JBC RESEARCH ARTICLE

A capped Tudor domain within a core subunit of the Sin3L/Rpd3L histone deacetylase complex binds to nucleic acid G-quadruplexes

Received for publication, August 9, 2021, and in revised form, December 28, 2021 Published, Papers in Press, January 1, 2022,
https://doi.org/10.1016/j.jbc.2021.101558

Ryan Dale Marcum, Joseph Hsieh, Maksim Giljen, Emily Justice, Nicolas Daffern, Yongbo Zhang, and Ishwar Radhakrishnan*

From the Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, USA

Edited by Patrick Sung

Chromatin-modifying complexes containing histone deacetylase (HDAC) activities play critical roles in the regulation of gene transcription in eukaryotes. These complexes are thought to lack intrinsic DNA-binding activity, but according to a well-established paradigm, they are recruited *via* protein–protein interactions by gene-specific transcription factors and post-translational histone modifications to their sites of action on the genome. The mammalian Sin3L/Rpd3L complex, comprising more than a dozen different polypeptides, is an ancient HDAC complex found in diverse eukaryotes. The subunits of this complex harbor conserved domains and motifs of unknown structure and function. Here, we show that Sds3, a constitutively-associated subunit critical for the proper functioning of the Sin3L/Rpd3L complex, harbors a type of Tudor domain that we designate the capped Tudor domain. Unlike canonical Tudor domains that bind modified histones, the Sds3 capped Tudor domain binds to nucleic acids that can form higher-order structures such as G-quadruplexes and shares similarities with the knotted Tudor domain of the Esa1 histone acetyltransferase that was previously shown to bind single-stranded RNA. Our findings expand the range of macromolecules capable of recruiting the Sin3L/Rpd3L complex and draw attention to potentially new biological roles for this HDAC complex.

Posttranslational modifications of core histones constitute a common molecular mechanism for regulating transcription by modulating DNA template accessibility to RNA polymerases, regulatory factors, and other effectors (1, 2). Among various posttranslational modifications, acetylation of lysine residues is not only abundant but also one that is characterized by high turnover, consistent with its central role in the dynamic induction and repression of genes (3). Deacetylation of histones in mammals is mediated in large part by histone deacetylases (HDACs) 1, 2, and 3 (4–6). These enzymes, found in at least six giant multiprotein complexes including the Sin3L/Rpd3L, Sin3S/Rpd3S, NurD, LSD1-CoREST, MiDAC, and SMART/NCoR complexes (7–11), exert their effects after recruitment to specific sites on the genome by DNA-bound transcription factors and/or specific histone modifications.

The Sin3L/Rpd3L complex is the prototypical HDAC complex found in organisms as diverse as yeast and human (8, 12). The complex plays fundamental roles in mammalian biology, regulating a wide array of genes involved in the cell cycle, differentiation, metabolism, and stem cell maintenance (13–15). The 1.2 to 2 MDa mammalian complex harbors at least ten constitutively associated subunits including Sin3A/B, HDAC1/2, RBBP4/7, Sds3/BRMS1/BRMS1L, SAP30/SAP30L, ING1b/ING2, SAP130a/b, ARID4A/B, FAM60A, and SAP25 (paralogous proteins in this list are separated by a '/'). The first five subunits on the list comprise the core complex because of their essential roles in complex assembly and stability (16–19); these subunits along with the ING subunits have orthologs in yeast. The RBBP, ING, and ARID4 subunits harbor WD-40, PHD, and Royal family domains that bind unmodified and modified histones, whereas the other subunits of the complex harbor conserved domains of unknown structure and function.

In the course of our studies to define the molecular roles of the key subunits of the Rpd3L/Sin3L complex, we previously described a novel zinc finger motif shared by the SAP30 and SAP30L subunits of this complex that we later showed turbocharges HDAC activity in response to small-molecule effectors such as inositol phosphates derived from membrane lipids (20, 21). We also showed how the Sds3 subunit provides a dimerization function for the complex that involves a region that assembles into a two-stranded antiparallel coiled-coil helix (22). We further showed that the subunit plays a critical role in core complex assembly by engaging directly and independently with Sin3 and HDACs; the subunit and its paralogs have been implicated in interactions with other subunits of the complex as well as with sequence-specific DNA transcription factors (22–24). At the cellular level, the subunit plays a critical role in the proper segregation of chromosomes during cell division by targeting the Sin3L/Rpd3L complex to pericentric heterochromatin (16). Recently, the subunit has also been implicated in resolving cotranscriptionally generated R-loops (featuring RNA–DNA hybrids and single-stranded DNA) that are a major source of genomic instability (25).

* For correspondence: Ishwar Radhakrishnan, i-radhakrishnan@northwestern.edu.

Two paralogs of Sds3 have been described including the breast cancer metastasis suppressor 1 (BRMS1) and a BRMS1-like protein called BRMS1L that share many domains found in Sds3 (24, 26). All three proteins are found in Sin3L/Rpd3L complexes and share certain key structural and functional features but are not functionally redundant. For example, disruption and downregulation of BRMS1 and BRMS1L is associated with metastasis of multiple types of cancers, whereas overexpression suppresses this effect through a mechanism involving the repression of several metastasis-associated protein-coding and microRNA genes (27–31). However, Sds3 overexpression fails to compensate for *BRMS1* deletion or epigenetic silencing in breast cancer and does not suppress metastasis (32); the molecular basis for this observation remains obscure (33, 34), warranting deeper structural and functional studies to understand the molecular roles of Sds3 and its paralogs.

Here, we describe the structure of another conserved domain of an unknown function in the Sds3 subunit that is shared with one of its paralogs, BRMS1L, but not with BRMS1. Our structural and biochemical analyses suggest that the domain broadly shares an SH3-like β-barrel fold found within many chromatin-binding transcription factors but instead of binding chromatin, the domain binds nucleic acids including both RNA and DNA with a strong preference for G-quadruplex forming sequences.

Results

A conserved capped Tudor domain of unknown structure & function in Sds3 and BRMS1L

Sequence analysis of Sds3 and BRMS1L orthologs from human to zebrafish revealed an ∼80 residue region at the C-termini of the respective proteins with a pattern of conservation that suggested a well-conserved, independently folded domain (Fig. 1*A*); the next well-conserved segment corresponding to the previously characterized Sin3-interaction domain resides ∼30 residues N-terminal of this domain (22). Searches conducted using the mouse Sds3 protein of the RCSB PDB database using BLAST and the repository of templates in the SWISS-MODEL homology modeling server failed to identify any *bona fide* structural homologs for this domain (*i.e.*, no suitable templates in the so-called safe zone for homology modeling).

Sds3 capped Tudor domain adopts a unique variation of a common fold

To gain insights into the structure and function of the capped Tudor domain (CTD), we expressed and purified a recombinant protein corresponding to residues 250 to 326 of mouse Sds3. A ^{1}H-^{15}N correlated NMR spectrum of this protein was characterized by narrow and well-dispersed resonances indicative of a folded domain (Fig. S1). The solution

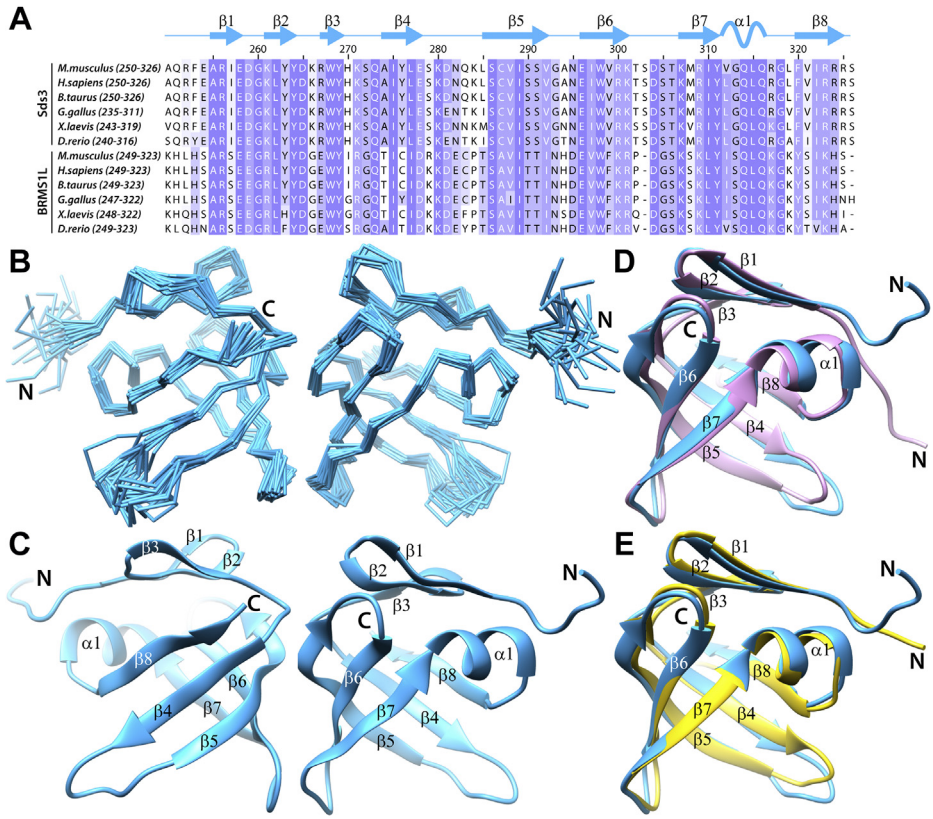

Figure 1. Solution structure of Sds3 capped Tudor domain and comparison with prediction. *A*, a CLUSTAL Ω-guided multiple sequence alignment of a ∼80-residue region at the C-terminus of Sds3 and BRMS1L orthologs from various eukaryotes. Two views differing by a 180° rotation around the vertical axis of (*B*), a best-fit backbone superposition of the ensemble of 20 conformers and (*C*), the corresponding representative structure. A best-fit backbone superposition of the representative NMR structure with (*D*), the highest-ranked Rosetta prediction and (*E*), the AlphaFold2 prediction. The cartoon on *top* of (*A*) identifies the locations of various secondary structural elements in the solution structure. BRMS1L, BRMS1-like.

