HEXIM1 homodimer binds two sites on 7SK RNA to release autoinhibition for P-TEFb inactivation

Yuan Yang, Maria Grazia Murrali, Yaqiang Wang, Sabrina Galvan, Neha Ajjampore, Juli Feigon*

Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA 90095-1569, USA

*Correspondence to Juli Feigon: feigon@mbi.ucla.edu

Abstract

Hexim proteins are RNA-dependent regulators whose main target is 7SK long non-coding RNA, a major regulator of eukaryotic mRNA transcription. 7SK RNPs control available intracellular concentrations of the kinase P-TEFb (Cdk9–CyclinT1/2) by sequestering it in an inactive form. Active P-TEFb phosphorylates NELF, DSIF, and the RNA polymerase II CTD to transition it from promoter-proximal pausing to productive elongation. P-TEFb associates with 7SK RNP via Hexim, which directly binds 7SK RNA. However, free Hexim is in an autoinhibited state that cannot inactivate P-TEFb, and how Hexim autoinhibition is released by 7SK remains unknown. Here, we show that one Hexim1 homodimer binds two sites on linear 7SK RNA in a manner that exposes the Cdk9 binding sites, which are otherwise masked within the autoinhibited dimer. These results provide mechanistic insights into Hexim–RNA specificity and explain how P-TEFb can be effectively regulated to respond to changing levels of transcriptional signaling.

Introduction

Hexim (HEXamethylene-bis-acetamide-Inducible protein in vascular smooth Muscle cells)[1] is a unique family of RNA-dependent regulatory proteins with no related protein family[2]. The best characterized and main functional target for Hexim is 7SK RNA, an abundant ~332 nt non-coding RNA that regulates eukaryotic transcription pause release. A few other RNAs interact with Hexim, including NEAT1[3] and several mRNAs[4]. Two paralogues of Hexim exist in mammals, Hexim1 and Hexim2, which are expressed in different tissue distribution patterns and can serve as compensatory replacements for each other in the 7SK RNP[5,6]. Hexim inhibits the kinase activity of positive transcription elongation factor b (P-TEFb) by sequestering it in the 7SK RNP, where 7SK RNA binding to Hexim is a prerequisite[7-9]. P-TEFb is a heterodimeric kinase comprising a catalytic subunit Cdk9 and regulatory subunit Cyclin T1/2 (CycT1 or CycT2, specific for Hexim1 and Hexim2, respectively[10]). Release from 7SK RNP allows P-TEFb to phosphorylate RNA polymerase II C-terminal domain (PolII-CTD) and several other transcription factors (NELF, DSIF, PAF and SPT6), a step required to trigger the promoter-proximal paused state to transition into productive elongation[11,12]. 7SK RNP is also an important host target for HIV-1 viral transcription and replication, where Tat (Trans-activator of transcription) protein hijacks the 7SK RNP to the HIV-1 promoter and extracts P-TEFb into a viral super elongation complex together with TAR (trans-activation response) RNA element of the nascent HIV-1 transcript[13-16].

Hexim1 has a C-terminal coiled-coil domain that mediates stable homodimerization in solution and serves as the interaction site for CycT1[17] (Fig. 1A). The N-terminal one third (1-149) of Hexim1 is poorly conserved among vertebrates (Supplementary Figs. 1 and 2) and is predicted to be disordered (Supplementary Fig. 3)[17]. The central part of Hexim1 (150-250) has not been characterized previously, with the exception of a small region (~20 aa nuclear localization signal stretch) that binds 7SK RNA, also referred to as Basic Region (BR, aa 150-169)[7,8]. The BR of Hexim1 and Hexim2 contains two Arg/Lys-rich stretches separated by a

proline-serine; these are often referred to as ARM (arginine-rich motif)[7,18] due to sequence similarity to the single ARM of HIV-1 Tat (Fig. 1B). Following the BR is a linker containing a conserved PYNT motif that has been shown to interact with Cdk9 and is required for inhibition of Cdk9 kinase activity by Hexim[7,19]. C-terminal to the PYNT motif is a highly negatively charged sequence, referred to as Acidic Region (AR, aa 211-249). Previous studies proposed a model where charge-charge interaction between Hexim BR and AR is required for the RNA-dependence of Hexim-mediated P-TEFb inactivation[9,20]. However, it is unclear how this charge-mediated Hexim autoinhibition prevents PYNT–Cdk9 interaction in the absence of RNA.

Hexim–7SK RNA binding site specificity and stoichiometry are also unsettled issues, with three different considerations from previous studies. One study proposed that Hexim1 is a promiscuous double-stranded RNA binder[21], while the majority of other in vitro studies[22-25] focus on an RNA-binding site specific for Hexim consisting of a signature palindrome of base-paired GAUC motifs (GAUC2 and GAUC3, nts 42-45 and 64-67, respectively) in the distal half of 7SK stem-loop 1 (SL1), which has been generally used as a search motif for 7SK genes[26,27] (Site 1 in Fig. 1C). Site 1 has also been identified biochemically and structurally as the RNA-binding site for HIV-1 Tat[24,28]. Importantly, one of these two studies[28] biochemically identified a second Hexim1 binding site in the middle region of SL1 that lacks GAUC motif (Site 2 in Fig. 1C). Furthermore, a single GAUC motif (GAUC1, nts 13-16) is present in proximal SL1 within one RNA helical turn of the binding/capping site for methylphosphate capping enzyme, or MePCE (G1-C108, G2-C107 and single-stranded nts 109-114)[29]. The issues of Hexim binding specificity and stoichiometry are further complicated by the recent identification of at least two different 7SK RNA conformations in cells[30]: a linear 7SK where a long SL1 forms the above Hexim binding site(s), and a circular 7SK where a shorter SL1alt forms (Fig. 1C). A rearranged GAUC palindrome forms in the circular 7SK, consisting of the first two GAUC motifs (GAUC1 and GAUC2; Site 3 in Fig. 1C). Both linear and circular 7SK conformations constitutively assemble with two core protein components, MePCE and Larp7[31,32]. We and others showed that switching

between the two 7SK RNP pools, likely corresponding to the two RNA conformations, is required for regulated P-TEFb release in response to stress[30,32,33].

Here, we systematically dissected the 7SK RNA binding sites of Hexim1 and characterized the Hexim1 binding stoichiometry in vitro. Using PRE NMR methods, we find that PYNT motif contacts both BR and AR in the inter-monomer interaction of Hexim1, and is unmasked upon RNA interaction with the BR. Combining NMR and ITC binding experiments, we show that the inter-monomer interaction in Hexim1 dimer enhances its specificity for linear 7SK RNA conformation over the circular conformation, through cooperative binding to the two high affinity sites in SL1 and weakening the affinity to the single RNA sites by sequestering BR in autoinhibition. These findings reveal an extensive Hexim1 intra-dimer interaction network formed by intrinsically disordered segments, provide mechanistic insights into how Hexim1 autoinhibition correlates with 7SK RNA-dependence of P-TEFb inactivation, and show how 7SK RNA conformational switching affects Hexim/P-TEFb sequestration in 7SK RNP.

Results

Hexim1 uses tandem arginine/lysine-rich motifs for RNA interaction

To evaluate the conformational preferences of the conserved central part, we expressed and purified a monomeric Hexim1 construct (amino acids 142-253) encompassing BR, linker (L) containing PYNT motif, and AR. This construct is referred to as BR-L-AR in the following text. We investigated the interaction of BR-L-AR with the previously characterized 7SK RNA binding site (Site 1 in Fig. 1C), using designed RNA constructs from the distal end of linear 7SK SL1 (SL1-dI and SL1-d$I_{\Delta U}$), by NMR (Fig. 2 and Supplementary Fig. 4) and electrophoretic mobility shift assay (EMSA; Supplementary Fig. 5). NMR 1D imino proton spectra of the RNA in the absence and presence of BR-L-AR (Fig. 2A) and ^{15}N-^{1}H HSQC spectra of BR-L-AR titrated with RNA (Fig. 2C) show that BR-L-AR forms a 1-to-1 complex with SL1-d$I_{\Delta U}$. The RNA imino

spectra exhibit slow to intermediate exchange on the NMR timescale, whereas the protein ^{15}N-^{1}H HSQC spectra exhibit intermediate to fast exchange.

Several additionally truncated monomeric Hexim1 constructs, BRhalf, BR and BR-L (Fig. 2B and Supplementary Fig. 6), were used to compare with BR-L-AR and to probe interactions with 7SK RNA in this study (Supplementary Fig. 5). We observed a solubility limit of ~70μM for BR-L-AR despite testing many conditions. However, BRhalf, BR and BR-L constructs are easily soluble at ~0.5–0.8 mM under similar conditions. This suggests that BR-L-AR is prone for aggregation mediated by intermolecular interactions that do not exist in constructs lacking the AR.

Despite the low BR-L-AR concentration we could achieve, we were able to assign $^{1}H_{N}$, ^{15}N, $^{1}H\alpha$, $^{13}C\alpha$, and $^{13}C\beta$ chemical shifts from multi-dimensional heteronuclear NMR experiments (see Methods) and calculate secondary structure propensity (SSP) scores[34] using $C\alpha$ and $C\beta$ chemical shifts (Fig. 2D). The SSP scores, ranging from +1 to -1, reflect the fraction of α- or β- structure formed by each residue. From the SSP scores, we identified an α-helix (aa 171-194) in the linker region, between the BR and PYNT motif. We refer to the two Arg/Lys-rich regions in BR as ARM1 and ARM2 from here on (Figs. 1B and 2B). The ARM1 residues 151-154 have a moderate propensity to form β-structure, with a minimum SSP score reaching -0.45. The rest of BR-L-AR is largely disordered, with SSP scores well within the range of +0.3 to -0.3. We next analyzed spectral changes of backbone amide resonances in BR-L-AR when SL1-$dI_{\Delta U}$ is titrated in (Fig. 2C). We observed significant chemical shift changes and severe peak broadening for residues 151-170 spanning the entire BR, including both ARM1 and ARM2 as well as the PYYK sequence following ARM2 (Fig. 2B,E). Interestingly, we also observed chemical shift changes in the C-terminal region of AR at residues 241-242 (Fig. 2C,E). The severe line broadening of BR due to RNA binding persists in the range of 0–300 mM KCl salt concentration, indicating chemical exchange between two or more RNA-bound protein

conformations. Possibilities include interaction with ARM1 or ARM2 or an ARM1-ARM2 β-hairpin. Consistent with two or more RNA-bound protein conformations, we observed two or three distinct sets of resonances in the ^{15}N-^{1}H HSQC spectrum of the SL1-dI bound BR-L performed in the presence of 3mM magnesium ions (Supplementary Fig. 6F). In summary, Hexim1 BR-L-AR is intrinsically disordered, contains a short α-helix in the linker following the BR, and both ARM1 and ARM2 in BR participate in RNA binding. Whether both ARMs bind simultaneously or independently is less clear. In contrast to Hexim, HIV-1 Tat protein has only one ARM that is necessary and sufficient for binding 7SK linear SL1-dI[23,24,28].

