RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation

Zuyao Ni1,8, Chao Xu2,8, Xinghua Guo1, Gerald O Hunter3, Olga V Kuznetsova4, Wolfram Tempe12, Edyta Marcon1, Guoqing Zhong1, Hongbo Guo1, Wei-Hung William Kuol5, Joyce Li1, Peter Young1, Jonathan B Olsen1,5, Cuihong Wan1, Peter Loppnau2, Majida El Bakkouri2, Guillermo A Senisterra2, Hao He2, Haiming Huang1, Sachdev S Sidhu1,5, Andrew Emili1,5, Shona Murphy4, Amber L Mosley3, Cheryl H Arrowsmith2,6, Jinrong Min2,7 & Jack F Greenblatt1,5

The RNA polymerase II (RNAPII) C-terminal domain (CTD) heptapeptide repeats (1-YSPTSPS-7) 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 RNAPII CTD repeats phosphorylated at S2 and S7. Crystal structures of the RPRD1A, RPRD1B and RPRD2 CIDs, alone and in complex with RNAPII CTD phosphoisoforms, elucidate the molecular basis of CTD recognition. In an example of cross-talk between 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.

The CTD of the largest subunit, POLR2A, of human RNAPII consists of multiple, sometimes degenerate, heptapeptide repeats with consensus sequence 1-YSPTSPS-7 (refs. 1,2). The CTD is phosphorylated during transcription on Y1, T4 and all three serine residues3. Different phosphorylation patterns, proline isomerization3,4 and modification of nonconsensus CTD residues5 create a ‘CTD code’6 that recruits various factors to regulate transcription, mRNA processing and histone modification5,7–10. There also appears to be cross-talk between different CTD modifications5.

General transcription factor TFIIH phosphorylates S5 and S7 (to form SSP and S7P) in promoter regions11–15, where SSP recruits mRNA-capping enzymes16–19, the yeast COMPASS complex for histone H3 K4 trimethylation by Set1 (MLL proteins in humans)20–22 and the yeast Sen1–Nrd1–Nab3 complex to terminate noncoding small nuclear RNAs (snRNAs), small nucleolar RNAs and cryptic unstable transcripts23. SSP also characterizes ‘poised’ RNAPII in transcriptionally inactive regions24–27. RNAPII escaping the promoter is phosphorylated at S2 and S7, where RPRD1A and/or 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.

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 controversial33,39. We set out here to characterize three human CTD-interacting proteins. Our structural and biophysical studies show that the CID-containing proteins RPRD1A, RPRD2 and RPRD1B (an oncogene)40, all of which associate with RNAPII41, associate preferentially as dimers with S2P- and, to a lesser extent, with S7P-containing CTD peptides, whereas SSP interferes with binding. We show that 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 SSP located in the intervening region to RPAP2 for dephosphorylation.

RESULTS

Recognition of specific CTD phosphoisoforms by RPRDs

RPRD1A, RPRD1B and RPRD2 (‘RPRDs’)41 were previously found to coprecipitate with phosphorylated RNAPII41. Here we used isothermal
titration calorimetry (ITC) to determine dissociation constants of recombinant RPRD CIDs (Fig. 1a) for CTD peptides containing two heptapeptide repeats without modification (UnM) or with serine phosphorylation at positions 2, 5 or 7 (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 (denoted S2,7P CTD) increased the RPRD1A and RPRD1B affinities by 1.6-fold and 2.6-fold, respectively, compared to S2P alone, whereas the RPRD2 CID affinity remained unchanged. In contrast, S5P in the same heptapetide 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 coprecipitate with S5P-containing RNAPII in cell extracts imply that S5P-containing repeats exist on the same CTDs as do S2- and/or S7-containing repeats that bind RPRDs. That S5P on a repeat adjacent to one with S2P (S7-S5 P CTD in Table 1) had no significant effect on CTD binding supports this idea.