NMR structure of the domain was determined using a combination of ^{1}H-^{1}H NOE-based distance and backbone chemical shift-based torsion angle restraints (Table 1). Structure determination and refinement resulted in an ensemble of 20 converged conformers with average RMSDs in the ordered regions of 0.52 Å and good agreement with experimental restraints and excellent backbone and covalent geometry (Fig. 1*B* and Table 1). The domain comprises eight strands and a short helix (Fig. 1*C*). Except at the N- and C-termini and the loop connecting β4 and β5, the conformers adopt highly similar backbone conformations. Strands β4 to β8 form an antiparallel five-stranded closed β-barrel fold, reminiscent of SH3-domains, with one mouth of the barrel capped by the three-stranded β-sheet formed by β1, β2, and β3.

To test the reliability of predictions by *de novo* methods, the mouse Sds3 sequence was submitted to the Robetta server for tertiary structure prediction using TrRefineRosetta (35). The backbone RMSD for the 65 C^{α} atom pairs involved in the best-fit superposition between the top solution and the representative structure from the NMR ensemble was 0.84 Å (Fig. 1*D*). The backbone RMSD between the AlphaFold2 prediction in the predicted protein structure database (Uniprot accession: Q8BR65) and the representative NMR structure over 67 C^{α} atom pairs involved in the best-fit superposition were marginally better at 0.81 Å (Fig. 1*E*). Thus, although homology modeling (aka comparative modeling) methods failed to detect a suitable template for modeling, both AlphaFold2 and Rosetta could predict the structure of this domain with high confidence.

To gain insights into the domain's function, we sought to establish the closest homolog at the structural level by searching the RCSB PDB database using DALI (36). Although DALI returned many hits, the one in the PDB25 database with the highest Z-score (6.9) corresponded to the bromo adjacent homology (BAH) domain of the *Zea mays* protein ZMET2, a DNA cytosine-5-methyltransferase ((37); Fig. 2*A*). A best-fit superposition of all the regular secondary structural elements shared by the two proteins, except for the slightly elongated helical segment in Sds3, yielded a backbone RMSD of 1.87 Å. The ZMET2 BAH domain binds to histone H3K9me$_2$, in a manner reminiscent of methyl lysine/methylarginine-binding by the so-called Royal domains including the Tudor, MBT, chromobarrel, and PWWP domains ((38); Fig. S2). Comparisons with a representative structure for each of these four members of the Royal family suggested that the Sds3 CTD might be evolutionarily closest to the Tudor domain with a best-fit superposition of backbone atoms of 1.39 Å (Fig. 2*B*). The Sds3 CTD harbors two features that distinguish it from regular Tudor domains: (1) a longer helical segment linking the two C-terminal strands of the β-barrel and (2) a three-stranded β-sheet at the N-terminus that closes one edge of the barrel, thereby capping it. Therefore, we refer to the Sds3 domain as a CTD.

Table 1
Nuclear magnetic resonance structure determination statistics

Restraint statistics	
NOE-based distance restraints	1456
Unambiguous NOE-based restraints	1295
Intraresidue	778
Sequential (| $i - j$ | = 1)	232
Medium-range (1 < | $i - j$ | ≤ 4)	59
Long-range (| $i - j$ | > 4)	226
Ambiguous NOE-based restraints	161
Hydrogen bonding distance restraints	32
Torsion angle restraints	(63 ϕ, 63 ψ)
Structure quality of NMR ensemble	
Restraint satisfaction	
RMS differences for distances (Å)	0.0079 ± 0.0009
RMS differences for torsion angles (°)	0.3133 ± 0.0790
Deviations from ideal covalent geometry	
Bond lengths (Å)	0.0035 ± 0.0001
Bond angles (°)	0.4361 ± 0.0146
Impropers (°)	1.3050 ± 0.1113
Ramachandran plot statistics (%)	
Residues in most favored regions	84.2
Residues in additional allowed regions	15.1
Residues in generously allowed regions	0.5
Residues in disallowed regions	0.3
Average atomic RMSDs from average structure (Å)	
All atoms	1.76
All atoms except in disordered regions[a]	1.40
Backbone atoms (N, C^{α}, C')	
All residues	1.04
All residues except disordered regions[a]	0.52
All residues in secondary structural elements	0.44

[a] Disordered regions include two non-native residues at the N-terminus in addition to residues 250 to 251, 279 to 283, and residue 326 at the C-terminus of the domain.

Sds3 CTD is not a histone-binding module

Unlike the aforementioned BAH and Royal family domains, the Sds3 CTD lacks an aromatic cage near the remaining edge of the β-barrel for binding methyl lysine residues (Figs. 2, *A*–*C* and S2). Indeed, the domain in this region is devoid of aromatic residues (Fig. 2*C*). To test whether the domain could bind to posttranslationally modified histones, we screened the MODified histone peptide array using purified protein. Although an anti-Myc antibody bound to the Myc peptide that was included in the array as a positive control, no binding was detected for Sds3 to any of the histone peptides, both modified and unmodified, in the array (Fig. 2*D*). To complement these findings, NMR titrations of ^{15}N-labeled Sds3 CTD were conducted with dimethyl lysine, trimethyl lysine, acetyl lysine, and dimethylarginine as well as with an unmodified histone H3 peptide (residues 1–42). However, none of these compounds produced any discernible perturbations in the NMR spectra. Collectively, these results suggest that the Sds3 CTD does not bind either the unmodified or any of the well-characterized posttranslationally modified histone tails.

Sds3 CTD surface properties suggest a role in nucleic acid binding

To gain clues into its molecular function, we then analyzed the surface properties of the Sds3 CTD. Given the considerably high levels of sequence conservation (Fig. 1*A*), mapping the information onto the molecular surface was not especially insightful. However, because Tudor domains have been implicated in functions other than histone binding, such as nucleic acid binding, we calculated the electrostatic potential using APBS (39) and mapped it onto the molecular surface of the domain. By doing so revealed an overwhelmingly electropositive or neutral surface with multiple, discrete patches that were strongly electropositive (Fig. 3*A*). Because the Sin3L/

