A structure-based mechanism for HEXIM displacement from 7SK

Vincent V. Pham¹, Michael Gao¹, Jennifer L. Meagher², Janet L. Smith²³ & Victoria M. D’Souza*¹

¹Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, 02138, USA. ²Life Sciences Institute, University of Michigan, Ann Arbor, MI, 48109, USA. ³Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, 48109, USA. *Correspondence and requests for materials should be addressed to V.M.D (email: dsouza@mcb.harvard.edu).

Productive transcriptional elongation of many cellular and viral mRNAs requires transcriptional factors to extract pTEFb from the 7SK snRNP by modulating the association between the HEXIM protein and the 7SK snRNA. Here we report the structure of the HEXIM arginine rich motif in complex with the apical stemloop-1 of 7SK (7SK-SL1apical) and detail how the HIV transcriptional regulator Tat from various subtypes overcome the structural constraints required to displace HEXIM. While the majority of interactions between 7SK and HEXIM and Tat are similar, critical differences exist that guide function. First, the conformational plasticity of 7SK enables the formation of three different base pair configurations at a critical remodeling site, which allows for the modulation required for HEXIM binding and its subsequent displacement by Tat. Furthermore, the specific sequence variations observed in various Tat subtypes all converge on remodeling 7SK at this region. Second, we show that HEXIM primes its own displacement by causing specific local destabilization upon binding — a feature that is then exploited by Tat to bind 7SK more efficiently. Overall, our study details the molecular environment presented by HEXIM and uncovers a destabilization-driven displacement strategy that increases the conformational sampling of 7SK-snRNP, which may allow diverse transcriptional factors to competitively regulate pTEFb.
Transcription of all class II genes is a highly regulated process within cells. Shortly after promoter clearance, RNA Polymerase II is inhibited by the negative elongation factors\(^1\text{-}^5\). Release from this stalled state requires all components to be phosphorylated by the positive elongation factor pTEFb, a heterodimeric complex consisting of Cyclin T1 and the cyclin-dependent kinase Cdk\(^9\text{-}^{11}\). However, most of the pTEFb is kept catalytically inactive in the nucleus by the 7SK small nuclear ribonucleoprotein (7SK snRNP) through its interactions with the HEXIM adapter protein and the 5’ stemloop-1 of the 7SK RNA\(^{12\text{-}^{18}}\) (7SK-SL1; Fig. 1a). Thus, productive transcriptional elongation of many genes requires transcriptional factors to extract pTEFb from the 7SK snRNP — a process that involves manipulating the interaction between HEXIM and 7SK. This association between 7SK and HEXIM tightly controls the balance between active and inactive pTEFb, and dysregulation of this interaction can have serious biological consequences including cardiac hypertrophy and breast and pancreatic cancers\(^{19\text{-}^{21}}\). Furthermore, as many viruses rely on the host transcriptional machinery to produce mRNA and genomes, they have also evolved mechanisms to capture pTEFb\(^{22\text{-}^{24}}\). One such unique case is the human immunodeficiency virus (HIV), which utilizes the viral Tat protein to extract pTEFb by binding to the same region of 7SK as HEXIM and directly displacing it\(^{24}\). Structural insights into the consequence of HEXIM binding to 7SK and how positive transcriptional factors like Tat compete with it are therefore important for understanding HEXIM’s potency as a critical negative regulator.

To date, two HEXIM proteins have been identified that are capable of carrying out the same function and both bind 7SK with identical regions of their Arginine Rich Motifs (ARMs) (residues 151-159 in HEXIM1 and 89-97 in HEXIM2)\(^{24\text{-}^{27}}\). Although HEXIM binds 7SK as a dimer, only one ARM directly contacts 7SK by engaging the apical region of stemloop-1 (G\(_{26}\) to C\(_{85}\), 7SK-SL1\(^{\text{apical}}\); Fig. 1a)\(^{28\text{-}^{33}}\). Both in vitro and in vivo studies have
shown that this represents the sole interaction between the two molecules that must be
modulated in order to release pTEFb\textsuperscript{27-29,31,32}.

Our previous work showed that 7SK-SL1\textsubscript{apical} is enriched in arginine sandwich
motifs (ASMs)\textsuperscript{34}. ASMs are defined by two nucleotides that stack in a manner that allows
for intercalation of arginine guanidinium moieties between the aromatic rings of the
bases\textsuperscript{34-37} (Fig. 2a). While a bulge pyrimidine forms the cap by engaging in a triple base
interaction with an n + 2 base pair in the stem, a Watson-Crick base paired nucleotide
preceding the bulge forms the base of the interaction. In the free 7SK-SL1\textsubscript{apical}, three such
bulges fold into preformed arginine sandwich motifs (ASM\textsubscript{1}, ASM\textsubscript{2} and ASM\textsubscript{4}) poised for
arginine guanidinium moieties to dock into them. A fourth bulge folds into a pseudo
configuration (pseudo-ASM\textsubscript{3}) where U\textsubscript{40} is able to form a triple base interaction with the
A\textsubscript{43}-U\textsubscript{66} base pair to form the cap, but the base of the sandwich is sequestered in a reverse
Hoogsteen interaction, excluding it from use as a classical ASM. Our work also showed
that HIV-1 Tat NL4-3 (Tat\textsuperscript{NL4-3}) uses its arginine-rich motif to intercalate arginines not only
into the three preformed ASMs, but also to remodel the pseudo-ASM into a classical
ASM\textsuperscript{34}. This structural remodeling of pseudo-ASM\textsubscript{3} is a key mechanism through which
Tat displaces HEXIM.

However, without the structure of the HEXIM:7SK-SL1\textsubscript{apical} interaction, it is
currently unclear what structural constraints Tat would need to overcome in order to
access pTEFb. Furthermore, while the Tat ARM is highly conserved, sequence variations
exist in different strains that allow for HEXIM displacement. For example, the ARM of
Tat Finland (Tat\textsuperscript{Fin}; KR\textsubscript{52}KHRRR) differs from HEXIM (KK\textsubscript{151}KHRRR) by only a single
amino acid and would lack one of the ASM interactions from the previously described
Tat NL4-3 strain (KR\textsubscript{52}RQRRR). Additionally, while Tat subtype G (Tat\textsuperscript{G}; KR\textsubscript{52}R\textsubscript{53}HRRR)
has an equivalent number of arginines as Tat\textsuperscript{NL4-3}, the critical linker sequence connecting
ASM\textsubscript{3}/ASM\textsubscript{4} and ASM\textsubscript{1}/ASM\textsubscript{2} interactions is the same as HEXIM. In this study, we present
the structure of the 7SK-SL1⁰ᵖⁱᶜᵃˡ in complex with the HEXIM, Tat⁰ⁿ, and Tat⁰ᵍ ARMs. Despite sequence variations, the structures show deep major groove intercalations of all ARMs, albeit with differential interactions with pseudo-ASM₃ and ASM₄. Furthermore, we show that HEXIM causes local destabilization of ASM₄, enhancing Tat’s affinity for 7SK. These studies thus uncover a feature in which HEXIM facilitates its own displacement by increasing conformational sampling, which may be a more general mechanism of pTEFb capture.

Results

Comparative binding affinities of HEXIM and Tat to 7SK.

