Published in final edited form as:
Nat Struct Mol Biol. 2014 August ; 21(8): 686–695. doi:10.1038/nsmb.2853.

RPRD1A and RPRD1B Are Human RNA Polymerase II C-Terminal Domain Scaffolds for Ser5 Dephosphorylation

Zuyao Ni[#1], **Chao Xu**[#2], **Xinghua Guo**[1], **Gerald O. Hunter**[3], **Olga V. Kuznetsova**[4], **Wolfram Tempel**[2], **Edyta Marcon**[1], **Guoqing Zhong**[1], **Hongbo Guo**[1], **Wei-Hung William Kuo**[5], **Joyce Li**[1,5], **Peter Young**[1], **Jonathan B. Olsen**[5], **Cuihong Wan**[1], **Peter Loppnau**[2], **Majida El Bakkouri**[2], **Guillermo A. Senisterra**[2], **Hao He**[2], **Haiming Huang**[1], **Sachdev S. Sidhu**[1,5], **Andrew Emili**[1,5], **Shona Murphy**[4], **Amber L. Mosley**[3], **Cheryl H. Arrowsmith**[2,6], **Jinrong Min**[2,7], and **Jack F. Greenblatt**[1,5]

[1]Donnelly Centre, University of Toronto, Toronto, Ontario, Canada

[2]Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada

[3]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana, USA

[4]Sir William Dunn School of Pathology, University of Oxford, Oxford, UK

[5]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

[6]Department of Medical Biophysics, University of Toronto, Ontario Cancer Institute, Campbell Family Cancer Research Institute, Toronto, Ontario, Canada.

[7]Department of Physiology, University of Toronto, Toronto, Ontario, Canada.

[#] These authors contributed equally to this work.

SUMMARY

The RNA polymerase II (RNAPII) carboxyl-terminal domain (CTD) heptapeptide repeats (Y1-S2-P3-T4-S5-P6-S7) undergo dynamic phosphorylation and dephosphorylation during the transcription cycle to recruit factors that regulate transcription, RNA processing and chromatin modification. We show here that RPRD1A and RPRD1B form homodimers and heterodimers through their coiled-coil domains and interact preferentially *via* CTD interaction domains (CIDs) with CTD repeats phosphorylated at S2 and S7. Our high resolution crystal structures of the RPRD1A, RPRD1B and RPRD2 CIDs, alone and in complex with CTD phosphoisoforms, elucidate the molecular basis of CTD recognition. In an interesting example of cross-talk between

Correspondence should be addressed to J.F.G. (jack.greenblatt@utoronto.ca) or J.M. (jr.min@utoronto.ca).

AUTHOR CONTRIBUTIONS Z.N. and J.F.G. conceived the project and designed the experiments. Z.N. and C.X. performed protein expression, purification and crystallization experiments and analyzed the structure data. W.T. and M.E.B. conducted crystallographic data collection and structure determination and refinement. Z.N., X.G., G.O.H, O.V.K., E.M., G.Z., H.G., W.-H.W.K., J.L., P.Y., J.B.O., C.W., P.L., G.A.S., H.H., and H.H. conducted experiments. S.S.S., A.E., S.M., A.L.M., C.H.A., and J.M. guided the experiments. All authors commented on the manuscript. Z.N. analyzed the data. Z.N. and J.F.G. wrote the manuscript. J.M. and J.F.G. supervised the project.

Accession codes. The crystal structures described in this paper have been deposited in the Protein Data Bank: RPRD2 CID (4FLB), RPRD1A CID-S7P CTD complex (4JXT), RPRD1B CID-S2P CTD complex (4Q94), RPRD1B CID-unmodified CTD complex (4Q96) and RPRD1B coiled-coil domain (4FLA) (See Supplementary Table 1).

COMPETING FINANCIAL INTERESTS Authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

different CTD modifications, our data also indicate that RPRD1A and RPRD1B associate directly with RPAP2 phosphatase and, by interacting with CTD repeats where phospho-S2 and/or phospho-S7 bracket a phospho-S5 residue, serve as CTD scaffolds to coordinate the dephosphorylation of phospho-S5 by RPAP2.

INTRODUCTION

The carboxyl-terminal domain (CTD) of the largest subunit, POLR2A, of human RNA polymerase II (RNAPII) consists of multiple, sometimes degenerate, heptapeptide repeats with consensus sequence Y1-S2-P3-T4-S5-P6-S7 (refs. 1,2). The CTD is phosphorylated during transcription on Y1, T4, and all three serine residues[3]. Different phosphorylation patterns, proline isomerization[3,4], and modification of non-consensus CTD residues[5] create a "CTD code"[6] that recruits various factors to regulate transcription, mRNA processing and histone modification[5,7-10]. There also appears to be crosstalk between different CTD modifications[5] .

General transcription factor TFIIH phosphorylates S5 and S7 (S5P, S7P) in promoter regions[11-15], where S5P recruits mRNA-capping enzymes[16-19], the yeast COMPASS complex for histone H3K4 trimethylation by Set1 (MLL proteins in human)[20-22], and the yeast Sen1-Nrd1-Nab3 complex to terminate non-coding snRNAs, snoRNAs and cryptic unstable transcripts[23]. S5P also characterizes "poised" RNAPII in metazoan promoter regions[24-27]. RNAPII escaping the promoter is phosphorylated on S2 by p-TEFb[28,29], CDK12 and/or CDK13 (ref. 30), and BRD4 (ref. 31). Transition from high S5P to high S2P is accompanied by partial dephosphorylation of S5P by Rtr1 in yeast[32]. The human Rtr1 homologue RPAP2 is similarly needed for dephosphorylation of S5P during snRNA gene transcription[33]. Further downstream, S7P[34] and the remaining S5P[35,36] are removed by Ssu72. S2P dephosphorylation is carried out by yeast Fcp1 and its mammalian homologue CTDP1 (refs. 28,37,38).

Despite extensive CTD studies, how the CTD is structurally organized and how CTD modifications regulate each other remain largely unknown. In addition, the phosphatase activity of RPAP2 is controversial[33,39]. We set out here to characterize three human CTD-interacting proteins. Our structural and biophysical studies show that the CTD interaction domain (CID)-containing proteins RPRD1A, RPRD2, and the oncoprotein RPRD1B[40], which associate with RNAPII[41], associate preferentially as dimers with S2P- and, to a lesser extent, S7P-containing CTD peptides, whereas S5P interferes with binding. We show RPAP2 is a substrate-selective phosphatase, whose interaction with RNAPII requires RPRD1A and/or RPRD1B. By binding two S2P- and/or S7P-CTD repeat-containing decameric sequences, RPRD1A-RPRD1B dimers act as scaffolds that organize the CTD to present S5P located in the intervening region to RPAP2 for dephosphorylation.

Europe PMC Funders Author Manuscripts

RESULTS

Recognition of specific CTD phosphoisoforms by RPRDs

RPRD1A, RPRD1B and RPRD2 (“RPRDs”) were previously found to co-precipitate with phosphorylated RNAPII[41]. Here we used isothermal titration calorimetry (ITC) to determine dissociation constants (K_d) of recombinant RPRD CIDs (Fig.1a) for CTD peptides containing two heptapeptide repeats without modification or with serine phosphorylation at positions 2, 5, or 7 (UnM, S2P, S5P, or S7P, respectively). The CTD-binding affinities of the three RPRD CIDs followed the order S2P (K_d from 6.8 to 8.4 μM) > S7P (K_d from 23.6 to 82.8 μM) > UnM (K_d from 114 to 355 μM) > S5P CTD (K_d >1,000 μM) (Table 1). Combining S2P and S7P on the same repeats (i.e. S2,7P CTD) increased the RPRD1A and RPRD1B affinities by 1.6-fold and 2.6-fold, respectively, compared to S2P alone, while the RPRD2 CID affinity remained unchanged. In contrast, S5P in the same heptapeptide repeat with either S2P or S7P (i.e. S2,5P CTD or S5,7P CTD, respectively, in Table 1) abolished detectable binding (K_d >1,000 μM) for all three CIDs. This result and our previous observation that RPRDs co-precipitate with S5P-containing RNAPII in cell extracts[41] imply that S5P-containing repeats exist on the same CTDs as S2- and/or S7-containing repeats that bind RPRDs. That S5P on a repeat adjacent to one with S7P (S7P-S5P-CTD in Table 1) had no significant effect on CTD binding supports this idea.

Thirty-one of 52 mammalian CTD repeats, mainly in the C-terminal half, deviate from consensus, predominantly at position 7 and most frequently by lysine substitution. With a CTD peptide containing S2P and lysine, rather than serine, at position 7 (S2P-K7 CTD in Table 1), RPRD1A bound 8.4-fold more strongly than to a consensus peptide, while RPRD2 CID-binding was about 1.6-fold stronger, and RPRD1B CID-binding was slightly weaker (0.8-fold). These data suggest RPRDs would bind many, if not most, non-consensus CTD repeats.

Crystal structures of RPRD CID-CTD complexes

To better understand how RPRDs recognize the CTD, we obtained crystal structures for various RPRD CID-CTD complexes, starting with a 1.9 Å resolution structure of the RPRD1A CID bound to a 19-mer S7P CTD peptide, SPSYSPTSPSYSPTSPSYS, containing more than two heptapeptide repeats bearing three S7P residues (underlined). Like the CIDs of its yeast homologues Rtt103, Pcf11, and Nrd1 (refs. 23,42,43), the RPRD1A CID contained 8 α-helices arranged in a right-handed superhelical manner (Fig. 1b). A concave channel accommodated the CTD decapeptide PSYSPTSPSY (S7P underlined) in a linear conformation such that two S7P residues, S7aP and S7bP, occupied the channel entrance and exit, respectively (Fig. 1c,d). Structure statistics for structures described here are shown in Table 2 and Supplementary Table 1.

