

Preformed protein-binding motifs in 7SK snRNA: structural and thermodynamic comparisons with retroviral TAR

Michael A. Durney¹ and Victoria M. D'Souza^{1,*}

¹Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA. 02138, USA

*Correspondence: dsouza@mcb.harvard.edu

KEYWORDS: 7SK; elongation; NMR; TAR; EIAV

ABBREVIATIONS: HMQC, HSQC, NOE, NOESY, RDC, TAR

RESEARCH HIGHLIGHTS

- The first high-resolution structural data for the key regulatory 7SK snRNA.
- The RNA is preformed leading to a high-affinity interaction with arginine.
- 7SK does not require base-triple formation unlike retroviral TAR.

The 7SK small nuclear RNA is a highly conserved non-coding RNA that regulates transcriptional elongation. 7SK utilizes the HEXIM proteins to sequester the transcription factor P-TEFb by a mechanism similar to that used by retroviral TAR RNA to engage Tat and P-TEFb. Tat has also recently been shown to bind 7SK directly and recruit P-TEFb to TAR. We report here the solution structures of the free and arginine bound forms of stem loop 4 of 7SK (7SK-SL4). Comparison of the 7SK-SL4 and TAR structures demonstrates the presence of a common arginine sandwich motif. However, arginine binding to 7SK-SL4 is mechanistically distinct and occurs via docking into a pre-organized pocket resulting in a 1000-fold increased affinity. Furthermore, whereas formation of the binding pocket in TAR requires a critical base-triple, hydrogen-bond formation between the equivalent bases in 7SK-SL4 is not essential and the pocket is stabilized solely by a pseudo base-triple platform. In addition, this theme of preformed protein binding motifs also extends into the pentaloop. The configuration of the loop suggests that 7SK-SL4 is poised to make ternary contacts with P-TEFb and HEXIM or Tat. These key differences between 7SK-SL4 and TAR present an opportunity to understand RNA structural adaptation and have implications for understanding differential interactions with Tat.

INTRODUCTION:

RNA polymerase II is highly regulated at all phases of transcription. The elongation phase, in particular, has recently received renewed attention resulting in the identification and functional characterization of numerous regulatory factors.^{1,2} One of the key negative regulators of this process is the non-coding 7SK small nuclear RNA (snRNA) that inhibits the cyclin-dependant kinase P-TEFb (positive transcription elongation factor b).³⁻⁸ 7SK sequesters P-TEFb by forming a ternary complex with the cyclin T1 subunit of P-TEFb and the HEXIM proteins.⁹ The significance of this process is further highlighted by P-TEFb interactions with retroviral TAR (transactivation response region) RNA and Tat protein, which have evolved strategies to mimic the 7SK RNA and HEXIM protein, respectively.^{10,11} In contrast to 7SK, however, TAR promotes retroviral transcription by antagonistic positive transactivation of P-TEFb and subsequent stabilization of the elongating Pol II complex. Recent work has elucidated interactions of Tat with various cellular host factors, one of which is a direct interaction between 7SK and Tat.¹²⁻¹³ More significantly, Tat has been shown to bind 7SK RNA by expelling HEXIM from the ternary complex, prompting a re-evaluation of our understanding of P-TEFb equilibrium between host and viral transcription.^{13; 14}

The unique mechanism by which TAR interacts with cyclin T1 and Tat has been outlined. Specifically, two distinct regions of TAR promote formation of the ternary complex: the flexible apical loop binds both cyclin T1 and Tat while the flexible bulge region recognizes Tat.^{15;16} Collectively these interactions proceed via a cooperative induced fit mechanism, the structural basis of which has yet to be determined. Tat contains an arginine-rich motif (ARM) in which a single arginine residue has been shown by biochemical and *in vivo* transactivation assays to bind to the TAR bulge region.^{17,18} Several NMR studies have elucidated the major

conformational transition that argininamide (a tight-binding analog of arginine) induces in the TAR structure upon complex formation.¹⁹⁻²³ In the free form the bulged residues and the terminal A22 and U40 residues are not involved in hydrogen bonding interactions but upon arginine binding U23 and A22 sandwich arginine by forming a critical U23:A27-U38 base-triple and a A22-U40 base pair (Figure 1a, dashed red lines) and leads to a global dampening of interhelical motions.^{22,23}

The conserved stem loop 4 of 7SK (7SK-SL4) presents significant sequence homology to the TAR RNA from HIV-1 (Figure 1a). As an initial step towards understanding the structural basis of 7SK function and to draw parallels with retroviral TAR, we determined the NMR solution structure of 7SK-SL4 both free and in complex with argininamide. Using this model system we show that 7SK-SL4 contains preformed binding motifs and has an affinity for arginine three orders of magnitude greater than that of TAR. We further rationalize these findings by examining the thermodynamic signatures of argininamide binding to mutant 7SK-SL4 constructs. Our data, in addition to highlighting structural diversity in RNA-protein recognition, contribute to understanding the distinct strategies for differential Tat binding to 7SK and TAR.