As a first step toward identifying the atomic determinants of 7SK recognition by HEXIM, we performed binding studies using isothermal titration calorimetry (ITC). Studies have shown that the first 108 nucleotides comprising stemloop-1 of 7SK (G₁₋C₁⁰⁸; 7SK-SL1⁰ⁿ) are required for binding full length HEXIM as a dimer both in vivo and in vitro²⁴,²⁷,³². ITC traces of full-length HEXIM₁ with 7SK-SL1⁰ⁿ corroborate these studies (N = 2.1 ± 0.2, Kᵢ = 209 ± 34 nM; Fig. 1b). Additionally, ITC traces of full-length HEXIM₁ to 7SK-SL1⁰ⁿ-AGU and 7SK-SL1⁰ᵖⁱᶜᵃˡ (G₂₆ to C₈₅), both containing an AGU triloop engineered to prevent low levels of dimerization and thus aid in increasing quality of NMR spectra, also reveal dimeric HEXIM binding to both constructs with comparable binding affinities to the full-length HEXIM₁: 7SK-SL1⁰ⁿ complex (N = 2 ± 0.1, Kᵢ = 200 ± 20 nM, N = 1.8 ± 0.03, Kᵢ = 206 ± 58 nM, and N = 2.1 ± 0.2, Kᵢ = 209 ± 34 nM, respectively; Fig. 1c, d), indicating that the loop does not play a significant role in HEXIM₁ binding and confirming that the HEXIM:7SK interaction is confined to the apical region stem of 7SK-SL1.
Previous studies have also shown that while HEXIM binds 7SK as a dimer, the ARM of only one HEXIM monomer is responsible for interacting with 7SK both in vivo and in vitro. ITC traces of the N-terminal ARM residues (R\textsubscript{146}QLGKKKHRRR\textsubscript{156}; HEXIM\textsuperscript{N-ARM}) to 7SK-SL\textsubscript{1}apical with the AGU triloop confirm these findings as the binding affinities between the full-length dimeric HEXIM\textsubscript{1}:7SK-SL\textsubscript{1}apical and HEXIM\textsuperscript{N-ARM}:7SK-SL\textsubscript{1}apical complexes are similar (N = 1.8 ± 0.03, K\textsubscript{d} = 206 ± 58 nM and N = 1 ± 0.1, K\textsubscript{d} = 238 ± 35 nM, respectively; Fig. 1d, e). NMR studies comparing these two complexes show that binding of either full-length HEXIM or the N-ARM gives rise to the same nuclear Overhauser effect (NOE) shifts in 7SK-SL\textsubscript{1}apical, indicating that the HEXIM\textsuperscript{N-ARM}:7SK-SL\textsubscript{1}apical interaction represents the biologically relevant mode of HEXIM binding to 7SK (Supplementary Fig. 1).

Finally, our previous work showed that the Tat\textsuperscript{NL4-3} (KR\textsubscript{52}RQRRR) ARM represents the interaction domain between Tat and 7SK-SL\textsubscript{1}apical and has an approximately two-fold increased affinity over the HEXIM\textsuperscript{N-ARM}. ITC traces show that Tat Subtype G’s ARM (KR\textsubscript{52}RHRRR), which also has two N-terminal arginines, binds 7SK-SL\textsubscript{1}apical with a K\textsubscript{d} of 91 ± 32 nM (N = 1.1 ± 0.1; Fig. 1f), which is an approximately 2.6-fold increased binding affinity over HEXIM\textsuperscript{N-ARM} (Supplementary Table 1). On the other hand, Tat Finland’s ARM (KR\textsubscript{52}KHRRR), despite have an additional N-terminal arginine compared to the HEXIM\textsuperscript{N-ARM} (R52 and K151, respectively), does not have a statistically significant increase in binding affinity (K\textsubscript{d} of 198 ± 16 nM; Fig. 1g). Overall, these results highlight the need for understanding the HEXIM-bound 7SK; while the increased Tat\textsuperscript{G} affinity would allow for HEXIM displacement, it is unclear how Tat\textsuperscript{Fin} can achieve the same biological output.
Preformed configurations of ASM\(_1\) and ASM\(_2\) provide a common mode of interaction with C-terminal residues.

To understand how the HEXIM\(^{N-ARM}\) and the various Tat ARMs interact with 7SK-SL\(_1^{\text{apical}}\), we utilized a combination of small-angle X-ray scattering (SAXS) and NMR. All reconstructed ab initio SAXS envelopes showed no major overall global changes between peptide-bound and free 7SK-SL\(_1^{\text{apical}}\) (Fig. 2b, c). Numerous intermolecular NOEs place both HEXIM and Tat arginine rich motifs into the major groove of the RNA and allow us to define their interactions with all ASM regions. Base pairs in the lower part of the stem-loop below the G\(_{79}\)-U\(_{32}\) base pair, as well as the CAGUG pentaloop do not give any intermolecular NOEs, indicating that the interactions are contained within a single turn of the helix (Fig. 2d-f).

In the free 7SK-SL\(_1^{\text{apical}}\), ASM\(_1\) and ASM\(_2\) are placed in a tandem orientation and upon titration of the various ARMs, all expected NOEs for such configurations are retained. Unlike a typical ASM where the following nucleotide after the bulge is in a canonical Watson-Crick base pair, in ASM\(_1\), the residue A\(_{77}\) is configured into an A\(_{34}\)-A\(_{77}\) pair. A NOE from the A\(_{77}\) H8 proton to the H1’ of C\(_{75}\) positions this residue under the C\(_{75}\) cap (Supplementary Fig. 2). This confirms a planar orientation of C\(_{75}\) with the C\(_{33}\)-G\(_{78}\) base pair and configures A\(_{77}\) in such a way that it is perfectly positioned to interact with the guanidinium moiety of R156 in HEXIM\(^{N-ARM}\) and R57 in Tat\(^{\text{Fin}}\) and Tat\(^{G}\), which intercalate between C\(_{75}\) and G\(_{74}\) in a manner identical to canonical ASMs (Supplementary Fig. 2-5).

Similarly, in ASM\(_2\), the C\(_{71}\)\(^+\) base also retains its protonation at the N3 position as evidenced by a downfield shift of the N4 amino protons (Supplementary Fig. 6). The guanidinium moiety of R155 in HEXIM\(^{N-ARM}\) and R56 of Tat\(^{\text{Fin}}\) and Tat\(^{G}\) interact with G\(_{73}\) by intercalating between the C\(_{71}\)\(^+\) cap and G\(_{70}\) base of the motif (Supplementary Fig. 3-5). Additionally, intermolecular NOEs from the aromatic protons of the C\(_{75}\) and C\(_{71}\)\(^+\) caps and
the G_{74} and G_{70} bases of ASM_1 and ASM_2 to the H\gamma and the H\delta protons confirm that consecutive arginines R156 and R155 interact in a ladder-like configuration with the tandem preformed motifs ASM_1 and ASM_2, respectively (Fig. 3a and Supplementary Fig. 4c). Such NOEs are also observed in both the Tat^Fin and the Tat^G-bound complexes, confirming the similar placement of the C-terminal R57 and R56 into the tandem ASM_1 and ASM_2, respectively (Fig. 3a and Supplementary Fig. 5a, b, d, e). Taken together, the structures reveal a common mode of interaction between the non-varying C-terminal arginines and the tandem ASMs.

Rearrangement of pseudo-ASM_3 allows for HEXIM N-terminal interactions.

In the free 7SK-SL1^apical, pseudo-ASM_3 and ASM_1 adopt a pseudo-symmetrical architecture where the two motifs are spatially opposed. Upon HEXIM^N-ARM binding, the pseudo-ASM_3 maintains its U_{40}:A_{43}-U_{66} triple base interaction although the base of the sandwich, A_{39}, rearranges from a reverse Hoogsteen interaction with U_{68} into a cis Hoogsteen/sugar interaction, giving rise to an alternate pseudo configuration. (Fig. 3b,f). This is evidenced both by NOEs from the U_{68} imino proton to the A_{39} amino protons and NOEs from the U_{68} H2' and H3' protons to the A_{39} H8 proton (Supplementary Fig. 6b). This frees up the U_{68} imino proton to engage the backbone carbonyl of K152 while simultaneously bringing the N1 proton acceptor of A_{39} into the major groove to hydrogen bond with the side chain H\epsilon protons of K151 (Fig. 3b). Thus, both K151 and 152 can enter deep into the major groove by remodeling the pseudo-ASM_3.

The amino side chain of K151 is within hydrogen-bonding distance of the A_{39} N1 nitrogen as evidenced by NOEs from the K151 H\gamma and H\beta protons to the C_{37} H6 and H5 protons, respectively and from the K151 H\epsilon protons to the C_{38} H6 and H5 protons (Supplementary Fig. 4f). Additionally, NOEs between the K152 H\beta protons with the U_{68}
H5 proton, the K152 Hδ protons with the C67 and U66 H5 protons, and the K152 Hε protons with the C67 H5 and H6 protons position the amino side chain of K152 within hydrogen-bonding distance of the C67 backbone (Fig. 3b and Supplementary Fig. 4e, f). This gives rise to a forked configuration of the two lysines, orienting the side chain amino groups towards the phosphate backbones on opposite ends of the groove.