Linear 7SK SL1 contains two high affinity Hexim1 monomer binding sites

RNA secondary structure probing by either SHAPE-MaP under high Mg^{2+} [35] or DANCE-MaP that deconvolutes multiple conformations[30] identified two conformations for the 5'-hairpin of 7SK: one is the SL1 identified in the original linear 7SK conformation[36]; the other is a complete rearrangement of SL1 into a shorter alternative SL1 (SL1alt) where the first two GAUC motifs (GAUC1 and GAUC2) pair (Figs. 1C and 3C). Additional RNA elements contributing to Hexim1 binding specificity for the distal SL1 (Site 1) are two flanking U-rich bulges (U40U41 and U63) that can dynamically form base triples with the A-U, U-A or first G-C base pairs; which of these base triples if any are observed is highly dependent on sample conditions[23,24,37,38]. Nucleotide substitutions that disrupt base-triple formation capabilities from the two U-rich bulges to the central A-U base-pairs in the GAUC2-GAUC3 palindrome (Fig. 3A) reduce Hexim1 binding[37]. In circular 7SK SL1alt, a UU asymmetric bulge (U40U41) is also present 3' to the alternative GAUC palindrome (Fig. 3E). However, it has not been established whether SL1alt confers equivalent specificity and affinity for Hexim1 binding when compared to linear 7SK SL1.

In order to resolve the outstanding questions of Hexim1–7SK RNA binding site location(s), stoichiometry, and specificity, we first mapped the binding sites for monomeric Hexim1 using NMR (Fig. 3) and determined the corresponding binding stoichiometry by EMSA

(Supplementary Fig. 5). We designed a series of RNA stem-loop constructs encompassing segments of distal (d), middle (m) and proximal (p) regions of linear 7SK SL1 as well as circular 7SK SL1alt (Fig. 3A and Supplementary Fig. 5). Using chemical shift perturbation (CSP) of RNA imino proton resonances in ^{1}H-^{1}H NOESY spectra in the absence versus presence of monomeric Hexim1 BR-L-AR or BR, we identified SL1 Sites 1 and 2 as two high affinity binding sites, and Sites 4 and 5 as two lower affinity binding sites (Fig. 3 and Supplementary Fig. 7). The relative order of binding affinities was determined based on the following observations: (1) SL1-dI (Site 1) is saturated earlier than SL1-dII (Site 4) by BR-L-AR in EMSA experiments and two sequential shifts are observed on SL1-$d_{\Delta U}$ and SL1-d (contain both Sites 1 and 4) corresponding to individual monomers binding to Site 1 followed by Site 4 (Supplementary Fig. 5B,C); (2) when using longer 7SK RNA constructs that include two adjacent binding sites, the higher affinity site of the two is occupied by stoichiometric addition of Hexim1 BR as shown by NMR CSP of RNA iminos, i.e. in SL-d Site 1 shows CSP instead of Site 4, in SL1-mdII Site 2 shows CSP instead of Site 4, and in SL1-pm Site 2 shows CSP instead of Site 5 (Fig. 3A and Supplementary Fig. 5A); (3) upon binding RNA, the ^{15}N-^{1}H HSQC spectrum of BR monomer shows intermediate exchange behavior on the NMR timescale for SL1-dI (Site 1), intermediate to fast exchange for SL1-mdII (Site 2) and fast exchange for SL1-pIIo (non-specific binding site) (Fig. 3A, B, Supplementary Fig. 6). In comparison, BR-L-AR and SL1alt RNA are in intermediate-to-slow exchange in NMR binding experiments, where SL1alt RNA remains unsaturated at 1:1.2 RNA:protein ratio (Supplementary Fig. 8).

We also acquired thermodynamic parameters for the individual RNA binding sites by ITC using monomeric Hexim1 NBLA (N-terminal region and BR-L-AR; aa 1-253; Fig. 3D and Table 1). The binding affinities determined by ITC recapitulated what we observed by NMR and EMSA binding experiments, where the K_D values are 83 ± 28 nM, 189 ± 25 nM and 234 ± 52 nM for Sites 1, 2 and 3, respectively. These findings indicate that linear 7SK SL1 contains two high-

affinity sites for monomeric Hexim1 and two additional low-affinity sites that could serve as docking sites during initial encounter with Hexim1. In contrast, circular 7SK SL1alt contains only one binding site for one Hexim1 monomer.

Since full-length Hexim1 is a stable dimer, the EMSA, NMR, and ITC results with monomeric Hexim1 constructs (BR, BR-L-AR, NBLA) discussed above suggest that it could simultaneously bind two sites on SL1, one for each monomer. Indeed, EMSAs of SL1 constructs with full-length Hexim1 show that Hexim1 *dimer* binds first to Sites 1 and 2 and subsequently to Sites 4 and 5 (Supplementary Fig. 5D,E). Further results for how the Hexim1 dimer interacts with SL1 will be presented in a later section.

Characterization of Hexim1 autoinhibited conformation by PRE

Previous mutational analysis indicated that a positive-negative charge-mediated autoinhibition between Hexim1 BR and AR prevents inactivation of Cdk9 in the absence of 7SK RNA[20]. In order to characterize the Hexim1 autoinhibited conformation, we made individual serine-to-cysteine substitutions in the AR (S226C, S233C and S237C) of BR-L-AR, and covalently attached the nitroxide spin label MTSL to the cysteine side chain[39]. The introduction of the paramagnetic MTSL label causes distance-dependent line-broadening effects termed paramagnetic relaxation enhancement (PRE)[40], which can capture distance information for even weak and transient interactions[41,42] commonly found in disordered protein regions[43]. Here, we use PRE to probe monomeric Hexim1 conformation from the peak intensity changes in the presence versus the absence of MTSL.

In ^{15}N-^{1}H HSQC spectrum of ^{15}N-labeled BR-L-AR(S226C^{MTSL}), we observed an extensive broadening network from the MTSL tag (Fig. 4A,B). We defined regions of primary broadening (red shading) and secondary broadening (orange shading) as stretches harboring <u>3 or more residues</u> with intensity ratios lower than (mean - σ) and lower than (mean - 0.5σ), respectively (Fig. 4B). With this analysis, we observed that S226C^{MTSL} is in closest proximity to both the beginning and the end of the linker helix (red), and the secondary broadened regions

involve BR, PYNT, and the C-terminal end of AR (orange). In the ^{15}N-^{1}H HSQC spectrum of ^{15}N BR-L-AR(S237C^{MTSL}), almost all the resonances are broadened more severely than S226C^{MTSL} (Supplementary Fig. 9), indicating a significant role of the C-terminal part of AR in the autoinhibited conformation(s), consistent with our observation that residues Phe241-Met242 have significant CSP upon RNA binding (Fig. 2E). Note that these experiments where the MTSL-tag is attached to the NMR-observed ^{15}N BR-L-AR capture both *intra*- and *inter*-monomer interactions, if existing.

To dissect the *intra*- and *inter*-monomer components, we performed mixed labeling experiments where MTSL-tag is attached to ^{14}N BR-L-AR (NMR invisible), which is then added to ^{15}N BR-L-AR without MTSL (NMR detected) in 1-to-1 ratio. In these mixed labeling experiments, only *inter*-monomer interactions will result in line-broadening of ^{15}N BR-L-AR resonances, although to a much less extent than the above ^{15}N BR-L-AR^{MTSL} experiments, due to the fact that statistically only 50% of the “dimer” population is the combination of ^{15}N and ^{14}N-MTSL labeled proteins. Using the same definition of primary and secondary broadened regions, we observed that S237C^{MTSL} is in close proximity to almost the entire BR (154-168), with ARM2 broadened more severely than the R154–156 of ARM1 (Fig. 4C). S233C^{MTSL} mainly broadens ARM2–PYYK (Fig. 4D) and S226C^{MTSL} mainly broadens ARM2–PYYK, beginning of linker helix and PYNT motif (Fig. 4E). Note that the PYNT motif is a secondary broadened region in both the S237C^{MTSL} and S233C^{MTSL} mixed labeling experiments, and the central part of the linker helix is involved in secondary broadened region in S237C^{MTSL} experiment. These results indicate an overall head-to-tail orientation for BR-L-AR within the *inter*-monomer interaction, since the primary broadened region (red) shifts from N- to C-terminal direction when we move the MTSL-tag from C- to N-terminal direction (237 to 233, and to 226).

When comparing the ^{15}N S226C^{MTSL} PRE profile with mixed labeling S226C^{MTSL} profile, it is striking that the primary broadened regions in the mixed labeling experiment match approximately the secondary broadened regions in ^{15}N S226C^{MTSL} (compare Fig. 4E and 4B).

This indicates that the secondary broadened regions in ^{15}N S226C^{MTSL} are likely the *inter*-monomer component, whereas the primary broadened regions (beginning and end of linker helix) are the *intra*-monomer component. In summary, we observed both *intra*- and *inter*-monomer interactions in the monomeric Hexim1 BR-L-AR by PRE NMR, and the *inter*-monomer interaction likely adopts a head-to-tail orientation (N→C to C→N). These results also suggest that the PYNT motif is sequestered in *inter*-monomer and exposed in *intra*-monomer interactions.

Hexim1 PYNT motif interaction with BR and AR is unmasked by RNA binding

The PYNT motif plays a critical role in directly binding and inactivating the Cdk9 kinase subunit of P-TEFb[7,19]. We next investigated the PYNT interactions using an A201C substitution in BR-L-AR such that an MTSL-tag is attached immediately N-terminal to the P^{202}YNT^{205} motif (NMR spectra in Supplementary Fig. 10). The PRE profile of ^{15}N BR-L-AR(A201C^{MTSL}) recapitulated the secondary broadened regions in BR-L-AR(S226C^{MTSL}) experiment (Figs. 5A and 4B), where the PYNT motif, BR and C-terminal part of AR are significantly broadened. To further verify this tripartite BR–PYNT–AR interaction network, we looked at S183C substitution located in the central part of the linker helix (NMR spectra in Supplementary Fig. 11), a region that is broadened in the S237C mixed labeling experiment (Fig. 4C orange). The PRE profile of ^{15}N BR-L-AR(S183C^{MTSL}) shows significant broadening of the entire linker helix and PYNT motif, but only secondary broadening for the C-terminal end of the linker helix, BR and the two ends of AR (Fig. 5B). This indicates that the *inter*-monomer BR– PYNT–AR interactions partially involve the central part of the linker helix.