#### 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      | SPSYSPTS SPSYSPTS | 339 ± 56         | 114 ± 2          | 355 ± 30         |
| S2P CTD      | p SS PS SS PS      | 8.4 ± 0.7        | 6.8 ± 0.2        | 8.3 ± 0.5        |
| S5P CTD      | p SS PS SS PS      | >1,000           | >1,000           | >1,000           |
| S7P CTD      | p SS PS SS PS      | 49.8 ± 13.7      | 23.6 ± 3.2       | 82.8 ± 28.7      |
| S2,7P CTD    | p p SS PS SS PS    | 5.2 ± 0.5        | 2.6 ± 0.2        | 7.2 ± 0.1        |
| S2,5P CTD    | p p p SS PS PS    | >1,000           | >1,000           | >1,000           |
| S5,7P CTD    | p p p p SS PS    | >1,000           | >1,000           | >1,000           |
| S7P-S5P CTD  | p p p SS PS      | 48.3 ± 5.5       | 30.0 ± 4.9       | 112 ± 13         |
| S2P-K7 CTD   | p p p p p p      | 1.0 ± 0.1        | 8.3 ± 0.5        | 5.2 ± 0.1        |

Dissociation constants were measured with ITC. 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 with ‘p’. K_d values were from three technical replicates (average ± s.d.). UnM CTD, unmodified CTD peptide.

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, SPSYSPTS SPSYSPTS, containing more than two heptapeptide repeats bearing three S7P residues (underlined). Like the CIDs of its yeast homologs, Rtt103, Pcf11 and Nrd1 (refs. 23,42,43), the RPRD1A CID contained eight α-helices arranged in a right-handed superhelical manner (Fig. 1b). A concave channel accommodated the CTD decapetptide SPSYSPTS (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 d.

We observed many CID-CTD contacts (Fig. 1c–e). CID N18 formed water-mediated
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-coil domain (171–304) |
|--------------------|-------------------|-------------------------|------------------------|---------------------|------------------------------------|
| **Data collection**|                   |                         |                        |                     |                                    |
| Space group        | P4_2_2            | P4_2_2                  | P2_1                   | P2_1; 2_1          | P4_2_2                             |
| 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               |
|                    | α, β, γ (%)       | 90, 90, 90              | 90, 90, 90             | 90, 106, 94, 90    | 90, 90, 90                         |
| α, β, γ (%)        |                   | 90, 90, 90              | 90, 90, 90             | 90, 106, 94, 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) |
| Rmerge (%)         | 0.093 (0.749)     | 0.119 (0.927)           | 0.090 (0.986)          | 0.057 (1.076)      | 0.116 (0.925)                      |
| 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    | 11,159            | 11,605                  | 64,761                 | 25,264             | 36,091                             |
| Rfree / Rwork (%)  | 0.175 / 0.216     | 0.184 / 0.237           | 0.238 / 0.271          | 0.178 / 0.221      | 0.232 / 0.258                      |
| No. atoms          | Protein           | 1,113                   | 1,083                  | 4,214              | 2,065                              |
|                    | Peptide           | –                       | 102                    | 310                | 183                                |
|                    | Water             | 72                      | 88                     | 228                | 82                                 |
|                    | Protein           | 20.5                    | 23.5                   | 24.9               | 32.5                               |
|                    | Peptide           | –                       | 31.1                   | 22.5               | 25.9                               |
|                    | Water             | 28.4                    | 26.7                   | 27.0               | 36.2                               |
| r.m.s. deviations  | Bond lengths (Å)  | 0.015                   | 0.014                  | 0.007              | 0.014                              |
|                    | Bond angles (°)   | 1.4                     | 1.5                    | 1.1                | 1.5                                |

Values in parentheses are for highest-resolution shell.

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 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 probably enable flexibility for S7 modification and substitution hindrance and charge repulsion, thus accounting for disruption of CTD binding by S5P. Conversely, the two S7 positions at the channel entrance and exit probably enable flexibility for S7 modification and substitution.

To address whether RPRD CIDs similarly recognize an S2P-containing CTD, we crystallized the RPRD1B CID with the 19-mer S2P CTD peptide SPSYSPTSPSYSPSYSP (Fig. 2a, b). This 1.85-Å-resolution structure exhibited intrahelical swapping of two C-terminal helices, a phenomenon known as domain swapping, 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 RPRD1A CID–S7P CTD complex (Fig. 1d, e). Importantly, the highly conserved residue R106 formed two hydrogen bonds with the S2bP phosphate group, thus explaining the enhanced affinity over the unmodified CTD peptide (Table 1). Other differences included N69 forming water-mediated hydrogen bonds with the 57b hydroxyl group and N18 not making contact with S7a. The domain-swapped structure is likely to be 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 SPSYSPTSPSYSPSYSPYS. The folding and binding mode in this structure were also similar to those of the RPRD1A–S7P CTD and RPRD1B–S2P CTD complexes, but they 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), thus suggesting that 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, we mutated various CTD-interacting and evolutionarily conserved residues in the CIDs (Fig. 3a) and examined their effects on
binding by ITC. Mutating R114 or D65 to alanine in the RPRD1A CID diminished binding to all three tested peptides (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 but had 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, in which R114 and D65 are most important for interaction independent of phosphorylation state, and R106 is important for specific interaction with an S2P peptide.