Many CID-CTD contacts were observed (Fig. 1c,d,e**).** CID N18 formed water-mediated hydrogen bonds with the S7a phosphate group and the backbone carbonyl group of Y1b and further hydrogen-bonds with the backbone amide of Y1b. N64 and D65 hydrogen-bonded to the Y1b phenolic hydroxyl group deeply buried inside the center of the channel. N64 also contacted the P3b carbonyl group via a hydrogen bond. Q20 formed hydrogen bonds with

the P6a carbonyl group, the CTD backbone amide of S2b, and the S5b hydroxyl group. R114 hydrogen-bonded with the carbonyl groups of both T4b and P6b. N69 formed water-mediated hydrogen bonds with both the backbone amide and phosphate group of S7b. In addition, I110 contacted both P3b and T4b through hydrophobic interactions. Finally, the two S7P phosphate groups pointed toward nearby positively charged areas of the CID surface, one of them close to R72 (Fig. 1c). Altogether, direct and indirect hydrogen bonds, hydrophobic interactions, and electrostatic forces contribute to the interaction network between the S7P CTD and the RPRD1A CID.

Of note, CTD residue S5b was deeply buried in a negatively charged environment inside the channel and formed hydrogen bonds with Q20 (Fig. 1c-e). A phosphate group at this position would result in steric hindrance and charge repulsion, thus accounting for disruption of CTD binding by S5P. Conversely, the two S7 positions at the channel entrance and exit likely enable flexibility for S7 modification and substitution by different amino acids, explaining CID-binding to CTD repeats with K instead of S at position 7 (8/31 non-consensus repeats).

To address whether RPRD CIDs similarly recognize an S2P-containing CTD, we crystallized the RPRD1B CID with the 19-mer S2P CTD peptide SPSYSPTSPSYSPTSPSYS (S2P underlined) (Fig. 2a). This 1.85 Å resolution structure exhibited intra-homodimeric swapping of two carboxy-terminal helices, a phenomenon known as domain swapping[44], but the overall architecture of the CID surface was similar to the non-swapped RPRD1A CID. Thus, folding of the RPRD1B CID and contact residues between the S2P CTD peptide and the CID (Fig. 2a,b) were highly similar to those observed in the the RPRD1A CID-S7P CTD complex (Fig. 1d,e). Importantly, the highly conserved residue R106 formed two hydrogen-bonds with the S2bP phosphate group, explaining the enhanced affinity over the unmodified CTD peptide (Table 1). Other differences included N69 forming water-mediated hydrogen bonds with the S7b hydroxyl group, and N18 not making contact with S7a. The domain-swapped structure is likely stabilized by a disulfide bond involving C100 in both RPRD1B polypeptides (Fig. 2a). Whether the domain-swapped structure actually forms *in vivo* is unclear.

We also determined a 1.85 Å resolution structure of the RPRD1B CID in complex with the 19-mer UnM CTD peptide SPSYSPTSPSYSPTSPSYS. The folding and binding mode in this structure were also similar to those of the RPRD1A-S7P CTD and RPRD1B-S2P CTD complexes, but lacked phosphate-specific interactions and N69 interactions with the CTD (Fig. 2c,d). Unlike the β-turn CTD conformation in complexes with Pcf11 and Rtt103 (refs. 42,43), the decamer CTD peptides in all three RPRD CID complexes exhibited similar extended conformations (Fig. 2e,f).

Collectively, our structures with both unmodified and phosphorylated CTD peptides showed a conserved CID-CTD interface, with phosphorylation of S2 and S7 enhancing the interactions. The RPRD2 CID apo-structure at 1.8 Å resolution was also similar to those of the RPRD1A and RPRD1B CIDs (Supplementary Fig. 1), suggesting it might interact similarly with CTD peptides.

The influence of RPRD CID contact residues on CTD binding

To verify the interaction surface between the RPRD CIDs and CTD peptides, various CTD-interacting and evolutionarily conserved residues in the CIDs (Fig. 3a) were mutated and their effects on binding examined by ITC. Mutating R114 or D65 to alanine in the RPRD1A CID diminished binding to all three tested peptides (i.e. S2P, S7P, and UnM), whereas N18A, Q20A and R72A mutations had little effect (Fig. 3b). Similarly, an RPRD1B CID R114A mutation abolished binding to both S2P and S7P CTDs, whereas Q20 and R72 mutations had little effect (Fig. 3c). As expected, the RPRD1A R106A mutation strongly reduced binding (~27-fold) to an S2P peptide, while having little effect on binding to S7P and UnM peptides (Fig. 3b). These results confirmed the conserved CTD recognition modes of the RPRD1A and RPRD1B CIDs in our crystal structures, where R114 and D65 are most important for interaction independent of phosphorylation state, and R106 is important for specific interaction with an S2P CTD peptide.

To examine effects of CID mutations on interaction with RNAPII *in vivo*, extracts from HEK293 cells stably expressing wild type and mutant RPRD1B proteins with VAP (versatile affinity purification)-tags (containing 3xFlag, 6xHis, and Strep) were immunoprecipitated with antibodies recognizing the Flag epitopes, total RNAPII, or various RNAPII phosphoisoforms, respectively. Western blotting revealed that, consistent with the ITC results, the R114A mutation abolished RPRD1B binding to total RNAPII (N20 antibody in Fig. 3d) and various phosphoisoforms, whereas the R72A and Q20A mutations had no such effects (Fig. 3d). Taken together, these data revealed a molecular basis for S5P interference and an intrinsic preference of the RPRD CIDs for the S2P CTD phosphoisoform.

Dimerization of RPRDs

The RPRD1A and RPRD1B C-terminal regions (and residues 189-317 of RPRD2) are predicted to contain coiled-coil domains (Fig. 1a), previously known to mediate protein oligomerization[45]. Indeed, size exclusion chromatography eluted full length RPRD1A and RPRD1B, as well as their coiled-coil domains, as apparent multimers, while their CIDs eluted as monomers (Supplementary Fig. 2a). Similarly, full length RPRD1A and RPRD1B behaved like multimers in dynamic light scattering (DLS) analysis, while the three RPRD CIDs behaved like monomers (Supplementary Fig. 2b). Because size exclusion chromatography and DLS over-estimate asymmetric protein sizes, we used analytical ultracentrifugation (AUC) to estimate molecular weights unbiased by protein geometry. The ratios of deduced molecular weights obtained from sedimentation equilibrium to the predicted monomer ones were 2.4, 2.2 and 1.0 for full length RPRD1A, its coiled-coil domain and CID, respectively (Supplementary Table 2). Similarly, ratios were 2.0, 1.9 and 1.0 for full length RPRD1B, its coiled-coil domain and CID, respectively (Supplementary Table 2).

Because RPRD1A and RPRD1B behaved similarly in the above experiments, we further verified these observations focusing on RPRD1B. Sedimentation velocity revealed that most full-length RPRD1B and its coiled-coil domain exist as dimers (82.7% and 89.6%, respectively), while its CID was almost entirely monomeric (94.0%) (Supplementary Table

3). Furthermore, upon crosslinking with a relatively low concentration (30-300 μM) of suberic acid-bis-(3-sulfo-N-hydroxysuccinimide ester) (BS3), substantial amounts of intact RPRD1B and its coiled-coil domain migrated in SDS-polyacrylamide gels at molecular masses approximately equivalent to dimers, while its CID migrated at the position equivalent to monomers until a higher BS3 concentration (300-3000 μM) was used (Supplementary Fig. 2c). Collectively, these data demonstrated that RPRDs behave mainly as dimers with their coiled-coil domains as the dimerization modules.

To better understand the mechanism of dimerization, we obtained a 2.2 Å structure of the RPRD1B coiled-coil domain (residues L176-P301). This domain crystallized as a homodimer with a head-to-tail association mode (Fig. 4a). Each of the two polypeptides contained two major helices (α1, α2) with the two α2 helices being extensively intertwined (Fig. 4a). N-terminal regions (L176 to A191) of the two α1 helices contacted each other in an anti-parallel mode, packed against the central region of the two α2 helices and located at the same side of the homodimer (Fig. 4a). This fold potentially serves as a pad for the ipsilateral presentation of two N-terminal CIDs. Because of the high sequence identity of the RPRD1A and RPRD1B coiled-coil domains (64.3%) and their similar biochemical and biophysical behaviours, we predict the RPRD1A coiled-coil domain would fold in the same way.

Homodimerization of the RPRDs should enhance their binding to dual RPRD-binding CTD peptides. To address this issue, S2P peptides with either one (S2-5P CTD) or two (S2-5-2P CTD) CID-binding sites were used in ITC experiments. The affinities of full length RPRD1A and RPRD1B for peptides with two CID-binding sites were approximately 6-fold greater than those for peptides with one binding site, whereas no significant difference was observed for their CIDs (Fig. 4b). Again, S5P in the neighbouring region had no effect on affinity (Fig. 4b). RPRD dimer-enhanced affinities may help account for the substantial co-purification of the RPRDs with RNAPII[41].

Given the high sequence identity of the RPRD1A and RPRD1B coiled-coil domains, and the substantial reciprocal co-purification of the two proteins[41], we thought they may form heterodimers. To test this possibility, we first asked whether co-precipitation of RPRD1A and RPRD1B requires RNAPII. When cells were grown in the presence of α-amanitin to eliminate endogenous RNAPII[46], as confirmed by western blotting using N20 antibodies, VAP-tagged RPRD1A and RPRD1B co-precipitated with each other, but not GFP (Fig. 4c). In addition, VAP-tagged RPRD1B-R114A, a mutant that does not bind RNAPII (Fig. 3d), also co-precipitated with RPRD1A (Fig. 4d). These results indicated an RNAPII-independent interaction between RPRD1A and RPRD1B. To determine whether the interaction is direct, we carried out glutathione-S-transferase (GST) pull-down experiments and found that recombinant GST-RPRD1A pulled down full-length RPRD1B and its coiled-coil domain, but not its CID (Fig. 4e, Supplementary Fig. 3), suggesting heterodimerization via their coiled-coil domains.