Unlike the other three ASMs, where the NOEs clearly define a single predominant structural configuration as described above, multiple dynamic states exist for ASM4 (see below). In the most abundant form, the preformed nature found in the free state is retained as evidenced by a direct imino to imino connectivity between U44 and U63, along with maintenance of the G46-C62 Watson-Crick base pair (Supplementary Fig. 6c). In fact, this interaction is stabilized by K150, which displays NOEs between the Hε protons with the U63 and the U40 H5 protons, positioning the amino side chain within hydrogen bonding distance of the O4 atoms of both U63 and U40 (Supplementary Fig. 4d). Additional intermolecular interactions between the U40 H5 proton and the U63 H5 and H1′ protons with the K150 Hδ, Hγ, and Hβ protons places the K150 directly under the U63 cap of ASM4 (Supplementary Fig. 4d, e). Taken together, these data show that despite the lack of arginines, the lysine-rich N-terminus of HEXIMN-ARM is able to be accommodated by 7SK: the Watson-Crick face of A39 turns from the minor into the major groove to interact with K151 and 152, which then positions the K150 to interact with the oxygen-rich environment of the U63 and U40 caps.

Conformational plasticity of the ASM3/ASM4 region provides differential mode of interactions with N-terminal and spacer residues.

Our previous study showed that TatNL4-3 displaces HEXIM by remodeling the pseudo-ASM3 into a canonical ASM3 to allow for arginine intercalation34. Furthermore,
an additional arginine docks into the preformed ASM. While the mechanism of remodeling pseudo-ASM is conserved upon binding of both TatFin and TatG ARMs (Fig. 3c, e, g, h; Supplementary Fig. 6), both the drivers of the conformational switch and the engagement of the ASM vary depending on differences in amino acid sequences.

While TatFin has two major differences from TatNL4-3 (K53 to R53 and spacer H54 to Q54, respectively), it only differs in a single amino acid from HEXIM (R52 and K151, respectively). Like TatNL4-3, R52 is responsible for remodeling pseudo-ASM (Fig. 3c; Supplementary Fig. 5a). However, while R53 in TatNL4-3 flips over R52 and engages ASM, the equivalent K53 stays in the spacer region between the ASM1/ASM2 and ASM3/ASM4 regions in a manner similar to HEXIM as evidenced by NOEs between the K53 Hβ protons with the U68 H5 proton, the K53 Hδ protons with the C67 and U66 H5 protons, and the K53 Hε protons with the C67 H5 and H6 protons, which position the amino side chain of K53 within hydrogen-bonding distance of the C67 backbone (Fig. 3j; Supplementary Fig. 5a).

Like HEXIM, ASM remains unoccupied upon binding TatFin and the structure shows that the K51 amino side chain is positioned to hydrogen-bond with the U63 ribose ring in a stabilizing interaction (Fig. 3c). This is evidenced by NOEs between the K51 Hδ protons with the U63 H5 and H1′ protons and the K51 Hε protons with the U63 2′ hydroxyl proton (Supplementary Fig. 5c). Furthermore, the N-terminal K50 exits near the apical loop with NOEs observed between the K50 Hδ and Hε protons with the C38 and C37 H5 and H1′ protons position the amino side chain of K50 to the C38 phosphate backbone (Supplementary Fig. 5a).

Finally, in evaluating the structural consequences of the spacer substitution, we see that H54 and R55 in TatFin remain near ASM and ASM2, similar to what is found in HEXIM. This is evidenced by NOEs between the H54 (H153 in HEXIM) Hβ protons with
the C35, C36, and C37 H5 protons, placing H54 near ASM2 whereas the R55 (R154 in HEXIM) Hδ protons display NOEs with the A34 H1′ proton and the C33 H1′, H5, and H6 protons, positioning this spacer residue near ASM1 (Fig. 3i, j; Supplementary Fig. 4, 5). This is in contrast with the binding mode of TatNL4-3 in which the intercalation of R53 into ASM4 drags both the Q54 and R55 spacer residues towards the apical ASMs.

The importance of the histidine H54 spacer is even more evident in the TatG strain where it represents the only difference between TatNL4-3. This single difference changes the identity of the arginine that remodels pseudo-ASM3. In this ARM, the positioning of H54 near ASM2 precludes R53 from reaching ASM4 to accomplish the inverse intercalation seen in NL4-3 (Fig. 3d, e, k and Supplementary Fig. 5d, e). The interactions with the apical ASMs thus occur in a ladder-like manner where R53 intercalates into the remodeled ASM3 whereas R52 intercalates into ASM4 (Fig. 3d, e and Supplementary Fig. 5a, d, e). K51 makes the final stabilizing interaction with NOEs seen between the He protons with the U63 2′ hydroxyl proton, indicating a hydrogen-bonding interaction between the K51 amino side chain with the U63 ribose ring (Fig. 3d and Supplementary Fig. 5f). Taken together, these studies show that arginine sandwich motifs provide mini domains that arginine rich motifs of proteins can differentially interact with in order to achieve deep major groove binding into the stem of in 7SK-SL1 apical.

**HEXIM allows for increased conformational sampling of apical ASMs.**

While titration of all four arginine rich motifs stabilize the majority of 7SK-SL1 apical into one predominant configuration, the HEXIM ARM is an outlier wherein binding causes ASM4 to become destabilized and exhibit multiple conformations. In such conformations, the NOEs between the imino protons of U63 and U44 disappear, indicating the disruption of the U63:U44-A65 triple and loss of ASM4 (Supplementary Fig. 6c). The
destabilization of this region is also indicated by the line-broadening of K150, which interacts with U63 in the folded configuration (Supplementary Fig. 4e).

The destabilization of 7SK-SL1 apical only by HEXIM is further evident when comparing the thermodynamic profiles between Tat and HEXIM. The binding of TatFin and TatG strains are enthalpically driven ($\Delta H = -7.5 \pm 0.2$ and $-8.9 \pm 2.2$ kcal mol$^{-1}$, respectively; Fig. 4a) with a modest entropic contribution ($-T\Delta S = -1.7 \pm 0.3$ and $-2 \pm 0.8$ kcal mol$^{-1}$, respectively; Fig. 4a). On the other hand, HEXIM binding is entropically enhanced by approximately 2.5-fold over both Tat strains ($-T\Delta S = -4.6 \pm 0.8$ kcal mol$^{-1}$, $\Delta H = -4.4 \pm 0.7$ kcal mol$^{-1}$; Fig. 4a).

To evaluate the implication of HEXIM’s ability to locally destabilize ASM4 in the context of its displacement required for transcriptional regulation, we compared TatFin and TatG binding to 7SK both free and in the presence of HEXIM. Due to the modest differences in binding energetics between the different ARMs, competition experiments using ITC were not tractable. As seen in Fig. 4b, a 1:1 titration of either HEXIM or Tat into 7SK in the NMR shows the presence of both bound and free forms of 7SK. However, upon titration of TatFin and TatG into the HEXIM:7SK complex, we not only observe completion of complex formation but also a total displacement of HEXIM in both cases. This is especially striking given that the binding affinities of TatFin and HEXIM for free 7SK are equivalent. Taken together, these data indicate that Tat has a greater affinity for a HEXIM-bound 7SK complex. Finally, ITC data of full-length HEXIM bound to full-length 7SK snRNA ($N=1.9 \pm 0.1$; Fig. 4b) show that an entropy-driven interaction is maintained, and in fact, is even more pronounced ($-T\Delta S = -6.4 \pm 1.3$ kcal mol$^{-1}$, $\Delta H = -2.6 \pm 1$ kcal mol$^{-1}$Fig. 4a), suggesting that HEXIM binding may globally increase the conformational space sampled by the 7SK-snRNP complex. These studies suggest that destabilization by HEXIM may play an important role in how transcription factors access 7SK for pTEFb capture.
Discussion

The 7SK snRNP represents a central biomolecule that a wide range of transcriptional factors need to interact with to access pTEFb to control transcriptional elongation. In particular, pTEFb extraction by HIV Tat from this complex requires manipulating the interaction between the 7SK snRNA and the HEXIM adapter protein.

In this study, we solved the structures of the RNA binding domains of HEXIM and Tat bound to 7SK and gain several insights into their functional significance, including the malleability of 7SK, the local destabilization by HEXIM, and the specific sequence variations of Tat.

