expected merged helix H3–H4. Interestingly, while the global fold of HP1dm, with the two central AU base pairs of the GAUC motif changed for two GC, is similar to wild-type, the A39–U68 pair not formed in the wild-type is now observed. This observation underlines how subtle changes in an RNA sequence can affect

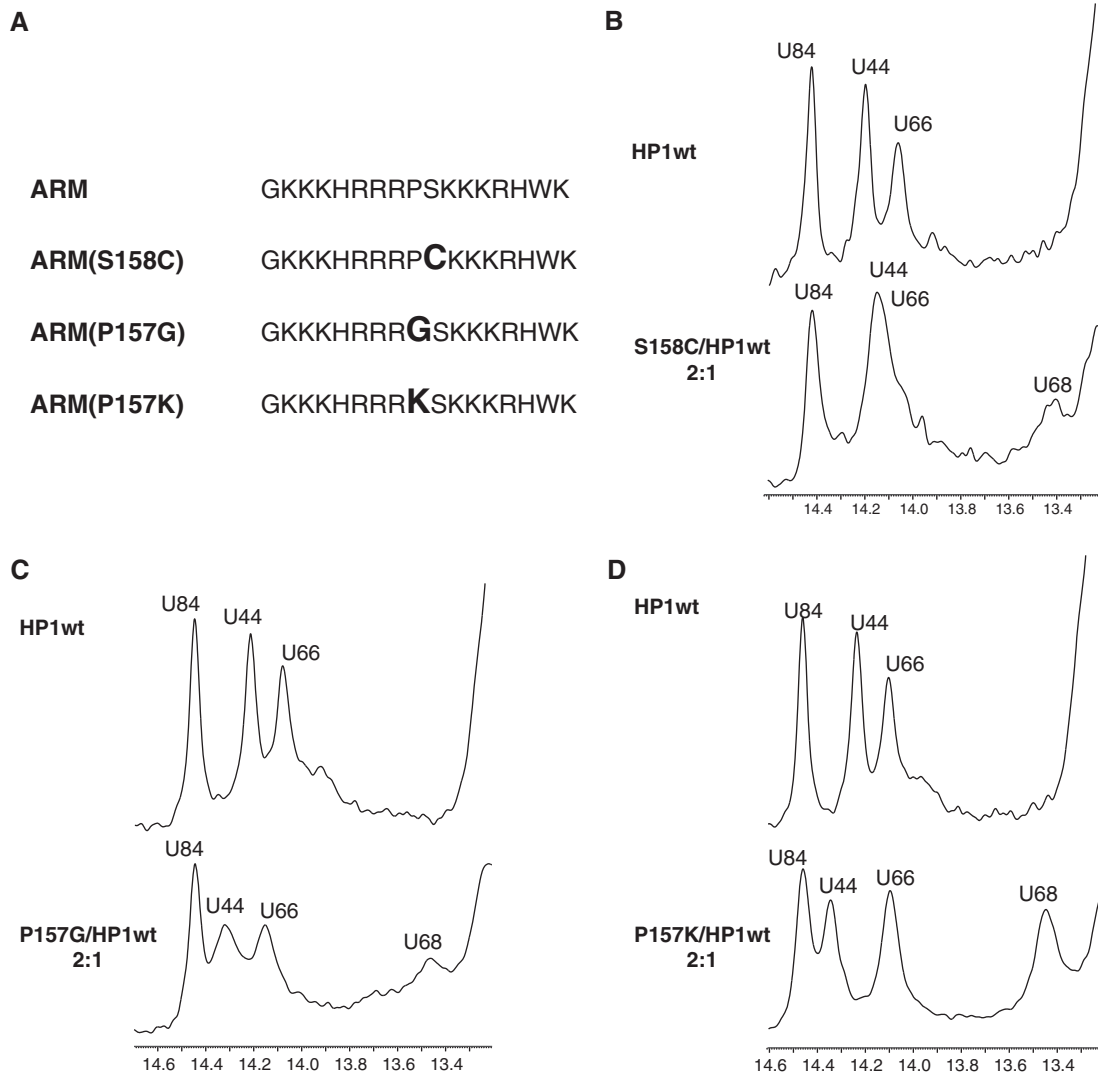


Figure 5. Effect of mutations in the ARM motif on the binding to HP1. (A) Sequences of ARM motif and its mutants. Mutations are indicated with bold letters. Regions corresponding to imino protons involved in AU pairs of HP1 are shown upon titration by S158C (B), P157G (C) and P157K (D).

dynamical properties of remote elements. Here, the stabilization of the core of H3 transmits into the neighboring H2, and produces an indirect, but effective, stabilization effect.

