Different views of the dynamic landscape covered by the 5'-hairpin of the 7SK small nuclear RNA

Karl Brillet, Denise Martinez-Zapien, Guillaume Bec, Eric Ennifar, Anne Catherine Dock Bregeon, Isabelle Lebars

Different views of the dynamic landscape covered by the 5'-hairpin of the 7SK small nuclear RNA

Karl Brillet[1], Denise Martinez-Zapien[2], Guillaume Bec[1], Eric Ennifar[1], Anne-Catherine Dock-Bregeon[3]* & Isabelle Lebars[1]*

[1] Université de Strasbourg, Architecture & Réactivité de l'ARN - CNRS UPR 9002, Institut de Biologie Moléculaire et Cellulaire, 2 allée Conrad Roentgen, F-67084 Strasbourg, France

[2] Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U964, Université de Strasbourg, 1 rue Laurent Fries, 67404 Illkirch Cedex, France

[3] Laboratory of Integrative Biology of Marine Models (LBI2M), Sorbonne University-CNRS UMR 8227, Station Biologique de Roscoff, Place Georges Teissier, CS90074, 29680 Roscoff Cedex, France

Shortened Title: Conformations of the 5'-hairpin of the 7SK snRNA

Key words: RNA, structure, 7SK, HEXIM, Tat

*To whom correspondence should be addressed:

Dr. Isabelle Lebars: i.lebars@ibmc-cnrs.unistra.fr (Tel.: +33 (0) 3 88 41 70 02)

Dr Anne-Catherine Dock-Bregeon: acdockbregeon@sb-roscoff.fr (Tel.: +33 (0) 2 98 29 23 32)

ABSTRACT

The 7SK small nuclear RNA (7SKsnRNA) plays a key role in the regulation of RNA polymerase II by sequestrating and inhibiting the positive transcription elongation factor b (P-TEFb) in the 7SK ribonucleoprotein complex (7SKsnRNP), a process mediated by interaction with the protein HEXIM. P-TEFb is also an essential cellular factor recruited by the viral protein Tat to ensure the replication of the viral RNA in the infection cycle of the human immunodeficiency virus (HIV-1). Tat promotes the release of P-TEFb from the 7SKsnRNP and subsequent activation of transcription, by displacing HEXIM from the 5'-hairpin of the 7SKsnRNA. This hairpin (HP1), comprising the signature sequence of the 7SKsnRNA, has been the subject of three independent structural studies aiming at identifying the structural features that could drive the recognition by the two proteins, both depending on Arginine Rich Motifs (ARM). Interestingly, four distinct structures were determined. In an attempt to provide a comprehensive view of the structure-function relationship of this versatile RNA, we present here a structural analysis of the models, highlighting how HP1 is able to adopt distinct conformations with significant impact on the compactness of the molecule. Since these models are solved under different conditions by NMR and crystallography, the impact of the buffer composition on the conformational variation was investigated by complementary biophysical approaches. Finally, using Isothermal Titration Calorimetry, we determined the thermodynamic signatures of the Tat-ARM and HEXIM-ARM peptide interactions with the RNA, showing that they are associated with distinct binding mechanisms.

INTRODUCTION

In human and other vertebrates, the 7SK RNA is an abundant small nuclear RNA (snRNA), acting as the scaffold of the ribonucleoprotein complex 7SKsnRNP (Diribarne and Bensaude 2009; Brogie and Price 2017; He et al. 2008; Xue et al. 2010), which includes the La-related protein LaRP7 and the methylphosphate capping enzyme MePCE. These proteins ensure the stability of a core 7SKsnRNP and promote further assembly of larger complexes (Xue et al. 2010; Jeronimo et al. 2007; Krueger et al. 2008; Markert et al. 2008). In particular, a complex with the HEXIM protein (HEXIM1 or HEXIM2 in human, with similar properties) recruits the positive transcription elongation factor b (P-TEFb) (Kobbi et al. 2016; Michels et al. 2004; Michels and Bensaude 2018a; Bigalke et al. 2011). P-TEFb, a heterodimer comprising the cyclin T1 and the cyclin-dependent kinase CDK9, plays an essential role in regulating the transcriptional activity of RNA polymerase II (RNApol II) in eukaryotes (Zhou and Yik 2006; Peterlin and Price 2006). Capture by the 7SKsnRNP inhibits P-TEFb kinase activity and prevents the phosphorylation of RNApol II and pausing factors. This reversible sequestration depends on the protein HEXIM, which binds the 7SK snRNA through its Arginine Rich Motif (ARM) (Dames et al. 2007; Michels and Bensaude 2018b; Nguyen et al. 2001; Schönichen et al. 2010; Schulte et al. 2005; Yang et al. 2001; Yik et al. 2005). P-TEFb is also an essential cellular factor for the regulation of the human immunodeficiency virus (HIV-1) replication and is hijacked to maximize HIV transcription (Mbonye and Karn 2017). The protein Tat binds the cyclin T1 using its activation domain, while its RNA-binding domain associates with the *trans*-activating responsive RNA element (TAR) located at the 5'-untranslated region of the viral genome (Muesing et al. 1987; Marciniak et al. 1990; Mancebo et al. 1997; Isel and Karn 1999; Karn 1999; Berkhout and van Wamel 2000). Tat has been shown to efficiently displace HEXIM1 from the 7SK RNA, resulting in the release of P-TEFb (Muniz et al. 2010; Rice 2017; Barboric et al. 2007). Early reports show that the ARM within the central region of HEXIM (residues 149-165 residues, in human HEXIM1) is fully functional *in vivo* and *in vitro* for 7SK binding and highlights the sequence similarity between the Tat and HEXIM peptides (Yik et al. 2004). The HEXIM ARM is bipartite, and comprises two stretches of arginines and lysines, separated by a proline (P) - serine (S) insertion in the middle. The first stretch

resembles closely the Tat peptide and ends at the proline residue. The second part contains a conserved tryptophan residue (Yik et al. 2004). Several studies demonstrate that HEXIM and Tat bind to the conserved 5'-terminal hairpin of the 7SK RNA, with the (24-87) region (named HP1 hereafter) being the minimal domain required for the interaction (Egloff et al. 2006; Belanger et al. 2009; Martinez-Zapien et al. 2015). Both HEXIM and Tat recognition sites comprise the signature of the 7SK RNA, a GAUC repeat framed by single stranded uridines (Marz et al. 2009). However, the nucleotides targeted by HEXIM1 and Tat are not strictly the same, suggesting distinct binding modes (Muniz et al. 2010; Martinez-Zapien et al. 2017; Bourbigot et al. 2016; Lebars et al. 2010). Recently, a structural investigation of the Tat-ARM motif binding to HP1 describes four pre-formed motifs in the free RNA promoting the penetration of the peptide meanders into a full turn of the major groove and folds back to form a hairpin, with a minimal conformational change upon recognition (Pham et al. 2018).

Aiming at identifying the structural features that could drive the recognition of HP1 by HEXIM1 and Tat, we previously determined the three dimensional structure of the HP1 RNA (with a UUCG apical tetraloop) using Nuclear Magnetic Resonance (NMR) and X-ray Crystallography (Bourbigot et al. 2016; Martinez-Zapien et al. 2017). Interestingly, the solution and crystal structures differ. Subsequently, an independent group solved a structure of the same region but with a different (GAGA) apical tetraloop by NMR (Pham et al. 2018).

In this work, we first provide a detailed analysis of the structural differences between the four models. Since the various structures were solved from very different environments in terms of salt content and pH, we used complementary biophysical approaches to investigate how the buffer conditions could lead to the observed variations. Finally, using Isothermal Titration Calorimetry (ITC), we determined the thermodynamic binding parameters of the Tat-ARM and HEXIM1-ARM peptides to the RNA in order to obtain a comprehensive view of the driving forces involved in the formation of the complexes. The present manuscript thus summarizes the results obtained while also providing a synthetic view of our present knowledge about this particularly versatile RNA.

RESULTS

Distinct conformations observed at high resolution illustrate the high flexibility of the RNA HP1-tetraloop molecules

Comparison of the four structures

Three independent structural studies of the 5'-terminal hairpin of the human 7SK RNA have been conducted using two different RNA constructs: one construct with a thermo-stable UNCG tetraloop (HP1-UUCG) and the second construct with a GNRA tetraloop (HP1-GAGA), both replacing the 11-nt apical loop of the HP1 hairpin. The secondary structure of HP1 comprises a long stem and several bulges (Fig. 1). The 7SK signature, a short helix formed by the repeat $(GAUC)_2$ framed by the single stranded uridines U40-U41 on one side and U63 on the other, is located in the apical part of the molecule (Fig. 1). Four distinct structures of HP1 have been observed (Fig. 1). Crystallographic determination from three different crystals (PDB Id: 5LYS, 5LYU, 5LYV) revealed that two major conformations of HP1-UUCG coexist in a single crystal (Fig. 1A) (Martinez-Zapien et al. 2017). They mainly differ at the position of the uridines framing the $(GAUC)_2$ sequence. A completely different, more extended, conformation of HP1-UUCG was determined simultaneously by NMR (Fig. 1B, PDB Id: 5IEM) (Bourbigot et al. 2016). More recently, a NMR study described a conformation of the HP1-GAGA version (Fig. 1C, PDB Id: 6MCI), which is compact and closer to the crystal structures than that of 5IEM conformation (Pham et al. 2018).