Lastly, we investigated the effect of 7SK RNA binding on these interactions. Addition of SL1-dI to ^{15}N BR-L-AR(S183C^{MTSL}) resulted in significant peak intensity recovery of the whole molecule with the exception of BR (Fig. 5B cyan). Note that this does not mean BR is still in contact with S183C^{MTSL} tag, since BR residues are severely broadened due to RNA binding such that their intensities will not recover regardless of PRE effects (Fig. 2E). Taken together,

the PRE data show that the PYNT motif is the interaction hub for *inter*-monomer interactions, where it simultaneously contacts both BR and AR. This *inter*-monomer interaction is released upon RNA binding to BR, resulting in unmasking of PYNT so it is available to bind and inactivate Cdk9 in the 7SK RNP.

Hexim1 autoinhibition enhances the specificity for linear 7SK SL1

In solution NMR spectra of full-length Hexim1, the majority of the BR-L-AR region and C-terminal coiled-coil domain (CC) residues are broadened beyond detection, although resonances from CC are visible at higher temperature and higher pH (Supplementary Fig. 12). The extent of line-broadening, which is likely due to chemical exchange of the BR-L-AR region and the CC in addition to increase in molecular weight, creates challenges for studying full-length dimeric Hexim1 conformations by traditional solution NMR methods including PRE-based methods discussed above. We therefore turned to ITC to obtain information regarding Hexim1–7SK RNA binding behavior and the role of Hexim1 autoinhibition in binding. We dissected RNA binding sites by using our shorter SL1 constructs (SL1-dI, SL1-m) and SL1alt versus the full-length SL1 that contains both Sites 1 and 2, and dissected the contribution from Hexim1 autoinhibition by using monomeric Hexim1 NBLA (aa 1-253) (mentioned in previous section) versus full-length dimeric Hexim1 (aa 1-359), where *inter*-monomer autoinhibition is enforced by dimerization in the full-length Hexim1 but not in the monomer NBLA (Table 1 and illustrated in Fig. 6). First, we systematically compared the N values to acquire information regarding binding stoichiometry (Table 1). We observed that the monomeric NBLA binds as one molecule per RNA to SL1-dI, SL1-m and SL1alt, which contain Sites 1, 2, and 3, respectively (N values close to 1) and binds as two molecules per full-length linear 7SK SL1, which contains Sites 1 and 2 (N value 0.4). Full-length Hexim1 (dimer) also binds as one monomer per SL1-dI, SL1-m and SL1alt (N values close to 1, using monomer concentrations when fitting titration curves), and binds as two monomers, i.e. one dimer, per linear SL1 (N value 0.4). We further verified the stoichiometry by reversing the titration order and switching Hexim1 from sample cell to the

titrating syringe, and again observed N values close to 1 for the short RNAs and 1.8 (close to 2) for linear 7SK SL1. These results are consistent with what we concluded from EMSA and NMR experiments (Fig. 3 and Supplementary Fig. 5) where both monomers within Hexim1 dimer are engaged in RNA binding, to Sites 1 and 2, simultaneously.

When we compared the dissociation constants K_D, we find that dimeric Hexim1 binding to linear 7SK SL1 shows cooperative binding with a much lower K_D (73nM) than the single RNA sites SL1-dI and SL1-m (~800nM for both). In contrast, no cooperativity was observed for monomeric Hexim1 NBLA, with K_D value for SL1 (335nM) slightly larger than those of SL1-dI and SL1-m. This cooperative effect is recapitulated in the experiments where the titration order is reversed. Importantly, the Hexim1 dimer binds 8–10 times weaker than the monomer for all three single RNA sites SL1-dI, SL1-m and SL1alt. Together, these data indicate that there is significant *inter*-monomer autoinhibition in Hexim1 dimer that weakens its binding to single RNA binding sites by sequestering the BR. This results in enhanced specificity for full-length linear 7SK SL1 over the single RNA sites.

We also looked at the weaker binding sites SL1-dII and SL1-p determined by NMR and EMSA (Sites 4 and 5, respectively, Supplementary Fig. 5), which exhibit much larger K_D values (1221 nM and 2108 nM, respectively) than Sites 1 and 2 for Hexim1 dimer. Interestingly, the N values for these weaker binding sites are slightly smaller than those for SL1-dI and SL1-m, indicating a partial release of autoinhibition such that only one monomer is available for binding within a population of the Hexim1 dimer. In contrast, SL1alt in circular 7SK conformation has weak binding affinity to Hexim1 dimer (K_D = 2035 nM), albeit with a N value closer to those of SL1-dI and SL1-m. This Hexim1 dimer–SL1alt affinity is unexpectedly much lower than the monomer–SL1alt, suggesting a significant kinetic component for the RNA binding affinity that differ among the single RNA sites. Consistent with these experiments, a SL1-d(api) construct used in a previous binding study [25] (similar to our SL1-d, Sites 1 and 4) has an enhanced

binding affinity (96nM) but a slightly smaller N value (1.4) than SL1, indicating a less stable release of Hexim1 autoinhibition.

In summary, Hexim1 dimer preferentially and cooperatively binds Sites 1 and 2 in linear 7SK SL1 with each monomer, respectively, and the *inter*-monomer autoinhibition within the dimer enhances the specificity for linear SL1 over single RNA binding sites, including the circular SL1alt, through cooperative binding to the two high affinity sites in SL1 and weakening the affinity to the single RNA sites. Together with the above PRE results, we conclude that Hexim1 binding to two sites in linear 7SK SL1 provides an effective switch to fully release Hexim1 autoinhibition in the dimer, which subsequently exposes two PYNT motifs for Cdk9 binding and inactivation of two P-TEFb heterodimers (illustrated in Fig. 6).

Discussion

Intrinsically disordered RNA-binding regions play significant roles in cellular function, but it has remained challenging to define their molecular mechanism when compared to their well-characterized folded RNA-binding domain counterparts[44,45]. Recent studies also showcase examples of disordered domains that can modulate the RNA- and DNA-binding of a folded domain within the same protein molecule[46,47]. Here, our studies of Hexim1 conformational preferences and its interaction with a dynamic non-coding RNA, 7SK, highlight the unique features of this protein and the complexity of IDR protein–RNA interactions. We show that Hexim1 uses a stand-alone disordered basic region (BR) to bind dsRNA sites with flanking U-rich bulge(s), while a regulatory disordered acidic region (AR) autoinhibits the RNA-binding to enforce specificity. Using a monomeric BR-L-AR construct, we show that linear 7SK SL1 contains two high affinity and two low affinity binding sites. Hexim1 dimer binds linear 7SK SL1 in a 1-to-1 stoichiometry, with each monomer BR engaged with one of the two high affinity sites. In the absence of RNA, the monomeric Hexim1 construct can adopt conformations with either *intra*- or *inter*-monomer interaction, but the PYNT motif is masked by a BR–PYNT–AR

interaction network only in the head-to-tail *inter*-monomer conformation (Fig. 6C). Strikingly, 7SK binding to the BR effectively releases *inter*-monomer PYNT masking. Thus, our data together with previous studies[9,20] support a model in which formation of the *inter*-monomer interaction in native Hexim1 dimer (Fig. 6D) results in simultaneous autoinhibition of RNA binding, by sequestering BR, and Cdk9 binding, by sequestering PYNT.

The regulatory role of the AR is potentially two-fold: (1) autoinhibition enhances specificity towards linear 7SK SL1 over circular 7SK; and (2) autoinhibition prevents Hexim1 from being trapped by non-functional decoy RNAs, such that searching process for functional 7SK RNA sites is accelerated, by analogy to the recently discovered role of disordered D/E patches accelerating DNA-target search[48]. In light of this particular feature of Hexim1, our study illustrates the importance of considering Hexim1 dimer interaction with a full-length linear 7SK SL1 and with circular 7SK SL1alt, which provides a more complete picture than previous seminal studies where only Site 1 (and Site 4) of linear 7SK SL1 was considered[23-25,37].

Based on the comprehensive binding studies presented here and our previous structural studies of MePCE–7SK–Larp7 core RNP[29,32,49], we conclude that the stoichiometry of the P-TEFb inhibitory 7SK RNP components is one linear 7SK RNA, one MePCE, one Larp7, one Hexim dimer, and two P-TEFb heterodimers (Fig. 6D). This model of two P-TEFb inactivated by Hexim binding simultaneously to two adjacent RNA binding sites is advantageous for modulating the amplitude of 7SK RNP response to altering levels of cellular signaling, where one or two P-TEFb units can be activated per unit signaling event. Based on our model, it is likely that 7SK–Hexim and Hexim–P-TEFb interactions cooperatively stabilize each other during 7SK RNP assembly, resulting in a kinetically advantageous regulatory step during the proposed PolII elongation switch-off via 7SK[50].

Recent studies have identified two different strategies for P-TEFb regulated release from 7SK RNP. One is the switch between two main 7SK RNA conformations, linear and circular[30,36], where SL1 in linear 7SK rearranges into a short SL1alt in circular 7SK (Fig. 1C). Given that

Hexim1 dimer binds circular 7SK SL1alt with micromolar affinity comparable to the two weak affinity sites on linear SL1, while Hexim1 dimer binds to linear 7SK SL1 with 73 nM affinity (Fig. 6B), RNA conformational switching indicates an effective strategy for P-TEFb binding and release. The other strategy involves N6 methylation of 8 adenines (m^6A) in 7SK by METTL3-METTL14, observed more prevalently in lung cancer cell lines than non-malignant cells[51] or in HeLa cells[52]. Specifically, m^6A43 and m^6A65 in GAUC2 and GAUC3 motifs[51] would disrupt base-triple formation involving these two A-U base pairs, potentially reducing Site 1 binding affinity given the importance of these base-triples for Hexim binding[37]. Thus, these two base modifications will likely change SL1 from two high affinity sites (Sites 1 and 2) into only one high affinity site (Site 2), resulting in significantly less stable binding to Hexim1 dimer. Together, our studies explain how 7SK conformational switching and m6A modifications affect Hexim1 binding and release. Finally, we note that Site 2 has GACA instead of GAUC in Site 1, but could still form a U•A-U triple at the second base pair from the flanking 5' bulge. In contrast, Sites 4 and 5 do not have GAUC palindrome and cannot form U•A-U triple, although Site 4 can form a C•G-C triple[23,24,37]. Thus, U•A-U triple containing sites from 5' U-bulge appear to bind with higher affinity to Hexim1, indicating structural specificity rather than strictly sequence specificity.