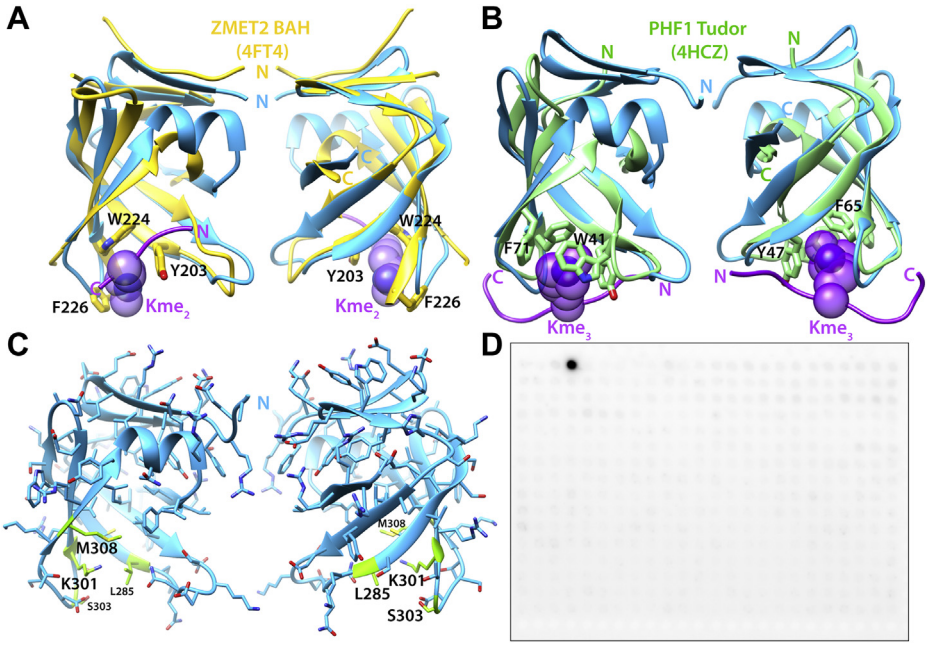

Figure 2. Structural relatives and insights into a potential function for the Sds3 CTD. *A*, a best-fit backbone superposition of Sds3 CTD (*blue*) with its relative ZMET2 BAH domain (*yellow*), as determined by DALI. The BAH domain binds to a dimethyl lysine-containing histone peptide with the modified residue (rendered transparently) binding to a 'cage' formed by the side chains of three aromatic residues (shown in *stick* representation). The BAH domain features two long insertions in the loop regions of Sds3 CTD connecting the β1 and β2 strands and the β6 and β7 strands; neither of them is shown for clarity. *B*, a best-fit backbone superposition of Sds3 CTD (*blue*) with the PHF1 Tudor domain (*green*) that binds to a trimethyl lysine-containing histone peptide. The modified lysine is rendered transparently, whereas the aromatic side chains of the four residues forming the 'cage' are rendered in *stick* mode. *C*, two views, identical to those shown in (*A*) and (*B*), of the Sds3 CTD with the side chains shown in *stick* representation to illustrate the general lack of aromatic side chains on one edge of the barrel that constitutes the canonical-binding pocket for modified lysines and arginines. The residues equivalent to those that form the aromatic cage in the ZMET2 BAH and PHF1 Tudor domains are colored in *green* and labeled. *D*, a binding screen conducted with purified His_6-tagged Sds3 CTD and the MODified histone peptide array. The results illustrate a complete lack of histone-binding activity for the CTD. The *sole dark spot* in the array corresponds to a Myc peptide, included in the array as a positive control, detected by an anti-Myc antibody. BAH, bromo adjacent homolog; CTD, capped Tudor domain.

Rpd3L complex functions, in part, by directly engaging with nucleosomes (40), we first asked whether the domain could bind to the well-characterized acidic patch on the surface of nucleosomes. Instead of using mononucleosomes, we used the histone H2A-H2B heterodimer (41), which is a well-established surrogate for the acidic patch in NMR titration experiments with Sds3 CTD. Once again, no discernible perturbations could be detected in the NMR spectra, ruling out a potential role for the CTD in nucleosome binding.

Because the knotted Tudor domain of Esa1 was previously shown to bind RNA (42), we asked whether the Sds3 CTD could have a similar function. To deduce potential RNA-binding motifs, systematic evolution of ligands by exponential enrichment (SELEX) experiments were performed starting with a 20mer randomized library (43). Samples of the RNA library after three rounds of selection and amplification were incubated with increasing amounts of maltose-binding protein (MBP)-tagged Sds3 CTD in electrophoretic mobility shift assays (EMSAs; Fig. 3*B*). Although no clear mobility shifts were observed in these experiments, samples of the library after six rounds of selection, amplification, and incubation with MBP-Sds3 CTD yielded a clear band whose mobility was significantly reduced compared to that of the free RNA. These bands were observed at micromolar concentrations of MBP-Sds3 CTD, implying a modest affinity interaction. After reverse transcription and amplification, the RNA library from this round was sent for next-generation sequencing. A total of 42×10^6 reads were obtained, out of which $\sim 10 \times 10^6$ were deemed to be of high quality for motif detection using the MEME suite. Because MEME can only handle a maximum of 500,000 sequences, the reads were randomly assigned to ten datasets, each comprising 500,000 sequences. The five most statistically significant motifs in each dataset reported by MEME were compiled (Table S1) and those that were found in more than three datasets are listed in Table 2. Somewhat unexpectedly and despite six rounds of enrichment, a single dominant motif did not emerge from these analyses. The most prevalent was a 7-residue motif (HGTGGTK; where H is A/C/T and K is G/T) found on average in 4.2% of the sequences. Remarkably, all the other motifs deduced from these analyses were significantly enriched in Ts and Gs (Table 2).

Sds3 CTD binds G-quadruplexes

Because Ts and Gs are commonly found in G-quadruplexes, given the high prevalence of the 7-residue motif in the library of selected sequences, we asked whether the sequence 5′-TGTGGTT-3′ could form a G-quadruplex by native PAGE analysis (note that we chose to perform these experiments with DNA rather than RNA because both molecules are capable of forming similar quadruplex structures; (44)). For these analyses, two controls comprising a 10mer self-

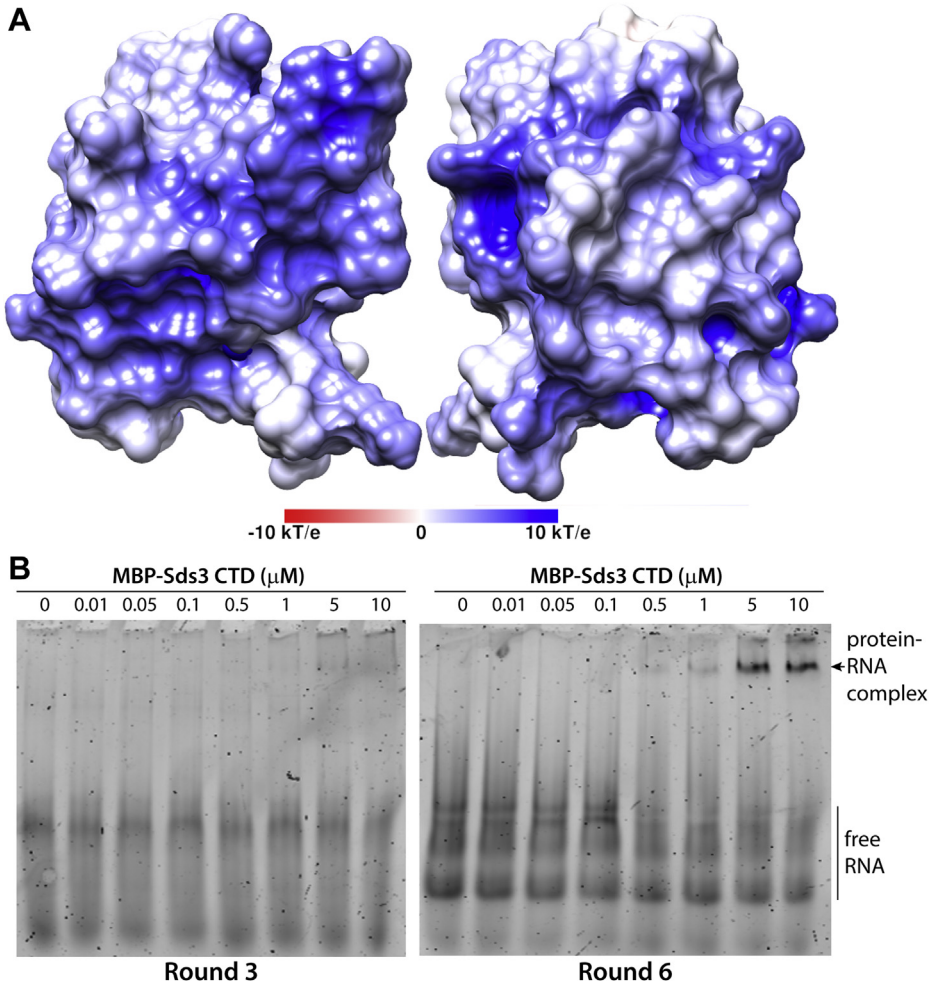

Figure 3. Analysis of the surface properties of Sds3 CTD suggests a nucleic acid-binding function. *A*, an electrostatic potential map calculated using APBS and projected on to the molecular surface of Sds3 CTD. The poses for the two views are identical to those shown in Figure 2. *B*, electrophoretic mobility shift assays conducted using MBP-tagged Sds3 CTD and the RNA library after round 3 (*left*) and round 6 (*right*) of the SELEX experiments. The library from round 6 was reverse transcribed and sent for NGS. The bands were stained with SYBR Gold and visualized using a Typhoon fluorescence imager. CTD, capped Tudor domain; MBP, maltose-binding protein; NGS, next generation sequencing; SELEX, systematic evolution of ligands by exponential enrichment.

complementary DNA duplex 5′-GCGAATTCGC-3′ and a 5′-TTGGGG-3′ (henceforth, T_2G_4) sequence that is known to form DNA quadruplexes (45). Both controls yielded bands with apparent sizes of 10 bp and 14.9 bp (lanes 2 and 3; Fig. 4*A*), respectively, consistent with duplex and quadruplex formation (note that for the latter, the expected size is 12 bp, disregarding differences in shape between the quadruplex and the molecular weight standards). By contrast, the 7mer sequence could not be detected, most likely because of its significantly elevated electrophoretic mobility arising from its single-stranded character and failure to form a higher-order structure (lane 4; Fig. 4*A*). Furthermore, titration of the 7mer produced no significant perturbations in the NMR spectrum of Sds3 CTD (Fig. S3). Collectively, these results establish that the 7mer motif was incapable of forming a G-quadruplex structure and the motif at the single-strand level was insufficient to engage with Sds3 CTD.

