A structure-based mechanism for displacement of the HEXIM adapter from 7SK small nuclear RNA

Vincent V. Pham¹, Michael Gao¹, Jennifer L. Meagher², Janet L. Smith ²,³ & Victoria M. D’Souza¹✉

Productive transcriptional elongation of many cellular and viral mRNAs requires transcriptional factors to extract pTEFb from the 7SK snRNP by modulating the association between HEXIM and 7SK snRNA. In HIV-1, Tat binds to 7SK by displacing HEXIM. However, without the structure of the 7SK-HEXIM complex, the constraints that must be overcome for displacement remain unknown. Furthermore, while structure details of the TatNL4-3-7SK complex have been elucidated, it is unclear how subtypes with more HEXIM-like Tat sequences accomplish displacement. Here we report the structures of HEXIM, TatG, and TatFin arginine rich motifs in complex with the apical stemloop-1 of 7SK. While most interactions between 7SK with 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.
transcription of all class II genes is a highly regulated process within cells. Shortly after promoter clearance, RNA Polymerase II is inhibited by negative elongation factors1–3. 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 Cdk96–12. 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′ stem-loop-1 of the 7SK RNA13–21 (7SK-SL1). 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 cancers24–27. Furthermore, as many viruses rely on the host transcriptional machinery to produce RNA and genomes they have also evolved mechanisms to capture pTEFb28–30. 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 it30–34. 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 can carry 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)34,35–38. Although HEXIM binds 7SK as a dimer, only one ARM directly contacts 7SK by engaging the apical region of stem-loop-1 (G26 to C85, 7SK-SL1)39,40. Both in vitro and in vivo studies have shown that this represents the sole interaction between the two molecules that must be modulated to release pTEFb37–39,41,42.

Our previous work showed that 7SK-SL1apical is enriched in arginine sandwich motifs (ASMs)43, 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 bases45–51. 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-SL1apical, three such bulges fold into preformed arginine sandwich motifs (ASM1, ASM2, and ASM3) poised for arginine guanidinium moieties to dock into them. A fourth bulge folds into a pseudo configuration (pseudo-ASM4) where U60 can form a triple-base interaction with the A43-U66 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 (TatNL4-3) 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 ASM45. This structural remodeling of pseudo-ASM4 is a key mechanism through which Tat displaces HEXIM.

However, without the structure of the HEXIM:7SK-SL1apical interaction, it is currently unclear what structural constraints Tat would need to overcome 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 (TatFin; KR52R53HRRR) differs from HEXIM (KR52R53HRRR) by only a single amino acid and would lack one of the ASM interactions from the previously described Tat NL4-3 strain (KR52R53RQQRR). Additionally, while Tat subtype G (TatG; KR52R53HRRR) has an equivalent number of arginines as TatNL4-3, the critical linker sequence connecting the ASM3 of ASM4 and ASM1/ASM2 interactions is the same as in HEXIM. In this study, we present the structure of the 7SK-SL1apical in complex with the HEXIM, Tat18, and Tat2 ARM. 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 comparative thermodynamic properties of 7SK recognition between HEXIM and Tat, we performed binding studies using isothermal titration calorimetry (ITC). ITC traces of the ARMs into 7SK-SL1apical-AGU produce significant nonspecific heats of binding as previously observed52. In a previous study by Brillot et al., high salt conditions (0.5 M NaCl) were used to abrogate such nonspecific interactions that stem from charge-charge interactions between the positively charged peptides and the negative RNA backbone52. While such a strategy is commonly used, it is not ideal for this system as the structuring of ASMs in 7SK is highly sensitive to ionic conditions and folds only around physiological salt conditions (Supplementary Figs. 1, 2)45. Therefore, to subtract the nonspecific heats of binding, we designed a control construct that lacks all ASMs (7SK-SL1apical-AGU). Indeed, the heats obtained from peptide titrations into this control construct completely accounted for the nonspecific heats, the subtraction of which allowed for experimental baselines to approach zero at saturation (Supplementary Fig. 2b). Titration of the N-terminal ARM residues of HEXIM (R142QLGKKKRRR156; HEXIMN-ARM) into 7SK-SL1apical-AGU containing an AGU triloop engineered to prevent low levels of dimerization gave a Kd of 229 ± 20 nM (N = 1 ± 0.1; Fig. 1a). The redesign of the previously used GAGA tetraloop45 to an AGU trioloop was done to prevent weak association between the tyrosine in the peptide and the tetraloop. Nevertheless, while the affinities obtained by AGU-trioloop are 2 to 3-fold weaker, the relative difference between HEXIM and Tat are similar (see below).