We compared Hexim1 with its paralogue Hexim2, and found that Hexim2 has a much shorter AR lacking negatively charged residues in the C-terminal AR (Supplementary Fig. 2). In contrast, Hexim1 AR has both an N-terminal and C-terminal cluster of D/E residues. Our studies of Hexim1 indicate a more significant role of the C-terminal D/E cluster in the BR–PYNT–AR interaction network. The absence of C-terminal D/E cluster in Hexim2 could reduce autoinhibition in Hexim2 vs Hexim1, leading to altered P-TEFb regulation in tissues primarily expressing Hexim2. HIV-1 Tat protein binds only linear 7SK SL1 Site 1[28] and does not contain an acidic patch that could inhibit its single ARM (Supplementary Fig. 13). Previous studies of Tat–Hexim competition only compared binding to Site 1, where Tat and monomeric Hexim1 had

similar binding affinities[24,25,38]. We postulate that Tat binding to linear 7SK SL1 Site 1 will result in only Site 2 being available for Hexim dimer binding. Our results indicate the importance of considering both SL1 Sites 1 and 2 as well as Hexim autoinhibition for understanding Tat–Hexim competition.

In summary, this work resolves three longstanding issues for Hexim function in 7SK RNP: how autoinhibition ensures Hexim1 specificity towards linear 7SK RNA; stoichiometry of Hexim–7SK binding and 7SK RNP components; and how Hexim1–7SK binding is coupled to release of the PYNT motif from autoinhibited Hexim1 to allow Cdk9 binding and inactivation in the 7SK RNP. Our studies of Hexim1 raise interesting questions for future investigations of the various P-TEFb release factors that interact with and/or remodel 7SK RNA. Characterization of other Hexim1 target RNAs including NEAT1[2,3] will also benefit from the details of autoinhibition in Hexim1 dimer and 7SK RNA binding sites presented here.

Methods

Plasmids and mutagenesis

Full-length human Hexim1 protein (aa 1–359) and monomeric Hexim1 BR-L-AR (aa 142–253) were cloned into pETDuet-1 vector with N-terminal His_6-tagged maltose binding protein (MBP) fusion followed by Tobacco Etch Virus (TEV) protease cleavage site between MBP and Hexim1. Further truncated monomeric Hexim1 constructs, BR-L (aa 136–209) and BRext (aa 149–179) were cloned into pETDuet-1 vector with N-terminal His_6-tagged 56-residue B1 domain of Streptococcal protein G (GB1) fusion[53] followed by Tobacco Etch Virus (TEV) protease cleavage site between GB1 and Hexim1 monomers. BR (aa 141–165) and BRhalf (aa 141–159) constructs were cloned by site-directed mutagenesis using the pETDuet-1 GB1-ts-BR-L plasmid as template. A non-native glycine remains at the N-terminus of the final full-length Hexim1 protein, BR-L-AR, BR-L, BR and BRhalf after TEV protease cleavage, whereas a native Glycine residue from BRext (aa 149) was a part of the TEV cleavage site (ts). Cysteine substitutions of BR-L-AR variants for PRE (S183C, A201C, S226C, S233C and S237C) were cloned by site-directed mutagenesis using the pETDuet1 MBP-ts-BR-L-AR as template.

Minimum linear 7SK (SL1-SL4^{ms2}) and linear SL1 RNA genes were cloned into pUC19 vector with a Hammerhead ribozyme gene at the 3′ end followed by a BamHI restriction enzyme cleavage sequence. All site-directed mutagenesis was performed using the Q5 kit (New England Biolabs).

Protein expression and purification

Hexim1 protein genes in either pETDuet-MBP or pETDuet-GB1 vectors were transformed into BL21-GOLD(DE3) competent cells for protein expression. Bacterial culture was grown at 37°C in minimum media to O.D.600 of 0.6, then transferred to 18°C for 1 hr prior to induction by 0.5mM IPTG for 18–20 hrs. For uniformly [^{15}N, ^{13}C] (or [^{15}N]) enriched Hexim1 proteins, M9 minimal media containing 1 g/L of $^{15}NH_4Cl$ and 3 g/L of [^{13}C-6]-D-glucose (or

natural abundance glucose) (Cambridge Isotope Laboratories) were used. Cells were pelleted by centrifugation, resuspended in lysis buffer (20 mM Tris, pH 7.5, 1 M NaCl, 5 mM imidazole and 1 mM PMSF) supplemented with protease inhibitor cocktail tablet (ThermoScientific Pierce A32965) and lysozyme. Cells were lysed by sonication on ice and clarified by centrifugation and filtration with 0.45 μm syringe filter. Lysates were loaded onto a 5-mL Ni Sepharose affinity column (HisTrap-HP; GE Healthcare) and a linear gradient of 5–500 mM imidazole was used to elute the His_6-tagged proteins. The eluted protein was added 1mg of TEV protease, dialyzed (20 mM Tris, pH 7.5, 300 mM NaCl, 5mM βME) and further purified with a second Ni affinity column to removed His_6-tagged fusion proteins (MBP or GB1) and TEV protease. For full-length Hexim1 protein, an additional purification step was carried out with a size-exclusion chromatography (SEC; HiLoad 26/600 Superdex 200 or HiLoad 16/600 Superdex 75; GE Healthcare) in binding buffer (20mM HEPES, pH 7.5, 150mM KCl, 1mM TCEP). Proteins were concentrated using Amicon devices (Millipore) and concentration was measured by absorbance at 280 nm. For monomeric Hexim1 BR-L-AR, concentrating and buffer exchanging using Amicon devices into NMR buffer (50mM sodium phosphate, pH 6.2, 0/150/300 mM KCl, 1mM TCEP, 0.02% NaN_3) was performed with caution to prevent precipitation at higher local protein concentration. For even shorter monomeric proteins, 1kDa MWCO dialysis tubing (for BRext) or desalting column (HiTrap Desalting 5mL, GE Healthcare; for BR and BRhalf) were used for buffer exchange into 5 mM NH_4HCO_3 buffer, freeze dried with a lyophilizer and dissolved directly in NMR buffer.

Nitroxide spin labeling of proteins

Protein modification with nitroxide spin label was carried out using established protocol[54]. Briefly, each purified protein solution, containing about 50 nmole of Hexim1 BR-L-AR cysteine variants, was buffer exchanged using a PD-10 desalting column with the gravity protocol to remove the reducing agent into PRE buffer (50mM sodium phosphate, pH 6.2, 150mM KCl, 0.02% NaN_3). The PD-10 elution was collected into a falcon tube, containing 10-fold molar excess of S-(1-oxyl-2,2,5,5,-tetramethyl-2,5,-dihydro-1H-pyrrol-3-yl) methylmethane-

sulfonothiolate (MTSL; Toronto Research Chemicals TRC-O875000) dissolved in 0.5 mL of PRE-buffer, covered with aluminium foil to avoid light. The reaction was kept at room temperature under gentle stirring in the dark for >15 minutes. Subsequently, an additional aliquot of MTSL was added to the reaction mixture to reach a final ratio of 1:20 protein:MTSL. The reaction was further incubated at room temperature under gentle stirring overnight in the dark.

The unreacted MTSL molecule was removed by running through a PD-10 desalting column equilibrated in PRE buffer twice using the gravity protocol. The MTSL-tagged protein was concentrated with caution using Amicon devices, and 5% v/v D_2O was added to prepare the final NMR sample.

In vitro transcription of RNA

RNA samples were prepared by in vitro transcription with T7 RNA polymerase P266L variant[55] as previously described[32]. Short RNA constructs, including linear 7SK SL1 dissected constructs SL1-dI, SL1-$dI_{\Delta U}$, SL1-dII, SL1-d, SL1-$d_{\Delta U}$ SL1-mdII, SL1-m, SL1-pm, SL1-pII, SL1-pIIo, SL1-p, and circular 7SK SL1alt, were transcribed from chemically synthesized DNA oligonucleotides (Integrated DNA Technologies). Transcription mixtures (40 mM Tris, pH 8.0, 1 mM spermidine, 0.01% Triton-X100, 2.5 mM DTT, 25 mM $MgCl_2$, 0.5 μM oligonucleotide DNA template, T7 RNAP P266L, 4 mM each rATP, rUTP, rGTP and rCTP) were incubated at 37°C for 4–6 hrs. Longer RNA constructs minimum linear 7SK (SL1-$SL4^{ms2}$) and linear 7SK SL1 were transcribed from plasmid DNA templates prepared with QIAGEN Plasmid Maxi Kit and linearized with BamHI-HF restriction enzyme (New England Biolabs). Transcription mixtures are the same as the short RNA constructs above, except from lower $MgCl_2$ concentration (15–20 mM) and overnight incubation at 37°C.

Transcription mixtures were loaded onto 10–20% denaturing polyacrylamide gel electrophoresis (Urea-PAGE; 19:1 crosslinking ratio) for RNA purification. Purified RNA

molecules were eluted out of gel pieces either with an Elutrap device (GE Whatman) or by crush-and-soak in 1X TBE buffer (90 mM Tris-borate, 2 mM EDTA, pH 8.3). Collected RNA eluents were concentrated and buffer exchanged into sterilized nanopure water using 3-kDa or 10-kDa MWCO Amicon devices (Millipore), supplied with counterion in high salt buffer (10 mM sodium phosphate, pH 7.6, 1 mM EDTA, 1.5 M KCl) and buffer exchanged back into sterilized nanopure water. Dilute RNA samples <100 μM were heated at 95°C for 5 min and snap cooled on ice for 1 hr prior to final buffer exchanging and concentrating with Amicon devices (Millipore). Stock RNA solutions were exchanged into sterilized nanopure water and stored in -20°C freezer.

Electrophoretic mobility shift assay

1μL of 50μM RNA and stock solutions of 25–50μM Hexim1 proteins were used to achieve varying protein-to-RNA molar ratios in a total volume of 5.5μL in binding buffer and added 1μL loading dye (30% glycerol and bromophenol blue). The binding mixtures were incubated at room temperature for 15 min prior to gel electrophoresis on 5.7% non-denaturing polyacrylamide gel (PAGE; 37.5:1 crosslinking ratio) or 0.8% Agarose gel (for minimum linear 7SK SL1-SL4^{ms2}) in 0.5X TBE buffer at room temperature. Polyacrylamide gels (BioRad Mini-PROTEAN handcast) and agarose gel (BioRad Mini-Sub 7x7 cm) were run at 70 V. Polyacrylamide gels were subsequently stained with Toluidine blue solution and destained in water, whereas the Agarose gel was made with SYBR™ Safe stain (Invitrogen) added prior to solidifying. Gels were imaged with a Pharos FX Plus scanner (Bio-Rad).

NMR spectroscopy

For NMR assignments of SL1-d$I_{\Delta U}$ RNA bound BR-L-AR, dilute BR-L-AR uniformly enriched with ^{15}N and ^{13}C were combined with d$I_{\Delta U}$ in 1:1.2 molar ratio, and the dilute complex was buffer exchanged in NMR buffer and concentrated down to a final concentration of 0.22 mM BR-L-AR and 0.25 mM SL1-d$I_{\Delta U}$. Backbone assignments were carried out using the following NMR experiments: HNCACB, CBCA(CO)NH, HNCA and C(CO)NH. For NMR assignments of

Hexim1 BR-L-AR, 70 μM sample uniformly enriched with ^{15}N and ^{13}C in NMR buffer was used. Backbone assignments were carried out using the CBCA(CO)NH and HNCA NMR experiments, which were collected with non-uniform sampling and processed with iterative soft thresholding reconstruction[56], and comparison to the RNA-bound BR-L-AR assignments.