RESULTS AND DISCUSSION

Structure of 7SK-SL4 identifies a preformed protein-binding bulge

Our choice of 7SK-SL4 for structural analysis was guided by previously reported biochemical and *in vivo* analyses that directly implicate this region of 7SK in P-TEFb capture.^{24,25} Based on these results we designed a 42-nucleotide RNA construct shown in Figure 1a that folded homogeneously and yielded excellent NMR spectra. The imino proton walk together with

assignments of the ^1H - ^{15}N HSQC (heteronuclear single quantum coherence) spectrum confirmed the predicted secondary structure (Figure 1b). Two residue-specific labeled samples were used to complete sequence-specific resonance assignments for 7SK-SL4 using an NOE (nuclear Overhauser effect) strategy defined previously (Figure 1c).²⁶ Our studies resulted in a high-resolution ensemble of 7SK-SL4 structures with a root-mean-square-deviation (RMSD) of 0.38 Å for the heavy atoms in base paired residues (Figure 2a, Table 1).

Unexpectedly, 7SK-SL4 is highly structured even in the free form and has pre-organized protein binding motifs both in the apical pentaloop and bulge regions. The bulge between stems 2 and 3 is structured and very closely resembles the arginine/peptide bound structure of TAR.^{20;}
^{21; 22; 27} The binding pocket is capped by the G305-U319 wobble while the bulge residue C320 forms the base of the pocket in the major groove (Figure 2b). Several unambiguous base-ribose NOE contacts place residue C320 in an approximately planar orientation to the G303-C323 Watson-Crick pair with a potential to form a C320:C323-G303 base-triple interaction (Figure 2c). For example, the C320 ribose H1' has NOE interactions with both the H4 amino and H5 aromatic protons of the pyrimidine base of C323. This orientation is confirmed by complementary NOE interactions between a C320 N4 amino and imino proton of G303. Furthermore, an NOE between the C320 amino proton and the 2' proton of residue G302 places the C320 bulge residue into the major groove. Interactions between both the ribose H2' and H3' of U321 with the G322 aromatic H8 places U321 in the minor groove facing the G305-U319 wobble base pair. Unlike the surrounding stem residues, both the bulge residues C320 and U321 have a C2' *endo* pucker. All residues in stem 2 and stem 3 are involved in base pairing interactions, Figure 1b. Most importantly, the terminal G305-U319 wobble base pair in stem 3 is formed as evidenced by typical imino proton NOEs between the two residues (Figure 1b). In

contrast, the equivalent A22-U40 base pair in TAR forms only after ligand binding.²³ No significant bend was observed between the two stems which are arranged in near-coaxial stacking as is evident by standard stacking NOEs between C304 to G305. Tertiary contacts in the bulge, for example a cross-strand NOE between a G305 N2 amino proton and the G322 H8 proton as well as our dipolar coupling data define a 60° inter-helical displacement between the two adjacent stems.

The protein-binding motif is stabilized by a pseudo base-triple platform

In HIV-1, the U23:A27-U38 base-triple formation is critical for arginine complex formation. Conclusive evidence for this has been obtained from a HIV-1 TAR mutant where an isomorphic C⁺23:G27-C38 base-triple leads to argininamide binding only at low pH.²¹ The structure of the preformed bulge in 7SK-SL4 prompted us to ask whether residue C320 can, in fact, form a base-triple with the C323-G303 Watson-Crick pair. Such a triple base has been previously observed in the 30S subunit of the ribosome.²⁸ In the 7SK-SL4 structures, however, the non-canonical cytosine is positioned deep into the major groove surrounded by a negative electrostatic potential and suggested that protonation of C320 could lead to formation of additional hydrogen bonds as shown in Figure 3a.