To examine the effects of CID mutations on interaction with RNAPII in vivo, we immunoprecipitated extracts from HEK293 cells stably expressing wild-type and mutant RPRD1B proteins with VAP (versatile affinity purification) tags (containing 3× Flag, hexahistidine (His6) and streptactin), using antibodies recognizing the Flag epitopes, total RNAPII or various RNAPII phosphoisoforms. Western blotting revealed that, consistently 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). When 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. Indeed, size-exclusion chromatography eluted full-length RPRD1A and RPRD1B, as well as their coiled-coil domains, as apparent multimers, whereas their CIDs eluted as monomers (Supplementary Fig. 2a). Similarly, full-length RPRD1A and RPRD1B behaved like multimers in dynamic light-scattering (DLS) analysis, whereas the three RPRD CIDs behaved like monomers (Supplementary Fig. 2b).

Because size-exclusion chromatography and DLS overestimate 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), but its CID was almost entirely monomeric (94.0%) (Supplementary Table 3). Furthermore, upon cross-linking 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-PAGE gels at molecular masses approximately equivalent to dimers, whereas its CID migrated at the position equivalent to monomers until a higher BS3 concentration (300–3,000 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 and α2) with the two α2 helices being extensively intertwined (Fig. 4a). N-terminal regions (L176–A191) of the two α1 helices contacted each other in an antiparallel mode.

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 as in Figure 1d, except that the two molecules of the RPRD1B CID in the swapped dimers are colored in blue and cyan. 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 as in Figure 1e, except that residues from the two RPRD1B CID molecules in swapped dimers are shown in blue and cyan. The interactions between the CID and the CTD are depicted as in Figure 1e. (c) Detailed interactions of RPRD1B CID with UnM CTD. Protein and CTD peptide are shown as in Figures 1d and 2a. (d) Interaction diagram for the RPRD1B CID bound to the UnM CTD peptide. Proteins and peptide are colored as in Figures 1e and 2b. (e) F2 – F1 electron density maps of the three peptides (yellow sticks) in the indicated complexes, contoured at 3.0σ. (f) CTD path in different CIDs. Indicated CTD peptides (ribbon) were superimposed in the indicated complexes.

Nature Structural & Molecular Biology
and were 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 coiled-coil domains, and the substantial reciprocal copurification of RPRD1A and RPRD1B. To determine whether the interaction is direct, we carried out glutathione-S-transferase (GST) pulldown experiments and found that recombinant GST-RPRD1A pulled down full-length RPAP2 and its coiled-coil domain but not its CID (Fig. 4e and Supplementary Fig. 3), a result suggesting heterodimerization via their coiled-coil domains.

RPRD1A and RPRD1B mediate RPAP2 association with RNAPII

Given that RPAP2 does not directly bind the SS5 CTD\(^3\), we thought other factors might be required for RPAP2 binding to RNAPII. The copurification of RPRD1A and RPRD1B with RPAP2 (ref. 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 coprecipitated with RPAP2 equally well after cells were grown in the presence or absence of α-amanitin (Fig. 5a), thus indicating an RNAPII-independent interaction between RPAP2 and RPRD1B. VAP-tagged RPRD1A also coprecipitated RPAP2 after cells were grown in the presence of α-amanitin, although to a lesser extent (Fig. 5a), a result indicating that most RPRD1A might not be associated directly with RPAP2 and would coprecipitate it only indirectly via RNAPII and RPRD1B. Next, we found that recombinant GST-RPAP2 pulled down both RPRD1A and RPRD1B (Fig. 5b), thus 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 that RPRDs might be required for RPAP2 binding to RNAPII in vivo. To address this issue, we carried out communoprecipitation with extracts from HEK293 cells depleted of an RPRD protein. Because RPRD1A and RPRD1B form heterodimers, knocking down either one likely would disrupt their function. Indeed, knocking down RPRD1A by stable expression of lentiviral-encoded short hairpin RNAs (shRNAs), or by transient expression of short interfering RNAs (siRNA) (Z.N. and X.G., unpublished data), substantially reduced the RPAP2 coprecipitated 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 coprecipitated by anti–S5P RNAPII CTD antibodies equally well after knockdown of GFP or RPAP2 (Fig. 5d), thus indicating that RPRD-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 knockdown of RPRD1A almost abolished RPAP2 association with various tested promoters, whereas 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. When taken together, these results demonstrate that RPRDs are required for RPAP2 association with RNAPII in vivo.