The four models are similar at the level of the helices and the apical loops, as well as for the U72 and U76 residues, which are swung outside the helix. They differ essentially at the predicted bulges (Fig. 1A, 1B). In the crystal structures, the stems are stabilized by coaxial stacking interactions, which extrude the single stranded residues. Most nest into the major groove forming base triples with base pairs often remarkably far in sequence. This leads to straight and compact structures, in contrast to the loosely packed NMR structures and results in a large difference of approximately 17 Å in length between the crystal structure (about 66 Å) and the extended NMR (about 83 Å) models. Alternative structures are observed at the internal loop in the central part of the molecule (Fig. 1 and Supplemental

Table S1). An interesting difference lies at the level of the U68 residue, a residue located just below the 7SK signature. In the extended conformation (5IEM), U68 is observed transiently involved in two possible hydrogen bonds with A39 and U40, in contrast with the Hoogsteen base pair formed in the compact solution structure (6MCI) and the Watson-Crick base pair formed with A39 in the crystal structures (Fig. S1C). A major variation was observed at the 7SK signature. Two different conformers are observed for U40 and U41 residues (PDB Id: 5LYS, 5LYV) within the same crystal. In the OUT conformation, U41 is flipped outwards to the solvent while U40 is involved in a base triple with the A43:U66 base pair of the $(GAUC)_2$ helix (Fig. 1A, Fig. S1B). In the IN conformation, both uridines are positioned within the major groove of the $(GAUC)_2$ helix, in which case U41 is involved in the triple with A43:U66, and U40 displaced towards the apex, making a base triple with the C45:G64 base pair. The OUT conformation is observed for 80% of the total conformers obtained from native, osmium or gold crystals, each comprising two molecules in the asymmetric unit and was shown to be more stable than the IN conformation in a molecular dynamics study (Röder et al. 2020). By contrast, in the extended NMR structure (5IEM), experimental data didn't show evidence for nucleobase flipping between U40 and U41 residues. U41 lies in the major groove and in proximity of A43-U66 and G42-C67 Watson-Crick base pairs, without establishing specific interactions (Fig. 1B, Fig. S1B). The recent NMR analysis performed with the HP1-GAGA construct (6MCI) shows a conformation similar with OUT, with U41 turned outwards, and U40 in a triple with A43:U66 (Fig. 1C). While this structure is compact, with a comparable length as in the crystal structures (66 Å), it shows alternative positions for several of the bulged residues. While C71, C75 and U63 make the same triples as in the crystal structures (Fig. 1), A34 and A77 residues are stacked inside, as in the extended NMR structure (5IEM).

Analysis of the models in light of NMR data

These differences led us to examine the compatibility of the models (IN and OUT conformations, and 6MCI) with our NMR data recorded with HP1-UUCG. Our reported assignment of HP1-UUCG was based on the analysis of a large set of homonuclear and heteronuclear experiments recorded on unlabeled and $^{15}N/^{13}C$ fully, segmentally or selectively labeled RNA constructs

(Bourbigot et al. 2016). The nucleotide specific labeling allowed the observation of NOEs resulting from the $^{15}N/^{13}C$ nucleotide alone or arising from the unlabeled other three nucleotides by recording 2D ^{13}C- edited/filtered NOESY, respectively. In addition, 3D heteronuclear experiments (^{1}H, ^{15}N, ^{13}C, ^{31}P) granted the determination of NMR restraints for each nucleotide. In particular, assignment of U40 and U41 residues stands on the analysis of 3D-HCP experiments (Marino et al. 1995) combined to ^{13}C-filtered/edited NOESY that were recorded on $^{13}C/^{15}N$ - G or - U or - A or - C selectively labeled RNA. These set of experiments provided unambiguous assignment and restraints for structure calculation.

First, to investigate the compatibility of the structures with the NMR data from HP1-UUCG (5IEM), we selected a series of protons as landmarks for specific conformations (stacking with adjacent bases or interaction with a nucleotide of the other chain). Then corresponding distances were calculated in crystal structures (HP1-UUCG: PDB Id 5 LYS, 5LYU, 5LYV) and NMR (HP1-GAGA: 6MCI) structure and we checked whether these could be observed in NOESY experiments (Supplemental Table S2). As NOE cross-peaks correlate protons within a distance of 5 Å, no peak can be observed when protons are more than 5 Å apart. For example, at the level of the U40-U41 bulge, a NOE involving U40-H6 and A39-H1' indicates that U40 is unpaired and stacked on residue A39 (Bourbigot et al. 2016). In the crystal structure, the corresponding distances are respectively 9.7 ± 0.1 Å and 8.0 ± 0.1 Å in IN and OUT conformations, which are too long to observe NOE correlations (Supplemental Table S2). On the other hand, the U41 imino proton exhibits weak NOE cross-peaks with both U66-H3 and C67-H5 protons in HP1-UUCG, which are not explained by the other three models (Bourbigot et al. 2016). Similar results are observable for A34, A77 and U63 (Table S2). Nucleotide C71, at the (C71-U72) bulge, shows the only example where distances (G73-H8 and C71-H1') are consistent with the observation of a NOE correlation in all structures. However, our NMR data didn't provide evidences of the involvement of C71 in a triple with G74:C35 as observed in crystals and HP1-GAGA structures.

Secondly, we compared all models with the measured one-bound N-H and C-H residual dipolar couplings (RDCs) from HP1-UUCG (5IEM) (Bourbigot et al. 2016). RDC data orient inter-nuclear vector with respect to the axis of the molecule and thus, provide long-range global restraints

improving the accuracy of the structure. We observe that the RDCs calculated from crystal and HP1-GAGA structures clearly deviate from the measured RDCs (Supplemental Table S3, Fig. 2). As RDCs are sensitive to both local environment and global fold, this indicates that the crystal and 6MCI structures are not observed under the conditions used in our NMR study. Undoubtedly, the NMR interpretation depends on the detection of cross-peaks with line shape and line width, which are directly impacted by the dynamic and the concentration of one or more conformations. The determination of models does not fully take into account the presence of different conformations with low percentages. Thus, we cannot exclude that the molecule HP1-UUCG (and most probably HP1-GAGA) fluctuates between different structures, not detectable by NMR.

The four experimental models were obtained from very different buffer conditions, raising questions about the impact of pH and magnesium concentrations on the major conformation in a given solution. In the following sections, we present investigations of the influence of the buffer composition (pH and Mg^{2+}) on the structure and recognition by Tat and HEXIM1 peptides.

Impact of the buffer composition on the HP1-UUCG structures

UV melting experiments were carried out to examine the thermal stability of HP1-UUCG at various pH values (5.2, 6.4 and 7.5) that were used for the different structural studies. HP1-UUCG exhibits a similar melting temperature (*Tm*) at all tested pH, suggesting that there is no significant stabilization resulting from the variation of pH by itself (Table 1 and Fig. S2). We next investigated the influence of the magnesium concentration, an ion well known to impact the stability and folding of RNA structures. Thermal denaturation experiments were recorded at pH 5.2, 6.4 and 7.5 with increasing magnesium concentrations in the range of 0.1 to 6 mM in a buffer containing 50 mM sodium phosphate (Fig. S2 and Table 1). At each pH, a clear increase of *Tm* is observed upon increasing magnesium concentration, indicating an enhancement in the thermal stability of the RNA (Table 1, Fig. S2). There is also a global decrease of *Tm* upon increasing pH (Fig. S2C), suggesting a combined effect of pH and magnesium on the stability of the RNA.

The Mg^{2+}-induced effects were previously analyzed by monitoring the imino proton region in the spectrum of HP1 (Bourbigot et al. 2016). As recalled in Supplemental Figure S3 and Fig. 3, only few changes were observable at low magnesium concentration (up to 10 mM), and in particular the stabilization of the base pairs involving G69, G70, G73 and G74 residues located in the central region. Further structural analysis based on NMR bidimensional experiments recorded at 3 mM Mg^{2+} showed a bending of the structure and the shift of residue U41 closer to the A43:U66 base pair. In the present study, the titration by monitoring of imino protons was continued until a concentration of 60 mM Mg^{2+} was reached, in order to attain the high magnesium concentration that was present in the crystallization drop (initially 50 mM Mg^{2+}). No significant changes were observed beyond a concentration of 10 mM Mg^{2+} (Fig. 3 and Fig. S3), at which the chemical shifts perturbation reaches a plateau. However, additional NOESY experiments recorded at 6 mM Mg^{2+} concentration show the disappearance of cross-peaks between U66-H3 and U41-H3 on one hand, and U40-H3 and G42-H1 on the other hand suggesting a further displacement of U40 and U41 residues (Fig. 3). Moreover, the concomitant appearance of a weak cross-peak between G70-H1 and G73-H1 indicates a stacking between residues G70 and G73 that was not observed at low magnesium concentration. The initiation of a structural compaction is thus observed at higher magnesium concentrations. Due to spectral broadening at concentrations above 6 mM, changes could not be followed by bidimensional NMR.

To overcome these limitations, we turned to small-angle X-ray solution scattering (SAXS), another independent structural method to observe molecules in solution that provides only low resolution information, but allows to work in various buffers. Concentrated solutions of the HP1-UUCG were studied by SEC-SAXS, where the diffusion data are recorded at the exit of a size exclusion chromatographic column, thus reducing aggregation. SAXS diffusion experiments were performed in three different buffers: (i) buffer "X", as close as possible to the crystallization condition (Martinez-Zapien et al. 2017) (50 mM TRIS pH 7.5, 50 mM NaCl, 50 mM $MgCl_2$) (ii) buffer "NMR", used in the original structure determination (Bourbigot et al. 2016) (10 mM sodium phosphate at pH 6.5) and (iii) buffer "SAXS" (10 mM sodium cacodylate at pH 6.5, 6 mM $MgCl_2$, 0.25 mM EDTA, 100 mM NaCl), which in earlier SAXS investigations of the 7SK RNA was found to provide the most reproducible and aggregation-free data (A.C. Dock-Bregeon, unpublished).

The SAXS diffusion curves obtained in the different buffers are similar and show only small differences. In particular, the scattering profiles obtained in the "X" and "SAXS" buffers, which contain magnesium, are very close (Fig. 4, red and yellow curves, respectively). Interestingly, the scattering profile from the solution in the "NMR" buffer, without magnesium, shows higher intensity at the smallest angles compared to the profile measured in "X" or "SAXS" buffers (Fig. 4, green curves above the others), and very close to the theoretical –blue- curve calculated with 5IEM coordinates (panel A). Indeed, the radius of gyration (R_g) deduced from the Guinier analysis is larger for the molecule in this Mg-free buffer (24.20 ± 0.03 Å), as compared with Rg in the case of the X buffer (23.70 ± 0.02 Å) or the SAXS buffer (22.90 ± 0.02 Å) (Fig. S4). These observations sustain that the magnesium tends to compact the HP1-UUCG structure. The Kratky plots (Fig. S4) indicate that the RNA is correctly folded in all three buffers. However, for the Mg-free buffer, the curve shows a shallower peak followed by a less strongly sloped tail as compared to the curves obtained with the Mg-containing buffers, which indicates that the conformation is more disordered (Plumridge et al. 2018).