To monitor base-triple formation we analyzed the behavior of the exchangeable amino protons by obtaining HSQC data under various pH and temperature conditions. At pH 5.2 and above, the Watson Crick C323 N4 amino protons have chemical shifts of 8.55 ppm and 6.55 ppm, which are typical for a hydrogen bonded/non-bonded proton pair respectively (Figure 3b). This suggests that the N4 amino group of C323 is not involved in base-triple formation under these conditions. Similarly, no evidence of N3 protonation of residue C320 was observed at these

conditions. However, on lowering the pH, we observed one of the N4 amino groups of C320 at chemical shift values of 8.52 ppm (^1H) and 104.9 ppm (^{15}N), which are typical of protonated cytosines (Figure 3c).²⁹ Simultaneously with protonation, an interesting correlation also occurs with the exchange behavior of the N4 amino protons of residue C323. The upfield shifted non-hydrogen bonded N4 amino proton of C323 is selectively exchange-broadened upon protonation of C320 and indicates participation in hydrogen bond formation with the O2 of C320 (Figure 3c). This correlated intensity change of the upfield C323 amino proton provides a spectroscopic signature for the formation of hydrogen bonds in the C⁺320:C323-G303 base-triple in a protonation-dependant manner. Importantly, analysis of the data at low pH showed no major structural rearrangements; all intra- and inter-residue NOEs were maintained. Including restraints for the C⁺320(O2)-C324(N4) and C⁺320(N3)-G303(O2) hydrogen bonds during structure calculations also confirmed the lack of structural adaptation upon base-triple formation. In these structures, the N4 proton of C320 was not within hydrogen bonding distance to N7 of G303, implying that only two of the three possible hydrogen bonds formed at low pH. Thus, although the arrangement of the three bases will allow for formation of a true base-triple under very high proton concentration, we show here that at physiological conditions the triple is not formed. We therefore refer to the C320:C323-G303 arrangement as a pseudo-triple platform, the nature of which is sufficient to define the preformed ligand binding pocket.

High-affinity arginine binding without conformational change in 7SK-SL4

To test the implication of a preformed structure we decided to quantify the strength of argininamide binding to 7SK-SL4 using isothermal titration calorimetry (ITC). Interestingly, ITC data obtained for titration of the wild type 7SK-SL4 tetraloop construct with argininamide

resulted in a dissociation constant (K_d) of $6.31 \pm 1.55 \mu\text{M}$ for the 1:1 complex (Figure 4a). However, the binding affinity of the HIV-1 TAR and argininamide has been quantified to be in the low millimolar range.¹⁸ We confirmed this apparent discrepancy by ITC experiments on a TAR construct and found that the interaction was indeed not amenable to standard ITC experiments due to the weak affinity. To verify our findings, we performed NMR titrations on the 7SK-SL4:argininamide complex. In contrast to TAR, intermolecular NOEs were observed at a stoichiometric 1:1 ratio providing further proof of tighter binding.²⁷ Comparison of the HMQC (heteronuclear multiple quantum coherence) spectra of free 7SK-SL4 and the argininamide bound complex revealed no major chemical shift changes suggesting lack of structural adaptation in the RNA structure on complex formation (Figure 4b). Furthermore, there is no change in the relative intensities of the bulge residues compared to the stem residues suggesting that, unlike TAR, the motions in the bulge are not significantly perturbed upon argininamide binding. In comparison to TAR, our structural and ITC data suggest that argininamide binds selectively to a localized and preformed region of the RNA. This reduced entropic barrier to complex formation increases the binding affinity of argininamide to 7SK by three orders of magnitude.

Structure of the 7SK-SL4:arginine complex

The structural similarities in the bulge regions of 7SK-SL4 and ligand-bound TAR prompted us to speculate on whether argininamide binds to this region. Our NMR studies on the complex resulted in a high-resolution ensemble of 7SK-SL4 structures with a RMSD of 0.42 Å for the heavy atoms in base paired residues (Table 1). An unambiguous set of intermolecular NOEs was observed for the complex that allowed us to precisely position the argininamide in the resulting structure ensemble (Figure 5a). The argininamide H β protons interact with both the

G303 H8 and C323 N4 amino protons while the H γ protons interact with the U319 H6 proton. The terminus of the argininamide side chain is precisely anchored by several critical NOEs such as the interactions of the H δ with a C304 N4 amino, the G322 imino and the C320 aromatic H5 and H6 protons. Two separate signals for the guanidinium H η protons gave weak interactions with the U27 base detectable at low temperatures and 500 ms NOE mixing time. Thus, in the complex, arginine docks onto the pseudo-triple and is sandwiched between the U319 cap and the C320 base of the pocket (Figure 5b, 5c). Arginine forms a planar structure with the C304-G322 base pair with the two amide groups of the guanidinium involved in an arginine fork interaction¹⁷: one is within hydrogen bonding distance of the O6 of G322 while the other interacts with the backbone phosphate O5 of C320 (Figure 5b). Together, these interactions create a stabilized hydrogen-bonding interaction for argininamide, which stacks on the C320:C323-G303 pseudo-triple platform and forms an intimate interface optimized by intermolecular electrostatic interactions. Importantly, both the sequential NOEs and chemical shifts of the residues in the binding pocket remain unperturbed confirming the lack of structural change in the complex.