The structures show that both HEXIM and Tat regulate pTEFb by directly binding the stem of 7SK-SL1 apical through intercalation of arginine rich motifs into an entire helical turn of the major groove. This is unusual as RNA major grooves are deep and narrow, making them generally inaccessible for protein binding. The architecture of the four sandwich motifs in 7SK allows for transcriptional regulators to differentially utilize their ARMs. The tandem preformed ASMs, ASM1 and ASM2, remain unchanged from their free configuration upon encountering C-terminal arginines of Tat and HEXIM. On the other hand, the apical pseudo-symmetrical ASMs, pseudo-ASM3 and ASM4, reconfigure depending on their binding partners. The structures show that the ASM3 region can adopt at least three different base pair interactions: a reverse Hoogsteen in the free state, a cis Hoogsteen/sugar interaction upon HEXIM binding, and a Watson-Crick base pair upon Tat binding. The cis Hoogsteen/sugar interaction is especially significant because it allows HEXIM to enter the major groove despite the lack of arginines in the N-terminus. Similarly, while ASM4 retains its preformed configuration found in the free state upon Tat binding, it can be destabilized in the presence of HEXIM and adopt multiple flexible
states. Taken together, these studies show that 7SK is adaptable in its ASM architecture, which can be modulated upon encountering different transcription factors.

Comparative analyses of HEXIM and Tat provide insights into how both positive and negative regulators can manipulate 7SK to carry out their transcription roles. Our studies implicate HEXIM as potentially having a dual structural role. On the one hand, it can bind with high affinity to the apical portion of 7SK-stemloop 1 to sequester pTEFb and on the other hand, it simultaneously causes local destabilization of this region, enhancing the binding affinities of positive regulators for pTEFb capture. In comparison to Tat, the thermodynamic profile and solution-state characteristics of HEXIM binding show an entropy-driven mode of interaction that is particularly attributed to the destabilization of U₆₃ in the ASM₄ region. Indeed, mutational studies have shown that deletion of U₆₃ significantly reduces HEXIM binding. This expansion in the dynamic state of 7SK surrounding the ASM₄ region is also supported both by in-vivo SHAPE analysis where U₆₃ becomes ultra-reactive upon HEXIM:pTEFb binding, and structural and molecular dynamics modeling. Furthermore, we show that Tat capitalizes on this increased dynamic state, binding with greater affinity to the HEXIM-bound complex than to free 7SK. While the use of a HEXIM-displacement mechanism for pTEFb capture by binding to 7SK-SL1 has yet to be discovered for cellular factors, the destabilization-driven preparation of 7SK snRNP for pTEFb extraction may be a general feature exploited by specialized transcriptional factors.

Comparative analysis of HEXIM and Tat also shed light on the sequence requirements of ARMs for pTEFb regulation. While N-terminal lysines of HEXIM allow for destabilization of ASM₄, the anchoring required to enter the major groove can only be provided by the stacking of C-terminal arginines within ASM₁ and ASM₂. Indeed, the importance of these C-terminal arginines for HEXIM binding is supported by their nearly complete conservation across metazoan species. On the other hand, the equivalent
arginines in HIV-1 Tat occur as a consecutive pair only in approximately 50% of reported strains, albeit with the strong requirement of at least one arginine. The structures show that these variations may be possible due to the anchoring provided by the arginines that intercalate into the apical ASMs. Nevertheless, when two arginines are present in Tat, the interactions with the tandem ASMs mirror HEXIM.

Furthermore, differences in N-terminal and spacer ARM residues orchestrate the structural modulations of the apical ASMs. In order to accommodate the continuation of the HEXIM ARM chain from the ASM1/ASM2 to the ASM3/ASM4 region required for U63 destabilization, K152 induces the reconfiguration of pseudo-ASM3 from a reverse Hoogsteen to a cis-Hoogsteen/sugar interaction. In all variations of N-terminal Tat sequences studied, binding is concomitant with the rearrangement of pseudo-ASM3 into a canonical ASM3 through the intercalation of an arginine.

The structures also provide insights into specific sequence variations that occur in the highly conserved Tat ARM to displace HEXIM. When two arginines are available in the N-terminal residues, both are involved in arginine sandwich interactions, providing a 2-fold increase in affinity; however, either R52 (TatNL4-3) or R53 (TatG) can act as the remodeler. This can be explained by the presence of either a glutamine or histidine spacer, respectively, which is the only amino acid difference between the two strains. As glutamine (75%) and histidine (15%) make up the majority of the sequence variation in this spacer, the structures show that these two spacer residues drive the differential positioning of the arginine remodeler. In the TatFin strain, which has a histidine spacer, it is the R52 that acts as a remodeler. In this case, the R53K substitution provides the stabilizing interactions to reposition the single R52 arginine near pseudo-ASM3.

Furthermore, it is also interesting to compare the mode of binding of TatFin to HEXIM. First, the single residue difference (R52 vs K151) provides TatFin with the
additional ASM intercalation required for displacement. Thus, Tat has evolved specific
sequence variations that allow for the reconfiguration of pseudo-ASMs, even in cases
where there is only a single variation from HEXIM. Second, despite both ARMs having
lysines positioned near ASMs, only HEXIM leads to local destabilization. Our studies
therefore provide HEXIM as an example of a negative regulator that primes its own
displacement by locally destabilizing 7SK (Fig. 4c). Overall, these studies have broader
implications for 7SK-snRNP mediated regulation. Given that the destabilization-driven
displacement is a robust mechanism, it is possible that other yet to be identified cellular
and viral transcriptional regulators recruit pTEFb through direct intercalation of ARMs
into 7SK-SL1 apical. Furthermore, as a destabilized state of 7SK snRNP is what is presented
to all transcriptional regulators, the mechanisms necessary to extract pTEFb may
converge on capitalizing on this conformational heterogeneity.

Methods

**RNA sample preparation:** RNA samples used for biophysical experiments were
synthesized by *in vitro transcription* using T7 RNA polymerase with either plasmid DNA
or with synthetic DNA templates containing 2'-O-methylated (Integrated DNA
Technologies) containing the T7 promoter and the desired sequences. Plasmid DNA for
7SK-SL1Full-WT and 7SK-SL1Full-AGU contain the T7 promoter, insert, and SmaI sequence
were cloned by Genscript in between the EcoRI and BamHI restriction sites of a puc19
vector. Plasmid DNA was prepared for *in vitro transcription* from a 5mL overnight culture
of NEB 5α Competent E.coli (C29871) transformed with the plasmid using Qiaprep Spin
Miniprep Kit (Qiagen 27104). 10 µL of purified DNA was combined with 25 µL of 2'-O-
methylated reverse primer at 100 µM (5'-mGmGAGCGGTGAGG GAGGAAG-3' where
m indicates 2’ O-methylated nucleotide), 2 µL of forward primer at 100 µM (5'-
GACAAGCCCGTCAGGG-3'), 2.44 mL of water, and two tubes of EconoTaq PLUS 2X Master Mix (Lucigen 30035-2). The 5 mL mixture was then aliquoted into 50 µL increments in a 96-well PCR plate and the template for in vitro transcription reactions were amplified using the following PCR protocol: 95 °C for 5 minutes, 34 cycles of (95 °C for 30 seconds, 50 °C for 1 minute, and 68 °C for 90 seconds), and 68 °C for 5 minutes. After PCR amplification, reactions were pooled into 5 mL volume in a 50 mL Falcon tube and 0.5 mL of 3M sodium acetate, pH 5 and 32 mL of 100% ethanol was added to the mixture and chilled at -80°C for at least 30 minutes before spinning down at 9000 x g for 10 minutes at 4 °C. The ethanol was decanted and the pellet was left to dry overnight before in vitro transcription use. Template preparation for 7SK-SL1apical using 2′-O-methylated reverse primers in order to suppress the heterogeneity at the 3′ end of the transcripts involved combining 15 mL of both forward (5′-TAATACGACTCACTATAGGGATCTGTCACCCCATTGATCGCCAGTGGCTGATCTGGCTGGCTAGGCGGGTCCC-3′) and reverse (5′-mGmGGACCCGCCTAGCCAGCCAGATCAGCCACTGGCGATCAATGGGGTGACAGATCCCTATAGTGAGTCGTATTA-3′ where m indicates 2′ O-methylated nucleotide) primers at 1 mM stock solution with 470 mL of water. The mixture was heated at 95 °C for five minutes and cooled at room temperature for 30 minutes before assembling the in vitro transcription reaction. Samples were either unlabeled, or residue-specifically labeled with 13C/15N- or 2H (Cambridge Isotope Laboratories, Inc.). After transcription, RNA samples were heat denatured and purified by using urea-denaturing polyacrylamide gels.