The experimental curves acquired in all three buffers were compared to the theoretical diffusion curves calculated using CRYSOL (Svergun et al. 1995) from three models: the crystal structure OUT (Martinez-Zapien et al. 2017), which gives a very similar curve as the IN conformation (not shown), the NMR structure (5IEM, Bourbigot et al. 2016) and the structure determined with the GAGA tetraloop (6MCI, Pham et al. 2018). As shown in Figure 4 (lower panels) the model that shows the best fit between calculated and experimental scattering profiles is the extended NMR model HP1-UUCG (Fig. 4A), while larger differences are observed with the model HP1-GAGA.

These results altogether highlight that HP1 is able to adopt several distinct conformations, the more compact being associated with magnesium ions. Whether these conformations are related to its function, especially to recognition by Tat or HEXIM1, is still an issue. We took the opportunity of this study evidencing buffer effects to question the binding of the peptides in the diverse pH conditions.

Investigation of binding of HEXIM1 and Tat ARM domains to HP1-UUCG by ITC

In order to determine the physical basis of HP1-UUCG binding by Tat and HEXIM1, we performed Isothermal Titration Calorimetry (ITC) experiments at various pH (5.2, 6.4 and 7.5). ITC is a widely used approach to directly study molecular interactions, and which provides insight into binding mechanisms.

To investigate the binding of ARM-HEXIM1 to HP1, we used a 17 aa peptide encompassing the HEXIM1-ARM bipartite sequence corresponding to residues 149-165 of the human protein with the proline (P) and serine (S) insertion of sequence GKKKHRRRPSKKKRHWK (Lebars et al. 2010). To analyze Tat binding, we used a peptide comprising the RNA binding motif (8 aa) and downstream amino acids (48-67) from the Tat sequence, GRKKRRQRRRPSQGGQTHQD, designed such as to perform ITC experiments with peptides of similar size.

Salt-Dependence of ARM-peptides binding to HP1

As ARM motifs contain a high density of positively charged lysines and arginines at neutral pH that could significantly contribute to nonspecific binding through electrostatic interactions with the RNA backbone, we examined the influence of salt on ARM/RNA interactions. As shown in Supplemental Figure S5, the data obtained at low range of salt concentrations (0.1 and 0.2 M NaCl) could not be fitted to a binding isotherm. This indicates a large contribution of nonspecific binding. The nonspecific binding decreases as the salt concentration increases and becomes negligible at 0.5 M NaCl. Similar observations have been reported for Rev-ARM/RNA complexes (Jayaraman et al. 2015). Consequently, all experiments were conducted in the presence of 0.5 M NaCl at pH 5.2, 6.4 and 7.5.

ITC reveals different binding mechanisms for Tat-ARM and HEXIM1-ARM to HP1

We first carried out ITC experiments by titrating the HP1-UUCG RNA by HEXIM1-ARM peptide on one hand, and by Tat-ARM peptide on the other hand at pH 6.4 (NMR conditions supplemented with 0.5 M NaCl). Both interactions display similar dissociation constants (K_d): 23.1 ± 0.6 µM with

HEXIM1-ARM and 21.9 ± 0.7 µM with Tat-ARM. However, the associated thermodynamic parameters show distinct features (Table 2, Fig. 5B and Fig. S6, middle). The Tat-ARM binding is characterized by a favorable enthalpy change (ΔH: -19.9 ± 0.4 kcal.mol^{-1}) and an unfavorable entropy change (-TΔS: 13.6 ± 0.4 kcal.mol^{-1}), while the HEXIM1-ARM binding is driven by a smaller negative ΔH (-6.3 ± 0.1 kcal.mol^{-1}) and negligible entropy change (-TΔS ≈ 0). The favorable enthalpy (ΔH<0) observed for both interactions indicates that the binding is driven by the formation of hydrogen and other non-covalent bonds. By contrast, the entropic difference suggests a loss of conformational freedom (-TΔS > 0) only for the Tat-ARM binding. Our previous NMR analysis showed that both peptides induce a stabilization of G69:C38 and G70:C37 base pairs and the closure of the base pair A39:U68, not formed in the free RNA (Lebars et al. 2010; Bourbigot et al. 2016). The appearance of new hydrogen bonds and base pairing correlates well with the observed favorable enthalpy upon peptide binding. A contrasting difference was that HEXIM1-ARM promoted an additional opening of two base pairs within the $(GAUC)_2$ helix (Lebars et al. 2010). The entropy value close to zero observed for HEXIM1-ARM binding could thus be explained by a compensation between the loss of conformational freedom common to the two peptides and the increased disorder associated with the unwinding at $(GAUC)_2$ induced only by the HEXIM1-ARM.

Next, we performed the same ITC experiments at pH 5.2 (Fig. 5A and Fig. S6, left). Upon HEXIM1-ARM binding, the dissociation constant (K_d) decreases to 11.6 ± 0.6 µM with favorable enthalpic and entropic contributions. The enthalpy value is similar to that observed at pH 6.4 (-4.0 ± 0.2 kcal.mol^{-1}) while the entropy is now negative (-2.8 ± 0.3 kcal.mol^{-1}) indicative of hydrophobic and stacking contributions. On the other hand, the titration conducted at pH 5.2 with the Tat-ARM peptide shows the same profile as at pH 6.4, characterized by an enthalpy-driven binding reaction with a dissociation constant of 22.5 ± 0.6 µM and ΔH and (-TΔS) values of -22.5 ± 0.6 and 16.0 ± 0.6 kcal.mol^{-1}, respectively. Thus, in contrast to HEXIM1-ARM, there is a minimal influence of the pH on Tat-ARM binding to the RNA. Finally, at pH 7.5 the thermodynamic signatures are similar as observed at pH 6.4, but with a globally weaker binding as demonstrated by a higher dissociation constant (Fig. 5C and Fig. S6).

We next performed the same experiments with the HP1-GAGA construct (Fig. S7). Similar results as those with HP1-UUCG were obtained for the Tat-ARM but they were different with the HEXIM1-ARM peptide. At pH 6.4, the favorable enthalpy is now accompanied by a weak, but clear, favorable entropy (Fig. S7). At pH 5.2, surprisingly, the data could only be fitted with a two sites binding model (see discussion).

In conclusion, our ITC experiments highlight different behaviors when Tat-ARM or HEXIM1-ARM bind to the HP1-tetraloop. Interestingly, Tat-ARM recognizes HP1 similarly at all tested pH, for both constructs, HP1-UUCG and -GAGA, with an enthalpy-driven reaction overwhelming the entropic cost. By contrast, the thermodynamics of the binding mode of HEXIM1-ARM depend on the pH and the nature of the construct.

DISCUSSION

NMR and Crystallography provide complementary information. On one side, NMR structure determination based mostly on short averaged proton-proton distances (<5 Å) leads to structural fluctuations. Nowadays, the accuracy of the solution structures has been improved with complementary long-range information (Hansen et al. 1998; Getz et al. 2007). On the other side, the crystal structures are directly deduced from the electron density, but are sensitive to the packing forces. These can nevertheless lead to observation of transient conformations, which are frozen in the crystal when they favor stable packing (Dock-Bregeon et al. 1988).

In the study of HP1-tetraloop, these techniques revealed the ability of a RNA molecule to adopt (at least) four conformations differing by their compactness and the local structures at the predicted bulges (Bourbigot et al. 2016; Martinez-Zapien et al. 2017; Pham et al. 2018). Crystal structures enabled to observe stabilized conformations in which helix-helix interactions are maximized by extruding most of the bulged nucleotides towards the major groove or the solvent. The equilibrium between these conformations was recently investigated by molecular dynamics techniques (Röder et al. 2020). This study demonstrated their coexistence at different relative abundances: a majority in the OUT-conformation (in the range of 64-69%), a minimal amount in the IN-conformation (0 to 6%) and a proportion of the extended conformation (5IEM) of about 25-29%, depending on the calculation strategy. The transition between these conformations may be important for HEXIM binding, since several point mutations leading to a loss of HEXIM-binding change the population of conformers in the ensemble (Röder et al. 2020), thus pointing towards a binding mechanism based on conformationnal selection. In the present work, we investigated the buffer conditions for their global impact on the structural landscape of the HP1-tetraloop.

Impact of pH and magnesium concentration

The HP1-UUCG solution structure observed at pH 6.4 in the absence of magnesium is strikingly more extended than the others, which all comprise protonated cytidines C71 and C75. In the HP1-UUCG crystal structures these residues form base triples, which contribute to lock the central region into a

tight architecture together with another triple involving A77 (Martinez-Zapien et al. 2017). This suggests that the compaction of the RNA originates from the protonation, thus the acidic pH. On the other hand, the SAXS experiment shows that the compaction is also favored by the presence of magnesium. Our NMR investigations highlight stabilization by Mg^{2+} ions for several base pairs in the central region (Fig. 3), including two which were involved in base triples with residues from the internal loop (A34 with G69-C38, C71 with C35-G74) in the crystal structures. Moreover, the crystal structures show a magnesium ion bound to the protonated C71 and to A34, which connects the central region to the upper part of HP1. On the whole, the emerging picture is that the compaction is driven by protonation and the resulting formation of base triples involving several internal loop residues (Fig. 1), but is also further stabilized by magnesium.

Mg^{2+}-induced effects were already reported in a study on the full-length 7SK RNA (Brogie and Price 2017; Krueger et al. 2010). The 5'-region of the human 7SK RNA contains a third GAUC sequence, which has been suggested to be involved in a "GAUC switch" with alternative base pairing with one GAUC of the HP1 domain and depending on magnesium concentration (Merino et al. 2005; Brogie and Price 2017). Since this requires melting of the $(GAUC)_2$ base pairs of HP1, the extended structure observed in solution could reflect a step within this process.

Conformational changes upon magnesium addition leading to distinct NMR and crystal structures have been described for other RNAs (Wu and Tinoco 1998). The ability of HP1 to switch like a spring oscillating between extended (neutral pH, low Mg) and compact (acidic pH, high Mg) forms could be one key feature of its biological function as regulator.