The preformed bulge structure of 7SK-SL4 raised the possibility that complex formation is not dependent on or is not promoted by base-triple formation as seen in TAR. We monitored the exchangeable spectra as a function of argininamide concentration. At a stoichiometric ratio the intensities of the two N4 amino protons of C323 remain equal and no exchange broadening occurs for the upfield non-hydrogen bonded proton (see above). In fact, we did not observe an intensity change for this proton even when the argininamide concentration was increased five-fold. We can thus conclude that the potential base-triple is not formed upon ligand binding. The complex was, however, still capable of forming the base-triple at pH values below 5.2. We

repeated the ITC experiments at low pH but did not see an influence on the binding constant further proving that tight arginine binding does not require base-triple formation (data not shown). Therefore, while an overall similarity between the bound structures of both 7SK-SL4 and TAR is apparent, our data show that the mechanisms of complex formation are strikingly distinct. Formation of the 7SK-SL4 complex does not occur via an induced fit mechanism but proceeds by docking into a preformed pocket. In addition, the preformed pocket is sufficient for binding and formation of a potential base-triple is not required for the high-affinity interaction.

Thermodynamic basis for arginine binding by 7SK-SL4 mutants

The 7SK-SL4:ARG structure allowed us to use ITC to delineate the thermodynamic bases of complex formation. We examined a series of mutants designed to probe the relative contributions of the bulge residues and the potential base-triple to argininamide binding (Figure 6a). Unexpectedly, all constructs examined have remarkably similar dissociation constants ($K_d = 6.91 \pm 1.49 \mu\text{M}$) and free energies of association ($\Delta G = 7.07 \pm 0.09 \text{ kcal/mol}$) (Figure 6b and 6c). The observed differences in the thermodynamic profiles of the wild type and mutant sequences can therefore be directly attributed to differing enthalpic and entropic contributions to the constant free energy of binding.

In the wild type construct, complex formation is primarily driven by a favorable entropic contribution and a relatively minor enthalpic contribution ($-T\Delta S = -5.68 \text{ kcal/mol}$ and $\Delta H = -1.44 \text{ kcal/mol}$ respectively). Since wobble base pairs are often critical determinants of specificity in RNA-protein recognition,³⁰ we designed a U319C construct to test the contribution of the G305-U319 base pair. This mutant is essentially identical to wild type and suggests that the presence of a terminal wobble base pair does not influence the interaction ($-T\Delta S = -5.53$

kcal/mol and $\Delta H = -1.60$ kcal/mol). A G305U/U319A double mutant representing the terminal base pair in TAR also yields the wild type profile (data not shown). We also probed if the nature of the triple platform influenced ligand-binding properties. A C320U substitution, which is not capable of protonation, also yielded essentially the same binding profile as wild type ($-T\Delta S = -5.97$ kcal/mol and $\Delta H = -0.96$ kcal/mol). However, when we replaced the pseudo-triple with an equivalent U:A-U triple designed to mimic the base-triple in TAR, a switch in the thermodynamic strategy for argininamide binding occurred. In contrast to wild type, the triple mutant binding interaction is predominantly enthalpic. Closer inspection reveals that the enthalpic and entropic contributions are almost exactly compensated to give identical binding constants ($-T\Delta S = -0.55$ kcal/mol and $\Delta H = -6.54$ kcal/mol, Figure 6b). In support of our structure, an increase of 5.10 kcal/mol could easily account for the two hydrogen bonds predicted. Our data thus suggests that an increase in enthalpic contribution for the triple mutant is due to true base-triple formation after ligand binding. The decrease in the entropic contribution could occur due to an enthalpic/entropic compensation in the system caused by a more structured binding interface due to hydrogen bond formation and/or improved stacking interactions in the arginine sandwich. Our data, however, excludes the latter since a sandwich formed by two uracils (C320U mutant) has the same profile as a sandwich formed by a cytosine and a uracil (wild type). The thermodynamic profiles and the dissociation constants observed in these studies provide evidence that the binding pocket is preformed in all constructs but addition of the ligand leads to hydrogen bond formation only in the triple mutant. Most importantly, our thermodynamic analysis conclusively shows that arginine binding does not lead to base-triple formation in 7SK-SL4.