**HEXIM ARM and Tat ARM peptide preparation:** Unlabeled HEXIM_{N-ARM} (GISYGRQLGKKKHRRRAHQ), Tat^{Fin ARM} (GISYGRKKRKHRRRAHQ), and Tat^{G ARM} (GISYGRKKRHRHRRRAHQ) peptides were purchased from Tufts University Core Facility at a 0.1 mmol scale. Tat adapters were placed around the HEXIM_{N-ARM} sequence to prevent non-physiological aggregation in solution state NMR studies. HEXIM_{N-ARM}
peptides containing selective $^{13}$C/$^{15}$N-labeled residues, underlined, (GISYGRQLGKKHRRRAHQ and GISYGRQLGKKHRRRAHQ) were purchased from New England Peptide.

**Full-length HEXIM1 preparation:** Synthetic DNA encoding HEXIM1 (2-359) was cloned into a bacterial pMCSG7 expression vector $^{42}$ [JLS1] encoding an N-terminal tobacco etch virus (TEV) protease-cleavable His$_6$ tag and was expressed in *E. coli* BL21 AI cells in an overnight culture at 20°C. Cells were lysed by sonication in buffer containing 50 mM Tris pH 8.0, 500 mM NaCl, 0.1% b-mercaptoethanol, 50 mM (NH$_4$)$_2$SO$_4$ and protease inhibitor aprotinin and leupeptin. His$_6$-HEXIM1 was purified from the cleared cell lysate using Ni-NTA resin (Qiagen) and the His$_6$ tag was cleaved with TEV protease. The HEXIM was run over a second Ni-NTA column, followed by anion exchange on a 5 mL HiTrap Q HP column (Cytiva) and gel filtration on a Superdex 200 16/60 column (Cytiva) in a final buffer containing 25 mM Hepes pH 7.5, 200 mM NaCl, 5% glycerol, 1 mM TCEP. HEXIM was flash frozen in liquid nitrogen and stored at -80°C.

**Isothermal titration calorimetry:** Binding constants for the interactions of 7SK-SL$_{1}^{apical}$ with the HEXIM$_{N}$-ARM and Tat$_{fin}$ and Tat$_{G}$ ARMs and full-length HEXIM1 with 7SK-SL$_{1}^{apical}$, 7SK-SL$_{1}^{Full-WT}$, and 7SK-SL$_{1}^{Full-AGU}$ were measured using an ITC-200 microcalorimeter (MicroCal). 68 µM HEXIM$_{N}$-ARM peptide was titrated into 5 µM solutions of 7SK-SL$_{1}^{apical}$ in 10 mM sodium phosphate, 70 mM NaCl, 0.1mM EDTA, pH 5.2 at 25°C. Titrations of Tat ARMs into 7SK-SL$_{1}^{apical}$ were also performed in the same buffer conditions as the HEXIM$_{N}$-ARM titration although the Tat ARM concentration was at 2.5 µM and the 7SK-SL$_{1}^{apical}$ concentration was at 45 µM. Titrations with full-length HEXIM1 were done at 100 µM of HEXIM1 into 3 µM of either 7SK-SL$_{1}^{apical}$, 7SK-SL$_{1}^{Full-WT}$, and 7SK-SL$_{1}^{Full-AGU}$ in a buffer of 25mM HEPES pH 7.5, 200mM NaCl, 5% glycerol, and 1mM TCEP. Titration curves were analyzed using ORIGIN (OriginLab) and all thermodynamic parameters are reported with n=3 experiments.
**Small angle X-ray scattering:** SAXS data for the 7SK-SL1\textsuperscript{apical}:HEXIM\textsuperscript{N-ARM}, 7SK-SL1\textsuperscript{apical}:Tat\textsuperscript{Fin} ARM, and 7SK-SL1\textsuperscript{apical}:Tat\textsuperscript{G} ARM complexes were obtained at SIBYLS beamline of Advanced Light Source at Lawrence Berkeley National Laboratory. Measurements were performed in a buffer containing 10 mM sodium phosphate, 70 mM NaCl, 0.1 mM EDTA, pH 5.2 and the background scattering was subtracted from the sample scattering to obtain the scattering intensity from the solute molecules. Data from three different concentrations (50, 75, and 100 µM) were compared with scattering intensities at q = 0 Å\textsuperscript{-1} [I(0)], as determined by Guinier analysis, to detect possible interparticle interactions. Data was analyzed by using ScÅtter software, and the ab initio envelope structures were reconstructed by using DAMMIF/DAMMIN software.

**NMR data acquisition, resonance assignment and structural calculations:** For NMR experiments, the Tat ARM/HEXIM\textsuperscript{N-ARM} :7SK-SL1\textsuperscript{apical} complexes were dissolved in a buffer containing 10 mM potassium phosphate, 70 mM NaCl, and 0.1 mM EDTA, pH 5.2 whereas the full-length HEXIM1:7SK-SL1\textsuperscript{apical} complex was in a buffer with 25 mM HEPES pH 7.5, 200 mM NaCl, 5% \textsuperscript{2}H-glycerol, and 1 mM TCEP. All NMR experiments were acquired by using Bruker 700 or 800 MHz instruments equipped with cryogenic probes. Spectra for observing non-exchangeable protons were collected at 298K in 99.96% D\textsubscript{2}O, whereas those for exchangeable protons were at 283K and 298K in 10% D\textsubscript{2}O. For NOESY experiments, mixing times were set to 200 ms. To help unambiguously assign the intermolecular NOEs of the HEXIM\textsuperscript{N-ARM} with 7SK-SL1\textsuperscript{apical}, we used both specifically protonated GA, AC, and GU samples of 7SK-SL1\textsuperscript{apical} and two HEXIM\textsuperscript{N-ARM} peptides synthesized by with different combinations of \textsuperscript{13}C/\textsuperscript{15}N-labeled amino acids. Samples of the 7SK-SL1\textsuperscript{apical}:HEXIM\textsuperscript{N-ARM} was prepared at 1:0.9 equivalents whereas the 7SK-SL1\textsuperscript{apical}:Tat\textsuperscript{Fin} ARM, 7SK-SL1\textsuperscript{apical}:Tat\textsuperscript{G} ARM, and full-length HEXIM1:7SK-SL1\textsuperscript{apical} complexes were prepared at a 1:0.3 equivalents to avoid any non-specific binding of the peptides or aggregation of the proteins to the RNA. Assignments for non-exchangeable
$^1$H, $^{13}$C, $^{15}$N signals of 7SK-SL1 apical in complex with HEXIM$^{N-ARM}$ and Tat ARMs were obtained by analyzing two-dimensional $^1$H-$^1$H NOESY recorded with non-labeled samples and two-dimensional $^{13}$C-HMQC and $^{15}$N-HSQC and three-dimensional $^{13}$C-edited HMQC-NOESY spectra for labeled samples.

Initial structural models were generated using manually assigned restraints in CYANA where upper-limit distance restraints of 2.7 Å, 3.3 Å, and 5.0 Å were employed for direct NOE cross-peaks of strong, medium and weak intensities, respectively. However, for crosspeaks pairs associated with the intra-residue H8/6 to H2' and H3', upper distance limits of 4.2 Å and 3.2 Å were employed for NOEs of medium and strong intensity, respectively. To prevent the generation of structures with collapsed major grooves, cross-helix P–P distance restraints (with 20% weighting coefficient) were employed for A-form helical segments. Standard torsion angle restraints were used for regions of A-helical geometry, allowing for ± 50º deviations from ideality ($\alpha$=−62°, $\beta$=180°, $\gamma$= 48°, $\delta$= 83°, $\epsilon$=−152°, $\zeta$=−73°) Standard hydrogen bonding restraints with an approximately linear NH–N and NH–O bond distances of 1.85 ± 0.05 Å and N–N and N–O bond distances of 3.00 ± 0.05 Å, and two lower-limit restraints per base pair (G–C base pairs: G-C4 to C-C6 ≥ 8.3 Å and G-N9 to C-H6 ≥ 10.75 Å; A–U base pairs: A-C4 to U-C6 ≥ 8.3 Å and A-N9 to U-H6 ≥ 10.75 Å) were employed in order to weakly enforce base-pair planarity (20% weighting coefficient).