Because G-quadruplexes comprising only two planar G-tetrads are not known to exist, we asked whether a repeating sequence such as 5′-TGGTGGTGGTGGT-3′ (henceforth, $T(GGT)_4$) might form a G-quadruplex structure. Native PAGE analysis of this sequence revealed several bands with apparent sizes of 12.7, 22.4, and 42.2 bp (Fig. 4*A*) with the former being the predominant species. Because $T(GGT)_4$ is a 13-nucleotide oligomer, the formation of these higher molecular weight species is attributed to self-association, consistent with the formation of higher-order structures, and most likely, G-quadruplexes. Indeed, previous studies involving multiple GGT repeats have demonstrated the formation of intramolecular G-quadruplexes comprising a stack of four planar G-tetrads, although these sequences were about twice in length compared to those used in this study (46–48). We interpret our native PAGE results for $T(GGT)_4$ to the formation of a dimer as well as higher-order multimers characterized by one or more stacks of four planar G-tetrads. Because our native PAGE experiments were performed in the presence of 100 mM NaCl, and because K+ ions are widely known to stabilize G-quadruplexes (49), we asked whether the addition of 100 mM KCl could alter the stability and the relative populations of the various species. Surprisingly, the banding

Table 2
Statistically significant sequence motifs identified by MEME from NGS data

Motif	Dataset	Number of sites	Probability	E-value
HGTGGTT	2	33,555	0.067110	6.10×10^{-16}
CGTGGTT	4	13,453	0.026906	8.10×10^{-11}
MGTGGTT	5	26,004	0.052008	1.60×10^{-14}
CGTGGTT	8	14,162	0.028324	5.50×10^{-06}
CGTGGTK	9	19,955	0.039910	1.30×10^{-08}
MGTGGTT	10	21,703	0.043406	1.60×10^{-12}
CCGTTTGTGGTGCGTTTTT	2	2766	0.005532	1.50×10^{-27}
	4	2851	0.005702	2.50×10^{-35}
	6	2828	0.005656	8.20×10^{-36}
	10	2916	0.005832	5.70×10^{-31}
CCGTTTGTGGTGCGTTTTTG	1	2703	0.005406	1.10×10^{-29}
	3	2761	0.005522	3.40×10^{-23}
	5	3339	0.006678	7.20×10^{-31}
	7	2347	0.004694	9.20×10^{-24}
	8	3217	0.006434	1.80×10^{-25}
	9	2797	0.005594	1.60×10^{-36}
GGCGTTGTCCGTGGTTTGTG	1	1753	0.003506	2.30×10^{-10}
	3	2616	0.005232	1.60×10^{-24}
	5	1851	0.003702	1.80×10^{-16}
	6	1896	0.003792	1.70×10^{-25}
	8	3858	0.007716	1.40×10^{-19}
	9	2551	0.005102	1.20×10^{-17}
	10	1566	0.003132	3.80×10^{-23}
GTCTGTGGTTGGTCTTGGCT	3	295	0.000590	5.30×10^{-10}
	6	319	0.000638	4.10×10^{-05}
	8	478	0.000956	5.20×10^{-07}

IUPAC codes: H: A/C/T; M: A/C; K: G/T.

patterns were mostly unchanged in the absence and presence of KCl (compare lanes 5 and 10; Fig. 4*A*).

We then interrogated the stabilities of the structures formed by the four DNA sequences by denaturing them in 8 M urea followed by heat treatment (95 °C for 2 min) before resolving the various species by PAGE. Interestingly, both T_2G_4 and TGTGGTT failed to yield any detectable bands, whereas the 10mer and $T(GGT)_4$ both yielded bands consistent with the formation of a duplex and a dimeric G-quadruplex, respectively (Fig. 4*B*). We attribute these bands to their ability to refold readily. To conclusively establish whether $T(GGT)_4$ formed G-quadruplex structures, we recorded the ^{1}H NMR spectrum for this sequence in water. Resonances in the 10 to 12 ppm range characteristic of imino protons involved in

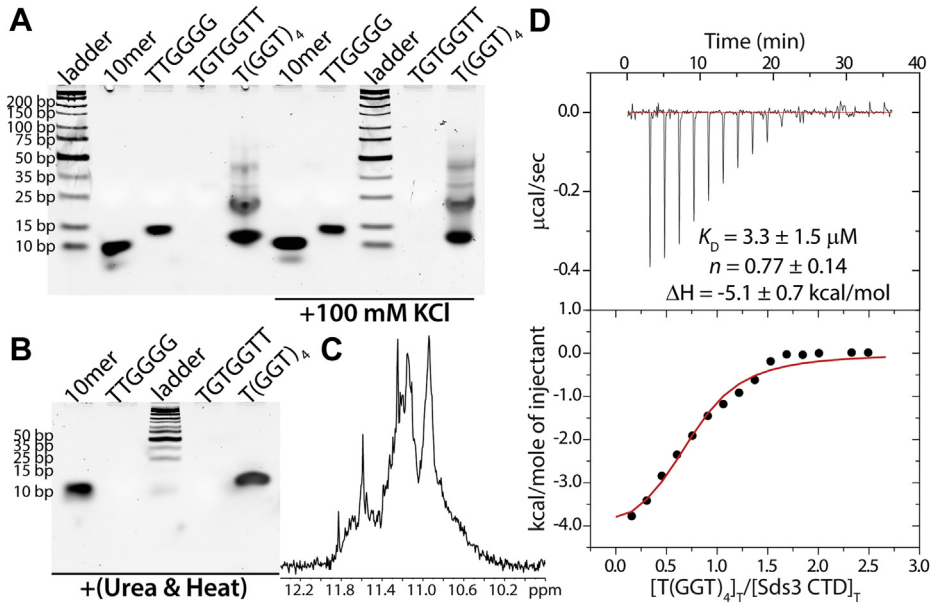

Figure 4. A DNA sequence-bearing multiple copies of a commonly occuring motif deduced from SELEX experiments forms a stable G-quadruplex and binds Sds3 CTD. *A*, native PAGE analyses of a 10mer model duplex and the T_2G_4 G-quadruplex along with two other DNA sequences bearing motifs from SELEX experiments. The gels were stained with SYBR Gold and visualized using a Sapphire fluorescence imager. *B*, polyacrylamide gel electrophoresis analysis after chemical and thermal denaturation of the same four DNA oligomers shown in (*A*). *C*, an expanded plot corresponding to the imino proton region in the 1D ^{1}H NMR spectrum of $T(GGT)_4$ characteristic of a G-quadruplex. *D*, isothermal titration calorimetry thermograms (*top*) and the binding isotherms (*bottom*) from a titration of Sds3 CTD with the $T(GGT)_4$ G-quadruplex. The experiment was performed in triplicate at 25 °C and the average and SD of the fitted values are shown. CTD, capped Tudor domain; SELEX, systematic evolution of ligands by exponential enrichment.

hydrogen-bonding interactions in G-tetrads could be readily observed, establishing the formation of G-quadruplex structures ((50); Fig. 4*C*).

To test whether Sds3 CTD could bind to the $T(GGT)_4$ G-quadruplex, the protein was titrated with the DNA in isothermal titration calorimetry experiments. Changes in heat could be readily detected during these titrations (Fig. 4*D*), whereas control titrations with the DNA performed in the absence of protein yielded little or no heat. These results collectively establish a direct physical interaction between the protein and the DNA. Nonlinear least-squares fitting of the isotherms established an equilibrium dissociation constant for the interaction in the low micromolar range (3.3 ± 1.5 μM; average and standard deviations are from three independent measurements; Fig. 4*D*). However, the efforts to generate a sample of the Sds3 CTD-$T(GGT)_4$ complex for mapping the G-quadruplex-interaction surface of the CTD using NMR were unsuccessful because of poor solubility of the protein–DNA complex at the concentrations (>100 μM) required for these studies.

We then asked whether the model T_2G_4 G-quadruplex used in PAGE analyses could interact with Sds3 CTD. We first recorded ^{1}H NMR spectra to confirm G-quadruplex formation for this sequence, which as expected, is characterized by the presence of four narrow, imino proton resonances in the 10 to 12 ppm region emanating from each of the four planar G-tetrads (Fig. S4; four additional resonances of reduced intensity are observed in the imino proton region suggesting the formation of two types of quadruplexes that most likely differ in strand direction). The addition of one equivalent of T_2G_4 to ^{15}N-Sds3 CTD induced significant perturbations in the CTD NMR spectrum (Fig. 5*A*), with only a few resonances shifting to new positions, whereas several others undergo various degrees of line broadening. The resulting protein–DNA complex is thus in intermediate exchange on the NMR timescale with lifetimes on the timescale of tens of milliseconds, characteristic of complexes with micromolar affinities. Interestingly, the minor quadruplex species is relatively unperturbed in the presence of Sds3 CTD, implying that the interaction with the major species is specific (Fig. S4). By contrast, titrations of Sds3 CTD with the 10mer DNA duplex used in PAGE analyses as control elicited substantially diminished perturbations in the CTD spectrum characterized by an average chemical shift deviation of 0.014 ± 0.012 ppm and a protein–DNA complex in fast exchange on the NMR timescale (*i.e.*, unlike the Sds3 CTD–T_2G_4 complex, the protein hops rapidly between the duplex DNA-bound and free states; Fig. S5). In addition, virtually no perturbations in the CTD spectrum were detected in the NMR titrations of the protein with a random RNA 8mer sequence (5′-AACUGUCG-3′). Collectively, our results indicate that Sds3 CTD preferentially associates with G-quadruplexes over double-stranded DNA, single-stranded DNA, and single-stranded RNA sequences.