Structure of 7SK-SL4 also identifies preformed protein-binding loop

The apical loop in TAR is the primary motif for ternary contacts in the Tat:TAR:P-TEFb complex. Currently, structural data for the ternary contacts are only available for the equine infectious anemia virus (EIAV).³¹ In this system, the loop is also known to undergo a major structural adaptation upon protein binding.^{31,32} Remarkably, similar to the bulge the well-defined free 7SK-SL4 pentaloop has significant common characteristics not with the free form but with the bound form of the EIAV loop. This allows us to propose a comparative structural alignment of the 7SK-SL4 pentaloop (A₃₁₀UGUG) with the bound form of EIAV TAR hexaloop (UC₁₂UGCG) (Figure 7).

The loop residues in 7SK-SL4 gives rise to many inter-residue NOEs which allowed us to confidently characterize the preformed nature of this motif, Figure 7a. The first residue A310 has continuous stacking interactions with both the stem residue C309 and the loop residue U311 and projects into the loop. In addition, standard long-range minor groove NOEs from the H2 proton of A310 to the H1' protons of U311 and G315 were observed confirming the A-helical environment of these residues. Although, there is a break in sequential connectivities after U311, the relative orientation of the last three loop residues is clearly defined by many intra- and inter-residue NOEs. The G312 residue exhibits NOEs to both the preceding U311 and the following U313 residues: the H2' interacts with the H3' of U311 while both the H1' and H4' protons interact with the H6/H5 protons of U313. We also observe long range NOEs from the aromatic H8 proton of G314 to the H2 of A310, the H1' of U311, and the H4' protons of both G312 and U313, which firmly defines the orientation of the G314 base. Furthermore, the position of G314 ribose was also described by NOEs to H2 proton of A310 and H8 proton of G315.

Apart from residue A310, all of the other riboses in the loop have a C2' *endo* pucker. Importantly, however, both G312 and G314 bases are maintained in a *syn* conformation as evidenced by intense H8 to H1' NOEs and characteristic downfield C8 chemical shifts (Figure 4b).³³ In contrast, both the equivalent G14 and G16 in EIAV are in an *anti* conformation in the free form but only adopt a *syn* conformation upon complex formation. Thus, the overall orientation of the loop residues in the free form of 7SK-SL4 and the bound form of EIAV are very similar; the first two residues are stacked in a helical environment while the last three residues loop out of the helix in the major groove. In EIAV, these looped out residues are engaged in intermolecular contacts with Tat and P-TEFb. Similar conformations and arrangements of the equivalent residues in 7SK-SL4, suggest that the loop is preformed and poised to make similar contacts.

SUMMARY AND IMPLICATIONS:

The solution structure of stem-loop 4 of human 7SK snRNA in complex with arginine represents the first high-resolution structural data for this key regulatory RNA and complements the recent characterization of protein structures within the snRNP.^{34,35} The structure explains sequence conservation and reveals motifs that may play roles in recognition of P-TEFb and the HEXIM proteins in addition to retroviral Tat. A preformed protein-binding pocket in the bulge is composed of a pseudo-triple platform and it is noteworthy that hydrogen bond formation is not required for complex formation. Furthermore, repeating the theme of preformed motifs, the capping pentaloop adopts a structure remarkably similar to that of the cyclin-bound form of TAR from EIAV.

Comparison of the 7SK-SL4 and HIV-1 TAR structures reveals how convergent evolution provided the virus with structurally similar yet mechanistically distinct pathways for P-TEFb capture. Tat binding to TAR causes a dramatic conformational transition in the RNA that leads to structuring of the flexible bulge. It is hence subjected to an entropic penalty that reduces its binding affinity by three orders of magnitude compared to 7SK-SL4. Although arginine is a simple model system representing Tat, the repeating theme of preformed versus flexible motifs both in the capping pentaloop and the bulge regions in 7SK-SL4 suggests that the relative difference in affinity will be maintained in Tat.

Evolution of 7SK-SL4 to function only in engaging P-TEFb and HEXIM may have favored preformed binding motifs. Why has TAR then specifically evolved to maintain flexibility rather than adapt for tighter binding to Tat? The answer may be related to the limited pools of P-TEFb available for Tat binding: one active and engaged in transcription and another in the inhibitory 7SK snRNP complex.³⁶ Recent experiments have shown that Tat can displace HEXIM from the 7SK snRNP to capture P-TEFb and can indeed interact directly with 7SK. Thus, the 7SK snRNP may be the only source of P-TEFb for Tat, and binding to 7SK may be required before committing to a retroviral transcript containing TAR. Since there is also a limited availability of Tat in the initial stages of HIV-1 transcription, a prematurely high affinity interaction with TAR would reduce the pool available for P-TEFb capture and lead to detrimental effects during latency³⁷. In summary, our results in addition to highlighting the role of RNA structure in RNA-protein recognition provide a basis for differential binding of Tat to 7SK and HIV-1 TAR RNA³⁸.