NMR experiments were then performed on complexes formed between the mutants HP1 and ARM peptide (Figure 6). With HP1dm, the titration induced a uniform broadening of resonances upon binding, while the chemical shifts remained almost unchanged characterizing a non-specific interaction without change of the RNA fold (data not shown). Previous works reported that HEXIM1 binding is sensitive to the nucleotide composition of this GAUC–GAUC motif. Reversing GAUC–GAUC to CUAG–CUAG had been shown to abolish the interaction (35). Our results obtained with the GGCC–GGCC mutation shows that even when changing only the two central base pairs and keeping the order of purines and pyrimidines, the specific binding is inhibited.

The ARM titrations performed on HP1ΔU4041 and HP1ΔU63 also induced a global broadening of imino protons resonances with little changes in chemical shifts. The base pairing of H3 is conserved upon interaction, as illustrated by the correlation between U44 and G64 imino protons that remains unchanged in NOESY spectra in comparison to free RNA (Figure 6D and E). Both deletions of the bulged pyrimidines affected the binding mode of the peptide, and suppressed the specific opening of the GAUC motif. These evidences for the importance of the bulges correlate with previous published works reporting that mutation of U40U41 to A40A41 abolishes the binding to HEXIM1. Mutation of U63 to A63 was shown to alter also the association, but more mildly (35). In summary, the association with ARM seems to be driven by the specific sequence composition GAUC–GAUC, with an essential contribution of the two AU base pairs and by the presence of the two bulges that flank the motif.