RPRDs stimulate CTD S5P dephosphorylation by RPAP2 in vitro

The Murphy33 and Tong39 groups have reported different results regarding the ability of RPAP2 to act as a phosphatase in vitro. To address this controversial issue, we incubated the conserved

---

**Figure 4** RPRD1A and RPRD1B dimerization through their coiled-coil domains. (a) Crystal structure of RPRD1B C-terminal coiled-coil domain. The two polypeptides are shown in darker and lighter shades. (b) ITC measurement of the $K_d$ between full-length RPRD1A and RPRD1B or their CID with CTD peptides bearing either one or two CID-binding sites. Ratios of ITC measurements for the interactions of the indicated RPRDs with the two indicated CID peptides are shown. Underlined are the CID-binding sites in the CTD peptides. Phosphorylated serines are labeled. Bars indicate ranges of two technical replicates. (c) IP/WB analysis with the indicated antibodies, showing the interaction between lentiviral-transduced VAP-RPRD1A and VAP-RPRD1B in HEK293 cells. (d) IP/WB analysis with the indicated antibodies, showing the interaction between lentiviral-transduced VAP-RPRD1A and VAP-RPRD1B in HEK293 cells. (e) GST pulldown experiment showing a direct interaction between GST-tagged RPRD1A and histidine-tagged RPRD1B purified from Escherichia coli. Uncropped images of gels in Figure 4c–e are shown in Supplementary Data Set.

---

**Figure 5** RPRDs interact with RPAP2 and mediate its interaction with RNAPII. (a) IP/WB analysis with 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 h. (b) GST pulldown experiments showing direct interaction between GST-tagged RPRD2 and histidine-tagged RPRD1A or RPRD1B purified from E. coli. (c) IP/WB with the indicated antibodies, showing the effects of shRNA-mediated RPRD1A knockdown on the association of RPAP2 with RNAPII in HEK293 cells. (d) IP/WB with the indicated antibodies, showing the effects of shRNA-mediated RPAP2 knockdown on the association of RPRD1A with RNAPII in HEK293 cells. (e) ChIP experiments with the indicated antibodies, showing the effects of siRNA-mediated RPRD1A knockdown (WB analysis shown in the inset) on 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. * $P < 0.05$ by two-tailed Student’s t test compared with a no-siRNA transfection control. Error bars indicate s.d. of three separate cell cultures, and the bar for β-ACTIN indicates the range of two separate cell cultures. Uncropped images of gels are shown in Supplementary Data Set.
and functional RPAP2 fragment, RPAP2(1–334), with the phosphatase substrate 6,8-difluoro-4-methylumbelliferyl phosphate (DiMUP). RPAP2(1–334), but neither RPRD1A nor RPRD1B, dephosphorylated DiMUP in a time- and dose-dependent manner (Fig. 6a). We obtained similar results 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). We detected no measurable phosphatase activity when using the canonical phosphatase substrate p-nitrophenylphosphate (pNPP) (G.O.H. and A.L.M., unpublished data), a result 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) in which the products of phosphatase reactions containing biotinylated CTD peptides were adsorbed to streptavidin-coated wells in ELISA plates. We assessed phosphorylation levels with antibodies that specifically recognize particular CTD phosphoisoforms. In agreement with previous observations, we detected no measurable phosphatase activity on SSP- or S2P-containing CTD peptides in reactions with RPAP2 alone (Fig. 6b–d).

Given that RPRDs tether RPAP2 to RNAPII (Fig. 5), and overexpression of RPAP2 decreases SSP at various promoters, 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 CTD or coiled-coil domain, dephosphorylated SSP in a time- (Fig. 6b) and dose-dependent manner (Fig. 6c). S2P was dephosphorylated to a much lesser extent (Fig. 6c). We obtained similar results for RPRD1B but not for its R114A mutant that fails to bind the CTD (Fig. 6d). RPRD1A-dependent SSP dephosphorylation by RPAP2 was also inhibited by vanadate and several metals (Supplementary Fig. 4d), thus suggesting that similar dephosphorylation mechanisms operate on DiMUP 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 $K_M$ 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 consistently with RPRD dimerization, a CTD peptide containing two S2P CID-binding sites bracketing an SSP residue located in the intervening sequence (S2-5-2P CTD) served as a
substrate in reactions containing RPRD1A or RPRD1B (Fig. 6c,d), whereas SSP located outside two S2P-containing CID-binding sites (S2-2-5P CTD) did not (Fig. 6c,d).