Binding to peptides

Formation of RNA complexes often involves conformational adaptation of partners, termed "induced-fit" (Williamson 2000), as described in several RNA-ARM complexes such as HIV Tat-TAR (Puglisi et al. 1992; Aboul-ela et al. 1995), HIV Rev-RRE (Battiste et al. 1996) or BIV Tat-TAR (Ye et al. 1995). Peptides corresponding to ARMs, often found in disordered regions of proteins, have been proven to be excellent models for characterizing fine specific interactions with their RNA targets (Smith et al. 2000). Here, we analyzed the binding of Tat-ARM and HEXIM1-ARM to HP1-UUCG

and HP1-GAGA using ITC. Our results highlight thermodynamic features compatible with induced-fit mechanisms for both peptides, and in-line with previous observations by NMR. In particular, the thermodynamic signatures for HEXIM1-ARM binding is compatible with the additional opening of two base pairs within the $(GAUC)_2$ helix (Lebars et al. 2010), not observed for Tat-ARM (Fig. S6). With the Tat-ARM peptide chosen for our studies, the thermodynamic signatures indicate an induced fit mechanism. This is contrasting with the interaction proposed by Pham *et al.*, which showed the Tat peptide, being folded as a hairpin, nesting in an open major groove (Fig. 6), and where most residues are in the same conformations as in the free RNA, except nucleotide A39. The difference may be ascribed to our extended version of the Tat-ARM peptide, which includes downstream residues with different physicochemical properties.

Detailed analysis of the ITC data reveals some differences in the peptides behavior. One relates to pH. While Tat-ARM behaves similarly at all pH, HEXIM1-ARM binding is different at acidic pH 5.2 as compared to pH 6.4 and 7.5. Moreover, at pH 5.2, HEXIM1-ARM binding shows an additional favorable entropic contribution, that suggests additional stacking or hydrophobic interactions. This observation tallies with our previous studies showing that Tat-ARM and HEXIM1-ARM peptides target globally the same site within HP1, but with an extent of the perturbation (as represented by the number of affected nucleotides) larger for HEXIM1. In particular, these include changes at the base pairs just below the apical loop.

A puzzling difference appears for the two RNA constructs. At pH 5.2, binding with HEXIM1-ARM is clearly different for HP1-UUCG (one site, favorable entropy and enthalpy, K_D ~12 μM) compared with HP1-GAGA, where a two-sites binding model is required to fit the experimental curve (Fig. S7). For Tat-ARM, the thermodynamic signatures are similar for HP1-UUCG and –GAGA. This suggests that the tetraloop sequence/structure influences HEXIM binding. The comparison of the GAGA *vs* UUCG HP1 structures does not reveal major differences, and contribution of UNCG and GNRA tetraloops to stem-loop stabilities have been shown to be similar (Sheehy et al. 2010 and our data). However, this does not exclude fine-tuning of the structural dynamics of the closing base pairs at the small helix, between the apical loop and the 7SK-motif. Interestingly, earlier NMR footprinting studies showed perturbation of this region upon HEXIM1-ARM peptide binding (Lebars et al. 2010).

The pH effect could, at least partly, be due to the histidine protonation, but since both peptides contain histidines, the major effect is probably due to the protonation of cytidines C71 and C75, which, as discussed above, has strong structural impact. Interestingly, the fit to a two-sites binding model suggests that the protonated structure offers additional possibilities for the HEXIM-ARM binding process, as compared with Tat-ARM.

Interestingly, most of the nucleotides, that we identified to be involved in the induced-fit mechanism, in particular U40, U41 and U63, undergo the largest changes within all models. The original structural description highlighted that the IN and OUT conformations (5LYS, 5LYU, 5LYV) have direct impact on the accessibility of the major groove in the apical region, suggesting a mechanism regulated by conformational changes of "gates" such as U41, U40 and U63 (Fig. 6B, 6D) (Martinez-Zapien et al. 2017; Röder et al. 2020). A structural investigation of the HEXIM1-ARM interaction with its bipartite sequence may help clarify the mechanism, but future investigations should take in account that HEXIM exists as a dimer. Moreover, HP1 is part of a 331-nt RNA composed of four domains. The second ARM binding site, hypothesized to be located in the HP1 hairpin (Martinez-Zapien et al. 2015), was initially localized outside of the domain (Muniz et al. 2010). The rest of the RNA possibly influences the selection of the "good" conformation of HP1, relatively to its various biological functions (Brogie and Price 2017). It cannot be excluded that bulged residues (among U41, U72 and U76), which were trapped in the grooves of symmetrical molecules in the monoclinic crystals, may be stabilized within the large 7SK molecule. The 7SK RNA comprises, in particular, the conserved 3'-hairpin HP4, which was also implicated in P-TEFb regulation (Egloff et al. 2006), and was shown to adopt a conformation similar to TAR (Durney and D'Souza 2010). In our opinion, the four conformations adopted by the 5'-terminal hairpin of 7SK could correspond to different states adapted to association with and dissociation from different partners. The influence of other proteins, such as LaRP7, required for 7SK stability cannot be excluded.

MATERIALS AND METHODS

Structures analysis and visualization

Structures were visualized and analyzed with MacPymol2 (PyMOL Molecular Graphics System, Version 1.2 Schrödinger, LLC) and 3DNA software packages (Colasanti et al. 2013). For distances measurements, hydrogen atoms were added to crystal structures using MOLPROBITY (Williams et al. 2018). Root-mean-square deviation (RMS) between experimental RDC values and RDCs calculated from NMR and X-ray structures were calculated with the PALES software (Zweckstetter et al. 2004).

RNA samples preparation

Milligram quantities of RNAs were prepared by *in vitro* transcription using in-house T7 RNA polymerase from oligonucleotide templates and from linearized plasmids (Milligan and Uhlenbeck 1989). Plasmids were obtained by cloning the HP1 sequence into pRZ or pHDV encoding for a hammerhead ribozyme in 3' of the target RNA (Walker et al. 2003). The DNA templates were purchased from IDT (Belgium). Transcription conditions were optimized according to previous protocols (Milligan and Uhlenbeck 1989; Wyatt et al. 1991).

All RNAs were purified on 8M urea denaturing polyacrylamide gels (Wyatt et al. 1991). After electroelution and ethanol precipitation, RNAs were dialyzed against water. Samples were dried using a SpeedVac and resuspended at a chosen concentration in buffers used for ITC and UV melting experiments. Each sample was refolded by heating at 95°C (2 min) and snap-cooled at 4°C. The concentration of RNA samples was measured using a Nanodrop Spectrometer and calculated with molar extinction coefficients.

UV melting experiments

Thermal denaturation of HP1-UUCG was monitored on a CARY3500 UV/vis spectrophotometer (Agilent) equipped with a 8-positions sample holder and a Peltier temperature control accessory. The experiments were performed at 1 μM final concentration in 50 mM sodium phosphate buffer, at pH

5.2, 6.4 and 7.5 at 20°C, in 100 µL micro quartz cuvettes. The RNA was refolded as described above and the magnesium was next added in order to have a monomeric form of HP1 as we previously determined. For each pH, magnesium was added at concentrations ranging from 0.1 to 6 mM. A cuvette that contained the buffer with no magnesium was used as a reference. Samples were overlaid with 200 µL of mineral oil to prevent evaporation at high temperature. An initial 15 minutes equilibrium time at 25°C was included prior to the temperature ramping. Denaturation of the samples was achieved by increasing the temperature at 1°C/min from 25 to 95°C and followed at 260 nm. The melting temperature (*Tm*) was determined as the maximum of the first derivative of the UV melting curves. Each experiments was repeated independently two or three times.

Peptide synthesis, purification and analysis

Tat-ARM and HEXIM-ARM 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, 21x250 mm, 300 A, 5 µM) with an acetonitrile gradient in 0.1% trifluoroacetic acid. The molecular weight and purity of the peptide were confirmed by mass spectroscopy. The sequences of Tat-ARM and HEXIM-ARM peptides are respectively **GRKKRRQRRR**PSQGGQTHQD and GKKKHRRRPSKKKRHWK. Peptides were dialyzed against water. After lyophilization, the concentration of Tat-ARM peptide was determined by weighting the resulting powder. The concentration of HEXIM-ARM was determined on a Nanodrop Spectrometer using a molar extinction coefficient calculated with the ExPASy Proteomics (https://web.expasy.org/protparam/).

NMR experiments

Magnesium titrations were performed at 700 MHz on a Avance III Bruker spectrometer equipped with a z-gradient TCI probe. NMR data were processed using TopSpin (Bruker). NMR experiments were performed in 50 mM sodium phosphate buffer (pH 6.4) at 15°C, in 90/10 H_2O/D_2O. The concentration of RNA sample was 100 µM. Sample volume was 150 µL in 3 mm NMR tube. Solvent suppression

was achieved using combined "Jump and Return" and WATERGATE sequences (Piotto et al. 1992; Plateau and Gueron 1982). Two-dimensional NOESY spectra were acquired at 15°C in 90/10 H_2O/D_2O. Base pairing was established as previously described (Lebars et al. 2010).

SEC-SAXS (Size exclusion Chromatography-Small angle X-ray Scattering) data acquisition and data reduction

SEC- SAXS data were collected at the SOLEIL Light Source on beamline SWING. Samples at a concentration of about 100 µM were loaded onto a size exclusion column, Agilent BioSEC3 with a pore size of 300 Å, previously equilibrated in the buffer chosen among Buffer "X" (50 mM TRIS pH 7.5, 50 mM NaCl, 50 mM $MgCl_2$), Buffer "NMR" (10 mM sodium phosphate at pH 6.5) or Buffer "SAXS" (10 mM sodium cacodylate at pH 6.5, 6 mM $MgCl_2$, 0.25 mM EDTA, 100 mM NaCl). The main advantage of SEC-SAXS is that it allows the separation of monodisperse samples from aggregates. The primary reduction of the SAXS data was performed using the Foxtrot software from the SWING beamline at SOLEIL synchrotron (https://www.synchrotron-soleil.fr/en/beamlines/swing). Briefly, buffer curves were averaged and the average buffer diffusion curve used to correct all the sample curves along the elution profiles for the solvent contribution. Then, an initial Guinier approximation was employed to obtain the radius of gyration (R_g) of each frame along the elution profiles. Curves showing a constant R_g were averaged to obtain a final SAXS data curve for each sample. The program PRIMUS from the ATSAS suite of programs was used for the calculation of the R_g indicated in the text (Konarev et al. 2003). The fit of the models with the diffusion data was estimated with the program CRYSOL (Svergun et al. 1995).