MATERIALS AND METHODS:

RNA sample preparation

Overlapping PCR primers were designed to include the T7 promoter, an insert sequence corresponding to nucleotides 296-331 of 7SK (NCBI reference sequence NR_001445.2) with three additional G-C base pairs to improve transcription efficiency and monomeric stem-loop formation, a *SmaI* linearization site and restriction sites for *BamHI* and *EcoRI* for insertion of the amplified product into pUC19. RNA samples, transcribed and purified as described²⁶, were annealed under dilute conditions followed by slow cooling to ensure homogeneous monomer formation as confirmed by native PAGE.

NMR data acquisition and analysis

NMR data were acquired using Bruker 600 MHz and 700 MHz spectrometers equipped with cryoprobes. Spectra were recorded at 298 K with the exception of data for the imino region for which data were also recorded at 278 K. Assignments for non-exchangeable ¹H and ¹³C signals were obtained from 2D NOESY, 2D HMQC and 3D HMQC-NOESY datasets recorded with unlabeled and nucleotide-specific (AU^{CN}-SL4 and GC^{CN}-SL4) selectively labeled samples²⁶. The exchangeable signals were assigned using 2D NOESY, HSQC and 3D HSQC-NOESY data recorded with unlabeled and uniformly ¹⁵N-labeled samples. Phage-aligned samples (~10 mg/mL) were used to measure dipolar coupling (RDC) data³⁹.

Structure calculations

Structures were calculated as described²⁶ using manually assigned restraints in CYANA.⁴⁰ Cross-helix distance restraints with a 20% weighting coefficient were employed in the helical

segments of the structure to limit the approach of the phosphate groups. Torsion angle restraints for helical stem residues were centered around published A-form values ($\alpha = -68^\circ$, $\beta = 168^\circ$, $\gamma = 54^\circ$, $\epsilon = -153^\circ$, $\zeta = -71^\circ$) with allowed deviations of $\pm 50^\circ$. The ribose torsion angle (δ) was restrained in the range 55° - 190° to reduce the influence of these restraints on the sugar pucker. No torsion angle restraints were applied for the pentaloop and bulge regions. Four restraints per observed hydrogen bond were employed to enforce approximately linear NH-N and NH-O distances of $1.85 (\pm 0.05) \text{ \AA}$. Axial and rhombic components of the orientation tensor were determined from grid search calculations. Molecular images were generated with PyMOL (www.pymol.org).

Isothermal titration calorimetry

Binding constants for the interaction of wild type and mutant 7SK-SL4 (G300-U326) and TAR GNRA tetraloop constructs with argininamide were measured using a ITC-200 microcalorimeter (MicroCal). Briefly, $\sim 0.40 \text{ mM}$ argininamide was titrated into $\sim 0.020 \text{ mM}$ solutions of RNA constructs in 10 mM sodium phosphate, pH 5.2, 10 mM NaCl at 25°C . Titration curves were analyzed using ORIGIN (OriginLab).

ACCESSION NUMBERS:

Coordinates and restraints for the final ensemble of 10 structures of the 7SK-SL4:ARG complex have been deposited in the Protein Data Bank with ID code 2KX8.

ACKNOWLEDGEMENTS: We wish to thank the MIT/Harvard Center for Magnetic Resonance (funded by NIH grant EB-002026) for NMR access.