To identify the region of Sds3 CTD involved in binding to the T_2G_4 G-quadruplex, we quantified the peak intensity ratios in the holo and apo HSQC spectra and mapped them on to the molecular surface of the CTD. The strongest perturbations were observed for a contiguous surface formed largely by residues in strands β6, β7, the loop preceding β6, and the sole helix connecting β7 and β8 (Fig. 5*B*). This surface is distinct from the one located at the edge of the barrel that is commonly used by the BAH domain and the Royal family domains to engage with chromatin targets (Figs. 2, *A* and *B*, and S2). Interestingly, the three-stranded β-sheet formed by the capping motif of the CTD does not show significant spectral perturbations, implying that this novel feature is not essential for binding nucleic acids, at least those that were tested in this study.

Discussion

Histone deacetylase containing chromatin-modifying complexes frequently contain many protein subunits with the nonenzymatic subunits widely thought to impart genome targeting specificity, especially because HDACs exhibit little sequence specificity themselves for acetylated targets. Two common mechanisms of genome targeting involve protein–protein interactions with sequence-specific DNA-binding factors and/or engagement with specific posttranslational modifications on histones. An especially intriguing feature of the Sin3L/Rpd3L complex is that the core subunits, including Sin3, Sds3, and SAP30, harbor the domains of unknown structure and function that are narrowly distributed and found only in the respective orthologs and paralogs. The Sin3 subunit performs a scaffolding function for the assembly of the complex by engaging directly with most of the subunits while also providing multiple surfaces for direct engagement with DNA-bound factors (15). The SAP30 subunit is involved in turbocharging the catalytic activity of HDAC1, whereas Sds3 is thought to impart stability to the complex while also providing a dimerization function and interaction sites for DNA-bound factors and other subunits of the complex.

The discovery of a type of Tudor domain in Sds3 was unexpected and initially suggested an unrecognized function for the subunit in chromatin binding. However, as our subsequent studies have shown, Sds3 CTD shares more in common with another type of Tudor domain found in the Esa1 histone acetyltransferase (HAT) previously shown to bind single-stranded RNA than with canonical Tudor domains (42). Although these noncanonical Tudor domains share a backbone RMSD of only 1.98 Å, with the highest level of structural similarity in the region spanning β4 to β7 of Sds3 CTD, both domains bind to nucleic acids on the body of the barrel (Fig. 6). The involvement of overlapping surfaces of the β-barrel in these distantly related domains is particularly striking. Even more interesting is the presence of conserved tryptophan and tyrosine residues at the protein-nucleic acid interface of both proteins (Fig. 6), although the exact locations of these residues are not conserved between these domains. The latter likely reflects the different specificities of the domains for their target(s). Both the involvement of aromatic residues and their location on the surface of β-sheets are defining features shared with RNA-recognition motifs (51). Thus, both knotted and

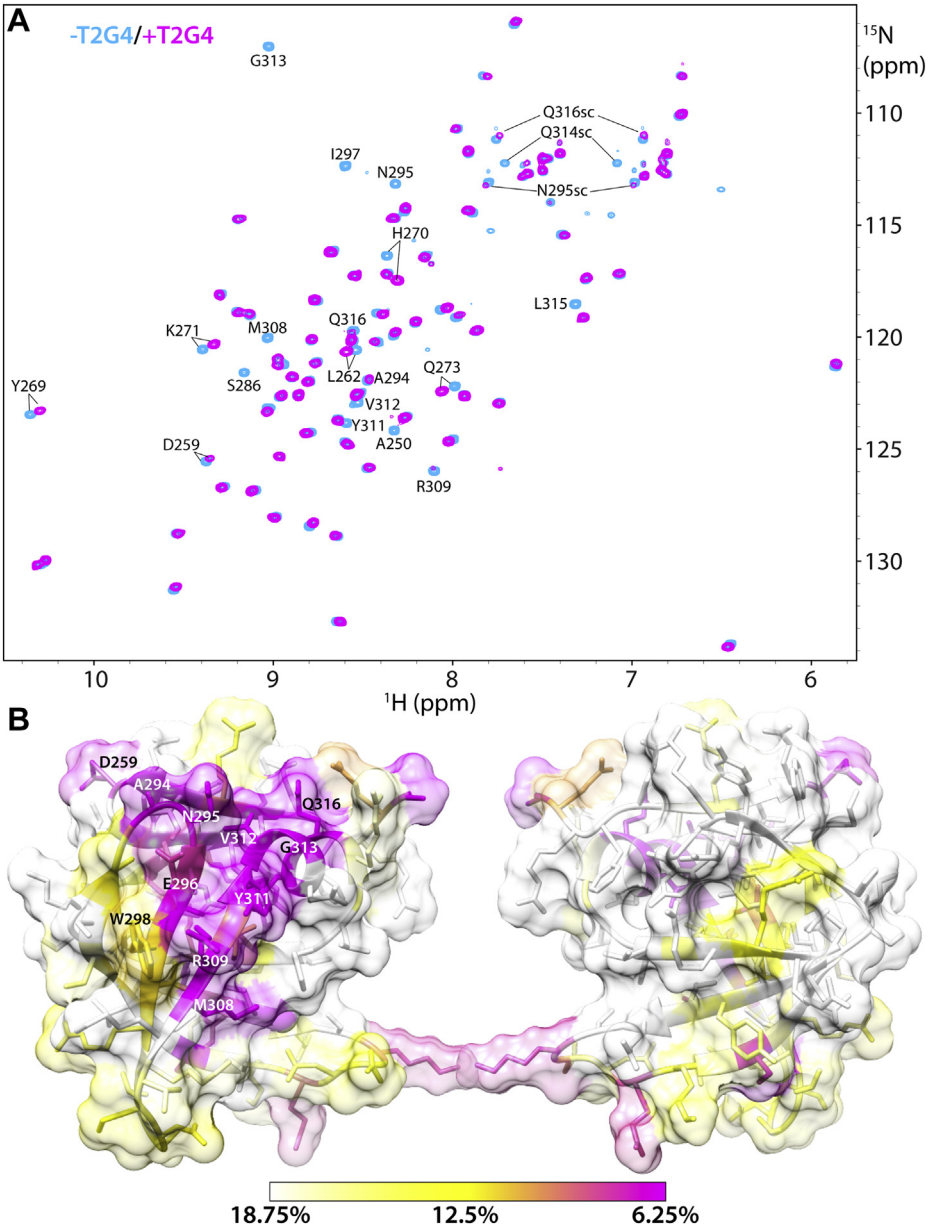

Figure 5. Sds3 CTD binds to a G-quadruplex. *A*, ^{1}H-^{15}N HSQC spectra of Sds3 CTD recorded in the absence (*blue*) and presence (*magenta*) of 1 equivalent of T_2G_4 G-quadruplex DNA. Strongly perturbed resonances are annotated. To facilitate an objective comparison between the holo and apo spectra, the contour thresholds were adjusted using the peak intensities of the 'unperturbed' resonances. *B*, front and back views of the molecular surface of Sds3 CTD colored according to the ratio of the raw peak intensities in the holo and apo spectra (*i.e.*, I_{holo}/I_{apo}). The surface is rendered semitransparently to help identify the underlying residue. Note that the peak intensities of all resonances were uniformly diminished because of the larger size of the resulting complex and due to sample dilution caused by the addition of the quadruplex. CTD, capped Tudor domain.

CTDs appear to have independently acquired nucleic acid-binding functions through a process of convergent evolution. Because Esa1 is a HAT and a member of the NuA4 HAT complex (52), it is intriguing that these noncanonical Tudor domains are found in complexes with opposing enzymatic activities that produce contrasting transcriptional outcomes.

Although both Sds3 and BRMS1L share a CTD, the ortholog proteins define separate clades consistent with their distinct patterns of sequence conservation (Fig. 1*A*), implying that the domains, although sharing a similar function, likely encode different specificities for their targets. Interestingly, the N-terminal three-stranded β-sheet that forms the capping motif that is unique to these Tudor domains does not seem to be involved in nucleic acid binding for the sequences that were tested. However, this structural motif also harbors conserved, solvent-exposed tyrosine and tryptophan residues (Y263 and W268) that could potentially be involved in binding other nucleic acid targets. Deeper biochemical and structural studies are required once their *in vivo* targets have been identified.