RPRD1A and RPRD1B mediate RPAP2 association with RNAPII

Given that RPAP2 does not directly bind the S5P CTD[33], we thought other factors might be required for RPAP2 binding to RNAPII. The co-purification of RPRD1A and RPRD1B with

RPAP2[41] suggested that RPRDs might be such factors. To test this possibility, we first asked whether association of RPAP2 with RPRDs *in vivo* requires RNAPII. VAP-tagged RPRD1B co-precipitated with RPAP2 equally well after cells were grown in the presence or absence of α-amanitin (Fig. 5a**),** indicating an RNAPII-independent interaction between RPAP2 and RPRD1B. VAP-tagged RPRD1A also co-precipitated RPAP2 after cells were grown in the presence of α-amanitin, although to a lesser extent (Fig. 5a**)**, indicating most RPRD1A might not be associated directly with RPAP2 and would only co-precipitate it indirectly via RNAPII and RPRD1B. Next, we found that recombinant GST-RPAP2 pulled down both RPRD1A and RPRD1B (Fig. 5b), suggesting a direct interaction. We also found that recombinant full length RPRD1A and RPRD1B, but not their CIDs or coiled-coil domains, interacted weakly with recombinant RPAP2 in ITC experiments (Supplementary Table 4). Collectively, these data demonstrated that RPRD1A and RPRD1B bind directly to RPAP2.

This direct interaction also suggested RPRDs might be required for RPAP2 binding to RNAPII *in vivo*. To address this issue, we carried out co-immunoprecipation with extracts from HEK293 cells depleted of an RPRD protein. Since RPRD1A and RPRD1B form heterodimers, knocking down either one seemed likely to disrupt their function. Indeed, knocking down RPRD1A by stable expression of lentiviral-encoded shRNAs, or by transient expression of siRNA (Z.N. and X.G., unpublished data), substantially reduced the RPAP2 co-precipitated by antibodies recognizing either S2P or S5P on the RNAPII CTD, whereas the amount of precipitated RNAPII remained unchanged (Fig. 5c). Conversely, and in contrast, RPRD1A was co-precipitated by anti-S5P RNAPII CTD antibodies equally well after knocking down GFP or RPAP2 (Fig. 5d), indicating RPRDs-RNAPII association does not need RPAP2. Chromatin immunoprecipitation (ChIP) carried out in HeLa cell extracts indicated that RPAP2 occupies various promoters that we tested , but not a *U2 snRNA* untranscribed region, and siRNA-mediated knock-down of RPRD1A almost abolished RPAP2 association with various tested promoters, while RNAPII levels were reduced only by up to 50% at those promoters (Fig. 5e). Disrupting RPAP2 recruitment by knocking down RPRD1A alone further supported the notion that RPRD1A and RPRD1B are functional as heterodimers. Taken together, these results demonstrated that RPRDs are required for RPAP2 association with RNAPII *in vivo*.

RPRDs stimulate CTD S5P dephosphorylation by RPAP2 *in vitro*

The Murphy and Tong groups have reported different results regarding the ability of RPAP2 to act as a phosphatase *in vitro*[33,39]. To address this controversial issue, we incubated the conserved and functional RPAP2 fragment, RPAP2 (1-334)[33], with the phosphatase substrate 6,8-difluoro-4-methylumbelliferyl phosphate (DiFMUP). RPAP2 (1-334), but neither RPRD1A nor RPRD1B, dephosphorylated DiFMUP in a time- and dose-dependent manner (Fig. 6a). Similar results were obtained for full length RPAP2 (G.O.H and A.L.M., unpublished data). RPAP2 activity was not affected by addition of RPRD1A or RPRD1B but was fully or partially inhibited by the phosphatase inhibitor vanadate and several metals (Supplementary Fig. 4a-c). No measurable phosphatase activity was detected using the canonical phosphatase substrate para-nitrophenylphosphate (pNPP) (G.O.H and A.L.M., unpublished data), indicating that RPAP2 phosphatase activity is substrate-specific.

To determine whether RPAP2 can dephosphorylate CTD substrates *in vitro*, we used an enzyme-linked immunosorbent assay (ELISA), where the products of phosphatase reactions containing biotinylated CTD peptides were adsorbed to streptavidin-coated wells in ELISA plates. Phosphorylation levels were assessed by antibodies that specifically recognize particular CTD phosphoisoforms. In agreement with previous observations[39], no measurable phosphatase activity was detected on S5P- or S2P-containing CTD peptides in reactions with RPAP2 alone (Fig. 6b-d).

Given that RPRDs tether RPAP2 to RNAPII (Fig. 5), and over-expression of RPRDs decreases S5P at various promoters[41], we hypothesized that RPRDs might stimulate RPAP2 activity on CTD substrates. Indeed, ELISA experiments showed that recombinant RPAP2 in conjunction with recombinant RPRD1A, but not its CID or coiled-coil domain, dephosphorylated S5P in a time- (Fig. 6b) and dose-dependent manner (Fig. 6c, left). S2P was dephosphorylated to a much lesser extent (Fig. 6c). Similar results were obtained for RPRD1B, but not for its R114A mutant that fails to bind the CTD (Fig. 6d left). RPRD1A-dependent S5 dephosphorylation by RPAP2 was also inhibited by vanadate and several metals (Supplementary Fig. 4d), suggesting similar dephosphorylation mechanisms are operating on DiFMUP and CTD substrates. The CTD dephosphorylation rate was low (Fig. 6b), probably because the RPAP2 concentration in our reactions (3 μM) was much lower than the K_d of the RPAP2-RPRD interaction (447-478 μM, Supplementary Table 4). Post-translational modifications or other factors may strengthen this interaction *in vivo* or, alternatively, compartmentalization in transcription factories may increase the local concentrations of RNAPII, RPRDs and RPAP2.

Interestingly, and consistent with RPRD dimerization, a CTD peptide containing two S2P CID-binding sites bracketing an S5P residue located in the intervening sequence (S2-5-2-P CTD) served as a substrate in reactions containing RPRD1A or RPRD1B (Fig. 6c,d, left), whereas S5P located outside two S2P-containing CID-binding sites (S2-2-5P CTD) did not (Fig. 6c,d, right).

We then asked whether S7P-CTD can also stimulate the RPRD-dependent CTD phosphatase activity of RPAP2. Since S7P next to S5P in a dual RPRD CID-bound S7P CTD peptide (S77-5-77P CTD) prevented antibody recognition of S5P in ELISA assays (Supplementary Fig. 4e), we used a peptide with an unmodified S7 adjacent to S5P (S77-5-7P CTD) for the assay. RPRD1A or RPRD1B stimulated RPAP2 activity towards S5P, but not S7P (Supplementary Fig. 4f), although less than was the case for S2-5-2P CTD peptides (Fig. 6c,d), perhaps because of the lower affinities of RPRD1A and RPRD1B for the S77-5-7P CTD peptide (20.0 ± 4 μM and 12 ± 2 μM, respectively) than for the S77-5-77P CTD peptide (7.6 ± 1.5 μM and 2.7 ± 0.5 μM, respectively), or the S2-5-2P CTD peptide (4.3 ± 2.4 μM and 1.7 ± 1.3 μM, respectively).

Next, we tested whether RPAP2 is needed to dephosphorylate S5P *in vivo*. ChIP experiments showed that knocking down RPAP2 in HEK293 cells with two independent shRNAs increased the levels of S5P, relative to RNAPII, at several tested promoter regions (Supplementary Fig. 5a), even though no obvious global change of either S5P or S2P phosphorylation could be detected by western blotting (Supplementary Fig. 5b).

Consistently, siRNA-mediated knock-down of RPRD1A in HeLa cells increased the S5P levels near the *ACTB* (*β-ACTIN*) promoter and other promoter regions (Supplementary Fig. 5c). These results were in line with our previous observation that over-expression of RPRDs leads to a decrease of S5P at various promoter regions[41].

In summary, multiple S2P- and/or S7P-CTD repeats interact with the two CIDs of homo- and/or hetero-dimeric RPRD1A and/or RPRD1B, which in turn recruit RPAP2 to dephosphorylate the S5P specifically located in the intervening region (Fig. 7a,b). Thus, RPRDs serve as scaffolds that recruit RPAP2 and organize the CTD into an appropriate conformation for the specific dephosphorylation of S5P by its substrate-selective phosphatase activity (Fig. 7a,b).

DISCUSSION

We have shown that the RPRD1A and RPRD1B coiled-coil domains enable them to homo- and/or heterodimerize and present a two-CID scaffold, most likely on the same side of the coiled-coil domain, for interaction with RNAPII CTD repeats (Fig. 7a.b). The consensus CTD peptide bound in the channel of each RPRD CID is the decamer sequence PSYSPTSPSY, rather than a simple CTD heptapeptide repeat (underlined). Thus, the consensus CTD sequence bound by the two CIDs in an RPRD dimer is the tetracosamer (24-mer) sequence PSYSPTSPSYSPTSPSYSPTSPSY, which contains two CID binding sites (underlined) that bracket an intervening tetramer sequence SPTS (S5 underlined). Although RPRD1A and RPRD1B homodimers exhibit different affinities for non-canonical CTD repeats, such differences might not be significant for RPRD1A-RPRD1B heterodimers. Because RPRD CIDs bind both canonical and non-canonical CTD repeats (Table 1), and each RPRD dimer occupies more than 3 heptapeptide repeats, each mammalian 52-repeat RNAPII CTD could theoretically accommodate up to 13 such structures (Fig. 7c). Interaction of CIDs with 10 rather than 7 CTD amino acids probably explains why the functional unit of the CTD appears to contain two heptapeptide repeats[47].

Dimerization of the RPRDs enhances CTD binding, which could be explained by the synergistic effect of multiple CID-CTD interactions. A similar synergistic effect was observed for binding of yeast Pcf11 and Rtt103 to S2P CTD, where these two proteins bind cooperatively with each other, resulting in enhanced CID-CTD interaction[43]. Pcf11 is also a subunit of CF1A in which two other subunits, Rna14 and Rna15, form a heterotetramer that also interacts with the CTD[48]. Thus, the phenomenon of CID cooperativity in CTD recognition and transcription regulation is likely conserved between yeast and human.

Our data indicates that RPAP2 is a substrate selective phosphatase. In CTD peptides, the SPTS tetramer bracketed by a pair of RPRD dimer-bound S2P/S7P CTD repeats is likely the appropriate substrate for RPAP2 (Fig. 7a,b). Therefore, a precise CTD conformation oriented by RPRD dimers presents an ideal CTD arrangement to RPAP2. We designate the CTD-RPRD dimer-RPAP2 complex as a “CTDsome” (Fig. 7a,b). This “CTDsome” model describes a novel organization of the CTD repeats, reveals how this helps the substrate selective RPAP2 phosphatase recognize its CTD target and precisely delineates a

mechanism for regulating S5P that uniquely requires “CTD code” crosstalk from S2P and S7P to S5P.