The CYANA structure with the lowest target function was used as the initial model for structure calculations Xplor-NIH to incorporate electrostatic constraints. First, structures were calculated using annealing from 2000°C to 25°C in steps of 12.5°C. Standard energy potential terms for bonds, angles, torsion angles, van der Waals interactions and interatomic repulsions were included. The statistical backbone H-bond potential was utilized for protein residues. Energy potentials for NOEs, hydrogen bonds, and planarity were incorporated with restraints derived from NMR data. All restraints
used in CYANA were included with the exception of phosphate-phosphate distances. The structures were sorted by energy using bond, angle, dihedral, and NOE energy potential terms, and the ten percent of the structures with the lowest sort energy were further minimized with SAXS terms to incorporate orientation restraints. For this step, minimization started at 1500°C to 25°C in steps of 12.5°C. The lowest ten percent of these were deposited in the RCSB databank.
Figure 1. Characterization of HEXIM and Tat binding to 7SK:
(a) Cartoon representation of the HEXIM dimer and pTEFb heterodimer binding to the
7SK snRNP (top). Upon introduction of Tat, HEXIM is displaced from the snRNP (below).
Not depicted are MEPCE and LARP7. Representative ITC data for: full-length HEXIM1
bound to 7SK-SL1\textsuperscript{full} (G\textsubscript{i}-C\textsubscript{108}) with a wild-type loop (b) or an AGU tri loop (c) show that
the loop does not play a significant role in dimeric HEXIM binding; full-length HEXIM1
shows a similar binding affinity to both 7SK-SL1\textsuperscript{full} (d) and 7SK-SL1\textsuperscript{apical} (G\textsubscript{26}-C\textsubscript{85}) (e); the
full-length HEXIM dimer (d) and HEXIM\textsuperscript{N-ARM} (e) bind with similar affinities and
stoichiometry to 7SK-SL1\textsuperscript{apical}, indicating that the HEXIM\textsuperscript{N-ARM}:7SK-SL1\textsuperscript{apical} complex
represents the minimal binding interaction. Representative ITC traces of the Tat\textsuperscript{FIN} (f) and
Tat\textsuperscript{G} (g) ARMs into 7SK-SL1\textsuperscript{apical} show an approximately 1.2 and 2.6-fold increased
binding affinity compared to HEXIM\textsuperscript{N-ARM}. All reported values are for n=3 replicates.
Figure 2. 7SK-SL1apical in complex with HEXIMN-ARM, TatFin, and TatG ARMss.

(a) Cartoon depicting an arginine sandwich motif. (b) Secondary structure of free 7SK-SL1apical with a modified AGU tri loop. The base and cap residues forming ASM1, ASM2, pseudo-ASM3, and ASM4 colored in orange, green, magenta, and blue, respectively. The dashed arcs represent the triple-base interactions from the bulge to the stem, giving rise to the caps of the sandwiches. (c) Overlay of reconstructed ab initio SAXS envelopes of free 7SK-SL1apical (gray) and bound to HEXIMN-ARM (orange), TatFin (blue), and TatG (red) ARMss, demonstrating the lack of global rearrangement of the RNA. Representative NMR structures of 7SK-SL1apical bound to (d) HEXIMN-ARM, (e) TatFin, and (f) TatG shows engagement with all ASMs.
Figure 3. Details of intermolecular interactions between HEXIM, Tat\textsuperscript{Fin}, and Tat\textsuperscript{G} ARMs with 7SK-SL\textsuperscript{apical} and the rearrangement of the U\textsubscript{68}-A\textsubscript{39} base pair.

C-terminal arginines of all ARMs dock into ASM\textsubscript{1} and ASM\textsubscript{2} with identical tertiary structures; representative ARM of Tat\textsuperscript{Fin} is shown in (a). K150 and K151 in HEXIM (b), R52 and K51 in Tat\textsuperscript{Fin} (c), and K51, R52 and R53 in Tat\textsuperscript{G} (d,e) interact with the apical ASMs. The U\textsubscript{68}-A\textsubscript{39} base pair rearranges into a cis-Hoogsteen/sugar interaction upon HEXIM binding (f), while Tat\textsuperscript{Fin} (g) and Tat\textsuperscript{G} (h) both remodel ASM\textsubscript{3} by rearranging the U\textsubscript{68}-A\textsubscript{39} base pair into a Watson-Crick interaction. (i-k) Spacer residues between the ASM\textsubscript{1}/ASM\textsubscript{2} and ASM\textsubscript{3}/ASM\textsubscript{4} regions are positioned near ASM\textsubscript{1} and ASM\textsubscript{2}. In the case of Tat\textsuperscript{Fin} (j), K53 also acts as a spacer residue in order to allow for the remodeling of ASM\textsubscript{3} by R52.
Figure 4. Comparative thermodynamic analyses and competition experiments between Tat and HEXIM:

(a) Comparison of enthalpic and entropic contributions between TatG, TatFIN, and HEXIMN-ARM in complex with 7SK-SL1apical, and full length HEXIM in complex with 7SK snRNA. Entropy values were calculated using a T value of 298K. The reversal in the entropic and enthalpic contribution for Tat ARM compared to HEXIM is evident with HEXIM having an entropically-driven binding profile. (b) Representative ITC data for full-length HEXIM bound to 7SK snRNP demonstrating expected stoichiometry and specific binding. (c) NMR competition titration analysis showing binding of 7SK by Tat concomitant with the total displacement of HEXIM. Data are shown for the A39 (left) and A65 (right) h2-c2 correlations. The increase in Tat affinity for the HEXIM-bound complex is evident by lack of free-RNA populations for the A65 resonance in the competition experiment compared to binding to free 7SK.
### Table 1. NMR restraints and structure statistics for HEXIM, Tat^Fin^, and Tat^G^ ARMs in complex with 7SK-SL1^{apical}

|                      | HEXIM | Tat^Fin^ | Tat^G^ |
|----------------------|-------|----------|--------|
| **NMR distance and dihedral constraints** |       |          |        |
| **Distance restraints** |       |          |        |
| Total NOE            | 560   | 571      | 570    |
| Intraresidue         | 229   | 229      | 229    |
| Inter-residue        | 331   | 342      | 341    |
| - Sequential (|i − j| = 1) | 152   | 152      | 152    |
| - Nonsequential (|i − j| > 1) | 179   | 190      | 189    |
| Hydrogen bonds       | 152   | 157      | 163    |
| Total dihedral-angle restraints | 376   | 376      | 376    |
| Base pair            | 99    | 99       | 99     |
| Sugar puckers        | 112   | 112      | 112    |
| Backbone             | 127   | 127      | 127    |
| **Structure statistics** |       |          |        |
| Violations (mean ± s.d.) | 323 ± 14 | 342 ± 14 | 348 ± 17 |
| Distance constraints (Å) | 0.28 ± 0.007 | 0.28 ± 0.01 | 0.26 ± 0.007 |
| Dihedral-angle constraints (°) | 0.22 ± 0.05 | 0.16 ± 0.05 | 0.31 ± 0.08 |
| Max. dihedral-angle violation (°) | 6.12 ± 1.01 | 7.89 ± 2.10 | 14.6 ± 1.96 |
| Max. distance-violation violation (Å) | 1.69 ± 0.13 | 1.62 ± 0.16 | 2.15 ± 0.28 |
| **Deviations from idealized geometry** |       |          |        |
| Bond lengths (Å)     | 0.006 | 0.006    | 0.006  |
| Bond angles (°)      | 1.04 ± 0.009 | 1.06 ± 0.02 | 1.06 ± 0.005 |
| Improper (°)         | 0.68 ± 0.02 | 0.79 ± 0.06 | 0.94 ± 0.03 |
| **Average pairwise r.m.s. deviation (Å)^a** |       |          |        |
| NOE-restrained RNA and protein residues | 0.44 ± 0.08 | 0.54 ± 0.06 | 0.38 ± 0.05 |

^aPairwise r.m.s. deviation was calculated among ten refined structures.