We then asked whether S7P CTD could also stimulate the RPRD-dependent CTD phosphatase activity of RPAP2. Because S7P next to SSP in a dual RPRD CID–bound SSP CTD peptide (S77-5-77P CTD) prevented antibody recognition of SSP in ELISAs (Supplementary Fig. 4e), we used a peptide with an unmodified S7 adjacent to SSP (S77-5-77P CTD) for the assay. RPRD1A or RPRD1B stimulated RPAP2 activity toward SSP, but not S7P (Supplementary Fig. 4f), although to a lesser extent than for S2-5-2P CTD peptides (Fig. 6c,d). Perhaps this was because of the lower affinities of RPRD1A and RPRD1B for the S77-5-77P 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 SSP in vivo. ChIP experiments showed that knocking down RPAP2 in HEK293 cells with two independent shRNAs increased the levels of SSP relative to RNAPII at several tested promoter regions (Supplementary Fig. 5a), even though no obvious global change of either SSP or S2P phosphorylation was detectable by western blotting (Supplementary Fig. 5b). Consistently with this, siRNA-mediated knockdown of RPRD1A in HeLa cells increased the SSP levels near the ACTB (β-actin) promoter and other promoter regions (Supplementary Fig. 5c). These results were in line with our previous observation that overexpression of RPRDs leads to a decrease of SSP at various promoter regions41.

In summary, multiple S2P-CTD and/or S7P-CTD repeats interact with the two CIDs of homo- and/or heterodimeric RPRD1A and/or RPRD1B, which in turn recruit RPAP2 to dephosphorylate the SSP 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 SSP 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, probably 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, 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 noncanonical CTD repeats, such differences might not be significant for RPRD1A–RPRD1B heterodimers. Because RPRD CIDs bind both canonical and noncanonical CTD repeats (Table 1), and each RPRD dimer occupies more than three heptapeptide repeats, each mammalian S2-repeat RNAPII CTD could theoretically accommodate up to 13 such structures (Fig. 7c).

Interaction of CIDs with ten rather than seven CTD amino acids probably explains why the functional unit of the CTD appears to contain two heptapeptide repeats47.

Dimerization of the RPRDs enhances CTD binding, an effect that could be explained by the synergistic effect of multiple CID–CTD interactions. A similar synergistic effect has been observed for binding of yeast Pcf11 and R11 to S2P CTD, in which these two proteins bind cooperatively with each other to result in enhanced CID–CTD interaction43. Pcf11 is also a subunit of CF1A in which two other subunits, Rna14 and Rna15, form a heterotetramer that also interacts with the CTD48. Thus, the phenomenon of CID cooperativity in CTD recognition and transcription regulation is probably conserved between yeast and humans.

Our data indicate that RPAP2 is a substrate-selective phosphatase. In CTD peptides, the SPTS tetramer bracketed by a pair of RPRD dimer–bound S2P and/or S7P CTD repeats is likely to be 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 new 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 cross-talk from S2P and S7P to S5P.

During transcription initiation, TFIIH phosphorylates both S5 and S7 on the CTD11–13,15. Subsequently, S7P facilitates association of RPRD1B involved in tumorigenesis is an important future goal. The high-resolution crystal structure of the RPRD1B CTD complex with S2P CTD may be useful for the design and development of anticancer compounds that regulate the CTD binding activity of RPRD1B.

RPRDs are capable of associating with transcription complexes not only in promoter regions but also in gene bodies and near mRNA 3’ ends40,41. This makes it likely that RPRD scaffolds could have additional roles other than S5 dephosphorylation in transcriptional regulation. Given that RPRD1B is an oncoprotein overexpressed in more than 80% of human tumors40, deciphering the particular activity of RPRD1B involved in tumorogenesis is an important future goal. The RPRD scaffolds could have additional roles other than S5 dephosphorylation in transcriptional regulation. Given that RPRD1B is an oncoprotein overexpressed in more than 80% of human tumors40, deciphering the particular activity of RPRD1B involved in tumorogenesis is an important future goal.

METHODS

Methods and any associated references are available in the online version of the paper.