During transcription initiation, TFIIH phosphorylates both S5 and S7 on the CTD[11-13,15]. Subsequently, S7P facilitates association of p-TEFb with the CTD[49], leading to phosphorylation of S2. In the transition from initiation to elongation, RPRD dimers bound to pairs of S2P and/or S7P-CTD decamers would recruit RPAP2 to dephosphorylate S5P. The timing of this cascade is highly similar to that of Rtr1-dependent S5P dephosphorylation that occurs in the transition from initiation to elongation in yeast[32].

RPRDs are capable of associating with transcription complexes not only in promoter regions but also in gene bodies and near mRNA 3′-ends[40,41]. This makes it likely that RPRD scaffolds could have additional roles in transcriptional regulation other than S5 dephosphorylation. Given that RPRD1B is an oncoprotein over-expressed in more than 80% of human tumors[40], deciphering the particular activity of RPRD1B involved in tumorigenesis is an important future goal. Importantly, the high resolution crystal structure of the RPRD1B CID in complex with S2P-CTD may be useful for the design and development of anti-cancer compounds that regulate the CTD binding activity of RPRD1B.

ONLINE METHODS

CTD Peptides

CTD peptides with a purity of 95% were purchased from PEPTIDE2 Inc.

Antibodies

β-ACTIN (cat. no. A-5441), Flag (F1804) (cat. no. F3165), and RPRD1B (cat. no. SAB1102247) are from Sigma-Aldrich. Gal4 (cat. no. 06-262) is from Millipore. GAPDH (cat. no. 398600) and IgG (cat. no. 10500C) are from Invitrogen. His (cat. no.71841) is from Novagen. RPAP2 (cat. no.17401-1-AP) is from ProteinTech Group Inc. RNAPII (N-20) (cat. no. SC-899) and RPRD1A (P15RS) (C-18, cat. no. SC-85089) are from Santa Cruz. S2P-RNAPII (cat. no. A300-654A) is from Bethyl Laboratories Inc. S5P-RNAPII (cat.no. ab5131) is from Abcam. S2PRNAPII (3E10), S5P-RNAPII (3E8), S7P-RNAPII (4E12) are from D. Eick laboratory. All antibodies were used with dilution of 1:5000 in 5% BSA in western blotting and ELISAs, and 2 μg in immunoprecipitation-western blotting and ChIP experiments. All the commercial antibodies have been validated for the relevant species and applications, as shown on the manufacturers’ websites. The antibodies from D. Eick laboratory have been validated as shown in the group’s publications[14].

Expression constructs, protein expression and purification

Full length RPRD1A and RPRD1B, the RPRD protein CIDs (RPRD1A S2-K137, RPRD1B S2-P135, and RPRD2 S15-K161) and coiled-coil domains (RPRD1A K134-D312, RPRD1B P137-D326 and RPRD1B 171-304), and full length RPAP2and its N-terminal segments (1-334) were cloned into the vectors pET15-MHL, pET28-MHL, pET28-Lic, pET28-GST-Lic, or pGEX-6P-1 by using the Infusion kit (In-FusionTM Dry-down Mix, Clontech S3533) according to the manufacturer’s instructions. Mutated clones were generated by

using infusion-mediated mutagenesis with the In-FusionTM Dry-down Mix (Clontech). The identities of all plasmid constructs were verified by sequence analysis. Plasmids were transformed into BL21 Star One Shot *E.coli* (Invitrogen), and protein expression was induced by the addition of 1 mM IPTG to the culture medium when the bacteria had reached an optical density (OD_{600}) of 0.4. Protein expression was allowed for 12 h at 14°C, after which the bacteria were pelleted and frozen at −80°C. The pellet was thawed on ice and the bacteria were lysed in binding buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM Imidazole, 5 % Glycerol, 1mM Protease inhibitor, 1 mM DTT and 6.25 units/ml of Benzonase). Undissolved debris was pelleted at 15000 rpm for 1 hour at 4°C. The 6xHis-tagged recombinant proteins were incubated with Ni-NTA resin (Qiagen). The beads were then washed four times with wash buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 30 mM Imidazole, and 5 % Glycerol) before elution with elution buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 250 mM Imidazole, and 5 % Glycerol). The 6xHis tag was removed by tobacco etch virus (TEV) protease before further purification by Superdex 75 or 200 gel filtration. GST-fused proteins were purified from *E. coli* by using GST resin (Novagen). All proteins were concentrated to 10 to 50 mg/ml in a buffer containing 10 mM HEPES (pH 7.5), 150 mM NaCl and stored at −80°C. The selenomethionyl protein sample was produced in *E. coli* grown in defined medium supplemented with selenomethionine using the M9 SeMET kit (Medicilon, Cat.MD045004). The purification procedure was the same as for the native protein.

Protein crystallization, data collection, structure determination and refinement

Crystals of the RPRD protein CIDs, the CID-CTD peptide complexes, and the coiled-coil domain were obtained with the sitting-drop vapour-diffusion method at 18° C in the buffers listed in Supplementary Table 1. CTD peptides were mixed with proteins in a molar ratio of 3:1. The crystals were cryo-protected by the reservoir solution, supplemented with 10% (v/v) glycerol, as indicated in Supplementary Table 1, and flash-frozen in liquid nitrogen.

Diffraction data were collected as described in Table 2 and Supplementary Table 1. The RPRD2-CID was solved by single wavelength anomalous diffraction (SAD)[50] with praseodymium derivative diffraction data collected on a rotating copper anode source. The other CID structures were derived from this model by molecular replacement. The structure of the RPRD1B C-terminal coiled-coil domain was solved by SAD with a selenomethionyl derivative[51]. For each model and in several iterations, manual rebuilding in COOT[52] was followed by atomic coordinate and temperature factor refinement (programs for final refinement listed in Supplementary Table 1) and geometry validation in MOLPROBITY[53]. Selected geometry restraints for model refinement were prepared with JLigand[54].

Cell cultures, lentivirus infections of VAP-tagged RPRD proteins and shRNAs, and siRNA transfection

HEK293 and HeLa cells were cultured in Dulbecco's modified Eagle's medium with 10% fetal bovine serum and antibiotics. Lentiviruses were produced and used to infect HEK293 cells at a multiplicity of infection <1 as described previously[55]. Each open reading frame (ORF) was cloned into a lentiviral expression vector so as to attach a VAP-tag (containing a 3xFLag-6xHis-Strep triple tag) as described previously[56,57] and sequence verified.

Lentivirus-encoded VAP-tagged ORFs and shRNAs were transduced into HEK293 cells as described[56,57]. Transduced cells were selected with puromycin (Sigma) at a concentration of 2 μg/ml for a minimum of 48 hours. A Gateway-compatible entry clone for RPRD1B (OHS1770-9385393) was obtained from the human ORFeome library (Open Biosystems). The following shRNA (obtained from the laboratory of Jason Moffat, University of Toronto) sequences were used for lentivirus-mediated knockdowns: shGFP: CGACCACATGAAGCAGCACGA; shRPRD1A: CGGCAGAAATAGATGATAGAA; shRPAP2-1: GCCAAGTTACGAGAATTTGAA; shRPAP2-2: GCAAGACTTTGTTTCCTCCAT.

Three μg of siRNA were transfected with 3 μl Lipofectamine2000 (Invitrogen) into each well of a 6-well plate for 24 hours and followed by transferring the cells to a 10 cm dish. Twenty-four hours later, the same amount of siRNA was transfected again with Lipofectamine2000. The pool of the following two siRNA (Purchased from Sigma-Aldrich) was used in transfection: siRPRD1A-1: GGAAGAAAGGUCUGUUUAUdTdT, siRPRD1A-2: GACUCUAGAUCUCGUUAGAdTdT. Buffer was used as a negative control.

Analytical ultracentrifugation (AUC)

For the sedimentation equilibrium experiments, the proteins at various concentrations (OD_{280} of 1, 0.5 and 0.25) in buffer containing 10 mM HEPES (pH7.5) and 150 mM NaCl were centrifuged at 13000 and 15000 rpm at 4 °C in a Beckman Optima XL-A analytical ultracentrifuge using an An-60 Ti rotor. Absorbance at 280 nm was monitored. Data analysis was done with the Origin MicroCal XL-A/CL-I Data Analysis Software Package version 4.0.

For sedimentation velocity AUC, proteins at OD of 1 were centrifuged at 55000 rpm at 4° 280 C in a Beckman Optima Model XL-A analytical ultracentrifuge equipped with an An-60 Ti rotor. The sedimentation data were fitted to a continuous distribution model c(s) using SEDFIT[58]. The sedimentation coefficients obtained from the fitting were corrected to the density (ρ) and viscosity (η) of the buffer at 4°C to obtain *s20, w*.

Western blots

Cells were subjected to three freeze-thaw cycles in high salt lysis buffer [10 mM Tris-HCl, pH 7.9, 10% glycerol, 420 mM NaCl, 0.1% Nonidet P-40, 2 mM EDTA, 2 mM DTT, 10 mM NaF, 0.25 mM Na_3VO_4, and 1x protease inhibitor mixture (Sigma)], followed by centrifugation at 14,000 rpm for 1 hour at 4 °C to remove insoluble materials. 20 to 100 μg proteins were separated by electrophoresis on a Tris 4–20 % SDS-polyacrylamide gel (BioRad) and transferred to nitrocellulose or PVDF membranes. Transferred samples were immunoblotted with primary antibodies, followed by incubation with horseradish peroxidase-conjugated goat anti-mouse or goat anti-rabbit secondary antibodies (Jackson Lab). Western blot detection was performed using enhanced chemiluminescence (GE Healthcare).

Immunoprecipitation

HEK293 cell lysates were incubated with 2 μg of antibody overnight at 4°C followed by adding 20 μl of Protein G beads (Sigma) for an additional 4 hours incubation. After washing with low salt buffer (10 mM Tris-HCl, pH 7.9, 100 mM NaCl, 0.1% Nonidet P-40), associated proteins were eluted into protein loading buffer and separated by Tris 4–20 % SDS-polyacrylamide (BioRad), followed by western blot analysis.