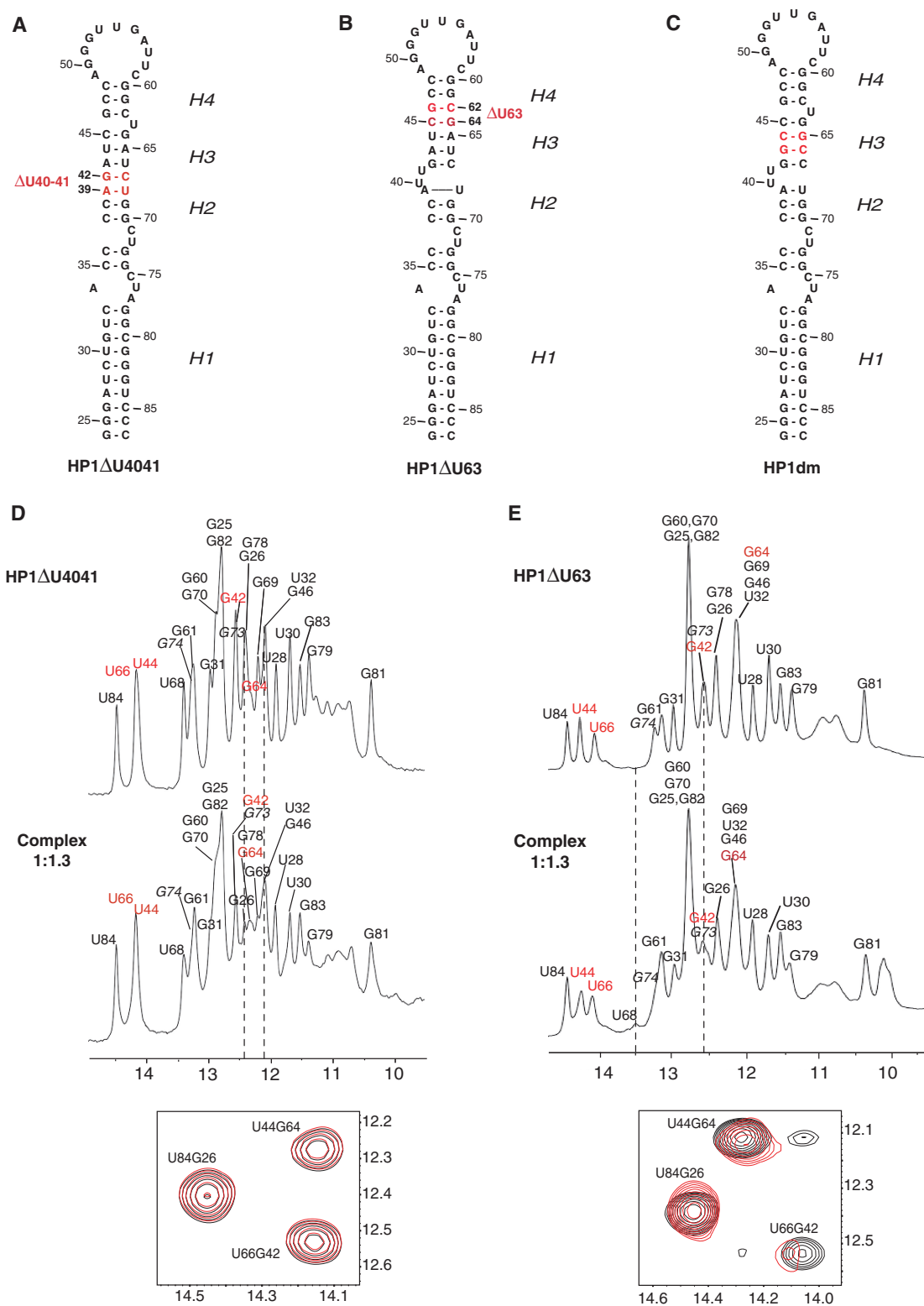


Figure 6. Role of bulges on the interaction with RNA-binding domain of HEXIM1. (A) Secondary structure of HP1ΔU4041 in which U40 and U41 have been deleted. (B) Secondary structure of HP1ΔU63 in which U63 had been deleted. (C) Secondary structure of HP1dm in which mutation GAUC/GAUC to GGCC/GGCC was introduced. Mutations are highlighted in red. (D) Titration of HP1ΔU4041 RNA by ARM domain. The assignment of imino proton based on the analysis of NOESY experiment is indicated for the free RNA (top) and the RNA:ARM complex (middle). Broken lines indicate imino proton that undergo chemical shift changes. Imino/imino protons region of a NOESY spectrum recorded at 10°C in 90/10 H₂O/D₂O with a mixing time of 300 ms is represented (bottom). Correlations U84G26, U44G64 and U66G42 are observable in the free HP1ΔU4041 (black) and remain unchanged in the bound RNA (red) (E) Titration of HP1ΔU63 RNA by the peptide. The assignment of imino proton based on the analysis of NOESY experiment is indicated for the free RNA (top) and the complex (middle). Broken lines indicate imino proton that undergo chemical shift changes. Imino/imino protons region of a NOESY spectrum recorded at 10°C in 90/10 H₂O/D₂O with a mixing time of 300 ms is represented (bottom). Correlations U84G26, U44G64 and U66G42 are observable in the free HP1ΔU63 (black) and remain unchanged in the bound RNA (red).

Functional investigations in the context of the full-length protein further assess the importance of the GAUC–GAUC sequence and its context

In order to gain insight into the 7SK:HEXIM1 mechanism of recognition, it was essential to study this interaction in the wider context of the full-length protein and RNA. Since this was not amenable by NMR, EMSA were used in that prospect. Mutants HP1 Δ U4041, HP1 Δ U63 and HP1dm were studied in the context of larger RNAs, as well as HP1AU43–44 and HP1AU65–66 where the GAUC motif was changed to GGCC on one strand only. EMSA experiments performed with the full-length 7SK RNA and the full-length recombinant HEXIM1 show that the binding of the protein is clearly affected by mutations in the GAUC region (Figure 7B). In order to compare the relative effect of mutations, experiments were next performed with a shorter RNA encompassing the full 5' terminal hairpin, HP1-long (1–108) (Figure 7C–G). Results are summarized in Figure 7H in which the percentage of shift is represented as a function of HEXIM1 concentrations in the range 0–1.2 μ M, compatible with the affinity recently estimated to 0.5 μ M (61). Large impact on HEXIM1 binding was again observed for modifications of the upper part of HP1 sequence, where deletion of the bulged uridines Δ U40–41, Δ U63, or change of the two central AU pairs of the conserved motif compromised strongly HEXIM1 binding (Figure 5C–E). Interestingly, the effect of GAUC change seemed less drastic on the 3'-side of the hairpin. Indeed, AU43–44GC inhibited binding, almost to the suppression observed with HP1dm, while with AU65–66GC, some affinity for HEXIM1 was maintained, even though the local structure is opened by the mutation.

Several mutations were also introduced to investigate the possible role of remote regions. In particular, several mutants were designed in the H1 stem (Figure 5A). The 3G–U wobble pairs, which seems a peculiar feature of mammalian 7SK were modified independently with Us to Cs and Gs to As, to build genuine Watson–Crick base pairs. Mutations U28C, G83A, U30C, G81A, U32C and G79A showed negligible effects, in line with the absence of contact deduced from the NMR experiment, and confirming that 7SK identity does not include the successive G–U pairs. A last mutation investigated the central region, changing it for a small stem by deletion of residues C71–U72, C75 and A77. The resulting structure (Δ loop71) of three base pairs (A34–U76, C35–G74 and C36–G73) should be much more rigid than the original (Figure 5A). This mutant showed medium-range effects (Figure 5G), in accordance with the previously published mutation analysis, showing this region to be involved in HEXIM1 recognition (35). Close comparison of the mutations, in the light of our experimentally determined secondary structure, shows that these similar effects are the result of quite different modifications, since the previous mutagenesis destroyed the two 35–74 and 36–73 base pairs, maintained here. Since our NMR experiments showed no specific binding by the ARM peptide at this lower part of the HP1 hairpin, these structural elements could constitute a secondary binding site, target of

another part of the HEXIM protein, as was hypothesized in a previous study, showing a crosslink of HP1 with the HEXIM1 210–220 region outside of the ARM (45). The effect of this Δ loop71 mutation could also be indirect, reflecting requirement for a dynamic, yet not destroyed, structure of the RNA at that level.