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

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank D.Y. Zhao, A. NRL and K. Liu for technical assistance, J. Moffat (University of Toronto) for providing the siRNAs and D. Eick (Helmholtz-Zentrum, Munich) for providing antibodies. 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), Edward Penley Abraham (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.). Oxford Stem Cell Institute 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, Jansen, Lilly Canada, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda and the Wellcome Trust (092869/Z/10/Z) (to C.H.A. and J.M.). Results in this paper are partially derived from work performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by University of Chicago, Argonne, for the US Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357.

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, 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. He and H. Huang 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. and J.F.G. wrote the manuscript. J.M. and J.F.G. supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Allison, L.A., Moyle, M., Shales, M. & Ingles, C.J. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell 42, 599–610 (1985).
2. Corden, J.L. Tails of RNA polymerase II. Trends Biochem. Sci. 15, 383–387 (1990).
3. Egloff, S., Dienstbier, M. & Murphy, S. Upgrading the RNA polymerase CTD code: adding gene-specific layers. Trends Genet. 28, 333–341 (2012).
4. Singh, N. et al. The Ess1 polyriol isomerase is required for transcription termination of noncoding RNAs via the Nrd1 pathway. Mol. Cell 36, 255–266 (2009).
5. Sims, R.J. et al. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science 332, 99–103 (2011).
6. Narayanswamy, S. The CTD code. Nat. Struct. Biol. 10, 679–680 (2003).
7. Corden, J.L. Transcription: seven ups the code. Science 318, 1736–1736 (2007).
8. Hsin, J.P., Sheth, A. & Manley, J.L. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3’ end processing. Science 334, 683–686 (2011).
9. Krogan, N.J. et al. Phosphorylation of the CTD by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol. Cell 11, 721–729 (2003).
10. Mayer, A. et al. CTD tyrosine phosphorylation impacts termination factor recruitment to RNA polymerase II. Science 336, 1723–1725 (2012).
11. Kim, M., Suh, H., Cho, E.J. & Buratowski, S. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. J. Biol. Chem. 284, 26421–26426 (2009).
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. 29, 5455–5464 (2009).
13. Akhtar, M.S. et al. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol. Cell 34, 387–393 (2009).
14. Chapman, R.D. et al. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science 318, 1780–1782 (2007).
15. Serizawa, H. et al. Association of Cdk-activating kinase subunits with transcription factor TFIIH. Nature 374, 280–282 (1995).
16. Ghosh, A., Shuman, S. & Lima, C.D. Structural insights to how mammalian capping enzyme reads the CTD code. Mol. Cell 43, 299–310 (2011).
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. 11, 3306–3318 (1997).
18. Cho, E.J., Takagi, T., Moore, C.R. & Buratowski, S. mRNA capping enzyme is recruited to the transcription complex by phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell. Biol. 318, 1549–1561 (2003).
19. Ng, H.H., Robert, F., Young, R.A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell 11, 709–719 (2003).
20. Krogan, N.J. et al. COMPASS, a histone H3 (lysine 4) methyltransferase required for telomeric silencing of gene expression. J. Biol. Chem. 277, 10753–10755 (2002).
21. Briggs, S.D. et al. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and DNA silencing in Saccharomyces cerevisiae. Genes Dev. 15, 3286–3295 (2001).
22. Vasileva, 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. 15, 795–804 (2008).
23. Mutskov, V.J., Farrell, C.M., Wade, P.A., Wolffe, A.P. & Felsenfeld, G. The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. Genes Dev. 16, 1540–1554 (2002).
24. Zettliger, J. et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat. Genet. 39, 1512–1516 (2007).
25. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77–88 (2007).
26. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322, 1845–1848 (2008).
27. Cho, E.J., Kobor, M.S., 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. 15, 3319–3329 (2001).
29. Zhou, Q., Li, T. & Price, D.H. RNA polymerase II elongation control. Annu. Rev. Biochem. 81, 119–143 (2012).
30. Bartkowiak, B. et al. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. 24, 2303–2316 (2010).
31. Devaiah, B.N. et al. BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. Proc. Natl. Acad. Sci. USA 109, 6927–6932 (2012).
32. Mosley, A.L. et al. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Mol. Cell 34, 168–178 (2009).
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 45, 111–122 (2012).
34. Bataille, A.R. et al. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol. Cell 45, 158–170 (2012).
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. 280, 37681–37688 (2005).
36. Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C. & Hampsey, M. Ssu72 Is an RNA polymerase II CTD phosphatase. Mol. Cell 14, 387–394 (2004).
37. Archambault, J. et al. FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxy-terminal domain of RNA polymerase II. J. Biol. Chem. 273, 27593–27601 (1998).
38. Kobor, M.S. et al. An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae. Mol. Cell 4, 55–62 (1999).
39. Xiang, K., Manley, J.L. & Tong, L. The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity. Nat. Commun. 3, 946 (2012).
40. Lu, D. et al. CREPT accelerates tumorigenesis by regulating the transcription of cell-cycle-related genes. Cancer Cell 21, 92–104 (2012).
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 2, 237–242 (2011).
42. Meinhart, A. & Cramer, P. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature 430, 223–226 (2004).
43. Lunde, B.M. et al. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. 17, 1195–1201 (2010).
44. Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. Protein Sci. 11, 1285–1299 (2002).
45. Nooren, I.M., Kaptein, R., Sauer, R.T. & Boelens, R. The tetramerization domain of the Mnt repressor consists of two right-handed coiled coils. Nat. Struct. Biol. 6, 755–759 (1999).
46. Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G. & Rutter, W.J. Specific inhibition of nuclear RNA polymerase II by α-amanitin. Science 170, 447–449 (1970).
47. Stiller, J.W. & Cook, M.S. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. Eukaryot. Cell 3, 735–740 (2004).
48. Noble, C.G., Walker, P.A., Calder, L.J. & Taylor, I.A. Rna14-Rna15 assembly mediates the RNA-binding capability of Saccharomyces cerevisiae cleavage factor IA. Nucleic Acids Res. 32, 3364–3375 (2004).
49. Czudnochowski, N., Bosken, C.A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nat. Commun. 3, 842 (2012).
ONLINE METHODS