Isothermal Titration Calorimetry experiments

ITC experiments were performed using an ITC-200 microcalorimeter (Malvern). RNA was synthesized using T7 RNA polymerase as described above. RNA samples were refolded by heating at 95°C (2 min) and snap-cooled at 4°C (5 min) in the buffer used for ITC experiments. Peptides were synthesized as described above. The concentration of RNA sample in the cell and peptide in the syringe were 50 µM and 600 µM respectively. Titration experiments were performed at 25°C in 50

mM sodium phosphate buffer (pH 5.2, 6.4 and 7.5) supplemented with 0.5 M NaCl, under constant stirring at 750 rpm, with 21 x 2 µL injections into 200 µL sample cell volume, with 3 min between injections. Each experiment was accompanied by the corresponding control experiment in which the peptide at the same concentration was injected into the buffer alone. Buffer corrected ITC profiles were then fitted with a one site model for binding using MicroCal PEAQ-ITC analysis software (Malvern). Each experiments was repeated independently two or three times.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was supported by "Agence Nationale des Recherches sur le SIDA et les hepatites virales" (ANRS) [Grant AO 2018-2 n°18366 to IL], CNRS and University of Strasbourg. We thank the NMR Platform at IGBMC supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05 and Instruct-ERIC. We thank Michal Gajda for initial experiments with SAXS at DESY, Pierre Roblin at the SWING beamline at SOLEIL Light Source for helpful assistance during data acquisition and analysis, Jonathan Dorival and Elizabeth Ficko-Blean for help with SAXS analysis, and Mirjam Czjzek for discussion of the results, and critical reading of the manuscript. We acknowledge SOLEIL for the provision of synchrotron radiation facilities (proposal ID: 20140786) in using SWING beamlines. We thank Samuela Pasquali for critical reading of the manuscript.

Funding for Open access charge: CNRS.

REFERENCES

Aboul-ela F, Karn J, Varani G. 1995. The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *Journal of Molecular Biology* **253**: 313–332.
Abu Almakarem AS, Petrov AI, Stombaugh J, Zirbel CL, Leontis NB. 2012. Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Research* **40**: 1407–1423.
Barboric M, Yik JHN, Czudnochowski N, Yang Z, Chen R, Contreras X, Geyer M, Matija Peterlin B, Zhou Q. 2007. Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. *Nucleic Acids Res* **35**: 2003–2012.
Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. 1996. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**: 1547–1551.
Belanger F, Baigude H, Rana TM. 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**: 1094–107.
Berkhout B, van Wamel JL. 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA* **6**: 282–295.
Bigalke JM, Dames SA, Blankenfeldt W, Grzesiek S, Geyer M. 2011. Structure and dynamics of a stabilized coiled-coil domain in the P-TEFb regulator Hexim1. *J Mol Biol* **414**: 639–653.
Bourbigot S, Dock-Bregeon AC, Eberling P, Coutant J, Kieffer B, Lebars I. 2016. Solution structure of the 5'-terminal hairpin of the 7SK small nuclear RNA. *RNA* **22**: 1844–1858.
Brogie JE, Price DH. 2017. Reconstitution of a functional 7SK snRNP. *Nucleic Acids Res* **45**: 6864–6880.
Colasanti AV, Lu X-J, Olson WK. 2013. Analyzing and building nucleic acid structures with 3DNA. *J Vis Exp* e4401.
Dames SA, Schönichen A, Schulte A, Barboric M, Peterlin BM, Grzesiek S, 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.
Diribarne G, Bensaude O. 2009. 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. *RNA Biol* **6**: 122–128.
Dock-Bregeon AC, Chevrier B, Podjarny A, Moras D, deBear JS, Gough GR, Gilham PT, Johnson JE. 1988. High resolution structure of the RNA duplex [U(U-A)6A]2. *Nature* **335**: 375–378.
Durney MA, D'Souza VM. 2010. Preformed protein-binding motifs in 7SK snRNA: structural and thermodynamic comparisons with retroviral TAR. *J Mol Biol* **404**: 555–567.
Egloff S, Van Herreweghe E, 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.
Getz M, Sun X, Casiano-Negroni A, Zhang Q, Al-Hashimi HM. 2007. Review NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers* **86**: 384–402.
Hansen MR, Mueller L, Pardi A. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* **5**: 1065–1074.
He N, Jahchan NS, Hong E, Li Q, Bayfield MA, Maraia RJ, Luo K, 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.
Isel C, 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.
Jayaraman B, Mavor D, Gross JD, Frankel AD. 2015. Thermodynamics of Rev–RNA Interactions in HIV-1 Rev–RRE Assembly. *Biochemistry* **54**: 6545–6554.
Jeronimo C, Forget D, Bouchard A, Li Q, Chua G, Poitras C, Thérien 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.
Karn J. 1999. Tackling Tat. *J Mol Biol* **293**: 235–254.
Kobbi L, Demey-Thomas E, Braye F, Proux F, Kolesnikova O, Vinh J, Poterszman A, Bensaude O. 2016. An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb. *Proc Natl Acad Sci USA* **113**: 12721–12726.
Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. 2003. *PRIMUS*□: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* **36**: 1277–1282.
Krueger BJ, Jeronimo C, Roy BB, Bouchard A, Barrandon C, Byers SA, Searcey CE, Cooper JJ, Bensaude O, Cohen EA, 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.
Krueger BJ, Varzavand K, Cooper JJ, Price DH. 2010. The Mechanism of Release of P-TEFb and HEXIM1 from the 7SK snRNP by Viral and Cellular Activators Includes a Conformational Change in 7SK ed. M.V. Blagosklonny. *PLoS ONE* **5**: e12335.
Lebars I, Martinez-Zapien D, Durand A, Coutant J, Kieffer B, Dock-Bregeon A-C. 2010. HEXIM1 targets a repeated GAUC motif in the riboregulator of transcription 7SK and promotes base pair rearrangements. *Nucleic Acids Res* **38**: 7749–7763.
Leontis NB. 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Research* **30**: 3497–3531.
Mancebo HS, Lee G, Flygare J, Tomassini J, Luu P, Zhu Y, Peng J, Blau C, Hazuda D, Price D, et al. 1997. P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev* **11**: 2633–2644.
Marciniak RA, Calnan BJ, Frankel AD, Sharp PA. 1990. HIV-1 Tat protein trans-activates transcription in vitro. *Cell* **63**: 791–802.
Marino JP, Schwalbe H, Anklin C, Bermel W, Crothers DM, Griesinger C. 1995. Sequential correlation of anomeric ribose protons and intervening phosphorus in RNA oligonucleotides by a 1H, 13C, 31P triple resonance experiment: HCP-CCH-TOCSY. *J Biomol NMR* **5**: 87–92.
Markert A, Grimm M, Martinez J, Wiesner J, Meyerhans A, Meyuhas O, Sickmann A, 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.
Martinez-Zapien D, Legrand P, McEwen AG, Proux F, Cragnolini T, Pasquali S, Dock-Bregeon A-C. 2017. The crystal structure of the 5′ functional domain of the transcription riboregulator 7SK. *Nucleic Acids Research* gkw1351.
Martinez-Zapien D, Saliou J-M, Han X, Atmanene C, Proux F, Cianférani S, Dock-Bregeon A-C. 2015. Intermolecular recognition of the non-coding RNA 7SK and HEXIM protein in perspective. *Biochimie* **117**: 63–71.
Marz M, Donath A, Verstraete N, Nguyen VT, Stadler PF, Bensaude O. 2009. Evolution of 7SK RNA and its protein partners in metazoa. *Mol Biol Evol* **26**: 2821–2830.
Mbonye U, Karn J. 2017. The Molecular Basis for Human Immunodeficiency Virus Latency. *Annu Rev Virol* **4**: 261–285.
Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am*

Chem Soc **127**: 4223–4231.
Michels AA, Bensaude O. 2018a. Hexim1, an RNA-controlled protein hub. *Transcription* **9**: 262–271.
Michels AA, Bensaude O. 2018b. Hexim1, an RNA-controlled protein hub. *Transcription* **9**: 262–271.
Michels AA, Fraldi A, Li Q, Adamson TE, Bonnet F, Nguyen VT, Sedore SC, Price JP, Price DH, 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.
Milligan JF, Uhlenbeck OC. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Meth Enzymol* **180**: 51–62.
Muesing MA, Smith DH, Capon DJ. 1987. Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. *Cell* **48**: 691–701.
Muniz L, Egloff S, Ughy B, Jády BE, Kiss T. 2010. Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator Tat. *PLoS Pathog* **6**: e1001152.
Nguyen VT, Kiss T, Michels AA, Bensaude O. 2001. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**: 322–325.
Peterlin BM, Price DH. 2006. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**: 297–305.
Pham VV, Salguero C, Khan SN, Meagher JL, Brown WC, Humbert N, de Rocquigny H, Smith JL, D'Souza VM. 2018. HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. *Nature Communications* **9**. http://www.nature.com/articles/s41467-018-06591-6 (Accessed September 12, 2019).
Piotto M, Saudek V, Sklenár V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **2**: 661–665.
Plateau P, Gueron M. 1982. Exchangeable proton NMR without base-line distorsion, using new strong-pulse sequences. *Journal of the American Chemical Society* **104**: 7310–7311.
Plumridge A, Katz AM, Calvey GD, Elber R, Kirmizialtin S, Pollack L. 2018. Revealing the distinct folding phases of an RNA three-helix junction. *Nucleic Acids Research* **46**: 7354–7365.
Puglisi JD, Chen L, Blanchard S, Frankel AD. 1995. Solution Structure of a Bovine Immunodeficiency Virus Tat-TAR Peptide-RNA Complex. *Science* **270**: 1200–1203.
Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**: 76–80.
Rice AP. 2017. The HIV-1 Tat Protein: Mechanism of Action and Target for HIV-1 Cure Strategies. *Curr Pharm Des* **23**: 4098–4102.
Röder K, Stirnemann G, Dock-Bregeon A-C, Wales DJ, Pasquali S. 2020. Structural transitions in the RNA 7SK 5' hairpin and their effect on HEXIM binding. *Nucleic Acids Res* **48**: 373–389.
Schönichen A, Bigalke JM, Urbanke C, Grzesiek S, Dames SA, Geyer M. 2010. A flexible bipartite coiled coil structure is required for the interaction of Hexim1 with the P-TEFB subunit cyclin T1. *Biochemistry* **49**: 3083–3091.
Schulte A, Czudnochowski N, Barboric M, Schönichen A, Blazek D, Peterlin BM, 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.
Sheehy JP, Davis AR, Znosko BM. 2010. Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA* **16**: 417–429.
Smith CA, Chen L, Frankel AD. 2000. Using peptides as models of RNA-protein interactions. *Meth Enzymol* **318**: 423–438.
Svergun D, Barberato C, Koch MHJ. 1995. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied*