CONCLUSION

Several recently published works (36–38) aiming at finding 7SK RNA in other organisms than higher eukaryotes by bioinformatics, reported the constant presence of two successive GAUC sequence patterns in the 5'-region of 7SK, as a characteristic of 7SK. In the most recent work (38), one of the top criteria for a wider search was that the two GAUC were base paired, separated by 6–30 nt, and formed a hairpin. In vertebrates and in lophotrochozoa, this short stem is flanked by bulged pyrimidines. The interacting region with the RNA-binding domain of HEXIM1 detected by our NMR experiments is centered exactly on this sequence/structure motif, a strong indication that this motif participates to the 7SK identity.

In the present work, NMR experiments and SHAPE probing firmly established the secondary structure of the 24–87 sub-domain of the eukaryotic riboregulator, 7SK. A base pair (A39–U68) predicted from standard folding programs does not form, resulting in a small internal loop. Interestingly, this base pair closes upon binding of the cognate ARM motif of the protein partner HEXIM. The 7SK-conserved sequence GAUC/.../GAUC, forming a short helix between two bulges, was proven to be determinant for HEXIM recognition. We show that the short helix opens when the ARM motif of HEXIM binds. Besides, the bulged U40, U41 and U63 seem to cluster in a structural motif which also changes upon HEXIM binding. This contribution provides a firm step towards understanding which features of 7SK RNA sustain its identity as HEXIM target. Future structural investigations will however be required to further unravel if these determinants are directly recognized by HEXIM or key elements impacting the local structure and dynamic of 7SK, and understand how HEXIM residues trigger the structural rearrangements.

In summary, we demonstrate that HEXIM recognition is not a static process, and, to phrase it as Frankel and Smith (62), 'more than a molecular handshake', picturing cases where both RNA and protein partners change conformation upon binding. Interestingly, co-folding, as an extreme manifestation of such induced-fit process, is found very often in RNA-complexes of viral origin, like Tat-TAR from HIV (63) or BIV (64), Rev-RRE from HIV (65), N peptide and BoxB RNA from lambda or P22 (66) phages [reviewed in (58)]. Moreover, these RNAs share some similarities with the HP1 subdomain of 7SK as they are built on a hairpin scaffold, with short helical stems interspersed with bulges and internal loops. Upon binding, the induced RNA structures generally exhibit novel features such as non-standard base pairs or triples. On the functional level, these viral systems are involved in regulation of transcription elongation, but by other