There is precedent for RNA mediating the recruitment of at least one other chromatin-modifying complex, the LSD1–CoREST complex (53). The complex harbors both a histone demethylase (LSD1) as well as HDACs 1 and 2 and is recruited

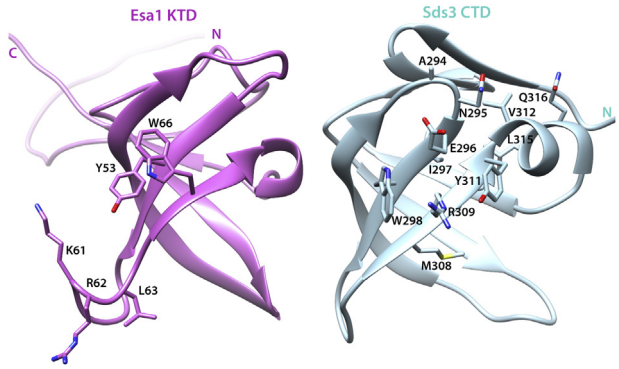

Figure 6. A side-by-side comparison following a best-fit backbone superposition of the Esa1 knotted Tudor domain (KTD; *left*; PDB ID: 2RO0) with the Sds3 capped Tudor domain (*right*). The views highlight the residues that form the RNA and G-quadruplex binding surfaces inferred from NMR titration experiments.

to the telomeric regions by a long noncoding RNA that associates with chromatin in these regions to facilitate telomere silencing and heterochromatin formation (54). The long noncoding RNA harbors many repeats of the 5′-UUAGGG-3′ sequence that forms intramolecular G-quadruplexes that in turn are critical for efficient interactions with LSD1 (55, 56). Because G-quadruplexes have been found to localize to heterochromatin and gene knockout studies implicate Sds3 in the proper establishment of pericentric heterochromatin (16, 57), it is tempting to speculate that the Sin3L/Rpd3L complex may be recruited to these regions through an analogous mechanism involving G-quadruplex structures to promote gene silencing and heterochromatin formation at centromeres. A similar molecular mechanism may be operative in the context of R-loops where single-stranded DNA are common, some of which have been found to harbor intramolecular G-quadruplex structures (44). Multiple subunits of the Sin3L/Rpd3L complex including Sds3 have been implicated in resolving these structures generated during transcription (25). We surmise that the CTD is involved in G-quadruplex-mediated recruitment of the Sin3L/Rpd3L complex to these regions.

Although our *in vitro* studies have provided intriguing hints regarding a potential role for Sds3 CTD in binding nucleic acids, most likely G-quadruplexes, it sets the stage for deeper and more focused studies to elucidate its precise role in Sin3L/Rpd3L biology. Although it appears unlikely that the CTD directly plays a key role in promoting Sin3L/Rpd3L complex stability, given that the domain is absent in the yeast ortholog, its conservation in diverse species nevertheless suggests a fundamental role in these organisms. Future studies aimed at globally identifying and characterizing nucleic acid targets of the CTD in cells could help clarify its molecular function and highlight the biological processes in which the domain plays a critical role.

In conclusion, we have described a new type of Tudor domain that appears to have evolved to perform noncanonical functions such as binding nucleic acids. If our prediction that Sds3 CTD has evolved to preferentially bind G-quadruplexes is confirmed by future studies, it would only be the second instance of an HDAC-containing chromatin-modifying complex implicated in direct recruitment by higher-order nucleic acid structures, expanding the repertoire of macromolecules that could function in this manner. Our findings thus draw attention to potentially new and underappreciated roles for both Sds3 and the Sin3L/Rpd3L HDAC complex in biology.

Experimental procedures

Construct generation, protein expression, and purification

The coding sequence of mouse Sds3 CTD (residues 250–326) was sub-cloned into the pMCSG7, pMCSG9, and pMCSG10 bacterial expression vectors (58). His_6-tagged-CTD encoded by the pMCSG7 vector was expressed at 16 °C in BL21(DE3) cells and subsequently purified *via* Ni^{2+}-affinity chromatography. Cell pellets were resuspended in lysis buffer (50 mM Tris, pH 8.0, 200 mM NaCl, 1 mM TCEP, 1 mM phenylmethylsulfonyl fluoride, 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100) and lysed by sonication. After centrifugation, the lysate was loaded onto a Ni^{2+}-affinity resin (Sigma), washed with high salt (800 mM NaCl), and eluted with 300 mM imidazole. The His_6-tag was removed by incubating the protein with TEV protease overnight at 4 °C, the samples concentrated and further purified *via* size exclusion chromatography using a Superdex 75 GL column (GE Healthcare) and a running buffer comprising 20 mM Tris (pH 8.0), 200 mM NaCl, and 1 mM DTT. Uniformly ^{15}N- and/or ^{13}C-labeled proteins were produced following the same procedure except they were grown in M9 minimal media supplemented with ^{15}N-ammonium sulfate and/or ^{13}C-glucose.

MBP- and GST-tagged proteins encoded by the pMCSG9 and pMCSG10 vectors were expressed and purified in a similar manner as the His_6-tagged Sds3 CTD with the following changes. GST-Sds3 CTD was purified with glutathione sepharose (GE Healthcare) and eluted using 25 mM glutathione, whereas MBP-Sds3 was purified with amylose resin (New England Biolabs) and eluted with 25 mM maltose. All proteins were stored at 4 °C until they were used.

Nuclear magnetic resonance spectroscopy

All NMR spectra were acquired at 25 °C on a 600 MHz Agilent DD2 spectrometer. Sds3 CTD samples in the range of 350 μM in 20 mM sodium phosphate buffer (pH 6.0) containing 50 mM NaCl, 1 mM DTT, and 10% D_2O were used to acquire NMR data. 3D HNCACB, CBCA(CO)NH, C(CO)NH-TOCSY, HNCO, and ^{15}N-NOESY-HSQC NMR spectra were acquired for sequence-specific backbone resonance assignments (59). Data processing was performed using Felix (Felix NMR), and the peaks in these spectra were picked in NMRFAM-Sparky (60) and submitted to I-PINE for peak assignment (61). All the assignments were checked manually for accuracy. Side chain assignments were performed manually using 3D HCCH-COSY and HCCH-TOCSY spectra acquired in D_2O; the sample for these experiments was generated by exchanging the buffer from H_2O to D_2O. Aromatic resonances were assigned based on a careful analysis of 2D ^{1}H-^{13}C

aromatic HSQC, (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE (62), and ^{1}H-^{1}H NOESY spectra recorded in D_2O.

1D ^{1}H NMR spectra for the $T(GGT)_4$ and T_2G_4 DNA samples at 200 to 300 μM concentration were acquired at 25 °C on an Agilent DD2 or Bruker Neo 600 MHz instrument. All the samples were prepared in 20 mM sodium phosphate buffer (pH 6) containing 50 mM NaCl and either 1 mM TCEP or DTT.

Nuclear magnetic resonance structure determination

Backbone ϕ and ψ dihedral angle restraints for structure calculations were derived from a combined analysis of the $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, and backbone ^{15}N chemical shifts using TALOS+ (63); only residues with reliability scores of 10 in secondary structural elements were restrained. ^{1}H-^{1}H NOE-based distance restraints were derived from three spectra, including 3D ^{15}N-edited NOESY (τ_m = 80 ms) recorded in H_2O, 3D ^{13}C-edited aliphatic NOESY (τ_m = 60 ms), and 2D ^{1}H-^{1}H NOESY (τ_m = 75 ms) recorded in D_2O.

Structures were determined using ARIA 1.2 in conjunction with CNS 1.1 starting from an initial structure with extended backbone conformation (64–66). All NOEs were calibrated automatically and were assigned iteratively by ARIA; the assignments were checked manually for errors after each run. Eighty conformers were calculated; 40 conformers with the lowest restraint energies were refined in a shell of water, and the 20 conformers with the lowest restraint energies and violations and ideal covalent geometry were selected. The final conformers were analyzed using CNS (64), PROCHECK (67), and scripts written in-house.

Histone peptide array

His_6-tagged Sds3 CTD was incubated with a MODified histone peptide array (Active Motif) at a concentration of 15 μM. After a 2 h incubation period, the array was washed and probed with an anti-His primary antibody (Thermo Fisher, MA121315, 1:1000 dilution), after which anti-mouse horseradish peroxidase-conjugated secondary antibody (Thermo Fisher Scientific, #OB617005, 1:1000 dilution) was used. The array was imaged using West Pico chemiluminescent substrate (Thermo Scientific, #34080) and a Syngene Pxi chemiluminescent imager. The screen was performed in duplicate as a test of reproducibility.

Nuclear magnetic resonance titrations

Sds3 CTD samples in the 150 to 230 μM range in 20 mM sodium phosphate buffer (pH 6.0) containing 50 mM NaCl, 1 mM DTT, and 10% D_2O were used for the NMR titration experiments. ^{1}H-^{15}N HSQC NMR spectra were acquired after the addition of excess dimethyl lysine, trimethyl lysine, acetyl lysine, and dimethylarginine (Sigma; all compounds were used without further purification). NMR titrations were also conducted with an unmodified histone H3 peptide (residues 1–42), purified histone H2A-H2B heterodimer, DNA and RNA oligonucleotides. The oligonucleotides were purchased from Integrated DNA Technologies and Dharmacon and used without further purification. Data processing and analysis were performed using Felix (Felix NMR) and NMRFAM-Sparky (60).

Systematic evolution of ligands by exponential enrichment

Systematic evolution of ligands by exponential enrichment (43) experiments were conducted with GST-tagged Sds3 CTD in 50 mM Tris buffer (pH 7.5) containing 150 mM NaCl, 1 mM $MgCl_2$, and 0.05% NP-50. An RNA library for selection experiments was obtained from TriLink Biotechnologies that consisted of random 20mer sequences, flanked by adapters of known sequence for reverse transcription, PCR amplification, and sequencing. To preclear the RNA library of nonspecific interactions with GST, the RNA library was initially incubated with purified GST immobilized on glutathione sepharose beads. The flow-through containing unbound RNA was then incubated with purified GST-Sds3 CTD immobilized on glutathione sepharose beads. The beads were washed extensively, and the protein was digested with proteinase K. Bound RNA was then purified using phenol/chloroform extraction. RNA was reverse transcribed and then PCR amplified. The PCR template was used to transcribe RNA, and the process was repeated six times to enrich Sds3 CTD binding sequences.