^bThese are residues 24:87 for the RNA and 150:157 for the peptides.
Supplementary Figure 1: 7SK-SL1<sub>apical</sub> construct design and comparison of 7SK-SL1<sub>apical</sub> binding to full-length HEXIM1 and the HEXIM N-ARM.

(a) Two-dimensional $^1$H-$^1$N HSQC spectra for 7SK-SL1<sub>apical</sub> with an AGU triloop (top) and a native loop (bottom) showing that all major motifs in the apical stem are unaffected by the change in the loop. (b) Two-dimensional $^1$H-$^{13}$C HMQC spectra of GU-labeled 7SK-SL1<sub>apical</sub> binding to 0.3 equivalents of full-length HEXIM1 (black) and 0.9 equivalents of HEXIM N-ARM (magenta). While there are slight differences in both carbon and proton chemical due to the salt concentration of the buffers, the HEXIM N-ARM makes same interactions with 7SK-SL1<sub>apical</sub> as the full-length HEXIM1.
Supplementary Figure 2: Non-exchangeable RNA proton assignments of the HEXIM-N-ARM: 7SK-SL1<sub>apical</sub> complex.

Portion of two-dimensional $^1$H-$^1$H NOESY spectra of the HEXIM-N-ARM: 7SK-SL1<sub>apical</sub> interaction using AC (top), GU (middle), and GA (lower) protonated samples with the other nucleotides being deuterated.
Supplementary Figure 3: Assignments of amino acid backbone and sidechain protons of the HEXIM N-ARM:7SK-SL1apical interaction.

HSQC spectra of the two selectively $^{13}$C/$^{15}$N- labeled HEXIM N-ARMs in complex with 7SK-SL1apical. Bolded residues in the sequences are the amino acids that have been selectively labeled. All backbone amide protons and the side chain amino protons were unambiguously assigned with the help of this selective labeling strategy.
Supplementary Figure 4: Assignments of the interactions between the HEXIM N-ARM and 7SK-SL1<sup>apical</sup>. Secondary structure of the 7SK-SL1<sup>apical</sup> showing connectivities to the HEXIM<sup>N-ARM</sup>. NOEs between: (a) A7 H8 proton and R156 guanidinium protons indicate R156 docks into the preformed ASM<sub>1</sub>; (b) G<sub>74</sub> and G<sub>73</sub> with the guanidinium protons of R156 and R155, respectively, and (c) 156 and 155 Hδ protons and H5 protons of the C<sub>75</sub> and C<sub>71</sub> caps support entry of these arginines into ASM<sub>1</sub> and ASM<sub>2</sub>; (d) K150 Hε protons and the U<sub>63</sub> and U<sub>40</sub> caps position K150 within hydrogen-bonding distance to the O4 of both residues. (e) K150 interacts with both the U<sub>63</sub> and U<sub>40</sub> caps of pseudo-ASM<sub>3</sub> and ASM<sub>4</sub> whereas interactions between the U<sub>66</sub> H1’ protons with the K152 Hδ protons positions K152 as a spacer residue. (f) Interactions of K152 and K151 with the C<sub>67</sub> and C<sub>38</sub> H5 protons, respectively, and NOEs between the R154 Hβ proton with C<sub>33</sub> determine the relative position of this spacer residue. (g) Positioning of the H153 Hβ protons with the H6 protons of C<sub>35</sub> and C<sub>36</sub>. 
Supplementary Figure 5: Interactions between 7SK-SL1apical and TatFin and TatG RBDs:

Secondary structure of the 7SK-SL1apical showing connectivities to the RBDs. NOEs between: (a) R57, R56, and R52 Hδ of TatFin protons and the C75, C71+, and U40 caps position then into ASM1, ASM2, and remodeled ASM3; K53 Hε and C67 and U66, and R55 Hδ and C33 position these spacer residues in-between the lower and apical ASMs; (b) K51 Hε and U63 and U66, and K50 Hε and C38 and C37 make up the final interactions as the peptide exits the groove; H54 Hβ and the C37 and C36 H5 provide information for the final spacer residue of the TatFin; (c) U63 2’ hydroxyl and K51 Hε indicate a hydrogen bonding interaction; (d) R56, R52 Hδ of TatG and the C71+ and U63 caps show interaction with ASM2 and ASM1; (e) R57, R52 Hδ and the C75 and U40 caps show interaction with ASM1 and the formation of ASM3 by R52; (f) U63 2’ hydroxyl proton and K51 Hε indicate K51 hydrogen bond with the U63 ribose ring.
Supplementary Figure 6: Characterization of the A₃⁹-U₆₈ base pair:

Portions of the ¹H–¹H 2D NOESY spectra. (a) The C₇₁⁺ amino protons are downshifted due to their participation in a triple base-interaction. (b) NOEs between the U₆₈ imino proton with the A₃⁹ amino protons and the U₆₈ H₂' and H₃' protons with the A₃⁹ H₈ proton confirms that the A₃⁹-U₆₈ base pair is in a cis Hoogsteen/sugar interaction. (c).

Portion of the ¹H–¹⁵N 2D HSQC spectrum for ¹⁵N, ¹³C-labeled 7SK-SL1 apical supports the assignments of the U₄₀ and U₆₃ cap imino protons.
**Supplementary Table 1.** ITC derived binding parameter:

| Complex                        | N-value   | $K_d$            |
|--------------------------------|-----------|------------------|
| HEXIM1:7SK-SL1\(^{\text{Full}}\) | 2.06 ± 0.15 | 209 ± 34.1 nM*\(^+\) |
| HEXIM1:7SK-SL1\(^{\text{Full-AGU}}\) | 1.98 ± 0.07 | 200.0 ± 20.0 nM* |
| HEXIM1:7SK-SL1\(^{\text{apical-AGU}}\) | 2.06 ± 0.15 | 206 ± 57.9 nM*\(^+\) |
| HEXIM-N-ARM:7SK-SL1\(^{\text{apical}}\) | 1.04 ± 0.05 | 238 ± 34.8 nM*\(^{\infty,E_\phi}\) |
| Tat-Finland-RBD:7SK-SL1\(^{\text{apical}}\) | 0.98 ± 0.03 | 198 ± 15.5 nM*\(^{E_\psi}\) |
| Tat-Subtype-G-RBD:7SK-SL1\(^{\text{apical}}\) | 0.96 ± 0.17 | 93 ± 45.9 nM*\(^{E_\psi}\) |

* $p=0.9996$
+ $p=0.9999$
\(^*=p=0.8942$
\(^\phi=p=0.7780$
\(^\psi=p=0.0051$
\(^\pi=p=0.0450$

One-way ANOVA, Tukey’s multiple comparisons test
Acknowledgements

This work was supported by HHMI grant 55108516 (to V.M.D’S) and NIH grant U54 AI50470 (formerly U54 GM103297) (to V.M.D. and J.L.S.). SAXS data was collected at the Advanced Light Source (ALS), SIBYLS beamline on behalf of US DOE-BER, through the Integrated Diffraction Analysis Technologies (IDAT) program. Additional support comes from the NIGMS project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483.

Accession codes

Atomic coordinates have been deposited in the Protein Data Bank under accession codes PDB XXXX (7SK-SL1<sup>apical</sup>:HEXIM<sup>N-ARM</sup>), PDB XXX (7SK-SL1<sup>apical</sup>:Tat<sup>Fin</sup>), and PDB XXX (7SK-SL1<sup>apical</sup>:Tat<sup>C</sup>). Chemical shifts have been deposited in the Biological Magnetic Resonance Data Bank under accession codes XXX (7SK-SL1<sup>apical</sup>:HEXIM<sup>N-ARM</sup>), XXX (7SK-SL1<sup>apical</sup>:Tat<sup>Fin</sup>), and XXXX (7SK-SL1<sup>apical</sup>:Tat<sup>C</sup>).

Author contributions

V.V.P, M.G., and V.M.D’S. conceived and designed the experiments. V.V.P., M.G. and J.L.M. purified the samples and V.V.P performed the NMR, ITC, and SAXS experiments. V.V.P., M.G., and V.M.D’S. performed the structural analyses, interpreted the data, and wrote the manuscript. V.V.P, M.G., J.L.M., J.L.S. and V.M.D. analyzed data and edited the manuscript.