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

Antibodies. Antibodies to β-actin (cat. no. A-5411), Flag (F1804) (cat. no. F3165), and RPRD1B (cat. no. SA1102247) are from Sigma-Aldrich. Antibody to Gal4 (cat. no. 06-262) is from Millipore. Antibody to GAPDH (cat. no. 386000) and IgG (cat. no. 10500C) are from Invitrogen. Antibody to RPAP2 (cat. no. 17401-1-AP) is from ProteinTech Group. Antibodies to RNAPII (N-20) (cat. no. A300-654A) is from Bethyl Laboratories. Antibody to S5P-RNAPII (C-18, cat. no. SC-85089) are from Santa Cruz. Antibody to S2P-RNAPII (3E8), and S7P-RNAPII (4E12) are gifts from D. Eick. All antibodies were used at a dilution of 1:5,000 in 5% BSA in western blotting and ELISAs, and 2 µg was used 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 have been validated as shown in the group’s publications14.

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 L171-304), and full-length RPA2 and its N-terminal segments (1–334) were cloned into the vectors pET15-MHL, pET28-MHL, pET28-Lic, pET28-GST-Lic, or pGEX-6P-1 with the Infusion kit (In-Fusion Dry-down Mix, Clontech) according to the manufacturer’s instructions. Mutated clones were generated by infusion-mediated mutagenesis with the In-Fusion dry-down mix (Clontech). The identities of all plasmid constructs were verified by sequence analysis. Plasmids were transferred 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 (OD600) 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 by binding buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM imidazole, 5% glycerol, 1 mM protease inhibitor, 1 mM DTT and 6.25 units/ml of benzonase). Undissolved debris was pelleted at 15,000 r.p.m. for 1 h at 4 °C. The His6-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, 150 mM NaCl). The His6-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 with GST resin (Novagen). All proteins were concentrated to 10 to 50 mg/ml in a buffer containing 10 mM HEPES, pH 7.5, and 150 mM NaCl and stored at –80 °C. The selenomethionyl protein sample was produced in E. coli grown in defined medium supplemented with selenomethionine with the M9 SeMET kit (Mediclon, 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 CTD–CTD peptide complexes with a purity of 95% were purchased from PEPTIDE2. The crystals were mixed with proteins in a molar ratio of 3:1. The crystals were cryoprotected in the reservoir solution, supplemented with 10% (v/v) glycerol, as indicated in Table 2 and Supplementary Table 1. All crystallographic data were collected at 18 °C 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)58 with preassembled-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 derivative59. For each model and in several iterations, manual rebuilding in Coot52 was followed by atomic-coordinate and temperature-factor refinement (programs for final refinement listed in Supplementary Table 1) and geometry validation in MolProbity53. Selected geometry restraints for model refinement were prepared with LigPrep54.