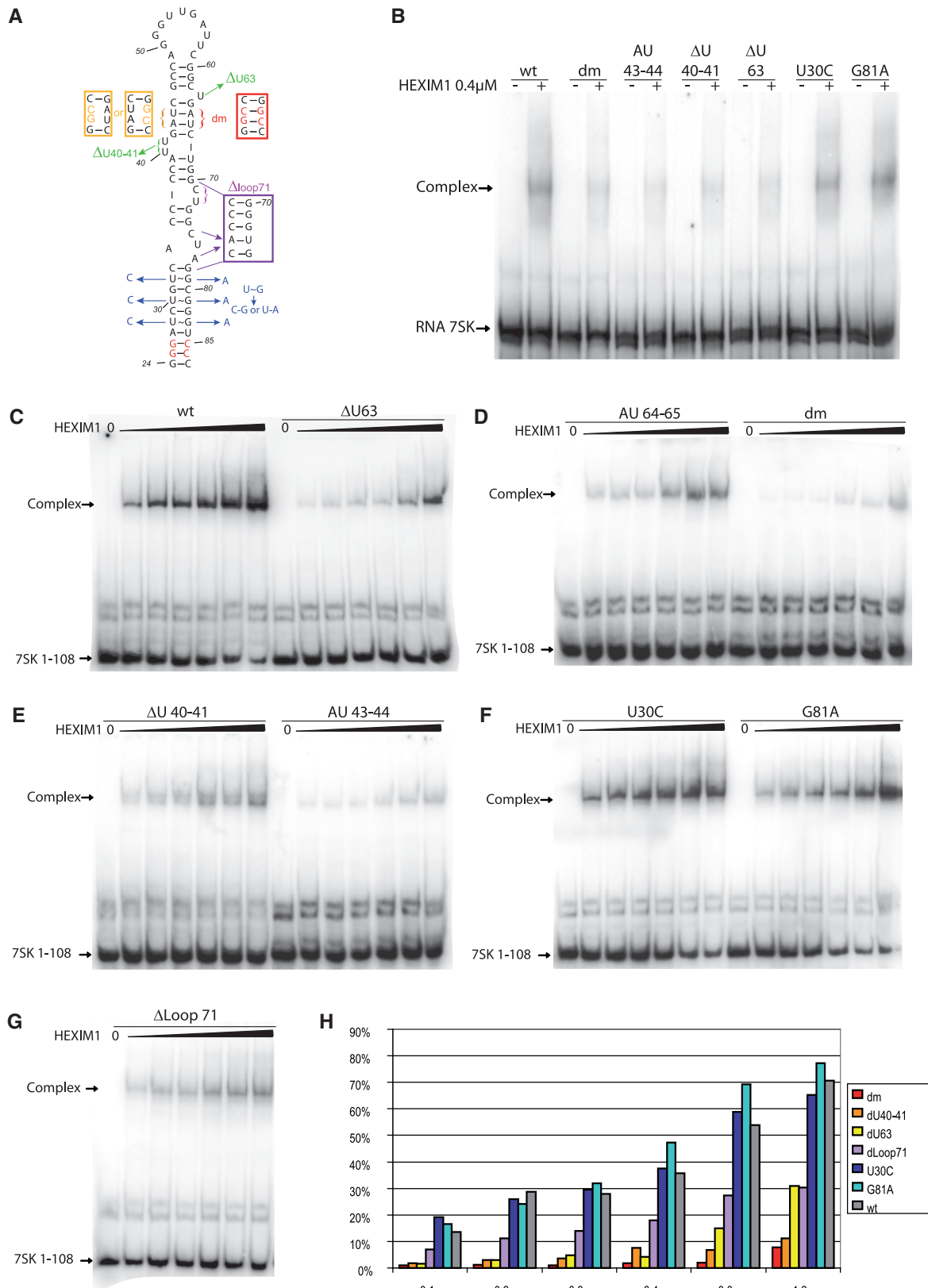


Figure 7. Functional impact of the 7SK mutations on the full-length HEXIM binding. (A) Mutations introduced into the 24–87 region of 7SK. (B) EMSA of several mutated full-length radioactively labeled 7SK at one concentration (0.4 μM) of full-length HEXIM1. (C–G) EMSA of mutated HP1-long (1–108) variants at increasing concentrations (0.1–1.2 μM) of full-length HEXIM1. The concentration of HEXIM1 was maintained low to avoid non-specific binding, a known tendency of HEXIM (70), that leads to the appearance of multiple bands. (H) Recapitulation of full length HEXIM1 binding of all HP1-long mutants. To simplify the figure some mutations were not represented. These are: HP1AU4344 and HP1AU6566 (which behave like HP1dm and $\Delta U63$, respectively), and U28C, G83A, U32C and G79A (which behave like WT).

mechanisms than 7SK (67). Another interesting feature shared by 7SK and these viral complexes, is that the proteins recognition devices are ARM motifs (57). ARM motifs, featuring mainly basic residues, may easily be pictured as positively charged, unstructured ribbons, ready to bind any RNA on the ribose-phosphate chain by pure electrostatic attraction. However, we show here in the HEXIM case that there is nonetheless specificity in the recognition. When, on the RNA side, we change the GAUC sequence or delete the bulges, the HEXIM ARM peptide is deprived of its capacity to bind specifically but not of its random, electrostatic binding. These differences are monitored finely by the NMR approach. On the ARM side, changing the proline or the serine also prevents the specific binding. HEXIM ARM is comprised of two segments, separated by this proline-serine sequence, which we show, is not only important for specificity but participates in the opening of the short helix. Incidentally, this indicates that such a simple organization bears already some functional meaning. HEXIM, as several of the above mentioned viral proteins, is mainly unfolded, apart from the coiled-coil at the C-terminus, far from the RNA-binding region (40). Interestingly, in the reported cases, ARM motifs can fold into various structures, and analogous peptides can employ distinct mechanisms of RNA recognition (68). The subtlety is that the RNA drives the protein folding, but does not do so as a static mould. Further investigations will be required to understand how the local information of recognition is transmitted in a mostly unstructured protein.

The function of 7SK is to inhibit, together with HEXIM, the kinase activity of P-TEFb. To explain why neither the free RNA, nor the free protein is able to perform that inhibition, the generally accepted model relies upon conformational changes. Previous works attributed to the HEXIM partner most of the change, as triggered by release of the acidic patch in the 200–250 region (69) upon 7SK binding. The present work clearly indicates that there are local changes already at the level of the 24–87 sub-domain of 7SK. Do these changes induce larger changes in the rest of the 7SK molecule, or are they locally transmitted from HEXIM ARM towards the farther coiled-coil and P-TEFb-binding domain, is still an open question. Our present results however, show that 7SK is more than a passive key to the regulation, but an active partner of the regulation mechanism.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dino Moras and Olivier Bensaude for their constant support and particularly Olivier for his useful suggestions about the article. The authors thank Jean-Louis Leroy for helpful discussions. They also thank the members of the Structural Biology and Genomics platform, CEGBS, Illkirch. The authors thank Claude Ling for excellent technical assistance at

the NMR spectrometer, Pascal Eberling for synthesizing the peptides and Meiggie Untrau for help with mutagenesis. The authors thank Greame Conn for the pHDV and pRZ plasmids, and Kevin Weeks for generous gift of 1M7, the SHAPE reactant and advice with the SHAPE method.