After the final round of selection, PCR products were submitted to the Northwestern NUSeq core facility. Sequencing reads were generated using an Illumina SR75 sequencer. The sequences were trimmed to remove adapters using Cutadapt (68) and filtered by quality (Fast QC) in the Galaxy bioinformatics suite (69). Ten datasets of 500,000 randomly selected sequences from $\sim 10^6$ high quality reads were extracted using the Galaxy bioinformatics suite for motif analysis. Motifs were identified and analyzed using the MEME suite (70). The top five motifs from each dataset were compiled and those that appeared in more than three datasets were deemed significant for inclusion in Table 2.

Polyacrylamide gel electrophoresis and EMSAs

Samples for EMSAs were prepared in 20 mM Tris buffer (pH 8), containing 200 mM NaCl and 1 mM DTT by mixing the increasing concentrations of MBP-tagged Sds3 CTD with 1 μM total RNA from rounds 3 and 6 (the final round) of SELEX. Electrophoretic mobility assays were performed using 5% native PAGE gels with 0.5× TB buffer (50 mM Tris, 50 mM boric acid). The gels were equilibrated for 30 min before the samples were loaded onto the gel. The samples were run at 4 °C for 90 min and then stained 30 min with SYBR Gold (ThermoFisher). The gels were imaged on a Typhoon fluorescence imager with the excitation and emission set to 480 nm and 520 nm, respectively.

DNA samples for native PAGE were prepared in 20 mM Tris, 20 mM sodium phosphate buffer (pH 6), containing 100 mM NaCl. Samples for 'denaturing PAGE' to test the chemical and thermal stability of various DNA were prepared in the same buffer as native PAGE but in the presence of 8 M urea and incubated at 95 °C for 2 min before being loaded on

to the gel. Both types of PAGE were performed using 10% PAGE gels with 0.5× TB buffer. The gels were equilibrated for 30 min before the samples were loaded. Samples were run at 4 °C for 45 min and then stained for 20 min with SYBR Gold before imaging using an Azure Biosystems Sapphire fluorescence imager with the excitation and emission set to 488 nm and 518 nm, respectively. The apparent molecular sizes of various species were inferred from the molecular weight standards that were run in parallel using the software provided by the vendor.

Isothermal titration calorimetry

Sds3 CTD was purified as described above although reducing agents in the buffer were omitted in the SEC step. Protein samples were prepared at 25 μM concentration in 20 mM Tris, 20 mM sodium phosphate buffer (pH 6), containing 100 mM NaCl and loaded into the sample cell. The protein was titrated with DNA samples including $T(GGT)_4$, T_2G_4, and the model 10mer duplex that were all prepared in the same buffer as the protein at 500 μM concentration and loaded into the syringe. All measurements were performed on a Malvern iTC200 instrument with the jacket temperature set to 25 °C.

Data availability

The ensemble of NMR structures along with the NMR-based restraints used in the structure calculations have been deposited with the wwPDB (accession ID 7SXI). The NMR chemical shifts for Sds3 CTD have been deposited with BMRB (accession ID 30969).

Supporting information—This article contains supporting information.

Acknowledgments—This work was supported by grants from the National Institutes of Health for upgrading the 600 MHz NMR console (S10 OD012016). We thank Arabela Grigorescu in the Keck Biophysics Facility for assistance with ITC measurements and members of the Radhakrishnan lab for critical comments. We are grateful to the Robert H. Lurie Comprehensive Cancer Center at Northwestern for supporting structural biology research. We thank Yawen Bai at the NCI for providing the plasmid encoding histone H2A-H2B heterodimer.

Author contributions—R. D. M. and I. R. conceptualization; R. D. M. and I. R. methodology; R. D. M., J. H., M. G., E. J., and N. D. investigation; R. D. M. writing–original draft; R. D. M. data curation; Y. Z. resources; I. R. funding acquisition; I. R. supervision; I. R. writing–review and editing.

Funding and additional information—This work was supported by grants from the American Heart Association to I. R. (17GRNT33680167) and R. D. M. (16PRE27260041) and to Northwestern University for providing J. H., M. G., and E. J. with funds for undergraduate research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: BAH, bromo adjacent homology; BRMS1, breast cancer metastasis suppressor 1; BRMS1L, BRMS1-like; CTD, capped Tudor domain; HAT, histone acetyltransferase; HDAC, histone deacetylase; MBP, maltose-binding protein; SELEX, systematic evolution of ligands by exponential enrichment.

References

1. Gardner, K. E., Allis, C. D., and Strahl, B. D. (2011) Operating on chromatin, a colorful language where context matters. *J. Mol. Biol.* **409**, 36–46
2. Bowman, G. D., and Poirier, M. G. (2015) Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* **115**, 2274–2295
3. Evertts, A. G., Zee, B. M., Dimaggio, P. A., Gonzales-Cope, M., Coller, H. A., and Garcia, B. A. (2013) Quantitative dynamics of the link between cellular metabolism and histone acetylation. *J. Biol. Chem.* **288**, 12142–12151
4. Moser, M. A., Hagelkruys, A., and Seiser, C. (2014) Transcription and beyond: The role of mammalian class I lysine deacetylases. *Chromosoma* **123**, 67–78
5. Yang, X. J., and Seto, E. (2008) The Rpd3/Hda1 family of lysine deacetylases: From bacteria and yeast to mice and men. *Nat. Rev. Mol. Cell Biol.* **9**, 206–218
6. Hayakawa, T., and Nakayama, J. (2011) Physiological roles of class I HDAC complex and histone demethylase. *J. Biomed. Biotechnol.* **2011**, 129383
7. Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Cote, J., and Wang, W. (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol. Cell* **2**, 851–861
8. Smith, K. T., Sardiu, M. E., Martin-Brown, S. A., Seidel, C., Mushegian, A., Egidy, R., Florens, L., Washburn, M. P., and Workman, J. L. (2012) Human family with sequence similarity 60 member A (FAM60A) protein: A new subunit of the Sin3 deacetylase complex. *Mol. Cell. Proteomics* **11**, 1815–1828
9. You, A., Tong, J. K., Grozinger, C. M., and Schreiber, S. L. (2001) CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1454–1458
10. Bantscheff, M., Hopf, C., Savitski, M. M., Dittmann, A., Grandi, P., Michon, A. M., Schlegl, J., Abraham, Y., Becher, I., Bergamini, G., Boesche, M., Delling, M., Dumpelfeld, B., Eberhard, D., Huthmacher, C., *et al.* (2011) Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat. Biotechnol.* **29**, 255–265
11. Guenther, M. G., Barak, O., and Lazar, M. A. (2001) The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol. Cell. Biol.* **21**, 6091–6101
12. Carrozza, M. J., Florens, L., Swanson, S. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005) Stable incorporation of sequence specific repressors Ash1 and Ume6 into the Rpd3L complex. *Biochim. Biophys. Acta* **1731**, 77–87. discussion 75-6
13. Silverstein, R. A., and Ekwall, K. (2005) Sin3: A flexible regulator of global gene expression and genome stability. *Curr. Genet.* **47**, 1–17
14. Grzenda, A., Lomberk, G., Zhang, J. S., and Urrutia, R. (2009) Sin3: Master scaffold and transcriptional corepressor. *Biochim. Biophys. Acta* **1789**, 443–450
15. Adams, G. E., Chandru, A., and Cowley, S. M. (2018) Co-repressor, co-activator and general transcription factor: The many faces of the Sin3 histone deacetylase (HDAC) complex. *Biochem. J.* **475**, 3921–3932
16. David, G., Turner, G. M., Yao, Y., Protopopov, A., and DePinho, R. A. (2003) mSin3-associated protein, mSds3, is essential for pericentric heterochromatin formation and chromosome segregation in mammalian cells. *Genes Dev.* **17**, 2396–2405