For NMR assignments of further truncated Hexim1 constructs, BR-L, BR, BRext and BRhalf, samples uniformly enriched with ^{15}N and ^{13}C in the range of 0.5–0.8 mM concentration in NMR buffer were used. Backbone assignments were carried out using the following NMR experiments: HNCACB, CBCA(CO)NH and C(CO)NH. Additional side-chain assignments were acquired for BR-L and BR using HCCH-TOCSY. BRhalf ^{15}N-^{1}H HSQC resonances were assigned based on almost complete overlapping with the BR ^{15}N-^{1}H HSQC peaks. For 1H imino sequential assignments of RNA samples in the absence and presence of monomeric Hexim1 proteins, the two-dimensional NOESY spectra were recorded using 5% D_2O samples at 283.15 K with NOE mixing time of 150 ms. For PRE NMR experiments, ^{15}N-^{1}H HSQC spectra were recorded at 290.15 K with Avance III HD 600MHz.

For the chemical shift perturbation (CSP) analysis of RNA constructs, observed imino chemical shift values (H3 from Us and H1 from Gs) were used to calculate the change in chemical shift Δ between the free and protein-bound states. For the CSP analysis for Hexim1 proteins, the overall change in chemical shift Δ were calculated for protein ^{15}N-^{1}H HSQC resonances between the free and RNA-bound states using the following equation:

$$\Delta = \sqrt{{\Delta\delta_H}^2 + 0.152 \times {\Delta\delta_N}^2}$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the differences between the $^{1}H_N$ and ^{15}N chemical shifts of the two states being compared.

All the NMR spectra were collected at 298.15K with Avance Neo 800MHz spectrometer equipped with TCI H&F cryoprobe, Avance III HD 600MHz spectrometer equipped with QCI HCNP cryoprobe and Bruker DRX 500 MHz with cryoprobe, except from when mentioned

specifically. Data were collected with TopSpin (Bruker), processed with NMRPipe [57], and analyzed in NMRFAM-Sparky[58]. Hexim1 protein CSP values and PRE intensity ratios were plotted and visualized with Igor-Pro 9 software.

Isothermal titration calorimetry

The dissociation constants (K_D) for binding between Hexim1 proteins and 7SK RNA constructs were determined using a MicroCal 200 ITC instrument (GE). RNA and protein were buffer exchanged using separate 3 kDa MWCO dialysis tubings into to the same bucket of ITC buffer (20mM HEPES, pH 7.5, 150mM KCl, 1mM TCEP). Dilute dialyzed RNA solutions <20 μM were heated at 95°C for 5 min and snap cooled on ice for 1 hr right before the first set of titration experiments. RNA final concentrations were determined by UV absorbance at 90°C at 260 nm with the Lambert-Beer law, using a Hewlett-Packard HP8453 diode-array UV/Visible spectrophotometer with Peltier temperature control. Extinction coefficients of melt RNA were calculated by the OligoAnalyzer Tool from Integrated DNA Technologies. RNA stock solutions were kept at room temperature to avoid cold-induced oligomerization for the 2–3 days duration of one complete titration set. RNA at concentration 107-195 μM in the syringe was titrated into 7-18 μM protein in the sample cell at 295K, or for reversed titration order protein at concentrations of 102-120 μM in the syringe was titrated into 6–12 μM RNA in the sample cell at 295K. Calorimetric data was fit using ORIGIN 7 (Micro-Cal). The binding parameters stoichiometry (N), entropy(ΔS), enthalpy (ΔH) and association constant (K_A) were kept as floating variables during each fit. Experiments were performed in three independent repeats with each experiment fit individually, of which binding parameters were averaged and standard deviations were calculated.

Multiple sequence alignment

Initial list of orthologs of Hexim1 and Hexim2 genes were extracted from NCBI gene cards (Gene ID: 10614 and 124790). Truncated and low-confidence genes were removed by manual

inspection, resulting in 294 genes for Hexim1 and 210 genes for Hexim2. Sequence alignments were performed by COBALT (NCBI) and visualized using Jalview[59]. Five sequences for each Hexim1 and Hexim2 were selected for the final display, while the sequence placements were kept from the full alignments. List of HIV-1 Tat protein sequences and alignment were acquired from UniProt, visualized using Jalview and one long isoform from each group and subtype of virus were selected for the final display.

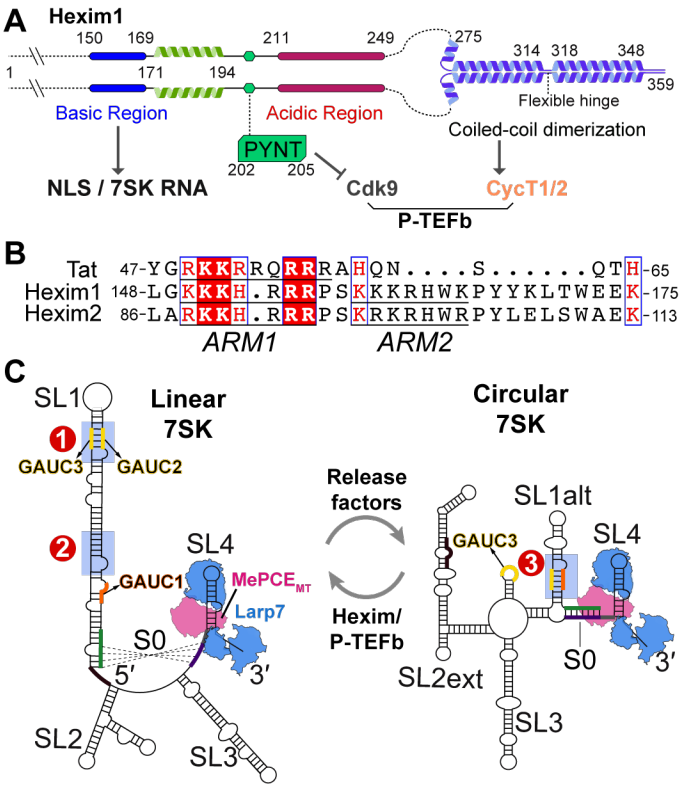

Fig. 1: 7SK RNA conformational switching affects available Hexim1 binding sites. A Domain schematic of Hexim1 homodimer. Interacting partners of Hexim1 structural motifs are indicated below the schematic. NLS is nuclear localization signal. **B** Sequence alignment between HIV-1 Tat (accession ID P04608, group M subtype B), Hexim1 and Hexim2 RNA binding domain. **C** The linear 7SK RNA conformation proposed by chemical probing[30,36,51] has two Hexim1 binding sites (Site 1 and Site 2), whereas the circular 7SK deconvoluted by DANCE-MaP[30] has one Hexim1 binding site (Site 3). MePCE and Larp7 proteins are illustrated as outlined magenta and blue domain cartoons, respectively, representing the shapes from 7SK core RNP cryo-EM structures[32].

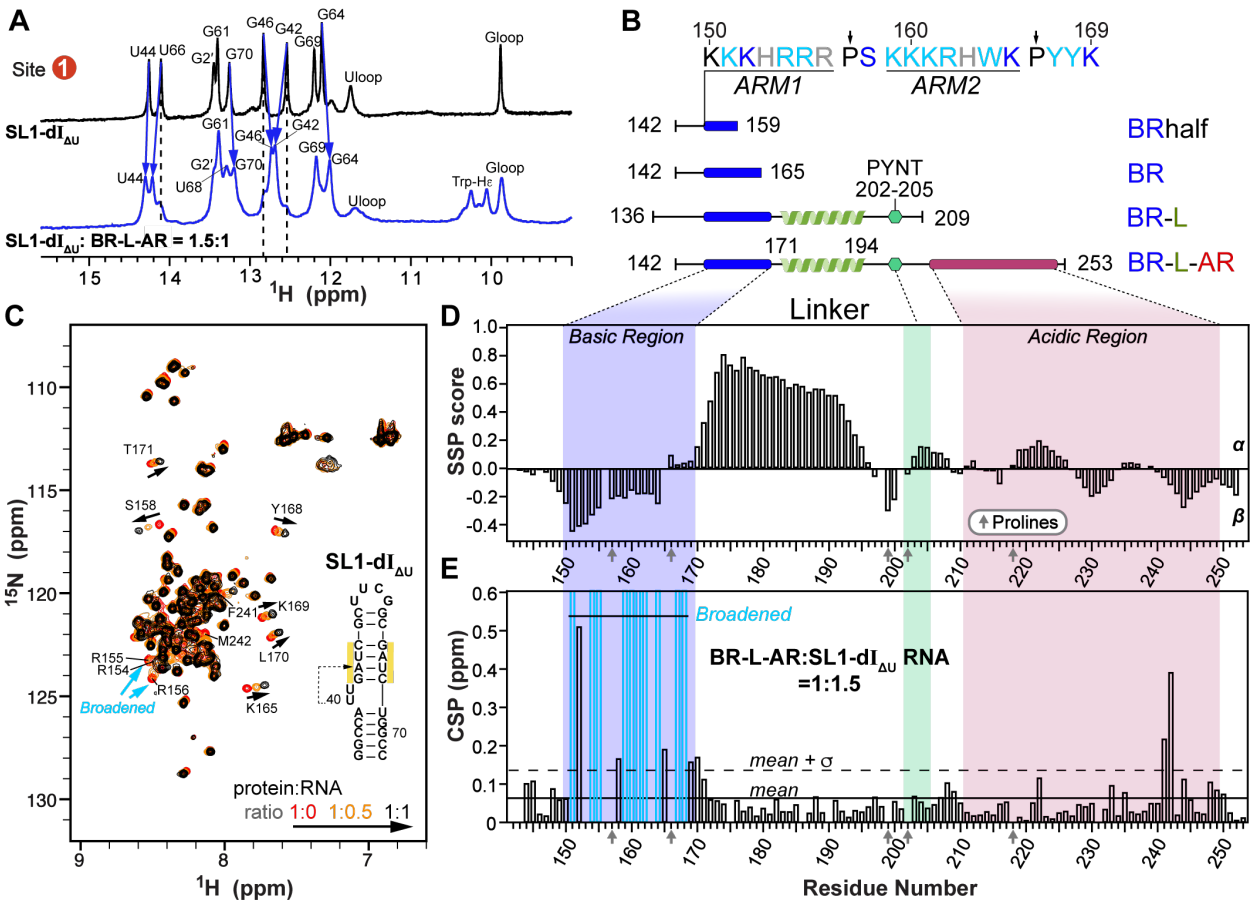

Fig. 2: NMR characterization of monomeric Hexim1 and its RNA interaction. A Proton 1D imino region of linear 7SK SL1-dI$_{\Delta U}$ in the absence and presence of monomeric Hexim1 BR-L-AR at 600MHz. **B** Various monomeric Hexim1 constructs used in this study are shown as domain schematics. The amino acid sequence of BR containing ARM1 and ARM2 is shown at top, where cyan residues indicate those with amide resonances broadened beyond detection upon RNA binding and blue residues indicate those with significant chemical shift changes. **C** ^{15}N-^{1}H HSQC spectral overlay at 500MHz of Hexim1 BR-L-AR (red) with addition of linear 7SK SL1-dI$_{\Delta U}$ at 1:0.5 (orange) and 1:1 (black). Representative resonances that shift significantly or broaden upon RNA binding are labeled with residue assignments. **D** Secondary structure propensity (SSP) scores, calculated from NMR assigned chemical shifts (see Methods), plotted as a function of residue number show that Hexim1 BR-L-AR is largely disordered, with a short α-helix between BR and PYNT motif. **E** Chemical shift perturbations (CSP) of Hexim1 BR-L-AR upon binding 7SK SL1-dI$_{\Delta U}$. Cyan bars indicate residues broadened beyond detection by RNA binding. Gray arrows in D and E indicate proline residues, which lack amide protons, resulting in lack of CSP values for the prolines and SSP values for the residues preceding prolines.