To confirm that interactions do not extend to the loop and are represented by these minimal constructs, we performed studies with full-length HEXIM into full-length 7SK-SL1Full (G1–C108), 7SK-SL1Full-AGU, and 7SK-SL1apical-AGU, all of which give rise to similar Kd’s (209 ± 30 nM, Kd = 200 ± 20 nM, and Kd = 206 ± 60 nM, respectively) and bound exactly as dimers (N = 2 ± 0.2, N = 2 ± 0.07, and N = 1.8 ± 0.03, respectively; Fig. 1b–d). Furthermore, NMR studies comparing full-length dimeric HEXIM1:7SK-SL1apical-AGU and HEXIMN-ARM:7SK-SL1apical-AGU complexes showed that binding of either full-length HEXIM or the N-ARM gives rise to the same chemical shifts in 7SK-SL1apical-AGU, indicating that the HEXIMN-ARM:7SK-SL1apical-AGU interaction represents the biologically relevant mode of HEXIM binding to 7SK (Supplementary Fig. 2).

Our previous work showed that the TatNL4-3 (KR52R53RQQRR) ARM represents the interaction domain between Tat and 7SK-SL1apical and has an approximately two-fold increased affinity over the HEXIMN-ARM, which provides an explanation for HEXIM displacement45. ITC traces show that Tat Subtype G’s ARM (KR52R53HRRR), which also has two N-terminal arginines, binds 7SK-SL1apical-AGU with a Kd of 81 ± 10 nM (N = 1.1 ± 0.1; Fig. 1e), which is an approximately 2.8-fold increased binding
On the other hand, Tat Finland’s ARM (KR52KHRRR), despite having an additional N-terminal arginine compared to the HEXIMN-ARM (R52 and K151, respectively), does not have a statistically significant increase in binding affinity over HEXIMN-ARM ($K_d$ of 172 ± 10 nM, $N = 1 ± 0.02$; Fig. 1f). Overall, these results highlight the need for understanding the HEXIM-bound 7SK; while the increased TatG affinity would allow for HEXIM displacement, it is unclear how TatFin can achieve the same biological output.

Preformed configurations of ASM1 and ASM2 provide a common mode of interaction with C-terminal arginines. To understand how the HEXIMN-ARM and the various Tat ARMs interact with 7SK-SL1apical-AGU, 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-SL1 apical-AGU (Supplementary Fig. 3). 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 stemloop below the G79-U32 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. 2 and Table 1).
In the free 7SK-SL1apical-AGU, ASM1 and ASM2 are placed in 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 ASM1, the residue A77 is configured into an A34-A77 base pair. A NOE from the A77 H8 proton to the H1′ of C75 positions this residue under the C75 cap (Supplementary Fig. 4). This confirms a planar orientation of C75 with the C33–G78 base pair and configures A77 in such a way that it is perfectly positioned to interact with the guanidinium moiety of R156 in HEXIMN-ARM and R57 in TatFin and TatG, which intercalate between C75 and G74 in a manner identical to canonical ASMs (Supplementary Figs. 4–7).

Similarly, in ASM2, the C71+ base also retains its protonation at the N3 position, as evidenced by a downfield shift of the N4 amino protons (Supplementary Fig. 8). The guanidinium moiety of R155 in HEXIMN-ARM and R56 of TatFin and TatG interact with G74 by intercalating between the C71+ cap and G70 base of the motif (Supplementary Figs. 5–7). Additionally, intermolecular NOEs from the aromatic protons of the C75 and C71+ caps and the G74 and G70 bases of ASM1 and ASM2 to the Hγ and the Hδ protons confirm that consecutive arginines R156 and R155 interact in a ladder-like configuration with the tandem performed motifs ASM1 and ASM2, respectively (Fig. 3a and Supplementary Fig. 6c). Such NOEs are also observed in both the TatFin and the TatG-bound complexes, confirming the similar placement of the C-terminal R57 and R56 into the tandem ASM1 and ASM2, respectively (Fig. 3a and Supplementary Fig. 7a, 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-ASM3 allows for HEXIM N-terminal interactions. In the free 7SK-SL1apical-AGU, pseudo-ASM3 and ASM4 adopt a pseudo-symmetrical architecture where the two motifs are spatially opposed. Upon HEXIMN-ARM binding, the pseudo-ASM3 maintains its U40:A43-U66 triple-base interaction although the base of the sandwich, A39, rearranges from a reverse Hoogsteen interaction with U68 into a cis-Hoogsteen/sugar interaction, giving rise to an alternate pseudo configuration (Fig. 3b). This is evidenced both by NOEs from the U68 imino proton to the A39 amino protons and NOEs from the U68 H2′ and H3′ protons to the A39 H8 proton (Supplementary Fig. 8b). This