Crystallography **28**: 768–773.
Walker SC, Avis JM, Conn GL. 2003. General plasmids for producing RNA in vitro transcripts with homogeneous ends. *Nucleic Acids Res* **31**: e82.
Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, et al. 2018. MolProbity: More and better reference data for improved all-atom structure validation: PROTEIN SCIENCE.ORG. *Protein Science* **27**: 293–315.
Williamson JR. 2000. Induced fit in RNA-protein recognition. *Nat Struct Biol* **7**: 834–837.
Wu M, Tinoco I. 1998. RNA folding causes secondary structure rearrangement. *Proc Natl Acad Sci USA* **95**: 11555–11560.
Wyatt JR, Chastain M, Puglisi JD. 1991. Synthesis and purification of large amounts of RNA oligonucleotides. *BioTechniques* **11**: 764–769.
Xue Y, Yang Z, Chen R, 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.
Yang Z, Zhu Q, Luo K, Zhou Q. 2001. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**: 317–322.
Ye X, Gorin A, Ellington AD, Patel DJ. 1996. Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nat Struct Biol* **3**: 1026–1033.
Ye X, Kumar RA, Patel DJ. 1995. Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem Biol* **2**: 827–840.
Yik JHN, Chen R, Pezda AC, Samford CS, 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.
Yik JHN, Chen R, Pezda AC, 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.
Zhou Q, Yik JHN. 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.
Zweckstetter M, Hummer G, Bax A. 2004. Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys J* **86**: 3444–3460.

TABLES

Table 1. Melting temperature, *Tm*, of 7SK HP1-UUCG

	Tm (°C)		
$[Mg^{2+}]$ (mM)	**pH 5.2**	**pH 6.4**	**pH 7.5**
0	63.7 ± 0.2	64.0 ± 0.0	65.0 ± 0.0
0.1	67.0 ± 0.5	65.5 ± 0.5	65.2 ± 0.2
0.3	69.2 ± 0.2	66.7 ± 0.2	66.5 ± 0.5
1	73.5 ± 0.5	71.0 ± 0.0	67.5 ± 0.5
3	78.0 ± 0.0	75.7 ± 0.2	71.0 ± 0.0
6	81.0 ± 0.0	80.0 ± 0.0	73.5 ± 0.0

Each *Tm* is represented by the mean ± standard error calculated from two or three independent measurements.

Table 2. Isothermal Titration Calorimetry of HEXIM1-ARM and Tat-ARM binding to HP1-UUCG

	pH	Kd (μM)	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)
HEXIM1-ARM	5.2	11.6 ± 0.6	-4.0 ± 0.2	-2.8 ± 0.3	-6.7 ± 0.1
	6.4	23.1 ± 0.6	-6.3 ± 0.1	-0.002 ± 0.012	-6.3 ±0.1
	7.5	73.9 ± 6.4	-5.6 ± 0.3	-0.005 ± 0.361	-5.6 ± 0.1
Tat-ARM	5.2	22.5 ± 0.6	-22.4 ± 0.6	16.0 ± 0.6	-6.3 ±0.2
	6.4	21.9 ± 0.7	-19.9 ± 0.4	13.6 ± 0.4	-6.4 ±0.0
	7.5	98.5 ± 21.9	-7.8 ± 0.7	2.3 ± 0.8	-5.5 ± 0.1

FIGURE LEGENDS

Figure 1: Comparison of the four structures from NMR and Crystallography highlighting the conformational landscape explored by HP1. (Top) Secondary structures showing nucleotides modified from the natural sequence of *homo* 7SK (outlined) and bulged nucleotides (colored). These are colored according to their situation: stacked inside the helix (green), swung outside the helix (red), or forming base triples (blue), with dashed lines connection to their coplanar base pairs. Base triples and the non-canonical pairs are further described using the Leontis-Westhof notation (Leontis 2002; Abu Almakarem et al. 2012). The small arrow in panel **(B)** indicates that U41 lies in the major groove in the vicinity of A43:U66 base pair in this structure. (Bottom) 3D structures in similar orientation emphasizing local dynamics, with the crystal structures colored by B-factors and NMR structures by local rmsd differences colored and represented as a thickness of the backbone. The color-scales are indicated. The indicated size of each conformer was estimated as the distances from G24(C1')- to $U2_{loop}$(C1') for the 5IEM and crystal structures and from G24(C1') to $A2_{loop}$(C1') for the 6MCI structure.

Figure 2: Correlation of experimental RDC values with RDCs calculated from **(A)** NMR solution structures (PDB Id: 5IEM) (Bourbigot et al. 2016) refined with distances, torsion angles and RDC constraints, **(B)** from X-Ray structures (PDB Id: 5LYS, 5LYU, 5LYV) (Martinez-Zapien et al. 2017) and **(C)** from NMR structures determined by Pham *et al.* (PDB Id: 6MCI) (Pham et al. 2018). For NMR structures, each color represents one conformer (Panels A and C). For X-ray structures, conformers used for calculations are indicated. The correlation coefficient (R^2) is indicated for each structure.

Figure 3: **(A)** Secondary structure of HP1-UUCG highlighting the nucleotides showing chemical shift perturbations (CSP) upon addition of magnesium at high concentration (60 mM). The extent of the CSP is represented as the thickness of the character. Nucleotides showing CSP at a threshold of 0.05 ppm are represented in blue, while nucleotides at threshold of 0.40 ppm are indicated in bold blue. **(B)**

Imino-protons region of 1D spectra recorded at 15°C in 50 mM sodium phosphate buffer (pH 6.4) with increasing magnesium concentrations, indicated on the left. Imino protons that undergo variations are indicated in blue. **(C)** Imino/Imino protons region of NOESY spectrum recorded at 15°C in 90/10 H2O/D2O with a mixing time of 400 ms in the absence of magnesium (black) and with 6 mM Mg^{2+} (blue). Cross-peaks that show up upon magnesium are highlighted in blue.

Figure 4: Small-angle X-Rays scattering (SAXS) analysis of the HP1-UUCG molecule in different buffers solutions. Top: Experimental scattering curves obtained in buffer "X" (50 mM NaCl, 50 mM $MgCl_2$, 50 mM Tris-HCl pH 7.5; red line), buffer "NMR" (10 mM Na phosphate pH 6.5; green line) or buffer "SAXS" (100mM NaCl, 10 mM NaCacodylate pH 6.5, 6 mM $MgCl_2$, 0.25 mM EDTA; yellow line) are represented together with the theoretical curves (blue line) back-calculated from the high-resolution structures with the program CRYSOL (Svergun et al. 1995). **(A)** Fit with the extended NMR structure (model 5IEM, blue line) (Bourbigot et al. 2016) **(B)** Fit with the crystal structure "OUT" (blue line) (Martinez-Zapien et al. 2017). Similar result was obtained with the "IN" crystal structure (not shown). **(C)** Fit with the compact NMR structure (model 6MCI, blue line) (Pham et al. 2018). Bottom: Differences between experimental and theoretical scattering curves.

Figure 5: ITC profiles for the binding of HEXIM-ARM and Tat-ARM to HP1-UUCG at 25°C in 50 mM sodium phosphate with 0.5 M NaCl, at pH 5.2 **(A)**, pH 6.4 **(B)** and pH 7.5 **(C)**. Top panels show ITC traces, and the bottom panels show integrated heat values as a function of the [peptide]/[RNA] ratio. Data were fitted using the single-site binding model.

Figure 6: Comparison of the conformations observed for HP1 with focus on the major groove of the $(GAUC)_2$ helix. **(A)** Surface view of the structure from PDB ID 6MCF, with the bound Tat peptide (yellow sticks) showing the two arginines (R53, left and R52, right) in the ASM4 and pseudo-ASM3, respectively. The colored surfaces correspond to the nucleotides shown to undergo major conformational changes in the four structures i.e. U40 (cyan), U41 (magenta) and U63 (orange). **(B)**

Similar view for the unbound conformation OUT (extracted from PDB Id. 5LYS and 5LYU) showing the major groove and the mouth open but not enough to accommodate arginine R52 in the conformation it takes in (A). **(C)** Extended conformation 5IEM showing a regular major groove, but filled partially by U41 (pink). **(D)** Conformation IN (extracted from PDB Id. 5LYS and 5LYU), showing the major groove tightly packed with the three uridines. The "mouth" is indicated with arrows.