REFERENCES

1. Goodrich, J. A. & Kugel, J. F. (2006). Non-coding-RNA regulators of RNA polymerase II transcription. *Nat Rev Mol Cell Biol* **7**, 612-6.
2. Saunders, A., Core, L. J. & Lis, J. T. (2006). Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* **7**, 557-67.
3. Peterlin, B. M. & Price, D. H. (2006). Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**, 297-305.
4. Weinberg, R. A. & Penman, S. (1968). Small molecular weight monodisperse nuclear RNA. *J Mol Biol* **38**, 289-304.
5. Zieve, G. & Penman, S. (1976). Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell* **8**, 19-31.
6. Nguyen, V. T., Kiss, T., Michels, A. A. & Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**, 322-5.
7. 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**, 317-22.
8. 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. & Bensaude, O. (2004). Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J* **23**, 2608-19.
9. Yik, J. H., Chen, R., Nishimura, R., Jennings, J. L., Link, A. J. & 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**, 971-82.
10. Barboric, M., Yik, J. H., 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**, 2003-12.
11. Tahirov, T. H., Babayeva, N. D., Varzavand, K., Cooper, J. J., Sedore, S. C. & Price, D. H. (2010). Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature* **465**, 747-51.
12. He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N. J., Alber, T. & Zhou, Q. (2010). HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Mol Cell* **38**, 428-38.
13. Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R. & Benkirane, M. (2010). HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Mol Cell* **38**, 439-51.
14. Sedore, S. C., Byers, S. A., Biglione, S., Price, J. P., Maury, W. J. & Price, D. H. (2007). Manipulation of P-TEFb control machinery by HIV: recruitment of P-TEFb from the large form by Tat and binding of HEXIM1 to TAR. *Nucleic Acids Res* **35**, 4347-58.
15. Richter, S., Ping, Y. H. & Rana, T. M. (2002). TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. *Proc Natl Acad Sci U S A* **99**, 7928-33.
16. Zhang, J., Tamilarasu, N., Hwang, S., Garber, M. E., Huq, I., Jones, K. A. & Rana, T. M. (2000). HIV-1 TAR RNA enhances the interaction between Tat and cyclin T1. *J Biol Chem* **275**, 34314-9.

17. Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991). Arginine-mediated RNA recognition: the arginine fork. *Science* **252**, 1167-71.
18. Tao, J. & Frankel, A. D. (1992). Specific binding of arginine to TAR RNA. *Proc Natl Acad Sci U S A* **89**, 2723-6.
19. Aboul-ela, F., Karn, J. & Varani, G. (1996). Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res* **24**, 3974-81.
20. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**, 76-80.
21. Puglisi, J. D., Chen, L., Frankel, A. D. & Williamson, J. R. (1993). Role of RNA structure in arginine recognition of TAR RNA. *Proc Natl Acad Sci U S A* **90**, 3680-4.
22. Brodsky, A. S. & Williamson, J. R. (1997). Solution structure of the HIV-2 TAR-argininamide complex. *J Mol Biol* **267**, 624-39.
23. Pitt, S. W., Majumdar, A., Serganov, A., Patel, D. J. & Al-Hashimi, H. M. (2004). Argininamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg²⁺-induced conformational stabilization. *J Mol Biol* **338**, 7-16.
24. Wassarman, D. A. & Steitz, J. A. (1991). Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol Cell Biol* **11**, 3432-45.
25. 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**, 630-42.
26. D'Souza, V., Dey, A., Habib, D. & Summers, M. F. (2004). NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* **337**, 427-42.
27. 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. *J Mol Biol* **253**, 313-32.
28. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonnrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* **407**, 327-39.
29. Nixon, P. L., Rangan, A., Kim, Y. G., Rich, A., Hoffman, D. W., Hennig, M. & Giedroc, D. P. (2002). Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J Mol Biol* **322**, 621-33.
30. Varani, G. & McClain, W. H. (2000). The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep* **1**, 18-23.
31. Anand, K., Schulte, A., Vogel-Bachmayr, K., Scheffzek, K. & Geyer, M. (2008). Structural insights into the cyclin T1-Tat-TAR RNA transcription activation complex from EIAV. *Nat Struct Mol Biol* **15**, 1287-92.
32. Hoffman, D. W. & White, S. W. (1995). NMR analysis of the trans-activation response (TAR) RNA element of equine infectious anemia virus. *Nucleic Acids Res* **23**, 4058-65.
33. Greene, K. L., Wang, Y. & Live, D. (1995). Influence of the glycosidic torsion angle on ¹³C and ¹⁵N shifts in guanosine nucleotides: investigations of G-tetrad models with alternating syn and anti bases. *J Biomol NMR* **5**, 333-8.

34. Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J., Chabot, B., Poirier, G. G., Hughes, T. R., Blanchette, M., Price, D. H. & Coulombe, B. (2007). Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol Cell* **27**, 262-74.
35. He, N., Jahchan, N. S., Hong, E., Li, Q., Bayfield, M. A., Maraia, R. J., 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**, 588-99.
36. Yang, Z., Yik, J. H., Chen, R., He, N., Jang, M. K., Ozato, K. & Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* **19**, 535-45.
37. Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **330**, 489-93.
38. Williamson, J. R. (2000). Induced fit in RNA-protein recognition. *Nat Struct Biol* **7**, 834-7.
39. Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* **5**, 1065-74.
40. Guntert, P., Mumenthaler, C. & Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**, 283-98.