Fig. 2 7SK-SL1apical in complex with HEXIMN-ARM, TatFin, and TatG ARMs. a Cartoon depicting an arginine sandwich motif. b Secondary structure of free 7SK-SL1apical-AGU with a modified AGU triloop. The base and cap residues forming ASM1, ASM2, pseudo-ASM3, and ASM4 are colored in orange, green, magenta, and blue, respectively. Dashed arcs represent triple-base interactions from the bulge to the stem, giving rise to the caps of the sandwiches. Representative NMR structures of 7SK-SL1apical-AGU bound to (c) HEXIMN-ARM, (d) TatFin, and (e) TatG show engagement with all ASMs.
frees up the U68 imino proton to engage the backbone carbonyl of K152 while simultaneously bringing the N1 proton acceptor of A39 into the major groove to hydrogen-bond with the side chain H6 protons of K151 (Fig. 3c). Thus, both K151 and 152 can enter deep into the major groove by remodeling the pseudo-ASM4.

The amino side chain of K151 is within hydrogen-bonding distance of the A39 N1 nitrogen as evidenced by NOEs from the K151 H5 and H6 protons to the C37 H6 and H5 protons, respectively, and from the K151 H5 protons to the C68 H6 and H5 protons (Fig. 3c and Supplementary Fig. 6). Additionally, NOEs between the K152 H5 protons with the U68 H5 proton, the K152 H6 protons with the C67 and U66 H5 protons, and the K152 H5 protons with the C67 H5 and H6 protons position the amino side chain of K152 within hydrogen-bonding distance of the C67 backbone (Fig. 3c and Supplementary Fig. 6b, 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-imino connectivity between U44 and U45 along with maintenance of the G46–C62 Watson–Crick base pair (Supplementary Fig. 8c). In fact, this interaction is stabilized by K150, which displays NOEs between the H6 protons with the U63 and U40 H5 protons, positioning the amino side chain within hydrogen-bonding distance of the O4 atoms of both U63 and U40 (Supplementary Fig. 6d). Additional intermolecular interactions between the U40 H5 proton and the U63 H5 and H1′ protons with the K150 H6, H5, and H6 protons places K150 directly under the U63 cap of ASM4 (Supplementary Fig. 6d, e). Taken together, these data show that despite the lack of arginines, the lysine-rich N-terminus of HEXIM\(^{NARM}\) can 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 K150 to interact with the oxygen-rich environment of the U63 and U40 caps.

Conformational plasticity of the ASM\(_1\)/ASM\(_4\) region provides a different mode of interactions with N-terminal and spacer residues. Our previous study showed that Tat\(^{NL4-3}\) displaces HEXIM by remodeling the pseudo-ASM\(_1\) into a canonical ASM\(_1\) to allow for arginine intercalation\(^{15}\). Furthermore, an additional arginine docks into the preformed ASM\(_4\). While the mechanism of remodeling pseudo-ASM\(_1\) is conserved upon binding of both Tat\(^{Fin}\) and Tat\(^{G}\) ARMs (Fig. 3b, f and Supplementary Fig. 6), both the drivers of the conformational switch and the engagement of the ASM\(_4\) vary depending on differences in amino acid sequences.