Chromatin immunoprecipitation (ChIP)

HEK293 (or HeLa cells, Fig. 5e and Supplementary Fig. 5c) cells were cross-linked with 1% formaldehyde (Sigma) at room temperature for 10 minutes, washed twice with ice-cold PBS, collected in 1 ml of PBS and centrifuged for 5 minutes at 5000 rpm. Cells were resuspended in 1 ml of lysis buffer [1% SDS, 10 mM EDTA, and 50 mM Tris-HCl (pH 8)] plus proteinase inhibitors (aprotinin, leupeptin, and pepstatin), incubated on ice for 10 min and sonicated to an average DNA size of approximately 500 bp. Chromatin was pre-cleared with 25 μl of Staph A (Calbiochem, Cat# 507862) at 4°C for 15 minutes (or 10 μl of Dynabeads at 4°C for 30 minutes, Fig. 5e and Supplementary Fig. 5c). A 100 μl aliquot of sonicated chromatin was immunoprecipitated (IP) with 2 μg antibodies overnight at 4°C. IP samples were centrifuged at 13,200 rpm and supernatant was incubated with 10 μl of Staph A at room temperature for 15 minutes (or 15 μl of Dynabeads at 4°C for 1 hour, Fig. 5e and Supplementary Fig. 5c). Precipitates were washed sequentially for 3 minutes in 1X dialysis buffer [2 mM EDTA, 50 mM Tris-HCl (pH 8), and 0.2% sarkosyl] twice, then IP wash buffer [1% Nonidet P-40, 100 mM Tris-HCl (pH 9), 500 mM LiCl 1% and deoxycholic acid] four times. Samples were extracted twice with 150 μl of elution buffer (1% SDS and 50 mM $NaHCO_3$), heated at 65°C overnight to reverse cross-links, and DNA fragments were purified with a QIAEX II Gel Extraction kit (catalog no. 20051). A 4-μl aliquot from a total of 200 μl was used in qPCR with the following primers: *RPRD1A* promoter forward: TGCTTCTTTGACAGAGTTTCCA, reverse: TGATGCCTGGGTTACTCTCA; *LEO1* promoter forward: ATACCCAGGCGAGAACAGGT, reverse: GACAGTGTCGCAAAGATTCG; *MED12* forward: TCTAGGCCAAAACCGAGCTA, reverse: GATGCAAACCAGGGAAGAAG; *CTR9* promoter forward: GGAGAGTCAGACGCCAGATG, reverse: GAAGCTTTTGTTGGCGTGTC; *ACTB* (*β-ACTIN)* forward: GGGCAACCGGCGGGGTCTTT, reverse: ACGCAGTTAGCGCCCAAAGG; *U2 snRNA* untranscribed region forward: CAGCTGTGGCTGGACAGGTTGGACC, reverse: CAGGGTTCCTGCACATTCACAGTTA. The yields of ChIP DNA were quantified with the SYBR Green kit (Applied Biosystems) using the 7300 Real Time PCR System (Applied Biosystems), or a Rotorgene RG-3000 (Corbett Research) in a 5 μl volume (or 10 μl volume, Fig. 5e and Supplementary Fig. 5c) in duplicates. PCRs consisting of 40 cycles of 95°C for 15 seconds and 55°C for 30 seconds were performed. Qt values were compared with a standard curve, the copy number was calculated, the amount of DNA precipitated by an irrelevant GAL4 or IgG antibody subtracted, and the percent ChIP DNA relative to input chromatin was calculated.

Isothermal titration calorimetry (ITC) measurements

ITC measurements were recorded at 25°C with a VP-ITC microcalorimeter (MicroCal Inc.). Various peptides were dissolved and dialyzed into the same buffer as that of the protein. 10 μl of peptide solution or proteins (500 μM) were injected into a sample cell containing 25 μM protein in 10 mM HEPES (pH 7.5) and 150 mM NaCl. A total of 25 injections were performed with a spacing of 180 seconds and a reference power of 15 μcal/second. Binding isotherms were plotted and analyzed with Origin Software (MicroCal Inc.). The ITC measurements were fitted to a one-site binding model.

Dynamic light scattering (DLS)

DLS experiments were performed using a DynaPro Titan instrument (Wyatt Technologies Corporation) equipped with a plate reader. 50 μl of 1 mg/ml proteins in buffer containing 10 mM HEPES (pH 7.5) and 150 mM NaCl were placed in 384 well optical bottom NUNC plates. The instrument laser power was adjusted to approximately 1×10^6 counts/second and data was collected with acquisition time of 10 seconds and averaged for 10 acquisitions. Analysis of the data was performed using Dynamics software

In vitro DiFMUP phosphatase assay

Recombinant proteins with or without the inhibitots were incubated with 10 μM 6,8-difluoro-4-methylumbelliferyl phosphate (DiFMUP, Molecular Probes) at 37°C in black, flat bottom, polystyrol 96-well plates in 100 μl reaction buffer (50 mM 3, 3-dimethylglutaric acid pH6.5 and 150 mM $NaCl_2$), followed by measuring the fluorescence every 20 min. for 2 hours using SpectraMax M5 Microplate reader. The relative fluorescence was compared against a standard curve of DiFMUP to determine the amount of products produced. Recombinant GST was reconstituted in the reaction buffer and used as a negative control throughout. Initial rates were determined by plotting the amount of product produced versus time.

In vitro phosphatase enzyme-linked immunosorbent assay (ELISA)

25 pmol biotinylated CTD peptides were added into NeutriAvidin (Thermo Scientific, Prod#31000)-coated Maxisorp plates (Thermo Scientific, Prod#464718) overnight at 4°C. Recombinant proteins were incubated in 30 μl reactions in buffer containing 50 mM Tris-HCl pH 6.5, 10 mM MgCl, 20 mM KCl and 5 mM DTT at 37 °C. After wash with PBS plus 0.5% Tween-20, antibodies were added and incubated at room temperature for 1 hour, followed by incubation with horseradish peroxidase-conjugated goat anti-mouse or goat anti-rabbit secondary antibodies (Jackson Lab), the addition of TMB peroxide substrate (Thermo Scientific Prod# 34021) and the measurement of absorbance signals at 450 nm.

Size exclusion chromatography

Size exclusion chromatography was carried out using a Superdex 75 HR 26/60 or Hiload16/60-Superdex 200 (GE Healthcare) column attached to an AKTA FPLC (GE Healthcare). The column was equilibrated with buffer containing10 mM HEPES, pH 7.5, 150 mM NaCl. Gel Filtration Standard (Bio-Rad, Cat. 151-1901) was used. All experiments were performed at 4°C. Proteins were detected using absorbance measured at 280 nm.

GST pull-down

GST-and His-tagged proteins expressed in *E. coli*. were sonicated and incubated for overnight at 4°C, followed by pull-down using glutathione beads. The beads were then washed four time with suspension buffer (20 mM Tris, pH 7.5, 400 mM NaCl, and 5% Glycerol), followed by the elution with elution buffer (20 mM Tris, pH 7.5, 400 mM NaCl, and 10 mM reduced Glutathione). Eluted materials were resolved on SDS-polyacrylamide gel and proteins were deteced by anti-His antibodies or anti-RPRD1A or anti-RPRD1B antibodies.

Protein cross-linking

Suberic acid-bis-(3-sulfo-N-hydroxysuccinimide ester) (BS3, ProteoChem Inc., Denver, US) was freshly dissolved in 25 mM sodium phosphate (pH 7.4) at a concentration of 50 mM. BS3 solution was incubated with proteins in 50 μl at room temperature for 1 hour, followed by quenching the unreacted BS3 with 40 mM Tris for 15 minutes at room temperature. Proteins were separated by Tris 4–20 % SDS-polyacrylamide gel (BioRad) and stained by the Coomassie blue reagent Instant Blue (Expedeon).

Statistical analysis

The p-values in ChIP, ELISA, ITC and DiFMUP vanadate inhibitor assays were analyzed by a two-tailed Student's *t*-test. The data analysis in DiFMUP metal inhibitor assays was performed using GraphPad Prism 6.0 for Windows, in which the p-value was determined using a one-way ANOVA Tukey's Multiple Comparison test.

Original images of the gels used in this study can be found in Supplementary Data Set 1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank D. Y. Zhao, X. Wu, and K. Liu for technical assistances, and J. Moffat (University of Toronto) for providing the shRNAs. This work was supported by grants from the Canadian Institutes of Health Research (CIHR MOP-258357) (to J.F.G. and A.E.), the Ontario Research Fund (to J.F.G. and A.E.), the US National Institutes of Health (GM099714) (to A.L.M.), the Wellcome Trust [092483/Z/10/Z] and EPA (to S.M.), and the Canada Foundation for Innovation (CFI) (to J.F.G. and A.E.), and by a CIHR Postdoctoral Fellowship (to Z.N.), OSCI Studentship (to O.V.K.) and Ontario Graduate Scholarship (to J.B.O.). The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from AbbVie, Boehringer Ingelheim, CFI, CIHR, Genome Canada through the Ontario Genomics Institute [OGI-055], GlaxoSmithKline, Janssen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda, and the Wellcome Trust [092809/Z/10/Z] (to C.H.A. and J.M.).