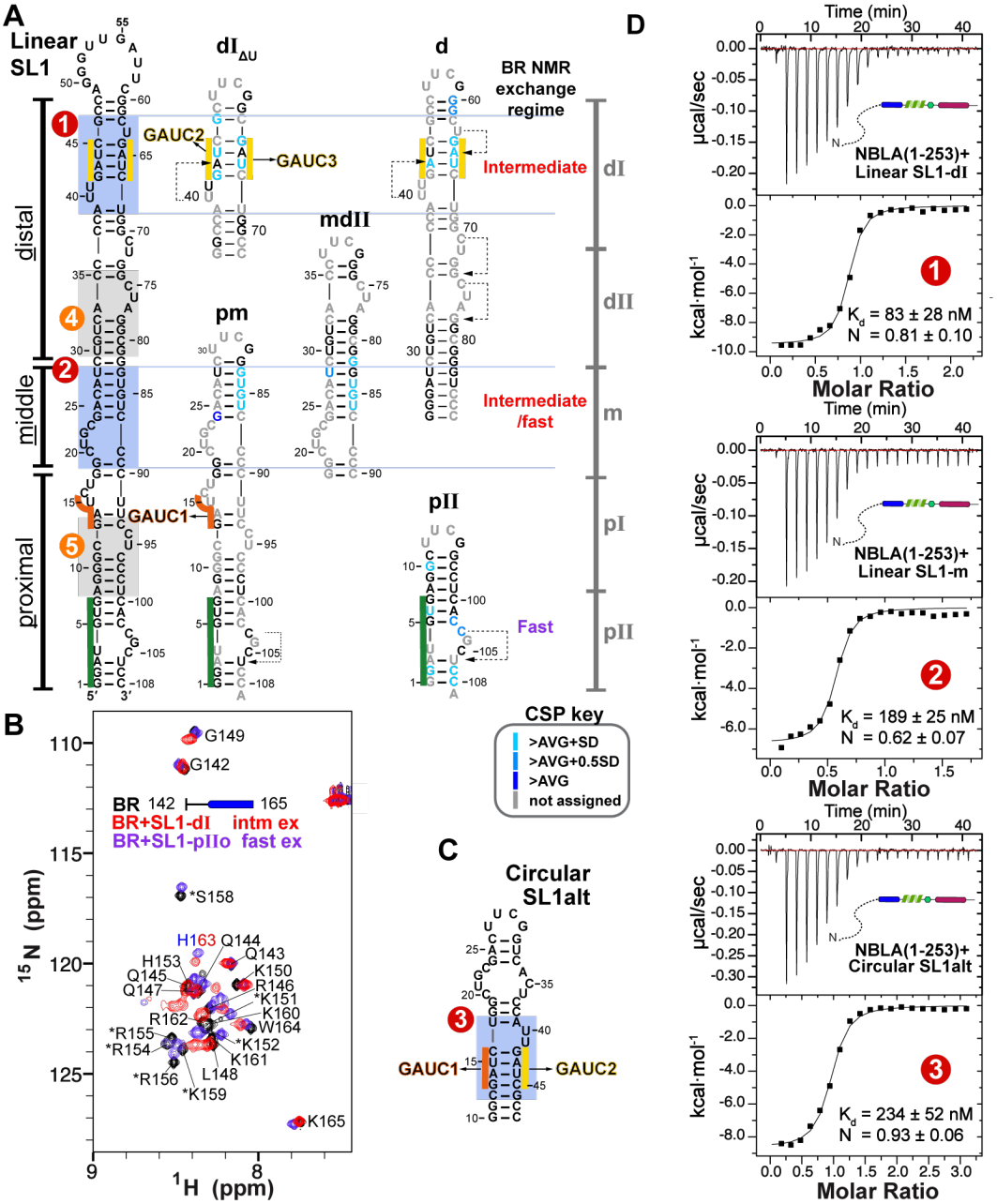

Fig. 3: Mapping and characterizing the RNA binding sites of monomeric Hexim1 by NMR and ITC. A Sequences and NMR-verified secondary structures of short linear 7SK SL1 constructs aligned with the full-length SL1 secondary structure. Nucleotides with significant chemical shift perturbation (CSP) upon binding monomeric Hexim1 are colored as indicated by the CSP key insert. **B** ^{15}N-^{1}H HSQC spectral overlay of Hexim1 BR (black) and with addition of 1:1 SL1-dI (red; high affinity, intermediate exchange at 800 MHz) or 1:1 SL1-pIIo RNA (purple; low affinity, fast exchange at 800 MHz). **C** Sequence and NMR-verified secondary structure of circular 7SK SL1alt construct. **D** ITC measurements showing affinities of individual 7SK RNA sites binding to monomeric Hexim1 NBLA. In **A** and **C,** determined high affinity and low affinity binding sites are indicated by light blue box (Sites 1, 2, 3) and gray boxes (Sites 4, 5), respectively. Regions of SL1 are indicated as proximal (pI and pII), middle (m), and distal (dI and dII) with gray bars to the right of panel A. GAUC motifs are highlighted with yellow and orange lines, and bases triples indicated by connecting dashed lines[23,24,37]

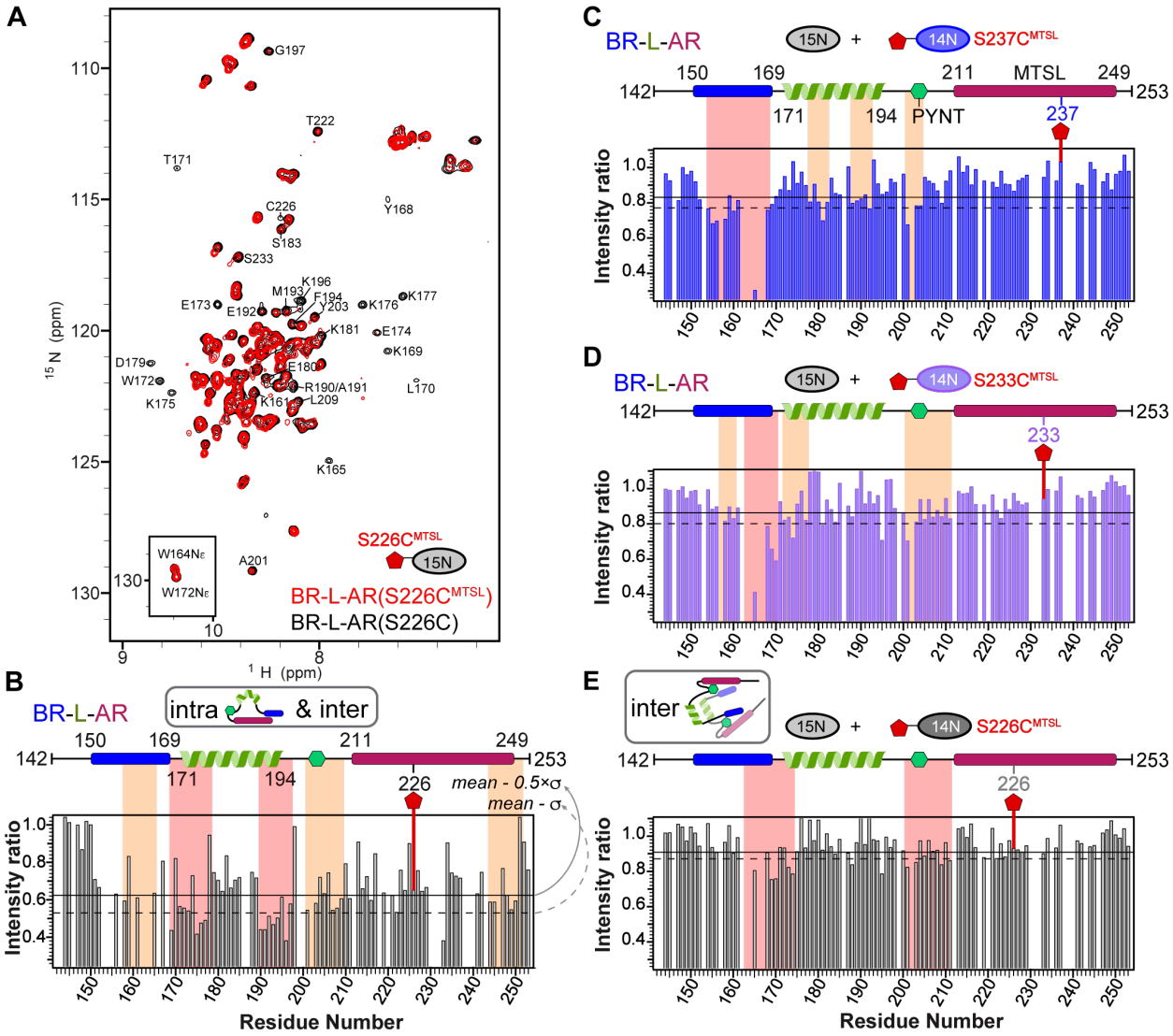

Fig. 4: Probing the intra- and inter-monomer components of Hexim1 autoinhibited conformation. **A** ^{15}N-^{1}H HSQC spectral overlay of Hexim1 BR-L-AR(S226C) (black) and BR-L-AR(S226C^{MTSL}) (red). Residues that are significantly broadened by MTSL are labeled on the spectrum. **B** PRE intensity ratios (I_{PRE}/I_{noPRE}) for BR-L-AR(S226C) without and with MTSL tag of resolved resonances plotted against residue number. Schematic of BR-L-AR with S226C^{MTSL} tag location indicated is shown at top. **C-E** Plots of intermolecular PRE intensity ratios of ^{15}N BR-L-AR to ^{14}N BR-L-AR(S237C^{MTSL}) (blue bars) (C), S233C^{MTSL} (purple bars) (D) and S226C^{MTSL} (dark gray bars) (E). Schematic of BR-L-AR construct with the S226C^{MTSL}, S233C^{MTSL} and S237C^{MTSL} tag location highlighted is shown at top for each plot. For B–E, solid horizontal line indicates value of (mean-0.5×σ) and dashed line indicates value of (mean-σ). Stretches of residues with intensity ratios below the dashed line (mean-σ) or solid line (mean-0.5×σ) are highlighted with red shaded boxes and orange shaded boxes, respectively.