While Tat\(^{Fin}\) has two major differences from Tat\(^{NL4-3}\) (K53 to R53 and spacer H54 to Q54, respectively), it only differs by a single amino acid from HEXIM (R52 and K151, respectively). Like Tat\(^{NL4-3}\), R52 is responsible for remodeling pseudo-ASM\(_1\) (Fig. 3c and Supplementary Fig. 7a). However, while R53 in Tat\(^{NL4-3}\) flips over R52 and engages ASM\(_4\), the equivalent K53 stays in the spacer region between the ASM\(_1\)/ASM\(_2\) and ASM\(_4\)/ASM\(_1\) regions in a manner similar to HEXIM as evidenced by NOEs between the K53 H5 protons with the U63 H5 proton, the K53 H6 protons with the U63 H6 protons, and the K53 H5 and H6 protons, positioning the amino side chain of K53 within hydrogen-bonding distance of the C68 backbone (Fig. 3c and Supplementary Fig. 7a).

As for HEXIM, ASM\(_4\) remains unoccupied upon binding Tat\(^{Fin}\) 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 of the K51 H5 protons with the U63 H5 and H1′ protons and the K51 H6 protons with the U63 H6 hydroxyl proton (Supplementary Fig. 7c). Furthermore, the N-terminal K50 exits near the apical loop, with NOEs observed of the K50 H5 and H6 protons with the C38 and C37 H5, and H1′ protons positions the amino side chain of K50 to the C38 phosphate backbone (Fig. 3c and Supplementary Fig. 7a).

Finally, in evaluating the structural consequences of the spacer substitution, we see that H54 and R55 in Tat\(^{Fin}\) remain near ASM\(_1\) and ASM\(_2\), similar to what is found in HEXIM. This is evidenced by NOEs of the H54 (H153 in HEXIM) H5 protons with the C35 and C36 H5 protons, placing H54 near ASM\(_2\), whereas the R55 (R154 in HEXIM) H6 protons display NOEs with the A34 H1′ proton and the C35, H1′, H5, and H6 protons, positioning this spacer residue near ASM\(_1\) (Fig. 3d and Supplementary Figs. 6, 7). This is in contrast with the binding mode of Tat\(^{NL4-3}\) in which the intercalation of R53 into ASM\(_4\) 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 Tat\(^{G}\) strain where it represents the only difference from Tat\(^{NL4-3}\). This single difference changes the identity of the arginine that remodels pseudo-ASM\(_3\). In this ARM, the positioning of H54 near ASM\(_2\) precludes R53 from reaching ASM\(_4\) to accomplish the inverse intercalation seen in NLI4-3 (Fig. 3d and Supplementary Fig. 7d, e). The interactions with the apical ASMs thus occur in a ladder-like manner where R53 intercalates into the remodeled ASM\(_3\), whereas R53 intercalates into ASM\(_4\) (Fig. 3c, d and Supplementary Fig. 7a, d, e). K51 makes the final stabilizing interaction with NOEs seen between

### Table 1 NMR and refinement statistics for HEXIM, Tat\(^{Fin}\), and Tat\(^{G}\) ARMs in complex with 7SK-SL1\(^{apical-AGU}\).

| NMR distance and dihedral constraints | HEXIM | Tat\(^{Fin}\) | Tat\(^{G}\) |
|--------------------------------------|-------|-------------|-------------|
| Distance constraints | | | |
| Total NOE | 584 | 589 | 588 |
| Intra-residue | 240 | 240 | 240 |
| Inter-residue | 344 | 349 | 348 |
| Sequential (| | | |
| Medium-range (| | | |
| Long-range (| | | |
| Intermolecular | 42 | 49 | 48 |
| Hydrogen bonds | 154 | 157 | 163 |
| Total dihedral angle constraints | 376 | 376 | 376 |
| Structure statistics | | | |
| Violations (mean and s.d.) | 323 ± 14 | 342 ± 14 | 348 ± 17 |
| Distance constraints (Å) | 0.28 ± 0.007 | 0.28 ± 0.001 | 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 constraint 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 angles (°) | 0.68 ± 0.02 | 0.79 ± 0.06 | 0.94 ± 0.03 |
| Average pairwise r.m.s. deviation (Å) | 1.55 ± 0.32 | 0.59 ± 0.06 | 2.25 ± 0.43 |
| Heavy** | 0.44 ± 0.08 | 0.54 ± 0.06 | 0.38 ± 0.05 |
| Backbone** | 0.44 ± 0.08 | 0.54 ± 0.06 | 0.38 ± 0.05 |

*Pairwise r.m.s. deviation was calculated among ten refined structures.