FIGURE LEGENDS:

Figure 1. Secondary structure and assignments of 7SK-SL4. (a) Secondary structure of HIV-1 TAR and 7SK-SL4. (b) Imino region of a 2D NOESY spectrum of 7SK-SL4 recorded at 278K. The sequential assignment pathways are indicated in red (stem 2) and blue (stem3). Most importantly, the terminal U319 and G305 residues are base paired. (c) Excerpt of ^{13}C - ^1H NOESY-HMQC sequential assignment data for 7SK-SL4 showing connectivities for the H1' and H6/H8 protons in the G318-C323 segment. The asterisk denotes a break in the canonical connectivity pattern between bulge residues U319 and C320.

Figure 2. Overall structure of 7SK-SL4 and detailed views of the preformed binding pocket. (a) NMR structure ensemble of 7SK-SL4 with cap residue U319 and base residue C320 colored magenta. (b) Close-up view of the electrostatic surface of the bulge region showing the preformed pocket. (c) Ensemble showing a close-up of the binding pocket and the potential C320:C323-G303 base-triple.

Figure 3. NMR characterization of the pseudo-triple platform. (a) Chemical structure of the potential C320⁺:C323-G303 base-triple. (b) Excerpt of hydrogen bonded (left column) and non-hydrogen bonded (right column) amino regions of ^{15}N - ^1H HSQC spectra at various pH values. (c) Amino region of ^{15}N - ^1H HSQC spectrum of 7SK-SL4 at pH 5.0 showing the N4 amino group signals of C320.

Figure 4. Tight ligand binding and preformed structure of 7SK-SL4. (a) ITC calorimetric output (top) and binding isotherm (bottom) for titration of minimal GNRA wild type 7SK-SL4

construct and argininamide. **(b)** Overlay of ^{13}C - ^1H HMQC spectra and assignments of free 7SK-SL4 (black) and complexed with argininamide (red).

Figure 5. Structural characterization of argininamide binding to 7SK-SL4. **(a)** Intermolecular NOEs observed in the complex of a minimal GNRA 7SK-SL4 construct titrated with a five-fold excess of argininamide. NOE interactions between the argininamide $\text{H}\beta$, $\text{H}\gamma$ and $\text{H}\delta$ protons to 7SK-SL4 residues in the binding pocket are indicated by dotted lines. **(b)** Close-up view of the binding pocket in the structure ensemble showing the C323-G303 base pair and residue G322 (grey), U319 and C320 (purple), and argininamide (dark blue). The lines indicate the arginine fork hydrogen bond interactions with the two argininamide $\text{N}\eta$ groups. **(c)** Shape complementarity of the complex; argininamide is in blue.

Figure 6. Thermodynamic analysis of argininamide binding to 7SK-SL4. **(a)** Secondary structure of the minimal GNRA 7SK-SL4 construct used for comparative ITC analysis with mutants indicated as U319C (blue), C320U (red) C320U:C323A-G303U (green). **(b)** Overlay of ITC calorimetric output (top) and binding isotherms (bottom) for wild type and mutant 7SK-SL4 constructs. **(c)** Bar graph representation of thermodynamic data for wild type and mutant 7SK-SL4 constructs.

Figure 7. Comparative structural alignment of 7SK-SL4 and EIAV TAR loops. **(a)** NMR structure ensemble showing the convergence of the pentaloop. **(b)** Comparison of the 7SK-SL4 pentaloop ($\text{A}_{310}\text{UGUG}$) with the bound form of the equine infectious anemia virus (EIAV) TAR hexaloop ($\text{UC}_{12}\text{UGCG}$) bound to cyclin T1 and Tat.

Table 1. NMR restraints and statistics for the final 7SK and 7SK-SL4:ARG structures

	7SK-SL4	7SK-SL4:ARG
Distance and dihedral restraints		
Total distance restraints	579	590
Intra-residue NOE	188	188
Inter-residue NOE	150	150
Inter-molecular NOE	-	11
Hydrogen bond	164	164
A-form geometry	77	77
Dihedral angle restraints	193	193
RDC restraints ^a	39	39
Structure statistics		
Target function (mean \pm SD, \AA^2)	0.58 \pm 0.10	0.16 \pm 0.00
Distance violations ($> 2 \text{\AA}$)	1	0
Dihedral angle violations ($> 5^\circ$)	0	1
van der Waals violations ($> 0.1 \text{\AA}$)	1	1
RDC violations ($> 0 \text{ Hz}$)	0	0
Structure convergence		
Base paired heavy atoms ^b (mean \pm SD, \AA)	0.38 \pm 0.13	0.42 \pm 0.10

^aThe axial (D_a) and rhombic (R) component of the alignment tensor used in the structure calculations are -32 Hz and 3%, respectively. ^bPairwise r.m.s deviation was calculated among 10 refined structures, which represent 5% of the starting structures.