Competing financial interests

The authors declare no competing financial interests.
References

1. Wada, T. et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12**, 343-356, doi:10.1101/gad.12.3.343 (1998).

2. Yamaguchi, Y. et al. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41-51, doi:10.1016/s0092-8674(00)80713-8 (1999).

3. Missra, A. & Gilmour, D. S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc Natl Acad Sci U S A* **107**, 11301-11306, doi:10.1073/pnas.1000681107 (2010).

4. Renner, D. B., Yamaguchi, Y., Wada, T., Handa, H. & Price, D. H. A highly purified RNA polymerase II elongation control system. *J Biol Chem* **276**, 42601-42609, doi:10.1074/jbc.M104967200 (2001).

5. Ehara, H. et al. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357**, 921-924, doi:10.1126/science.aan8552 (2017).

6. Marshall, N. F. & Price, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* **270**, 12335-12338, doi:10.1074/jbc.270.21.12335 (1995).

7. Peng, J., Zhu, Y., Milton, J. T. & Price, D. H. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev* **12**, 755-762, doi:10.1101/gad.12.5.755 (1998).

8. Kim, J. B. & Sharp, P. A. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J Biol Chem* **276**, 12317-12323, doi:10.1074/jbc.M010908200 (2001).

9. Fujinaga, K. et al. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol* **24**, 787-795, doi:10.1128/mcb.24.2.787-795.2004 (2004).

10. Vos, S. M. et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560**, 607-612, doi:10.1038/s41586-018-0440-4 (2018).

11. Yamada, T. et al. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol Cell* **21**, 227-237, doi:10.1016/j.molcel.2005.11.024 (2006).

12. Yang, Z., Zhu, Q., Luo, K. & Zhou, Q. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**, 317-322, doi:10.1038/35104575 (2001).
Nguyen, V. T., Kiss, T.,Michels, A. A. & Bensaude, O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**, 322-325, doi:10.1038/35104581 (2001).

Michels, A. A. *et al.* MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol Cell Biol* **23**, 4859-4869, doi:10.1128/mcb.23.14.4859-4869.2003 (2003).

Yik, J. H. *et al.* Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol Cell** **12**, 971-982, doi:10.1016/s1097-2765(03)00388-5 (2003).

Chen, R., Yang, Z. & Zhou, Q. Phosphorylated positive transcription elongation factor b (P-TEFb) is tagged for inhibition through association with 7SK snRNA. *J Biol Chem* **279**, 4153-4160, doi:10.1074/jbc.M310044200 (2004).

Michels, A. A. *et al.* Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO J* **23**, 2608-2619, doi:10.1038/sj.emboj.7600275 (2004).

Kobbi, L. *et al.* An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb. *Proc Natl Acad Sci U S A* **113**, 12721-12726, doi:10.1073/pnas.1612331113 (2016).

Sano, M. *et al.* Activation and function of cyclin T-Cdk9 (positive transcription elongation factor-b) in cardiac muscle-cell hypertrophy. *Nat Med* **8**, 1310-1317, doi:10.1038/nm778 (2002).

Kretz, A. L. *et al.* CDK9 is a prognostic marker and therapeutic target in pancreatic cancer. *Tumour Biol* **39**, 101428317694304, doi:10.1177/1010428317694304 (2017).

Schlafstein, A. J. *et al.* CDK9 Expression Shows Role as a Potential Prognostic Biomarker in Breast Cancer Patients Who Fail to Achieve Pathologic Complete Response after Neoadjuvant Chemotherapy. *Int J Breast Cancer* **2018**, 6945129, doi:10.1155/2018/6945129 (2018).

Cho, W. K., Jang, M. K., Huang, K., Pise-Masison, C. A. & Brady, J. N. Human T-lymphotropic virus type 1 Tax protein complexes with P-TEFb and competes for Brd4 and 7SK snRNP/HEXIM1 binding. *J Virol* **84**, 12801-12809, doi:10.1128/JVI.00943-10 (2010).

Zhou, M. *et al.* Tax interacts with P-TEFb in a novel manner to stimulate human T-lymphotropic virus type 1 transcription. *J Virol* **80**, 4781-4791, doi:10.1128/JVI.80.10.4781-4791.2006 (2006).

Muniz, L., Egloff, S., Ughy, B., Jady, B. E. & Kiss, T. Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator Tat. *PLoS Pathog* **6**, e1001152, doi:10.1371/journal.ppat.1001152 (2010).

Byers, S. A., Price, J. P., Cooper, J. J., Li, Q. & Price, D. H. HEXIM2, a HEXIM1-related protein, regulates positive transcription elongation factor b through
association with 7SK. J Biol Chem 280, 16360-16367, doi:10.1074/jbc.M500424200 (2005).

26 Yik, J. H., Chen, R., Pezda, A. C. & Zhou, Q. 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, 16360-16367, doi:10.1074/jbc.M500912200 (2005).

27 Egloff, S., Van Herreweghe, E. & Kiss, T. Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. Mol Cell Biol 26, 630-642, doi:10.1128/MCB.26.2.630-642.2006 (2006).

28 Li, Q. et al. Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. J Biol Chem 280, 28819-28826, doi:10.1074/jbc.M502712200 (2005).

29 Dulac, C. et al. Transcription-dependent association of multiple positive transcription elongation factor units to a HEXIM multimer. J Biol Chem 280, 30619-30629, doi:10.1074/jbc.M502471200 (2005).

30 Schonichen, A. et al. A flexible bipartite coiled coil structure is required for the interaction of Hexim1 with the P-TEFB subunit cyclin T1. Biochemistry 49, 3083-3091, doi:10.1021/bi902072f (2010).

31 Blazek, D., Barboric, M., Kohoutek, J., Oven, I. & Peterlin, B. M. Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. Nucleic Acids Res 33, 7000-7010, doi:10.1093/nar/gki997 (2005).

32 Martinez-Zapien, D. et al. Intermolecular recognition of the non-coding RNA 7SK and HEXIM protein in perspective. Biochimie 117, 63-71, doi:10.1016/j.biochi.2015.03.020 (2015).

33 Lebars, I. et al. HEXIM1 targets a repeated GAUC motif in the riboregulator of transcription 7SK and promotes base pair rearrangements. Nucleic Acids Res 38, 7749-7763, doi:10.1093/nar/gkq660 (2010).

34 Pham, V. V. et al. HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. Nat Commun 9, 4266, doi:10.1038/s41467-018-06591-6 (2018).

35 Chen, L. & Frankel, A. D. An RNA-binding peptide from bovine immunodeficiency virus Tat protein recognizes an unusual RNA structure. Biochemistry 33, 2708-2715, doi:10.1021/bi00175a046 (1994).

36 Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. Science 257, 76-80, doi:10.1126/science.1621097 (1992).
Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science* **270**, 1200-1203, doi:10.1126/science.270.5239.1200 (1995).

Martinez-Zapien, D. *et al.* The crystal structure of the 5 functional domain of the transcription riboregulator 7SK. *Nucleic Acids Res* **45**, 3568-3579, doi:10.1093/nar/gkw1351 (2017).

Bourbigot, S. *et al.* Solution structure of the 5'-terminal hairpin of the 7SK small nuclear RNA. *RNA* **22**, 1844-1858, doi:10.1261/rna.056523.116 (2016).

Brogie, J. E. & Price, D. H. Reconstitution of a functional 7SK snRNP. *Nucleic Acids Res* **45**, 6864-6880, doi:10.1093/nar/gkx262 (2017).

Marz, M. *et al.* Evolution of 7SK RNA and its protein partners in metazoa. *Mol Biol Evol* **26**, 2821-2830, doi:10.1093/molbev/msp198 (2009).

Stols, L. *et al.* A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr Purif* **25**, 8-15, doi:10.1006/prep.2001.1603 (2002).

Guntert, P., Mumenthaler, C. & Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**, 283-298, doi:10.1006/jmbi.1997.1284 (1997).

Tolbert, B. S. *et al.* Major groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement. *J Biomol NMR* **47**, 205-219, doi:10.1007/s10858-010-9424-x (2010).