** These are residues 24:87 for the RNA and 150:157 for the peptides.
the Hε protons and the U63 2′ hydroxyl proton, indicating a hydrogen-bonding interaction between the K51 amino side chain and the U63 ribose ring (Fig. 3d and Supplementary Fig. 7f). Taken together, these studies show that arginine sandwich motifs provide mini domains that arginine-rich motifs of proteins can differentially interact with to achieve deep major groove binding into the stem of 7SK-SL1apical-AGU.

**HEXIM allows for increased conformational sampling of apical ASMs.** While titration of all four arginine-rich motifs stabilize the majority of 7SK-SL1apical-AGU into one predominant configuration, the HEXIM ARM is an outlier wherein binding causes ASM1 and ASM4 to become destabilized and exhibit multiple conformations (Fig. 4a, b and Supplementary Fig. 4). 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. 8c). The destabilization of this region is also indicated by the line-broadening of K150, which interacts with U63 in the folded configuration (Supplementary Fig. 6e).

The destabilization of 7SK-SL1apical-AGU only by HEXIM is further evident when comparing the thermodynamic profiles between Tat and HEXIM. The binding of TatFin and TatG strains is enthalpically driven (ΔH = −7.5 ± 0.2 and −8.9 ± 2.2 kcal mol⁻¹, respectively; Fig. 4c) with a modest entropic contribution (−TΔS = −1.7 ± 0.3 and −2 ± 0.8 kcal mol⁻¹, respectively; Fig. 4c). On the other hand, HEXIM binding is entropically enhanced by ~2.5-fold over both Tat strains (−TΔS = −4.6 ± 0.8 kcal mol⁻¹, ΔH = −4.4 ± 0.7 kcal mol⁻¹; Fig. 4c). The Brillet et al. study performed in 0.5 M salt saw an unfavorable entropic contribution for Tat and a negligible entropic for HEXIM binding, underscoring the importance of maintaining native ASM folding for a mechanistic understanding of this biological process. Nevertheless, the overall observation that HEXIM binding is comparatively more entropic than Tat agrees with our results.