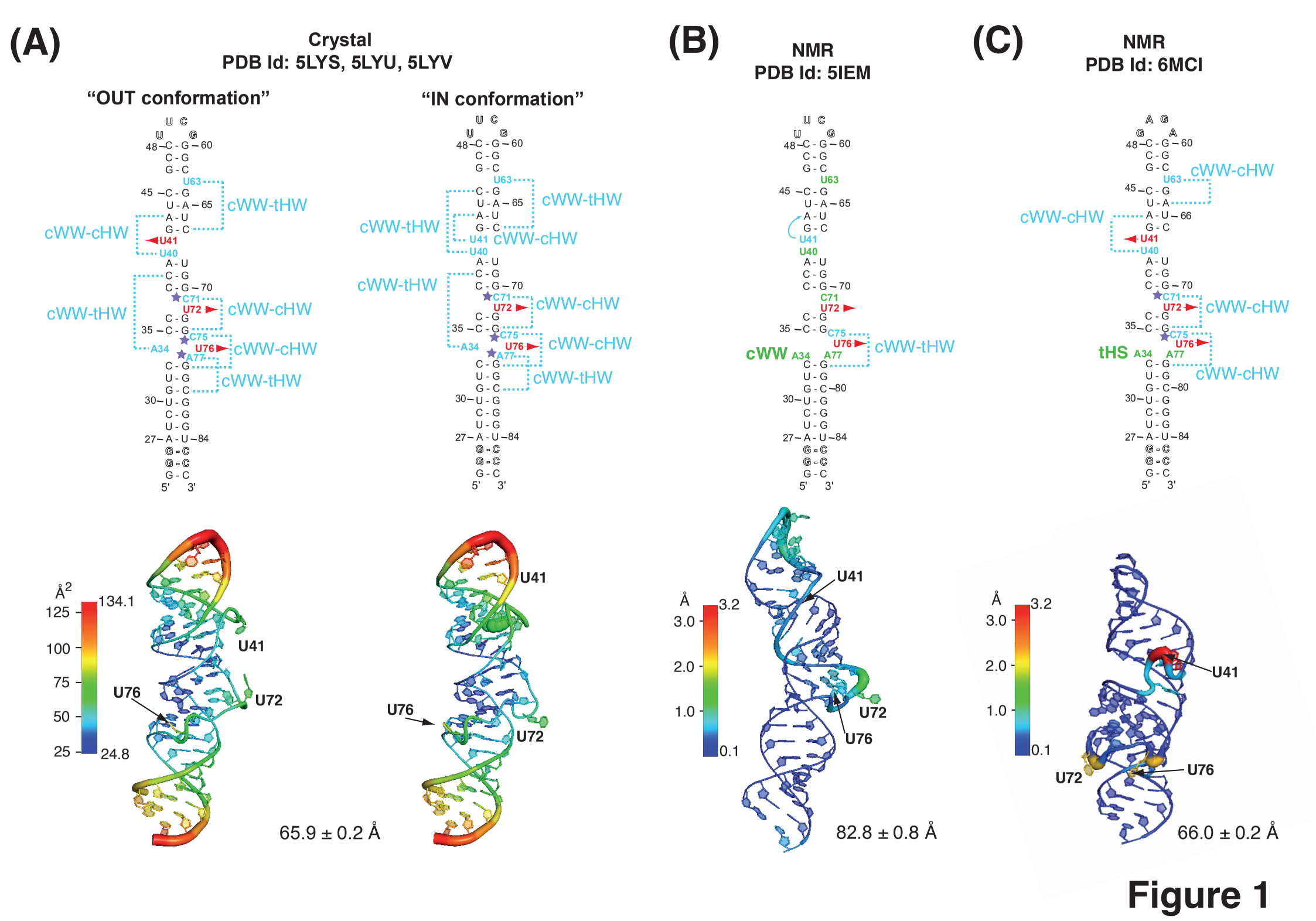
(A)
Crystal
PDB Id: 5LYS, 5LYU, 5LYV
"OUT conformation"
"IN conformation"
cWW-cHW
cWW-tHW
U41
U72
U76
Å2
134.1
125
100
75
50
25
24.8
65.9 ± 0.2 Å
(B)
NMR
PDB Id: 5IEM
cWW
tHS
Å
3.2
3.0
2.0
1.0
0.1
82.8 ± 0.8 Å
(C)
NMR
PDB Id: 6MCI
66.0 ± 0.2 Å

Figure 1

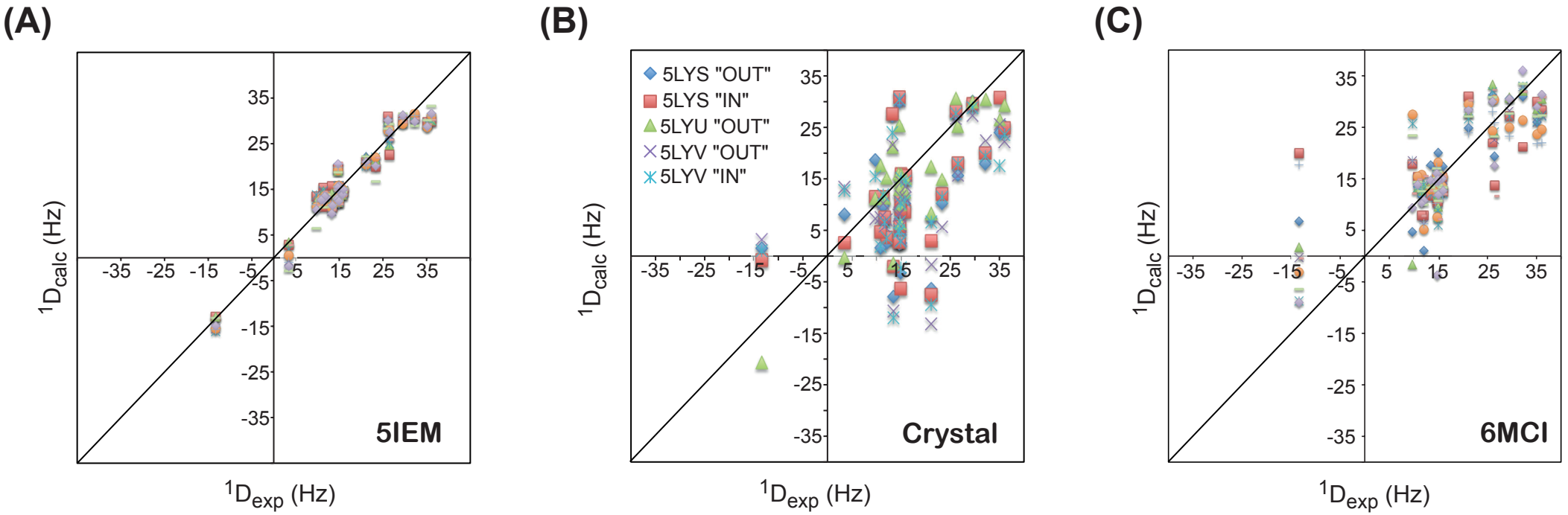
(A)
(B)
(C)
5LYS "OUT"
5LYS "IN"
5LYU "OUT"
5LYV "OUT"
5LYV "IN"
5IEM
Crystal
6MCI
$^1D_{calc}$ (Hz)
$^1D_{exp}$ (Hz)
R^2 = 0.95 ± 0.01
5LYS: R^2 = 0.53 ± 0.04
5LYU: R^2 = 0.80 ± 0.09
5LYV: R^2 = 0.43 ± 0.01
R^2 = 0.76 ± 0.13