REFERENCES

1. Allison LA, Moyle M, Shales M, Ingles CJ. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell. 1985; 42:599–610. [PubMed: 3896517]
2. Corden JL. Tails of RNA polymerase II. Trends Biochem Sci. 1990; 15:383–7. [PubMed: 2251729]
3. Egloff S, Dienstbier M, Murphy S. Updating the RNA polymerase CTD code: adding gene-specific layers. Trends Genet. 2012; 28:333–41. [PubMed: 22622228]

4. Singh N, et al. The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. Mol Cell. 2009; 36:255–66. [PubMed: 19854134]
5. Sims RJ 3rd, et al. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science. 2011; 332:99–103. [PubMed: 21454787]
6. Buratowski S. The CTD code. Nat Struct Biol. 2003; 10:679–80. [PubMed: 12942140]
7. Corden JL. Transcription. Seven ups the code. Science. 2007; 318:1735–6. [PubMed: 18079391]
8. Hsin JP, Sheth A, Manley JL. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3′ end processing. Science. 2011; 334:683–6. [PubMed: 22053051]
9. Krogan NJ, et al. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell. 2003; 11:721–9. [PubMed: 12667454]
10. Mayer A, et al. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science. 2012; 336:1723–5. [PubMed: 22745433]
11. Kim M, Suh H, Cho EJ, Buratowski S. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. J Biol Chem. 2009; 284:26421–6. [PubMed: 19679665]
12. Glover-Cutter K, et al. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Mol Cell Biol. 2009; 29:5455–64. [PubMed: 19667075]
13. Akhtar MS, et al. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol Cell. 2009; 34:387–93. [PubMed: 19450536]
14. Chapman RD, et al. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science. 2007; 318:1780–2. [PubMed: 18079404]
15. Serizawa H, et al. Association of Cdk-activating kinase subunits with transcription factor TFIIH. Nature. 1995; 374:280–2. [PubMed: 7885450]
16. Ghosh A, Shuman S, Lima CD. Structural insights to how mammalian capping enzyme reads the CTD code. Mol Cell. 2011; 43:299–310. [PubMed: 21683636]
17. McCracken S, et al. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. 1997; 11:3306–18. [PubMed: 9407024]
18. Cho EJ, Takagi T, Moore CR, Buratowski S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. 1997; 11:3319–26. [PubMed: 9407025]
19. Fabrega C, Shen V, Shuman S, Lima CD. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol Cell. 2003; 11:1549–61. [PubMed: 12820968]
20. Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell. 2003; 11:709–19. [PubMed: 12667453]
21. Krogan NJ, et al. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J Biol Chem. 2002; 277:10753–5. [PubMed: 11805083]
22. Briggs SD, et al. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev. 2001; 15:3286–95. [PubMed: 11751634]
23. Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol. 2008; 15:795–804. [PubMed: 18660819]
24. Mutskov VJ, Farrell CM, Wade PA, Wolffe AP, Felsenfeld G. The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. Genes Dev. 2002; 16:1540–54. [PubMed: 12080092]
25. Zeitlinger J, et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet. 2007; 39:1512–6. [PubMed: 17994019]

26. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007; 130:77–88. [PubMed: 17632057]
27. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008; 322:1845–8. [PubMed: 19056941]
28. Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes Dev. 2001; 15:3319–29. [PubMed: 11751637]
29. Zhou Q, Li T, Price DH. RNA polymerase II elongation control. Annu Rev Biochem. 2012; 81:119–43. [PubMed: 22404626]
30. Bartkowiak B, et al. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. 2010; 24:2303–16. [PubMed: 20952539]
31. Devaiah BN, et al. BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. Proc Natl Acad Sci U S A. 2012; 109:6927–32. [PubMed: 22509028]
32. Mosley AL, et al. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Mol Cell. 2009; 34:168–78. [PubMed: 19394294]
33. Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. Mol Cell. 2012; 45:111–22. [PubMed: 22137580]
34. Bataille AR, et al. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol Cell. 2012; 45:158–70. [PubMed: 22284676]
35. Hausmann S, Koiwa H, Krishnamurthy S, Hampsey M, Shuman S. Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. J Biol Chem. 2005; 280:37681–8. [PubMed: 16148005]
36. Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. Ssu72 Is an RNA polymerase II CTD phosphatase. Mol Cell. 2004; 14:387–94. [PubMed: 15125841]
37. Archambault J, et al. FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase IIO. J Biol Chem. 1998; 273:27593–601. [PubMed: 9765293]
38. Kobor MS, et al. An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae. Mol Cell. 1999; 4:55–62. [PubMed: 10445027]
39. Xiang K, Manley JL, Tong L. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. Nat Commun. 2012; 3:946. [PubMed: 22781759]
40. Lu D, et al. CREPT accelerates tumorigenesis by regulating the transcription of cell-cycle-related genes. Cancer Cell. 2012; 21:92–104. [PubMed: 22264791]
41. Ni Z, et al. Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B. Transcription. 2011; 2:237–42. [PubMed: 22231121]
42. Meinhart A, Cramer P. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature. 2004; 430:223–6. [PubMed: 15241417]
43. Lunde BM, et al. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol. 2010; 17:1195–201. [PubMed: 20818393]
44. Liu Y, Eisenberg D. 3D domain swapping: as domains continue to swap. Protein Sci. 2002; 11:1285–99. [PubMed: 12021428]
45. Nooren IM, Kaptein R, Sauer RT, Boelens R. The tetramerization domain of the Mnt repressor consists of two right-handed coiled coils. Nat Struct Biol. 1999; 6:755–9. [PubMed: 10426954]
46. Lindell TJ, Weinberg F, Morris PW, Roeder RG, Rutter WJ. Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. Science. 1970; 170:447–9. [PubMed: 4918258]
47. Stiller JW, Cook MS. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. Eukaryot Cell. 2004; 3:735–40. [PubMed: 15189994]

48. Noble CG, Walker PA, Calder LJ, Taylor IA. Rna14-Rna15 assembly mediates the RNA-binding capability of Saccharomyces cerevisiae cleavage factor IA. Nucleic Acids Res. 2004; 32:3364–75. [PubMed: 15215336]
49. Czudnochowski N, Bosken CA, Geyer M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nat Commun. 2012; 3:842. [PubMed: 22588304]
50. Wang BC. Resolution of phase ambiguity in macromolecular crystallography. Methods Enzymol. 1985; 115:90–112. [PubMed: 4079800]
51. Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. 1990; 9:1665–72. [PubMed: 2184035]
52. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
53. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010; 66:12–21. [PubMed: 20057044]
54. Lebedev AA, et al. JLigand: a graphical tool for the CCP4 template-restraint library. Acta Crystallogr D Biol Crystallogr. 2012; 68:431–40. [PubMed: 22505263]
55. Moffat J, et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell. 2006; 124:1283–98. [PubMed: 16564017]
56. Mak AB, et al. A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. Mol Cell Proteomics. 2010; 9:811–23. [PubMed: 20305087]
57. Ni Z, Olsen JB, Emili A, Greenblatt JF. Identification of mammalian protein complexes by lentiviral-based affinity purification and mass spectrometry. Methods Mol Biol. 2011; 781:31–45. [PubMed: 21877275]
58. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J. 2000; 78:1606–19. [PubMed: 10692345]

Europe PMC Funders Author Manuscripts

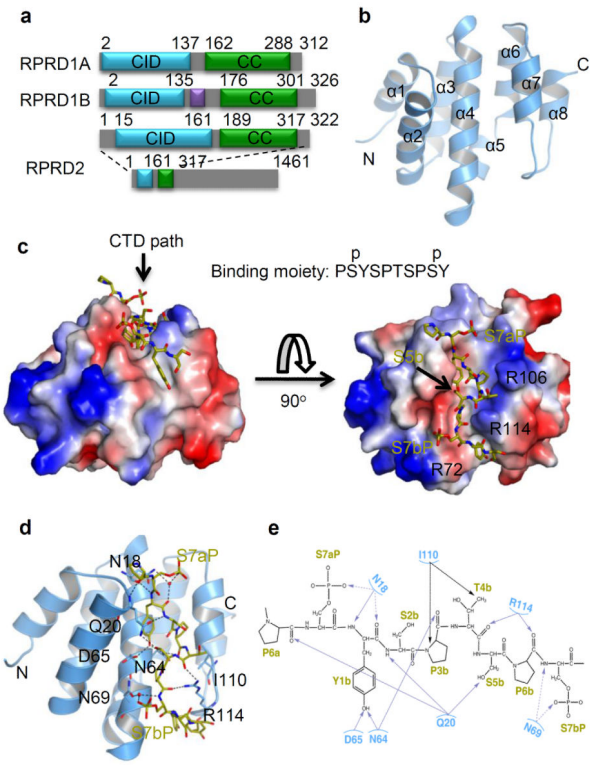

Figure 1.

Crystal structure of the RPRD1A CID with S7P CTD.

(**a**) Schematic of RPRD primary sequences. CID: CTD interaction domain. CC: Coiled-coil domain. Gray color: Predicted disordered sequence. Purple color: hypothetical ordered linker region of RPRD1B. (**b**) Crystal structure of RPRD1A CID. α: Alpha-helix. (**c**) Electrostatic surface representation of RPRD1A CID-S7P CTD complex. Yellow sticks: CTD peptide structure. Red: negative potential. Blue: positive potential. The two phosphate groups of the CTD peptide are labelled as S7aP and S7bP, respectively. R72, R106 and R114 in the CID and S5 of the CTD are labelled. (**d**) Detailed interactions of RPRD1A CID with S7P CTD. Yellow sticks: CTD peptide structure. Blue cartoon: CID structure. Residues involved in CTD binding are shown in stick model. Hydrogen bonds are indicated with dashed lines and water molecules are shown in red spheres. (**e**) Interaction diagram between RPRD1A CID and S7P CTD peptide. Interacting residues in the CID are contained within ovals. Direct hydrogen bonds, water mediated hydrogen bonds and hydrophobic interactions are indicated with purple solid, purple dash and black arrows, respectively.

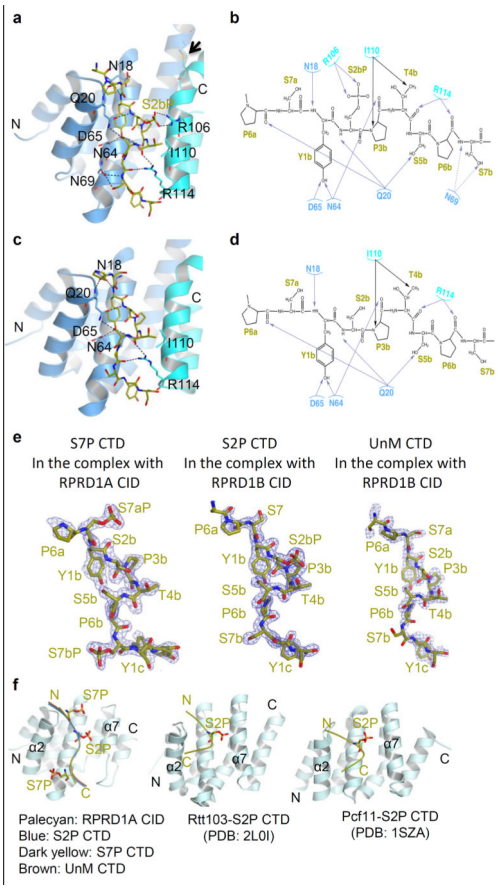

Figure 2.