To evaluate the implication of HEXIM’s ability to locally destabilize ASM1 and ASM4 in the context of its displacement required for transcriptional regulation, we compared TatFin and TatG ARM binding to 7SK-SL1apical both free and in the presence of...
HEXIM<sub>N-ARM</sub>. Due to the modest differences in binding energetics between the different ARMs, competition experiments using ITC were not tractable. A 1:1 titration of both Tat<sup>G</sup> and Tat<sup>F</sup> into 7SK in the NMR shows the ability to completely engage ASM2 and ASM3, while a significant fraction of ASM1 and ASM4 shows the presence of free configurations, indicating reduced access for the termini. However, upon titration of both Tats into the HEXIM-bound 7SK complex, we observe not only complete engagement of all ASMs but also a total displacement of HEXIM (Fig. 4a, b). This is especially striking given that the binding affinities of Tat<sup>F</sup> and HEXIM for free 7SK are equivalent. Taken together, these data indicate that Tat can better engage 7SK that is destabilized by HEXIM at the outer ASMs. Finally, ITC data of full-length HEXIM bound to full-length 7SK snRNA (<i>n</i> = 1.9 ± 0.1; Fig. 4d) show that an entropy-driven interaction is maintained and, in fact, is even more pronounced (<i>−TΔS</i> = −6.4 ± 1.3 kcal mol<sup>−1</sup>), 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 needs 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 gained 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 directly bind the stem of 7SK-SL^apo 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. On the one hand, the tandem preformed ASMs, ASM1 and ASM2, remain unchanged from their free configuration upon encountering the 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, and on the other hand, it simultaneously causes local destabilization of this region, enhancing the binding of a positive regulator such as Tat. 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 ASM1 and ASM4 regions, indicating a mechanism in line with conformational selection. Indeed, mutational studies have shown that deletion of U63 significantly reduces HEXIM binding.37,43 This expansion in the dynamic state of 7SK surrounding the ASM1 and ASM4 region is also supported both by in vivo SHAPE analysis where U63 and C95 become ultra-reactive upon HEXIM:pTEFb binding.53 Such increased conformational sampling was also demonstrated by structural and molecular dynamics modeling.45,52–57 Furthermore, we show that Tat capitalizes on this increased dynamic state, binding to more motifs 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 allows for transcriptional regulators to differentially utilize their ARMs. On the one hand, 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-ASM3, even in cases where there is only a single variation from HEXIM. Second, despite both ARMs having lysines positioned near ASM4, 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. 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-SL1apo. 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 77 promoter and the desired sequences. Plasmid DNA for 7SK-SL^WT and 7SK-SL^AGU containing the 77 promoter, insert, and Smal recognition sequence were cloned by GenScript in between the EcoRI and BamHI restriction sites of a puc19 vector. Plasmid DNA was prepared using the Qiagen miniprep kit according to the manufacturer’s instructions. Competent E.coli (C29871) transformed with the plasmid using Qiaprep Spin Miniprep Kit (Qiagen 27104). 10 μL of purified DNA were combined with 25 μL of 2′-O-methylated reverse primer at 100 μM (5′-GmGCGGTGGAGGAAG-3′), where m indicates 2′-O-methylated nucleotides, 25 μL of forward primer at 100 μM (5′-GACAAGCCCGTCAGGG-3′), and 2.4 μL of water, and two tubes of EconoTaq PLUS 2X Master Mix (Lucigen 10005-2). The 5 mL mixture was then aliquoted into 50 μL increments in a 96-well PCR plate and the templates for in vitro transcription reactions were amplified using the following PCR protocol: 95°C for 5 min, 34 cycles of (95°C for 30 s, 50°C for 1 min, and 68°C for 90 s), and 68°C for 5 min. After PCR amplification, reactions were pooled into 5 mL volume in a 50 mL Falcon tube and 0.5 mL of 3 M sodium acetate, pH 5 and 32 mL of 100% ethanol were added to the mixture and chilled at −80°C for at least 30 min before spinning down at 9000×g for 10 min at 4°C. The ethanol was decanted, and the pellet was left to dry overnight before in vitro transcription use. Template preparation for 7SK-SL^AGU was 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′-TAATACGACTCACTATAG-3′) and reverse (5′-GmGCGGTGGAGGAAG-3′) oligos at 10 μM in 50 μL of 1X nuclease-free water.
TAGGGATCTGTCACCCCACTATGCGGCTGCTATGCGGCTGAC
AGGGGGCTCC-3') and reverse (5'-GCAAGCTTCCCAGGACCA
ATCAGCCACTGGCGATCTTGAAGTGGTTGAT-3'), and the 7SK
TATTA-3' where m indicates 2° O'-methylated nucleotides) primers at 1 mM stock
solution with 470 mL of water. The mixture was heated at 95 °C for 5 min and
cooled at room temperature for 30 min before assembling the in vitro transcription
reactions. Samples were either unlabelled or residue-specifically labelled with 13C/
15N- or 2H-oligo (Cambridge Isotope Laboratories, Inc.). After transcription, RNA
samples were heat denatured and purified by using urea-denaturing poly-
acrylamide gels. The same in vitro transcription reaction protocol was done for
T7SK-SL1apical-AGU or T7SK-SL1apical
specificially protonated GA, AC, and GU samples of 7SK-SL1apical-AGU and two
HExiN4-ARM peptides synthesized by different combinations of 13C-labeled amino acids. Samples of the 7SK-SL1apical-AGU:HExiN4-ARM, the 7SK-SL1apical-AGU:TatFin ARM, and 7SK-SL1apical-AGU:TatG ARM were prepared at 1:0.9 equivalents, whereas the full-length HExiN1:7SK-SL1apical-AGU complex was prepared at 1:0.3 equivalents to avoid any nonspecific binding or aggregation of the protein to the RNA. Assignments for non-exchangeable 1H, 13C, and 15N signals of 7SK-SL1apical-AGU in complex with HExiN4-ARM and Tat ARMs were obtained by analyzing two-dimensional 1H-15N-NOESY recorded with non-labeled samples and two-dimensional 13C-HMQC and 13N-HSQC and three-dimensional 13C-edited HMBC NOESY spectra for labeled samples.