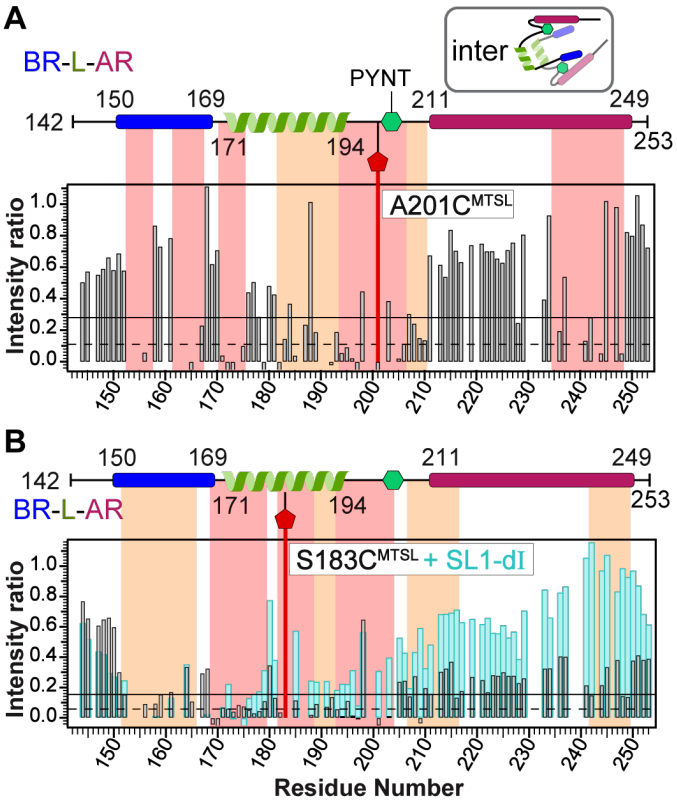

Fig. 5: PYNT^{205} is the interaction hub for Hexim1 autoinhibited conformation, and is unmasked upon RNA binding to BR. A Schematic and PRE intensity ratios (I_{PRE}/I_{noPRE}) for BR-L-AR(A201C) without and with MTSL tag of resolved resonances plotted against residue number. **B** Schematic and PRE intensity ratios (I_{PRE}/I_{noPRE}) for BR-L-AR(S183C) without and with MTSL tag of resolved resonances plotted against residue number in the absence (black/gray bars) and presence (cyan bars) of SL1-dI RNA (1:1.2 protein:RNA ratio). For A and B, solid horizontal line indicates value of (mean-0.5×σ) and dashed line indicates value of (mean-σ). Stretches of residues with intensity ratios below the dashed line (mean-σ) or solid line (mean-0.5×σ) are highlighted with red shaded boxes and orange shaded boxes, respectively.

Table 1: Thermodynamic parameters for binding of Hexim1 to 7SK RNA from ITC

	Sites	K_D (nM)	N
NBLA monomer			
SL1-dI	1	83 ± 28	0.81 ± 0.10
SL1-m	2	189 ± 25	0.62 ± 0.07
SL1alt	3	261 ± 68	0.93 ± 0.06
SL1	1,2 (4,5)	336 ± 134	0.36 ± 0.01
Hexim1 dimer			
SL1-dI	1	782 ± 35	0.76 ± 0.01
SL1-m	2	885 ± 58	0.79 ± 0.07
SL1alt	3	2035 ± 359	0.68 ± 0.04
SL1	1,2 (4,5)	73 ± 30	0.39 ± 0.05
SL1-dII	4	1221 ± 129	0.47 ± 0.05
SL1-p	5	2108 ± 197	0.49 ± 0.03
Hexim1 dimer (syringe)			
SL1-dI	1	458 ± 28	1.12 ± 0.05
SL1-m	2	782 ± 26	0.87 ± 0.04
SL1alt	3	748 ± 15	1.01 ± 0.09
SL1	1,2 (4,5)	183 ± 18	1.75 ± 0.01
SL1-d(api)	1,4	96 ± 22	1.40 ± 0.08

Concentrations of monomeric NBLA (1–253) proteins in the cell are 7–18 μM (see Methods). Sequence and secondary structures of RNA are shown in Fig. 3 and Supplementary Fig. 5. Sites (4,5) indicate that these are present in the RNA construct but are unoccupied in the titration experiment.

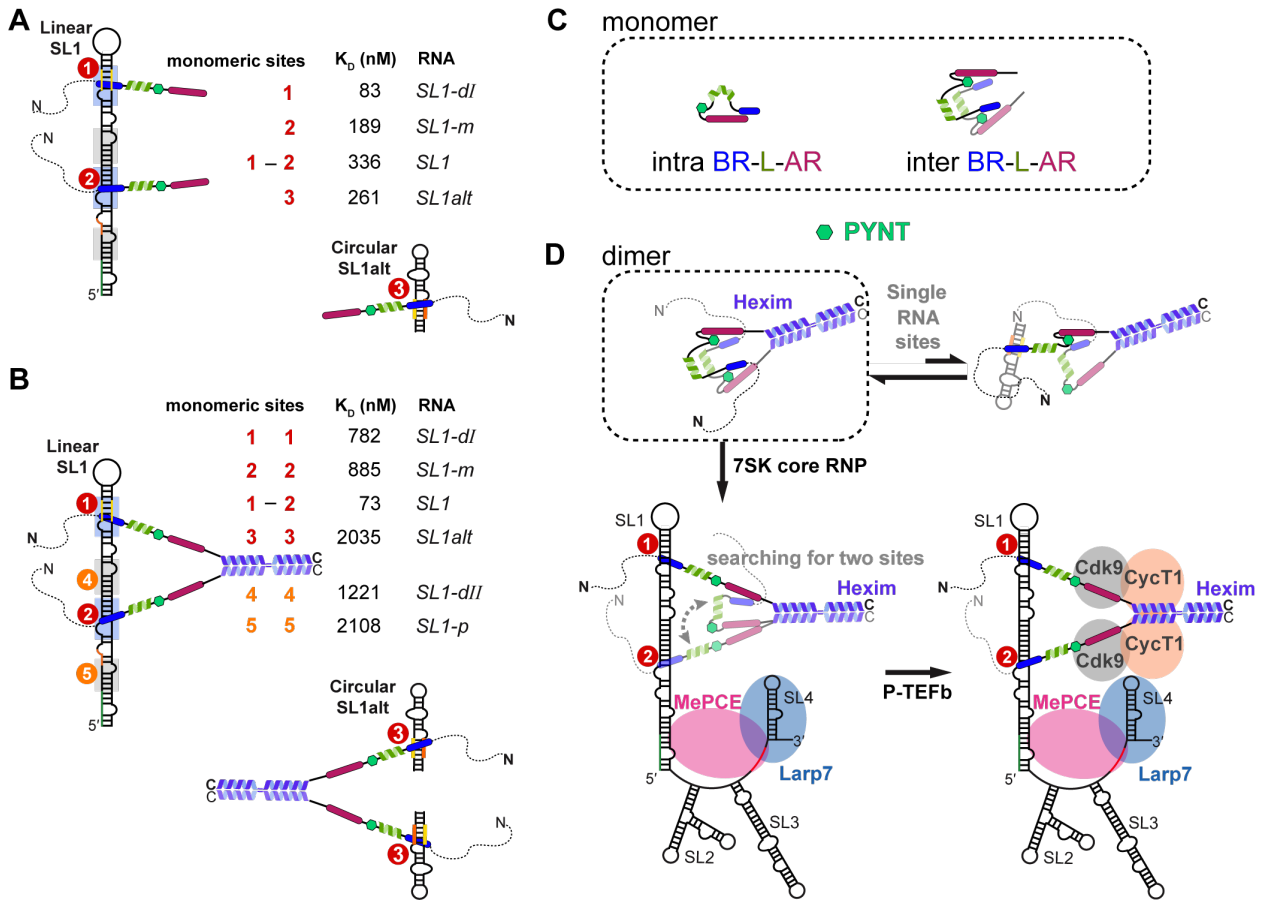

Fig. 6: Schematics illustrating how autoinhibited Hexim1 conformation can enhance the specificity for linear 7SK SL1 binding in order to expose PYNT motif. A Summary of ITC results showing non-cooperative binding of monomeric Hexim1 (NBLA) to sites 1 and 2, comparing to Site 3. **B** Summary of ITC results showing cooperative binding of Hexim1 dimer to Sites 1 and 2, comparing to the weaker affinity binding Sites 3, 4 and 5. Note that for single RNA sites, both monomers of Hexim1 dimer are bound with two identical copies of the single RNA site molecule at the end of the titrations, with an example illustrated for SL1alt (Site 3) at bottom. **C** In the monomeric Hexim1, both *intra*- and *inter*-monomer interactions exist. **D** In the full-length Hexim1 homodimer, *inter*-monomer interactions mask the PYNT motif using both BR and AR (top). Non-functional RNA sites do not stably release the autoinhibited Hexim1 dimer. A complete set of two high-affinity RNA binding sites in linear 7SK releases the autoinhibited Hexim1 dimer and exposes PYNT motifs for P-TEFb inactivation inside the assembled 7SK RNP.

Acknowledgment

This work was supported by the National Institutes of Health (NIH) R01AI155170 and R35GM131901 to J.F., American Heart Association Postdoctoral Fellowship 20POST35210850 to Y.Y. and American-Italian Cancer Foundation Postdoctoral Fellowship to M.G.M. The UCLA-DOE NMR core facility is supported in part by NIH instrumentation grants S10OD016336 and S10OD025073 and Department of Energy DE-FC02-02ER63421. We thank Catherine D. Eichhorn and Yanjiao Wang for helpful discussions.