17. Fleischer, T. C., Yun, U. J., and Ayer, D. E. (2003) Identification and characterization of three new components of the mSin3A corepressor complex. *Mol. Cell. Biol.* **23**, 3456–3467
18. Banks, C. A. S., Thornton, J. L., Eubanks, C. G., Adams, M. K., Miah, S., Boanca, G., Liu, X., Katt, M. L., Parmely, T. J., Florens, L., and Washburn, M. P. (2018) A structured workflow for mapping human Sin3 histone deacetylase complex interactions using halo-MudPIT affinity-purification mass spectrometry. *Mol. Cell. Proteomics* **17**, 1432–1447
19. Banks, C. A. S., Zhang, Y., Miah, S., Hao, Y., Adams, M. K., Wen, Z., Thornton, J. L., Florens, L., and Washburn, M. P. (2020) Integrative modeling of a Sin3/HDAC complex sub-structure. *Cell Rep.* **31**, 107516
20. He, Y., Imhoff, R., Sahu, A., and Radhakrishnan, I. (2009) Solution structure of a novel zinc finger motif in the SAP30 polypeptide of the Sin3 corepressor complex and its potential role in nucleic acid recognition. *Nucleic Acids Res.* **37**, 2142–2152
21. Marcum, R. D., and Radhakrishnan, I. (2019) Inositol phosphates and core subunits of the Sin3L/Rpd3L histone deacetylase (HDAC) complex up-regulate deacetylase activity. *J. Biol. Chem.* **294**, 13928–13938
22. Clark, M. D., Marcum, R., Graveline, R., Chan, C. W., Xie, T., Chen, Z., Ding, Y., Zhang, Y., Mondragon, A., David, G., and Radhakrishnan, I. (2015) Structural insights into the assembly of the histone deacetylase-associated Sin3L/Rpd3L corepressor complex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3669–E3678
23. Shi, X., Seldin, D. C., and Garry, D. J. (2012) Foxk1 recruits the Sds3 complex and represses gene expression in myogenic progenitors. *Biochem. J.* **446**, 349–357
24. Meehan, W. J., Samant, R. S., Hopper, J. E., Carrozza, M. J., Shevde, L. A., Workman, J. L., Eckert, K. A., Verderame, M. F., and Welch, D. R. (2004) Breast cancer metastasis suppressor 1 (BRMS1) forms complexes with retinoblastoma-binding protein 1 (RBP1) and the mSin3 histone deacetylase complex and represses transcription. *J. Biol. Chem.* **279**, 1562–1569
25. Salas-Armenteros, I., Perez-Calero, C., Bayona-Feliu, A., Tumini, E., Luna, R., and Aguilera, A. (2017) Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. *EMBO J.* **36**, 3532–3547
26. Nikolaev, A. Y., Papanikolaou, N. A., Li, M., Qin, J., and Gu, W. (2004) Identification of a novel BRMS1-homologue protein p40 as a component of the mSin3A/p33(ING1b)/HDAC1 deacetylase complex. *Biochem. Biophys. Res. Commun.* **323**, 1216–1222
27. Gong, C., Qu, S., Lv, X. B., Liu, B., Tan, W., Nie, Y., Su, F., Liu, Q., Yao, H., and Song, E. (2014) BRMS1L suppresses breast cancer metastasis by inducing epigenetic silence of FZD10. *Nat. Commun.* **5**, 5406
28. Zhang, S., Lin, Q. D., and Di, W. (2006) Suppression of human ovarian carcinoma metastasis by the metastasis-suppressor gene, BRMS1. *Int. J. Gynecol. Cancer* **16**, 522–531
29. Seraj, M. J., Samant, R. S., Verderame, M. F., and Welch, D. R. (2000) Functional evidence for a novel human breast carcinoma metastasis suppressor, BRMS1, encoded at chromosome 11q13. *Cancer Res.* **60**, 2764–2769
30. Hurst, D. R. (2012) Metastasis suppression by BRMS1 associated with SIN3 chromatin remodeling complexes. *Cancer Metastasis Rev.* **31**, 641–651
31. Smith, P. W., Liu, Y., Siefert, S. A., Moskaluk, C. A., Petroni, G. R., and Jones, D. R. (2009) Breast cancer metastasis suppressor 1 (BRMS1) suppresses metastasis and correlates with improved patient survival in non-small cell lung cancer. *Cancer Lett.* **276**, 196–203
32. Silveira, A. C., Hurst, D. R., Vaidya, K. S., Ayer, D. E., and Welch, D. R. (2009) Over-expression of the BRMS1 family member SUDS3 does not suppress metastasis of human cancer cells. *Cancer Lett.* **276**, 32–37
33. Hurst, D. R., and Welch, D. R. (2011) Unraveling the enigmatic complexities of BRMS1-mediated metastasis suppression. *FEBS Lett.* **585**, 3185–3190
34. Zimmermann, R. C., and Welch, D. R. (2020) BRMS1: A multifunctional signaling molecule in metastasis. *Cancer Metastasis Rev.* **39**, 755–768
35. Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020) Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1496–1503
36. Holm, L. (2020) Using dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42
37. Du, J., Zhong, X., Bernatavichute, Y. V., Stroud, H., Feng, S., Caro, E., Vashisht, A. A., Terragni, J., Chin, H. G., Tu, A., Hetzel, J., Wohlschlegel, J. A., Pradhan, S., Patel, D. J., and Jacobsen, S. E. (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**, 167–180
38. Xu, C., Cui, G., Botuyan, M. V., and Mer, G. (2015) Methyllysine recognition by the royal family modules: Chromo, Tudor, MBT, chromo barrel, and PWWP domains. In: Zhou, M. M., ed. *Histone Recognition*, Springer International Publishing, Berlin: 49–82
39. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., *et al*. (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128
40. Vermeulen, M., Carrozza, M. J., Lasonder, E., Workman, J. L., Logie, C., and Stunnenberg, H. G. (2004) *In vitro* targeting reveals intrinsic histone tail specificity of the Sin3/histone deacetylase and N-CoR/SMRT corepressor complexes. *Mol. Cell. Biol.* **24**, 2364–2372
41. Zhou, Z., Feng, H., Hansen, D. F., Kato, H., Luk, E., Freedberg, D. I., Kay, L. E., Wu, C., and Bai, Y. (2008) NMR structure of chaperone Chz1 complexed with histones H2A.Z-H2B. *Nat. Struct. Mol. Biol.* **15**, 868–869
42. Shimojo, H., Sano, N., Moriwaki, Y., Okuda, M., Horikoshi, M., and Nishimura, Y. (2008) Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. *J. Mol. Biol.* **378**, 987–1001
43. Manley, J. L. (2013) SELEX to identify protein-binding sites on RNA. *Cold Spring Harb. Protoc.* **2013**, 156–163
44. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., and Balasubramanian, S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474
45. Wang, Y., and Patel, D. J. (1992) Guanine residues in d(T2AG3) and d(T2G4) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* **31**, 8112–8119
46. Maity, A., Winnerdy, F. R., Chang, W. D., Chen, G., and Phan, A. T. (2020) Intra-locked G-quadruplex structures formed by irregular DNA G-rich motifs. *Nucleic Acids Res.* **48**, 3315–3327
47. Chung, W. J., Heddi, B., Schmitt, E., Lim, K. W., Mechulam, Y., and Phan, A. T. (2015) Structure of a left-handed DNA G-quadruplex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2729–2733
48. Do, N. Q., Chung, W. J., Truong, T. H. A., Heddi, B., and Phan, A. T. (2017) G-quadruplex structure of an anti-proliferative DNA sequence. *Nucleic Acids Res.* **45**, 7487–7493
49. Saintome, C., Amrane, S., Mergny, J. L., and Alberti, P. (2016) The exception that confirms the rule: A higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. *Nucleic Acids Res.* **44**, 2926–2935
50. Adrian, M., Heddi, B., and Phan, A. T. (2012) NMR spectroscopy of G-quadruplexes. *Methods* **57**, 11–24
51. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131
52. Avvakumov, N., and Cote, J. (2007) The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene* **26**, 5395–5407
53. Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., Regev, A., Lander, E. S., and Rinn, J. L. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11667–11672
54. Porro, A., Feuerhahn, S., and Lingner, J. (2014) TERRA-reinforced association of LSD1 with MRE11 promotes processing of uncapped telomeres. *Cell Rep.* **6**, 765–776
55. Patel, D. J., Phan, A. T., and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order

DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **35**, 7429–7455
56. Hirschi, A., Martin, W. J., Luka, Z., Loukachevitch, L. V., and Reiter, N. J. (2016) G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme. *RNA* **22**, 1250–1260
57. Hoffmann, R. F., Moshkin, Y. M., Mouton, S., Grzeschik, N. A., Kalicharan, R. D., Kuipers, J., Wolters, A. H., Nishida, K., Romashchenko, A. V., Postberg, J., Lipps, H., Berezikov, E., Sibon, O. C., Giepmans, B. N., and Lansdorp, P. M. (2016) Guanine quadruplex structures localize to heterochromatin. *Nucleic Acids Res.* **44**, 152–163
58. Eschenfeldt, W. H., Lucy, S., Millard, C. S., Joachimiak, A., and Mark, I. D. (2009) A family of LIC vectors for high-throughput cloning and purification of proteins. *Methods Mol. Biol.* **498**, 105–115
59. Ferentz, A. E., and Wagner, G. (2000) NMR spectroscopy: A multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33**, 29–65
60. Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327
61. Lee, W., Bahrami, A., Dashti, H. T., Eghbalnia, H. R., Tonelli, M., Westler, W. M., and Markley, J. L. (2019) I-PINE web server: An integrative probabilistic NMR assignment system for proteins. *J. Biomol. NMR* **73**, 213–222
62. Lohr, F., and Ruterjans, H. (1996) Novel pulse sequences for the resonance assignment of aromatic side chains in ^{13}C-labeled proteins. *J. Magn. Reson. B* **112**, 259–268
63. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223
64. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921
65. Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003) ARIA: Automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**, 315–316
66. Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2004) Correction of spin diffusion during iterative automated NOE assignment. *J. Magn. Reson.* **167**, 334–342
67. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486
68. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12
69. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Gruning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., *et al.* (2018) The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544
70. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208