Initial structural models were generated using manually assigned restraints in CYANA, where upper-limit distance restraints of 2.7, 3.3, and 3.0 Å were employed for direct NOE cross-peaks of a strong, medium, and weak intensities, respectively. However, for cross-peaks pairs associated with the intra-residue H6/H2' and H3', upper limit distances of 4.2 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 ±5° deviations from ideality (α = 62°, β = 180°, γ = 180°, δ = 83° and ε = 152°, ζ = 73°). Standard hydrogen-bonding restraints with approximately linear NH–N and NH–O bond distances of 1.85 ± 0.05 Å and N–N and O–O bond distances of 3.00 ± 0.05 Å, and two lower-limit restraints per base pair (G–C base pairs: G–C ≥ 3.3 Å and A–T ≥ 3.5 Å). These were used in order to weaken base-pair planarity (20% weighting coefficient).

The CYANA structure with the lowest target function was used as the initial model for the structure calculations using X-PLOR-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 NOE hydrogen bonds, and planarity were incorporated with restraints derived from NMR data. All restraints used in CYANA were included except for 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 SALS 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.

Reporting summary. Further information on data availability is found in the Nature Research Reporting Summary linked to this article.

Data availability. Atomic coordinates have been deposited in the Protein Data Bank under accession codes PDB 7T1O (7SK-SL1apical-AGU:TatG), PDB 7T1P (7SK-SL1apical-AGU:TatF5), and PDB 7T1J (7SK-SL1apical-AGU:TatFP). PDB 7T1O (7SK-SL1apical-AGU:TatG) and PDB 7T1P (7SK-SL1apical-AGU:TatF5) were also deposited in the Biological Magnetic Resonance Data Bank under accession codes 39071 (7SK-SL1apical-AGU:HExiN4-ARM), 39073 (7SK-SL1apical-AGU:TatFin), and 39072 (7SK-SL1apical-AGU:TatG). SAXS data were submitted to and validated by SASBDD (https://www.sasbdd.org/#/ under accession codes SADSMF (7SK-SL1apical-AGU:HExiN4-ARM), SADSMF (7SK-SL1apical-AGU:TatFin), and SADSMF (7SK-SL1apical-AGU:TatG)). Source data are available through Dryad (https://doi.org/10.5061/dryad.12jm6z17) and upon reasonable request from the corresponding author.

Received: 9 February 2021; Accepted: 19 July 2022.
Published online: 15 August 2022

References.

1. Wada, T. et al. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Sp4 and Sp5 homologs. *Genes Dev.* 12, 343–356 (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 (1999).
3. Mysza, A., & Gilmour, G. Interactions between DSIF (DDB1 inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc. Natl Acad. Sci. USA* 107, 11301–11306 (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 (2001).

Data availability.