FUNDING

French National Agency for Research (grant TrscrREGsnRNP ANR-06-BLAN-0072); PEPS 2010 (Projects Exploratoires/Premier Soutien, CNRS); CNRS and INSERM; University of Strasbourg, and the SPINE 2 European Project (Contract N° LSHG-CT-2006-031220); Doctoral fellowship from CONACyT (Mexico, to D.M.Z.). Funding for open access charge: Centre National de la Recherche Scientifique.

Conflict of interest statement. None declared.

REFERENCES

- Zieve,G., Benecke,B.J. and Penman,S. (1977) Synthesis of two classes of small RNA species in vivo and in vitro. *Biochemistry*, **16**, 4520–4525.
- Gurney,T. Jr and Eliceiri,G.L. (1980) Intracellular distribution of low molecular weight RNA species in HeLa cells. *J. Cell Biol.*, **87**, 398–403.
- Murphy,S., Di Liegro,C. and Melli,M. (1987) The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell*, **51**, 81–87.
- Kruger,W. and Benecke,B.J. (1987) Structural and functional analysis of a human 7S K RNA gene. *J. Mol. Biol.*, **195**, 31–41.
- Wassarman,D.A. and Steitz,J.A. (1991) Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol. Cell Biol.*, **11**, 3432–3445.
- Nguyen,V.T., Kiss,T., Michels,A.A. and Bensaude,O. (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature*, **414**, 322–325.
- Yang,Z., Zhu,Q., Luo,K. and Zhou,Q. (2001) The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature*, **414**, 317–322.
- Marshall,N.F. and Price,D.H. (1995) Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.*, **270**, 12335–12338.
- Baumli,S., Lolli,G., Lowe,E.D., Troiani,S., Rusconi,L., Bullock,A.N., Debreczeni,J.E., Knapp,S. and Johnson,L.N. (2008) The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO J.*, **27**, 1907–1918.
- Zhou,Q. and Yik,J.H. (2006) The Yin and Yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. *Microbiol. Mol. Biol. Rev.*, **70**, 646–659.
- Peterlin,B.M. and Price,D.H. (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell*, **23**, 297–305.
- Marshall,N.F., Peng,J., Xie,Z. and Price,D.H. (1996) Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.*, **271**, 27176–27183.
- Michels,A.A., Nguyen,V.T., Fraldi,A., Labas,V., Edwards,M., Bonnet,F., Lania,L. and Bensaude,O. (2003) MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol. Cell Biol.*, **23**, 4859–4869.
- Dulac,C., Michels,A.A., Fraldi,A., Bonnet,F., Nguyen,V.T., Napolitano,G., Lania,L. and Bensaude,O. (2005) Transcription-dependent association of multiple positive transcription elongation factor units to a HEXIM multimer. *J. Biol. Chem.*, **280**, 30619–30629.