Crystal structures of the RPRD1B CID with S2P and with UnM CTD.

(**a**) Detailed interactions between RPRD1B CID and S2P CTD. RPRD1B and CTD peptide are shown in the same way as in Fig. 1d, except that the two molecules of the RPRD1B CID in the swapped dimers are coloured in blue and cyan, respectively. An arrow indicates the disulfide bond formed between two RPRD1B CIDs. (**b**) Interaction diagram for the RPRD1B CID bound to the S2P CTD peptide. Proteins and peptide are colored the same as in Fig. 1e, except that residues from the two RPRD1B CID molecules in swapped dimers are shown in blue and cyan, respectively. The interactions between the CID and the CTD are depicted in the same way as in Fig. 1e. (**c**) Detailed interactions of RPRD1B CID with UnM CTD. Protein and CTD peptide are shown in the same way as in Fig. 1d and Fig. 2a. (**d**) Interaction diagram for the RPRD1B CID bound to the UnM CTD peptide. Proteins and peptide are colored the same as in Fig. 1e and Fig. 2b. (**e**) The |Fo|-|Fc| electron density maps of the three peptides (yellow sticks) in the indicated complexes are contoured at 3.0 δ. **(f)** CTD path in different CIDs. Indicated CTD peptides (ribbon) were superimposed in the indicated complexes[42,43].

Europe PMC Funders Author Manuscripts

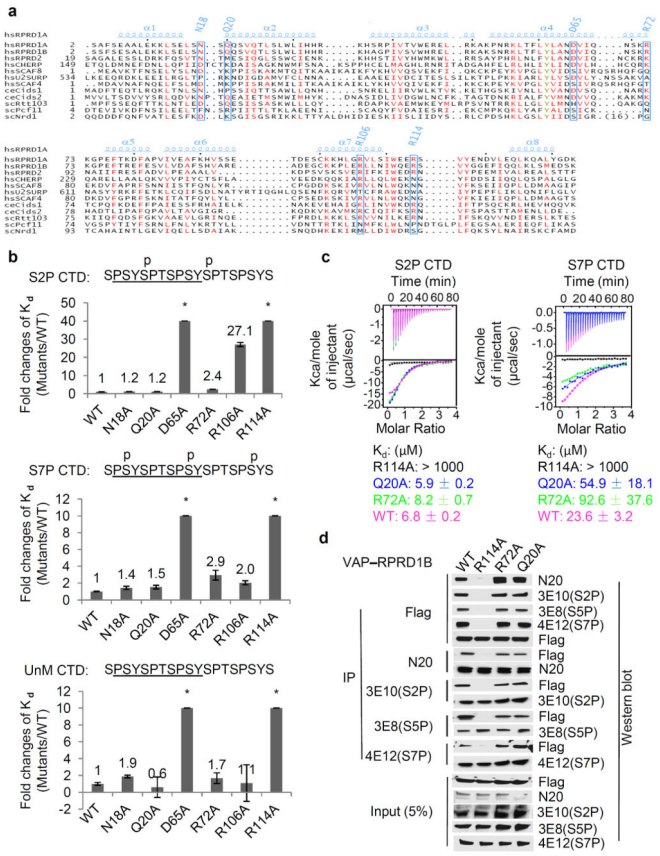

Figure 3.

Effects of RPRD CID mutations on CTD binding.

(**a**) Sequence alignment of CID domains using ESPript3. The absolutely conserved residues are shown in dark yellow, while the similar residues are shown in red. The secondary structures of RPRD1A CID domains are shown at the top. The RPRD1A and RPRD1B residues involved in recognizing CTD peptides are labeled in blue. (**b**) Isothermal titration calorimetry (ITC) measurement of the binding affinity (K_d) between the indicated CTD peptides and the indicated RPRD1A CID mutant recombinants in comparison with that of wild type (WT). Synthetic CTD peptides contained the underlined RPRD CID-binding site. Phosphorylated serine residues are marked. *: No measurable interaction was detected for the mutated CIDs, and thus the fold changes for mutant/WT are really infinite. Bars indicate ranges of two technical replicates. (**c**) ITC measurement of the binding affinity (K_d) between the indicated CTD peptides and the indicated RPRD1B CID mutant recombinants in comparison with that of WT. (**d**) Immunoprecipitation-western blot (IP-WB) analysis of the interaction between the indicated VAP-tagged-RPRD1B mutants and RNAPII in extracts from HEK293 cells, using the indicated antibodies. VAP-tag contains 3xFlag, 6xHis, and Strep. Uncropped images of gels are shown in Supplementary Data Set 1.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

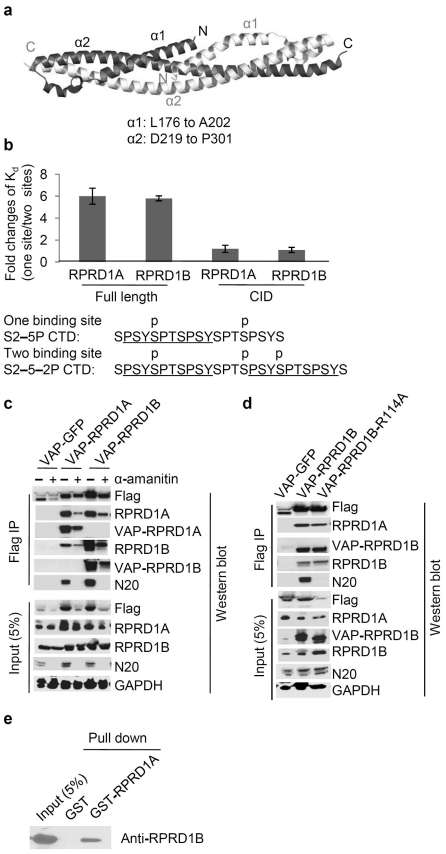

Figure 4.

RPRD1A and RPRD1B dimerization through their coiled-coil domains.

(**a**) Crystal structure of RPRD1B C-terminal coiled-coil domain. The two interacting polypeptides are shown in darker and lighter shades. α: alpha-helix. (**b**) ITC measurement of the binding affinity (K_d) between full length RPRD1A and RPRD1B or their CIDs with CTD peptides bearing either one or two CID-binding site(s). Ratios of ITC measurements for the interactions of the indicated RPRDs with the two indicated CTD peptides are shown. Underlined are the CID-binding sequences in the CTD peptides. Phosphorylated serines are labeled. Bars indicate ranges of two technical replicates. (**c**) IP-WB analysis using the indicated antibodies showing the interaction between lentiviral transduced VAP-RPRD1A and VAP-RPRD1B in HEK293 cells grown in the presence or absence of 2 μg/ml α-amanitin for 72 hr. VAP: versatile affinity purification tag. (**d**) IP-WB analysis showing the effect of the R114 mutation on the interaction between lentiviral transduced VAP-RPRD1B and RPRD1A in HEK293 cells. (**f**) Glutathione-S-transferase (GST) pull-down experiment showing a direct interaction between GST-tagged RPRD1A and His-tagged RPRD1B purified from *E. coli*. Uncropped images of gels in Fig. 4c-e are shown in Supplementary Data Set 1.

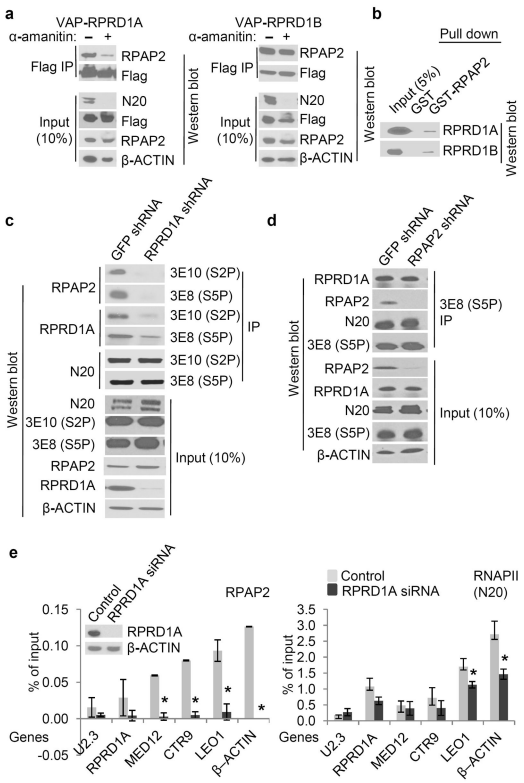

Figure 5.

RPRDs interact with RPAP2 and mediate its interaction with RNAPII.

(**a**) IP-WB analysis using the indicated antibodies showing the interaction between RPAP2 and lentiviral transduced VAP-RPRD1A or VAP-RPRD1B in HEK293 cells grown with or without 2 μg/ml α-amanitin for 48 hr. (**b**) GST pull-down experiments showing direct interaction between GST-tagged RPAP2 and His-tagged RPRD1A or RPRD1B purified from *E. coli*. (**c**) IP-WB using the indicated antibodies showing the effects of shRNA-mediated RPRD1A knock-down on the association of RPAP2 with RNAPII in HEK293 cells. (**d**) IP-WB using the indicated antibodies showing the effects of shRNA-mediated RPAP2 knock-down on the association of RPRD1A with RNAPII in HEK293 cells. (**e**) Chromatin immunoprecipitation (ChIP) experiments using the indicated antibodies showing the effects of siRNA-mediated RPRD1A knock-down (WB analysis shown in the insert) on the recruitment of RPAP2 at the various indicated promoters in HeLa cells. *β-ACTIN* primers are located at +300 bp. U2.3: *U2 snRNA* untranscribed region. *: Statistically significant ($p < 0.05$) using a two-tailed Student's *t*-test compared with no siRNA transfection control. Error bars, s.d. of three biological replicates. The bar for *β-ACTIN* indicates the range of two biological replicates. Uncropped images of gels are shown in Supplementary Data Set 1.