Atomic coordinates have been deposited in the Protein Data Bank under accession codes PDB 7T1O (7SK-SL1apical-AGU:TatG), PDB 7T1P (7SK-SL1apical-AGU:TatF5), and PDB 7T1J (7SK-SL1apical-AGU:TatFP). The chemical shifts have also been deposited in the Biological Magnetic Resonance Data Bank under accession codes 39071 (7SK-SL1apical-AGU:HExiN4-ARM), 39073 (7SK-SL1apical-AGU:TatFin), and 39072 (7SK-SL1apical-AGU:TatG). SAXS data were submitted to and validated by SASBDD (https://www.sasbdd.org/#/ under accession codes SADSMF (7SK-SL1apical-AGU:HExiN4-ARM), SADSMF (7SK-SL1apical-AGU:TatFin), and SADSMF (7SK-SL1apical-AGU:TatG)). Source data are available through Dryad (https://doi.org/10.5061/dryad.12jm6z17) and upon reasonable request from the corresponding author.
5. Ehara, H. et al. Structure of the complete elongation complex of RNA polymerase II with basal factors. Science 357, 921–924 (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 (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 (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 (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 (2004).
10. Vos, S. M. et al. Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature 560, 607–612 (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 (2006).
12. Krueger, B. J. et al. LARP7 is a stable component of the 7SK snRNP while P-TEFB is required for the transition into productive elongation. Mol. Cell 77, 118–131 (2019).
13. Yang, Z., Zhu, Q., Luo, K. & Zhou, Q. The 7SK small nuclear RNA inhibits the transactivation response element. Proc. Natl Acad. Sci. USA 113, 12721–12726 (2016).
14. Fujinaga, K., Luo, Z. & Peterlin, B. M. Genetic analysis of the structure and function of 7SK small nuclear ribonucleoprotein (snRNP) in cells. J. Biol. Chem. 289, 21181–21190 (2014).
15. Prasanth, K. V. et al. Nuclear organization and dynamics of 7SK RNA in regulating gene expression. Mol. Biol. Cell 21, 4184–4196 (2010).
16. Kohoutek, J., Blazek, D. & Peterlin, B. M. Hexim1 sequesters positive transcription elongation factor b from the class II transactivator on MHC class II promoters. Proc. Natl Acad. Sci. USA 103, 17349–17354 (2006).
17. Barboric, M. et al. Interplay between 7SK snRNA and oppositely charged molecules containing Cdk9 phosphorylated at threonine 186. J. Biol. Chem. 280, 28819–28826 (2005).
18. Scholzinger, 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 (2010).
19. 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 (2005).
20. Martinez-Zapnen, D. et al. Intermolecular recognition of the non-coding RNA 7SK and HEXIM1 protein in perspective. Biochimie 117, 63–71 (2015).
21. Chavali, S. S. et al. Arginine-fork motif in the riboregulator 7SK. Nucleic Acids Res. 37, 389 (2008).
22. Brown, J. A. Unraveling the structure and biological functions of RNA triple repeats is critical for processive transcription elongation. Mol. Cell 77, 118–131 (2019).
23. Schölzel, A. J. et al. TAX interaction with 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. Nucleic Acids Res. 33, 7000–7010 (2005).
24. Sano, M. et al. Activation and function of cyclin T-Cdk9 positive transcription elongation factor b in cardiac muscle-cell hypertrophy. Nat. Med. 10, 1310–1317 (2004).
25. Kretz, A. L. et al. CDK9 is a prognostic marker and therapeutic target in pancreatic cancer. Tumour Biol. 39, 1010428317693403 (2017).
26. Scholzinger, 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 (2018).
27. Shao, H. et al. HEXIM1 controls P-TEFB processing and regulates drug sensitivity in triple-negative breast cancer. Mol. Biol. Cell 31, 1867–1878 (2020).
28. Pico, W. K., Jiang, 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 (2010).
29. 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 (2006).
30. Muniz, L., Egloff, S., Ughy, B., Jady, B. E. & Kiss, T. Controlling cellular P-TEFB activity by the HIV-1 transactivation transactivator Tat. PLoS Pathog. 6, e1001152 (2010).
31. Sobhian, B. et al. HIV-1 Tat assembles a multifunctional transcriptional complex and stably associates with the 7SK snRNP. Mol. Cell 38, 439–451 (2010).
32. Krueger, B. J., Varzavand, K., Cooper, J. J. & Price, D. H. The mechanism of release of P-TEFB and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. PLAN ONE 5, e12335 (2010).
33. Barboric, M. et al. Tat competes with HEXIM1 to increase the active pool of P-TEFB for HIV-1 transcription. Nucleic Acids Res. 35, 2003–2012 (2007).
34. Yik, J. H., Chen, R., Perda, A. C., Samford, C. S. & Zhou, Q. A human immune deficiency virus Tat-like arginine-rich RNA binding domain is essential for HEXIM1 to inhibit RNA polymerase II transcription through 7SK snRNA-mediated inactivation of P-TEFB. Mol. Cell Biol. 24, 5094–5105 (2004).
35. 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 snRNA. J. Biol. Chem. 280, 16360–16367 (2005).
36. Yik, J. H., Chen, R., Perda, 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, 16588–16576 (2005).
37. Egloff, S., Van Herreweghe, E. & Kiss, T. Regulation of polymerase II transcription by 7SK small nuclear RNA. RNA 7, 1075–1086 (2001).
59. 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 (2002).

60. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).

61. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).

62. Guntert, P., Mumenthaler, C. & Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).

63. 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 (2010).

64. Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M. & Svergun, D. I. SASBDB: towards an automatically curated and validated repository for biological scattering data. *Protein Sci.* **29**, 66–75 (2020).

**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’S. and J.L.S.). SAXS data were 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.

**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’S. analyzed data and edited the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03734-w.

Correspondence and requests for materials should be addressed to Victoria M. D’Souza.

Peer review information Communications Biology thanks Shintaro Aibara and the other, anonymous reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Anam Akhtar and Christina Karlsson Rosenthal.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.