15. Blazek,D., Barboric,M., Kohoutek,J., Oven,I. and Peterlin,B.M. (2005) Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. *Nucleic Acids Res.*, **33**, 7000–7010.
16. Yik,J.H., Chen,R., Pezda,A.C. and Zhou,Q. (2005) Compensatory contributions of HEXIM1 and HEXIM2 in maintaining the balance of active and inactive positive transcription elongation factor b complexes for control of transcription. *J. Biol. Chem.*, **280**, 16368–16376.
17. Yang,Z., Yik,J.H., Chen,R., He,N., Jang,M.K., Ozato,K. and Zhou,Q. (2005) Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell*, **19**, 535–545.
18. Yang,Z., He,N. and Zhou,Q. (2008) Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. *Mol. Cell Biol.*, **28**, 967–976.
19. Jeronimo,C., Forget,D., Bouchard,A., Li,Q., Chua,G., Poitras,C., Therien,C., Bergeron,D., Bourassa,S., Greenblatt,J. *et al.* (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell*, **27**, 262–274.
20. Xue,Y., Yang,Z., Chen,R. and Zhou,Q. (2010) A capping-independent function of MePCE in stabilizing 7SK snRNA and facilitating the assembly of 7SK snRNP. *Nucleic Acids Res.*, **38**, 360–369.
21. Krueger,B.J., Jeronimo,C., Roy,B.B., Bouchard,A., Barrandon,C., Byers,S.A., Searcey,C.E., Cooper,J.J., Bensaude,O., Cohen,E.A. *et al.* (2008) LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. *Nucleic Acids Res.*, **36**, 2219–2229.
22. Markert,A., Grimm,M., Martinez,J., Wiesner,J., Meyerhans,A., Meyuhas,O., Sickmann,A. and Fischer,U. (2008) The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. *EMBO Rep.*, **9**, 569–575.
23. He,N., Jahchan,N.S., Hong,E., Li,Q., Bayfield,M.A., Maraia,R.J., Luo,K. and Zhou,Q. (2008) A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. *Mol. Cell*, **29**, 588–599.
24. Barboric,M., Lenasi,T., Chen,H., Johansen,E.B., Guo,S. and Peterlin,B.M. (2009) 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. *Proc. Natl Acad. Sci. USA*, **106**, 7798–7803.
25. Barrandon,C., Bonnet,F., Nguyen,V.T., Labas,V. and Bensaude,O. (2007) The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. *Mol. Cell Biol.*, **27**, 6996–7006.
26. Hogg,J.R. and Collins,K. (2007) RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA*, **13**, 868–880.
27. Van Herreweghe,E., Egloff,S., Goiffon,I., Jady,B.E., Froment,C., Monsarrat,B. and Kiss,T. (2007) Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *EMBO J.*, **26**, 3570–3580.
28. Isel,C. and Karn,J. (1999) Direct evidence that HIV-1 Tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation. *J. Mol. Biol.*, **290**, 929–941.
29. Karn,J. (1999) Tackling Tat. *J. Mol. Biol.*, **293**, 235–254.
30. Dingwall,C., Ernberg,I., Gait,M.J., Green,S.M., Heaphy,S., Karn,J., Lowe,A.D., Singh,M., Skinner,M.A. and Valerio,R. (1989) Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc. Natl Acad. Sci. USA*, **86**, 6925–6929.
31. Berkhout,B., Silverman,R.H. and Jeang,K.T. (1989) Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell*, **59**, 273–282.
32. Kim,Y.K., Bourgeois,C.F., Isel,C., Churcher,M.J. and Karn,J. (2002) Phosphorylation of the RNA polymerase II carboxyl-terminal domain by CDK9 is directly responsible for human immunodeficiency virus type 1 Tat-activated transcriptional elongation. *Mol. Cell Biol.*, **22**, 4622–4637.
33. Muesing,M.A., Smith,D.H. and Capon,D.J. (1987) Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. *Cell*, **48**, 691–701.
34. Gursoy,H.C., Koper,D. and Benecke,B.J. (2000) The vertebrate 7S K RNA separates hagfish (*Myxine glutinosa*) and lamprey (*Lampetra fluviatilis*). *J. Mol. Evol.*, **50**, 456–464.
35. Egloff,S., Van Herreweghe,E. and Kiss,T. (2006) Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. *Mol. Cell Biol.*, **26**, 630–642.
36. Gruber,A.R., Koper-Emde,D., Marz,M., Tafer,H., Bernhart,S., Obernosterer,G., Mosig,A., Hofacker,I.L., Stadler,P.F. and Benecke,B.J. (2008) Invertebrate 7SK snRNAs. *J. Mol. Evol.*, **66**, 107–115.
37. Gruber,A.R., Kilgus,C., Mosig,A., Hofacker,I.L., Hennig,W. and Stadler,P.F. (2008) Arthropod 7SK RNA. *Mol. Biol. Evol.*, **25**, 1923–1930.
38. Marz,M., Donath,A., Verstraete,N., Nguyen,V.T., Stadler,P.F. and Bensaude,O. (2009) Evolution of 7SK RNA and its protein partners in metazoa. *Mol. Biol. Evol.*, **26**, 2821–2830.
39. Yik,J.H., Chen,R., Pezda,A.C., Samford,C.S. and Zhou,Q. (2004) A human immunodeficiency virus type 1 Tat-like arginine-rich RNA-binding domain is essential for HEXIM1 to inhibit RNA polymerase II transcription through 7SK snRNA-mediated inactivation of P-TEFb. *Mol. Cell Biol.*, **24**, 5094–5105.
40. Schulte,A., Czudnochowski,N., Barboric,M., Schonichen,A., Blazek,D., Peterlin,B.M. and Geyer,M. (2005) Identification of a cyclin T-binding domain in Hexim1 and biochemical analysis of its binding competition with HIV-1 Tat. *J. Biol. Chem.*, **280**, 24968–24977.
41. Yik,J.H., Chen,R., Nishimura,R., Jennings,J.L., Link,A.J. and Zhou,Q. (2003) Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol. Cell*, **12**, 971–982.
42. Barboric,M., Kohoutek,J., Price,J.P., Blazek,D., Price,D.H. and Peterlin,B.M. (2005) Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb. *EMBO J.*, **24**, 4291–4303.
43. Dames,S.A., Schonichen,A., Schulte,A., Barboric,M., Peterlin,B.M., Grzesiek,S. and Geyer,M. (2007) Structure of the Cyclin T binding domain of Hexim1 and molecular basis for its recognition of P-TEFb. *Proc. Natl Acad. Sci. USA*, **104**, 14312–14317.
44. Anand,K., Schulte,A., Vogel-Bachmayr,K., Scheffzek,K. and Geyer,M. (2008) Structural insights into the cyclin T1-Tat-TAR RNA transcription activation complex from EIAV. *Nat. Struct. Mol. Biol.*, **15**, 1287–1292.
45. Belanger,F., Baigude,H. and Rana,T.M. (2009) U30 of 7SK RNA forms a specific photo-cross-link with Hexim1 in the context of both a minimal RNA-binding site and a fully reconstituted 7SK/Hexim1/P-TEFb ribonucleoprotein complex. *J. Mol. Biol.*, **386**, 1094–1107.
46. Walker,S.C., Avis,J.M. and Conn,G.L. (2003) General plasmids for producing RNA in vitro transcripts with homogeneous ends. *Nucleic Acids Res.*, **31**, e82.
47. Wyatt,J.R., Chastain,M. and Puglisi,J.D. (1991) Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques*, **11**, 764–769.
48. Plateau,P. and Gueron,M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.*, **104**, 7310–7311.
49. Piotto,M., Saudek,V. and Sklenar,V. (1992) Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions. *J. Biomol. NMR*, **2**, 661–665.
50. Sklenar,V., Piotto,M., Leppik,R. and Saudek,V. (1993) Gradient-tailored water suppression for H-1-N-15 Hsqc experiments optimized to retain full sensitivity. *J. Magnet. Reson. Ser. A*, **102**, 241–245.
51. Farjon,J., Boisbouvier,J., Schanda,P., Pardi,A., Simorre,J.P. and Brutscher,B. (2009) Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. *J. Am. Chem. Soc.*, **131**, 8571–8577.
52. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by

- selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
53. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
 54. Clarkson, J. and Campbell, I.D. (2003) Studies of protein-ligand interactions by NMR. *Biochem. Soc. Trans.*, **31**, 1006–1009.
 55. Leroy, J.L., Gao, X.L., Gueron, M. and Patel, D.J. (1991) Proton-exchange and internal motions in 2 chromomycin dimer DNA oligomer complexes. *Biochemistry*, **30**, 5653–5661.
 56. Michels, A.A., Fraldi, A., Li, Q., Adamson, T.E., Bonnet, F., Nguyen, V.T., Sedore, S.C., Price, J.P., Price, D.H., Lania, L. *et al.* (2004) Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J.*, **23**, 2608–2619.
 57. Weiss, M.A. and Narayana, N. (1998) RNA recognition by arginine-rich peptide motifs. *Biopolymers*, **48**, 167–180.
 58. Patel, D.J. (1999) Adaptive recognition in RNA complexes with peptides and protein modules. *Curr. Opin. Struct. Biol.*, **9**, 74–87.
 59. Weeks, K.M. and Crothers, D.M. (1991) RNA recognition by Tat-derived peptides: interaction in the major groove? *Cell*, **66**, 577–588.
 60. Weeks, K.M. and Crothers, D.M. (1993) Major groove accessibility of RNA. *Science*, **261**, 1574–1577.
 61. Czudnochowski, N., Vollmuth, F., Baumann, S., Vogel-Bachmayr, K. and Geyer, M. (2010) Specificity of Hexim1 and Hexim2 complex formation with cyclin T1/T2, importin alpha and 7SK snRNA. *J. Mol. Biol.*, **395**, 28–41.
 62. Frankel, A.D. and Smith, C.A. (1998) Induced folding in RNA-protein recognition: more than a simple molecular handshake. *Cell*, **92**, 149–151.
 63. Puglisi, J.D., Chen, L., Blanchard, S. and Frankel, A.D. (1995) Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science*, **270**, 1200–1203.
 64. Ye, X., Kumar, R.A. and Patel, D.J. (1995) Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.*, **2**, 827–840.
 65. Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D. and Williamson, J.R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, **273**, 1547–1551.
 66. Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A. and Patel, D.J. (1998) Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. *Nat. Struct. Biol.*, **5**, 203–212.
 67. Weiss, M.A. (1998) RNA-mediated signaling in transcription. *Nat. Struct. Biol.*, **5**, 329–333.
 68. Smith, C.A., Calabro, V. and Frankel, A.D. (2000) An RNA-binding chameleon. *Mol. Cell*, **6**, 1067–1076.
 69. Barboric, M., Kohoutek, J., Price, J.P., Blazek, D., Price, D.H. and Peterlin, B.M. (2005) Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb. *EMBO J.*, **24**, 4291–4303.
 70. Li, Q., Cooper, J.J., Altweger, G.H., Feldkamp, M.D., Shea, M.A. and Price, D.H. (2007) HEXIM1 is a promiscuous double-stranded RNA-binding protein and interacts with RNAs in addition to 7SK in cultured cells. *Nucleic Acids Res.*, **35**, 2503–2512.