Europe PMC Funders Author Manuscripts

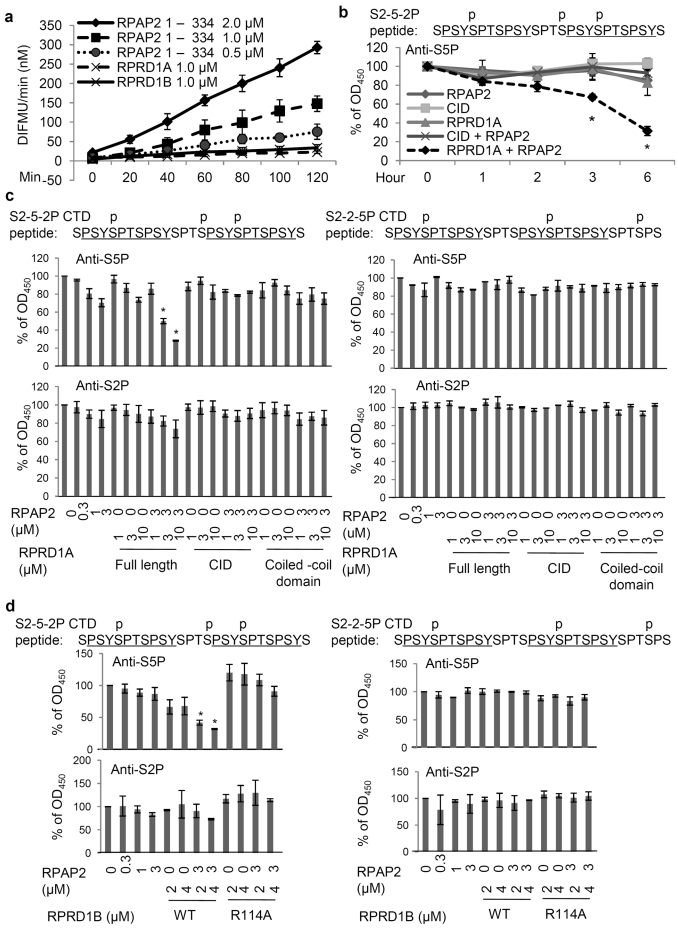

Figure 6.

RPRDs stimulate RPAP2 phosphatase activity for the CTD *in vitro*.

(**a**) *In vitro* assays showing the phosphatase activity of RPAP2 (1–334) on the substrate 6,8-difluoro-4-methylumbelliferyl phosphate (DiFMUP). Results are average ± s.d. from 5 technical replicates. (**b**) *In vitro* phosphatase ELISAs showing the time course of RPRD1A-stimulated RPAP2 phosphatase activity on S5P on the indicated biotinylated CTD peptides. S5P monoclonal antibody: 3E8. RPAP2: 3 μM. RPRD1A or RPRD1A-CID: 10 μM each. The results at 0 hour were set as 100%. *: $p<0.01$ as compared with the results at 0 hour (two-tailed student's *t*-test). Results are average ± s.d. from 3 technical replicates. (**c**) *In vitro* phosphatase ELISAs showing the effects of RPRD1A on RPAP2 phosphatase activity on S5P located in the region between two S2P-containing CTD decamer repeats. The assays contained biotinylated CTD peptides S2-5-2P CTD (left) or S2-2-5P CTD (right), bearing S5P located in the region between (left) or outside of (right) two CID-binding sites (underlined), respectively. (**d**) *In vitro* phosphatase ELISAs showing RPRD1B, but not its R114A mutant, stimulates RPAP2 phosphatase activity on the CTD *in vitro.* In Fig. 6c,d, no protein controls were set as 100%. Error bars, s.d. of three technical replicates. *: $p<0.01$ as compared with the no protein controls (two-tailed Student's *t*-test).

Europe PMC Funders Author Manuscripts

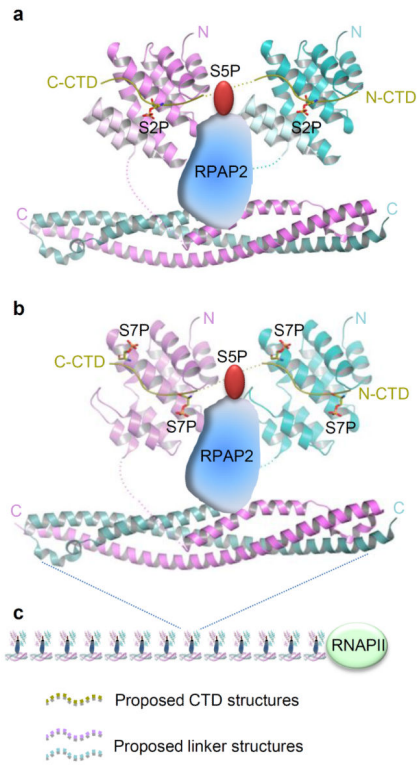

Figure 7.

Proposed “CTDsome” models for RPRD-dependent RPAP2 dephosphorylation of RNAPII S5P CTD. (**a**) Model of S2P-bound RPRD1B stimulating dephosphorylation of S5P by RPAP2. CID structures are from the RPRD1B CID in complex with a S2P CTD peptide. Coiled-coil domain structure is from RPRD1B. CTD is shown in dark yellow ribbon with S2P shown in red sticks. Two RPRD1B molecules are coloured as teal and magenta, and the relevant domain-swapped regions are coloured as magenta and light magenta, teal and palecyan, respectively. (**b**) Model of S7P-bound RPRD1A stimulating dephosphorylation of S5P by RPAP2. Two molecules are coloured as light magenta and palecyan. CID structures are from the RPRD1A CID. Coiled-coil domain structure is from RPRD1B. CTD is shown in dark yellow ribbon with S7P shown in red stick model. In **Fig. 7a.b**, the linker structures between CIDs and coiled-coil domains, and the CTD structures between two adjacent CIDs are undetermined. The RPAP2 structure is schematic. (**c**) Proposed multiple ”CTDsomes” on the RNAPII CTD (see text for description).

Table 1

RNAPII CTD binding affinities of RPRD CIDs

Peptide names	Peptide sequences	RPRD1A CID (K_d)	RPRD1B CID (K_d)	RPRD2 CID (K_d)
UnM CTD	SPSYSPTSPS YSPTSPSYS	339±56	114±2	355±30
S2P CTD	p p SPSYSPTSPS YSPTSPSYS	8.4±0.7	6.8±0.2	8.3±0.5
S5P CTD	p p SPSYSPTSPS YSPTSPSYS	>1,000	>1,000	>1,000
S7P CTD	p p p SPSYSPTSPS YSPTSPSYS	49.8±13.7	23.6±3.2	82.8±28.7
S2,7P CTD	p p p p SPSYSPTSPS YSPTSPSYS	5.2±0.5	2.6 ± 0.2	7.2±0.1
S2,5P CTD	p p p p SPSYSPTSPS YSPTSPSYS	>1,000	>1,000	>1,000
S5,7P CTD	p p p p SPSYSPTSPS YSPTSPSYS	>1,000	>1,000	>1,000
S7P-S5P CTD	p p p SPSYSPTSPS YSPTSPSYS	48.3±5.5	30.0±4.9	112±13
S2P-K7 CTD	p p SPKYSPTSPK YSPTSPKYS	1.0 ±0.1	8.3± 0.5	5.2±0.1

Dissociation constants were measured using isothermal titration calorimetry. Synthetic CTD peptides contained two consecutive heptapeptide repeats (underlined) with additional SPS and YS residues before and after the two repeats, respectively. Phosphorylated serine residues are marked. Dissociation constants (K_d) were from a minimum of two technical replicates (average ± s.d.). UnM CTD: unmodified CTD peptides.

Table 2

Data collection and refinement statistics

	RPRD2 CID (15–161)	RPRD1A CID (2–137) + S7P	RPRD1B CID (2–135) + UnM	RPRD1B (2–135) + S2P	RPRD1B coiled coiled domain (171–304)
Data collection					
Space group	$P4_32_12$	I4	$P2_1$	$P2_12_12_1$	$P42_12$
Cell dimensions					
a, *b*, *c* (Å)	40.34, 40.34, 145.73	93.26, 93.26, 36.03	55.65, 134.71, 55.71	39.29, 70.45, 108.86	100.22, 100.22, 142.89
α, *β*, *γ* (°)	90, 90, 90	90, 90 ,90	90, 106.64, 90	90, 90, 90	90, 90, 90
Resolution (Å)	28.52–1.80 (1.90–1.80)*	30.00–1.90 (2.00–1.90)	44.90–1.85 (1.89–1.85)	43.07–1.85 (1.89–1.85)	47.63–2.20 (2.32–2.20)
R_{merge}	0.093 (0.749)	0.119 (0.927)	0.090 (0.986)	0.057 (1.076)	0.116 (0.925)
$I / \sigma I$	25.0 (4.3)	14.8 (2.3)	10.1 (1.4)	21.7 (1.9)	15.5 (2.8)
Completeness (%)	98.9 (97.1)	98.3 (96.3)	99.9 (100.0)	100.0 (100.0)	100.0 (100.0)
Redundancy	12.5 (12.7)	7.3 (7.2)	3.8 (3.8)	7.1 (7.3)	9.6 (9.8)
Refinement					
Resolution (Å)	28.54–1.80	29.51–1.90	44.65–1.85	43.11–1.85	40.00–2.20
No. reflections	11159	11605	64761	25264	36091
R_{work} / R_{free}	0.175/0.216	0.184/0.237	0.238/0.271	0.178/0.221	0.232/0.258
No. atoms					
Protein	1113	1083	4214	2065	3945
Peptide	n/a	102	310	183	n/a
Water	72	88	228	82	148
B factors					
Protein	20.5	23.5	24.9	32.5	45.9
Peptide	n/a	31.1	22.5	25.9	n/a
Water	28.4	26.7	27.0	36.2	35.6
r.m.s. deviations					
Bond lengths (Å)	0.015	0.014	0.007	0.014	0.013
Bond angles (°)	1.4	1.5	1.1	1.5	1.3

*Values in parentheses are for highest-resolution shell.