Author contributions

Y.Y., M.G.M. and J.F. conceptualized the study and designed the experiments. Y.Y. and M.G.M. performed most of the experiments and analyzed the data. Y.W. designed SL1 proximal constructs, performed NMR assignments and EMSA binding tests. S.G. and N.A. helped with sample preparation for NMR and ITC experiments. Y.Y. and J.F. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests

Correspondence and requests for materials should be addressed to Juli Feigon.

References

1 Kusuhara, M. *et al.* Cloning of Hexamethylene-bis-acetamide-inducible Transcript, HEXIM1, in Human Vascular Smooth Muscle Cells. *Biomedical Research* **20**, 273-279 (1999). https://doi.org/10.2220/biomedres.20.273

2 Michels, A. A. & Bensaude, O. Hexim1, an RNA-controlled protein hub. *Transcription* **9**, 262-271 (2018). https://doi.org/10.1080/21541264.2018.1429836

3 Morchikh, M. *et al.* HEXIM1 and NEAT1 Long Non-coding RNA Form a Multi-subunit Complex that Regulates DNA-Mediated Innate Immune Response. *Mol Cell* **67**, 387-399.e385 (2017). https://doi.org/10.1016/j.molcel.2017.06.020

4 Fujimoto, Y., Nakamura, Y. & Ohuchi, S. HEXIM1-binding elements on mRNAs identified through transcriptomic SELEX and computational screening. *Biochimie* **94**, 1900-1909 (2012). https://doi.org/10.1016/j.biochi.2012.05.003

5 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**, 16368-16376 (2005). https://doi.org/10.1074/jbc.M500912200

6 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 (2005). https://doi.org/10.1074/jbc.M500424200

7 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 (2004). https://doi.org/10.1038/sj.emboj.7600275

8 Yik, J. H., Chen, R., Pezda, A. C., Samford, C. S. & Zhou, Q. A human immunodeficiency virus type 1 Tat-like arginine-rich RNA-binding domain is essential for HEXIM1 to inhibit RNA polymerase II transcription through 7SK snRNA-mediated inactivation of P-TEFb. *Mol Cell Biol* **24**, 5094-5105 (2004). https://doi.org/10.1128/mcb.24.12.5094-5105.2004

9 Czudnochowski, N., Bösken, C. A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun* **3**, 842 (2012). https://doi.org/10.1038/ncomms1846

10 Czudnochowski, N., Vollmuth, F., Baumann, S., Vogel-Bachmayr, K. & Geyer, M. Specificity of Hexim1 and Hexim2 complex formation with cyclin T1/T2, importin alpha and 7SK snRNA. *J Mol Biol* **395**, 28-41 (2010). https://doi.org/10.1016/j.jmb.2009.10.055

11 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). https://doi.org/10.1074/jbc.270.21.12335

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

13 D'Orso, I. & Frankel, A. D. RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. *Nat Struct Mol Biol* **17**, 815-821 (2010). https://doi.org/10.1038/nsmb.1827

14 Lu, H. *et al.* AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation. *Proc Natl Acad Sci U S A* **111**, E15-24 (2014). https://doi.org/10.1073/pnas.1318503111

15 D'Orso, I. *et al.* Transition step during assembly of HIV Tat:P-TEFb transcription complexes and transfer to TAR RNA. *Mol Cell Biol* **32**, 4780-4793 (2012). https://doi.org/10.1128/mcb.00206-12

16 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). https://doi.org/10.1093/nar/gkm063

17 Schulte, A. *et al.* Identification of a cyclin T-binding domain in Hexim1 and biochemical analysis of its binding competition with HIV-1 Tat. *J Biol Chem* **280**, 24968-24977 (2005). https://doi.org/10.1074/jbc.M501431200

18 Bayer, T. S., Booth, L. N., Knudsen, S. M. & Ellington, A. D. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *Rna* **11**, 1848-1857 (2005). https://doi.org/10.1261/rna.2167605

19 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 (2016). https://doi.org/10.1073/pnas.1612331113

20 Barboric, M. *et al.* Interplay between 7SK snRNA and oppositely charged regions in HEXIM1 direct the inhibition of P-TEFb. *Embo j* **24**, 4291-4303 (2005). https://doi.org/10.1038/sj.emboj.7600883

21 Li, Q. *et al.* HEXIM1 is a promiscuous double-stranded RNA-binding protein and interacts with RNAs in addition to 7SK in cultured cells. *Nucleic Acids Res* **35**, 2503-2512 (2007). https://doi.org/10.1093/nar/gkm150

22 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 (2010). https://doi.org/10.1093/nar/gkq660

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

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

25 Pham, V. V., Gao, M., Meagher, J. L., Smith, J. L. & D'Souza, V. M. A structure-based mechanism for displacement of the HEXIM adapter from 7SK small nuclear RNA. *Commun Biol* **5**, 819 (2022). https://doi.org/10.1038/s42003-022-03734-w

26 Gruber, A. R. *et al.* Arthropod 7SK RNA. *Mol Biol Evol* **25**, 1923-1930 (2008). https://doi.org/10.1093/molbev/msn140

27 Gruber, A. R. *et al.* Invertebrate 7SK snRNAs. *J Mol Evol* **66**, 107-115 (2008). https://doi.org/10.1007/s00239-007-9052-6

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

29 Yang, Y., Eichhorn, C. D., Wang, Y., Cascio, D. & Feigon, J. Structural basis of 7SK RNA 5'-γ-phosphate methylation and retention by MePCE. *Nat Chem Biol* **15**, 132-140 (2019). https://doi.org/10.1038/s41589-018-0188-z

30 Olson, S. W. *et al.* Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP. *Mol Cell* **82**, 1708-1723.e1710 (2022). https://doi.org/10.1016/j.molcel.2022.02.009

31 Xue, Y., Yang, Z., Chen, R. & Zhou, Q. A capping-independent function of MePCE in stabilizing 7SK snRNA and facilitating the assembly of 7SK snRNP. *Nucleic Acids Res* **38**, 360-369 (2010). https://doi.org/10.1093/nar/gkp977

32 Yang, Y. *et al.* Structural basis of RNA conformational switching in the transcriptional regulator 7SK RNP. *Mol Cell* **82**, 1724-1736.e1727 (2022). https://doi.org/10.1016/j.molcel.2022.03.001

33 Van Herreweghe, E. *et al.* Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *Embo j* **26**, 3570-3580 (2007). https://doi.org/10.1038/sj.emboj.7601783

34 Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* **15**, 2795-2804 (2006). https://doi.org/10.1110/ps.062465306

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

36 Wassarman, D. A. & Steitz, J. A. Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol Cell Biol* **11**, 3432-3445 (1991). https://doi.org/10.1128/mcb.11.7.3432-3445.1991

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

38 Brillet, K. *et al.* Different views of the dynamic landscape covered by the 5'-hairpin of the 7SK small nuclear RNA. *Rna* **26**, 1184-1197 (2020). https://doi.org/10.1261/rna.074955.120

39 Berliner, L. J., Grunwald, J., Hankovszky, H. O. & Hideg, K. A novel reversible thiol-specific spin label: papain active site labeling and inhibition. *Anal Biochem* **119**, 450-455 (1982). https://doi.org/10.1016/0003-2697(82)90612-1

40 Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355-5365 (2000). https://doi.org/10.1021/bi000060h

41 Clore, G. M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **109**, 4108-4139 (2009). https://doi.org/10.1021/cr900033p

42 Tang, C., Ghirlando, R. & Clore, G. M. Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *J Am Chem Soc* **130**, 4048-4056 (2008). https://doi.org/10.1021/ja710493m

43 Morris, O. M., Torpey, J. H. & Isaacson, R. L. Intrinsically disordered proteins: modes of binding with emphasis on disordered domains. *Open Biol* **11**, 210222 (2021). https://doi.org/10.1098/rsob.210222

44 Zeke, A. *et al.* Deep structural insights into RNA-binding disordered protein regions. *Wiley Interdiscip Rev RNA* **13**, e1714 (2022). https://doi.org/10.1002/wrna.1714

45 Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**, 327-341 (2018). https://doi.org/10.1038/nrm.2017.130

46 Qiu, C. *et al.* Intra- and inter-molecular regulation by intrinsically-disordered regions governs PUF protein RNA binding. *Nat Commun* **14**, 7323 (2023). https://doi.org/10.1038/s41467-023-43098-1

47 Chong, P. A., Vernon, R. M. & Forman-Kay, J. D. RGG/RG Motif Regions in RNA Binding and Phase Separation. *J Mol Biol* **430**, 4650-4665 (2018). https://doi.org/10.1016/j.jmb.2018.06.014

48 Wang, X. *et al.* Negatively charged, intrinsically disordered regions can accelerate target search by DNA-binding proteins. *Nucleic Acids Res* **51**, 4701-4712 (2023). https://doi.org/10.1093/nar/gkad045

49 Eichhorn, C. D., Yang, Y., Repeta, L. & Feigon, J. Structural basis for recognition of human 7SK long noncoding RNA by the La-related protein Larp7. *Proc Natl Acad Sci U S A* **115**, E6457-e6466 (2018). https://doi.org/10.1073/pnas.1806276115

50 Prasanth, K. V. *et al.* Nuclear organization and dynamics of 7SK RNA in regulating gene expression. *Mol Biol Cell* **21**, 4184-4196 (2010). https://doi.org/10.1091/mbc.E10-02-0105

51 Wang, Y. *et al.* N(6)-methyladenosine in 7SK small nuclear RNA underlies RNA polymerase II transcription regulation. *Mol Cell* **83**, 3818-3834.e3817 (2023). https://doi.org/10.1016/j.molcel.2023.09.020

52 Perez-Pepe, M. *et al.* 7SK methylation by METTL3 promotes transcriptional activity. *Sci Adv* **9**, eade7500 (2023). https://doi.org/10.1126/sciadv.ade7500

53 Cheng, Y. & Patel, D. J. An efficient system for small protein expression and refolding. *Biochem Biophys Res Commun* **317**, 401-405 (2004). https://doi.org/10.1016/j.bbrc.2004.03.068

54 Sjodt, M. & Clubb, R. T. Nitroxide Labeling of Proteins and the Determination of Paramagnetic Relaxation Derived Distance Restraints for NMR Studies. *Bio Protoc* **7** (2017). https://doi.org/10.21769/BioProtoc.2207

55 Guillerez, J., Lopez, P. J., Proux, F., Launay, H. & Dreyfus, M. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proc Natl Acad Sci U S A* **102**, 5958-5963 (2005). https://doi.org/10.1073/pnas.0407141102

56 Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H. & Wagner, G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J Biomol NMR* **52**, 315-327 (2012). https://doi.org/10.1007/s10858-012-9611-z

57 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995). https://doi.org/10.1007/bf00197809

58 Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325-1327 (2015). https://doi.org/10.1093/bioinformatics/btu830

59 Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **20**, 426-427 (2004). https://doi.org/10.1093/bioinformatics/btg430