Figure 2

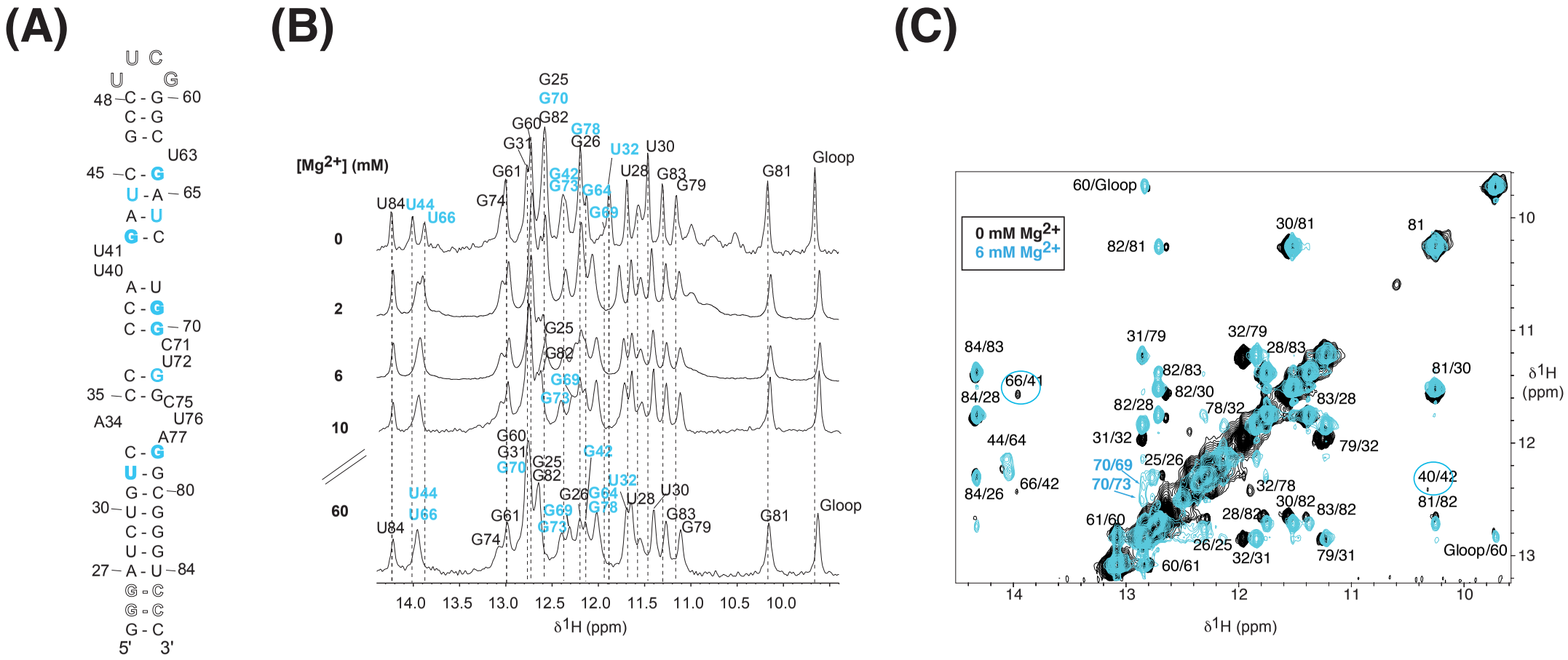

Figure 3

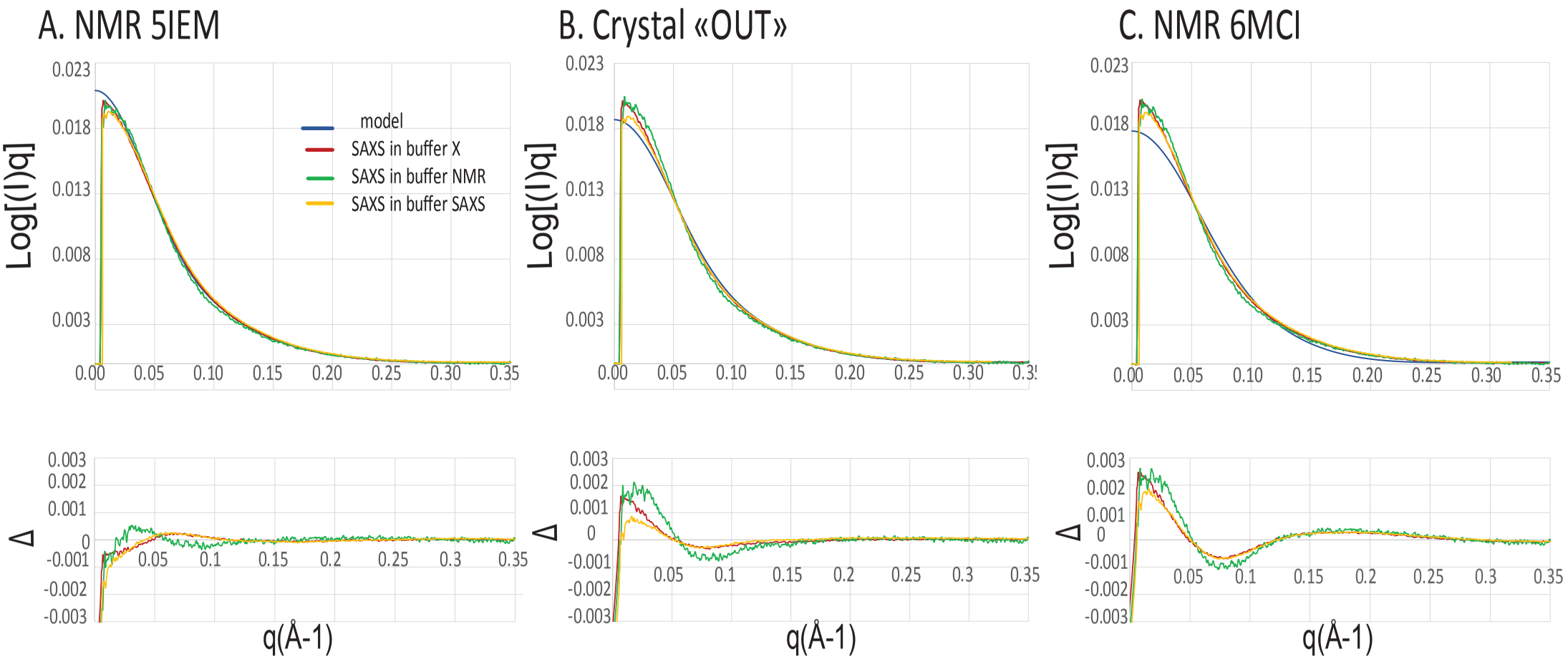
A. NMR 5IEM
B. Crystal «OUT»
C. NMR 6MCI
Log[(I)q]
model
SAXS in buffer X
SAXS in buffer NMR
SAXS in buffer SAXS
Δ
q(Å-1)

Figure 4

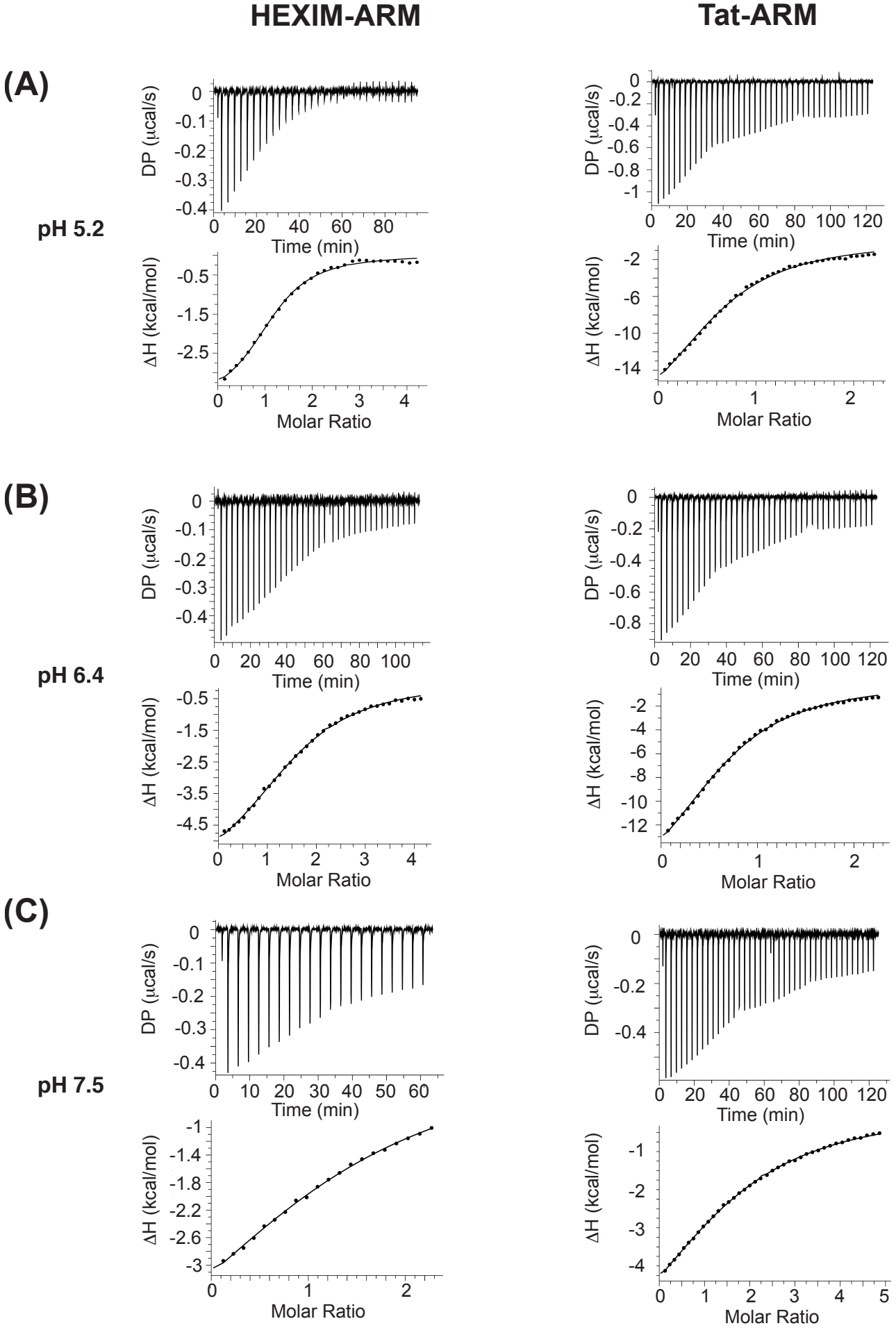
HEXIM-ARM
Tat-ARM
(A)
pH 5.2
(B)
pH 6.4
(C)
pH 7.5
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
Molar Ratio

Figure 5

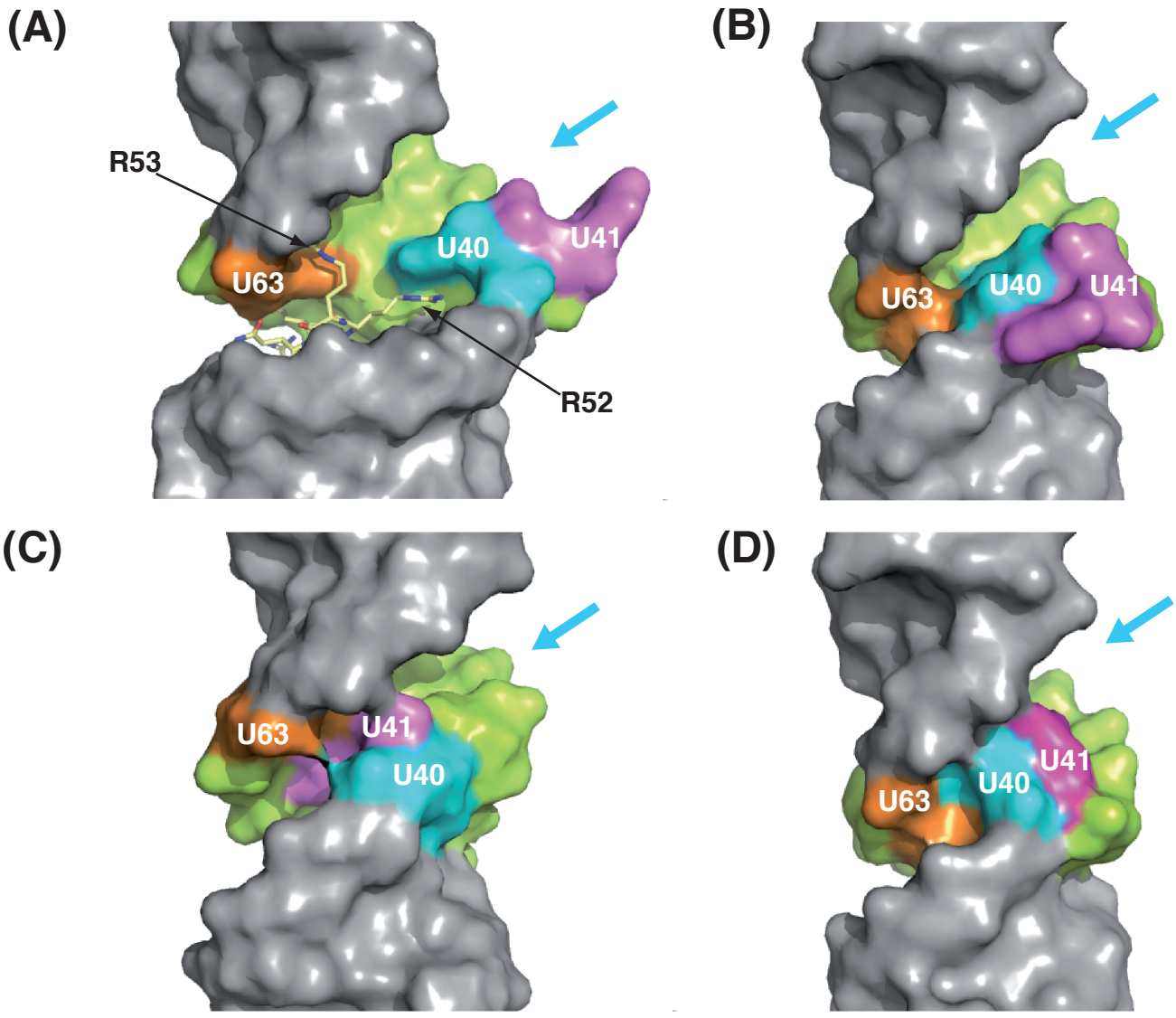

Figure 6