Cell cultures, lentivirus infections of VAP-tagged RPRD proteins and siRNAs, and siRNA transfection. HEK293 and HeLa cells were cultured in Dulbecco’s modified Eagle’s medium with 10% FBS and antibiotics. Lentiviruses were produced and used to infect HEK293 cells at a multiplicity of infection <1, as described previously55. Each open reading frame (ORF) was cloned into a lentiviral expression vector so as to attach a VAP tag (containing a 3x Flag–His6–streap triple tag) as described previously56,57 and sequence verified. Lentivirus-encoded VAP-tagged ORFs and siRNAs were transduced into HEK293 cells as described previously56,57. Transduced cells were selected with puromycin (Sigma) at a concentration of 2 µg/ml for a minimum of 48 h. A Gateway-compatible entry clone for RPRD1B (OHS1770-983593) was obtained from the human ORFeome library (Open Biosystems). The following shRNA (obtained from J. Moffat) sequences were used for lentivirus-mediated knockdowns: GFP shRNA, CGACCACTATGAGCAGCAGA; RPRD1A shRNA, CGGGCAAAATAGTAGTATAGAA; RPAP2-1 shRNA, GC CAATTTCAGAGAATTTGAA; RPAP2-2 shRNA, GCAAAGCTTGGTTGTCTCCAT.

Three micrograms of siRNA was transected with 3 µL Lipofectamine 2000 (Invitrogen) into each well of a six-well plate for 24 h; the cells were then transferred to a 10-cm dish. Twenty-four hours later, the same amount of siRNA was transfected again with Lipofectamine 2000. The pool of the following two siRNAs (purchased from Sigma-Aldrich) was used in transfection: RPRD1A-1 siRNA, GGAAGAAAGGUCUGUUUAUdTdT; RPRD1A-2 siRNA, GACUCUGACUGCUUGAAGAGdIdT; Buffer was used as a negative control.

Analytical ultracentrifugation (AUC). For the sedimentation equilibrium experiments, the proteins at various concentrations (OD280 of 1.0, 0.5, and 0.25) in buffer containing 10 mM HEPES, pH 7.5, and 150 mM NaCl were centrifuged at 13,000 and 15,000 r.p.m. at 4 °C in a Beckman Optima XL-A analytical ultracentrifuge with 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 OD280 of 1 were centrifuged at 55,000 r.p.m. at 4 °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) with SEDFIT58. 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 Na2O2, 0.25 mM Na3VO4, and 1× protease inhibitor mixture (Sigma)). This was followed by centrifugation at 14,000 r.p.m. for 1 h at 4 °C to remove insoluble materials. 20 to 100 µg proteins were separated by electrophoresis on a Tris 4–20% SDS-PAGE gel (BioRad) and transferred to nitrocellulose or PVDF membranes. Transferred samples were immunoblotted with primary antibodies followed by horseradish peroxidase-conjugated goat anti-mouse or goat anti-rabbit secondary antibodies (Jackson Lab). Western blot detection was performed with enhanced chemiluminescence (GE Healthcare).

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

Chromatin immunoprecipitation (ChIP). HEK293 (or HeLa cells, Fig. 5e and Supplementary Fig. 5e) cells were cross-linked with 1% formaldehyde (Sigma) at room temperature for 10 min, washed twice with ice-cold PBS, collected in 1 ml of PBS and centrifuged for 5 min at 5,000 r.p.m. 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, 507862) at 4 °C for 15 min (or 10 µl of Dynabeads at 4 °C for 30 min, Fig. 5c and Supplementary Fig. 5c). A 100-µl aliquot of sonicated chromatin was immunoprecipitated (IP) with 2 µg peptides overnight at 4 °C. IP samples were centrifuged at 13,200 r.p.m., and supernatant was incubated with 10 µl of staph A at room temperature for 15 min (or 15 µl of Dynabeads at 4 °C for 1 h, Fig. 5e and Supplementary Fig. 5e). Precipitates were washed sequentially for 3 min in 1× dialysis buffer (2 mM EDTA, 50 mM Tris-HCl, pH 8, and 0.2% sarkosyl) twice, then with 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 NaHCO₃) and then heated at 65 °C overnight to reverse cross links, and DNA fragments were purified with a QIAEX II Gel extraction kit (Qiagen, 2005)). A 4-µl aliquot from a total of 200 µl was used in qPCR with the following primers: RPRD1A promoter forward, TGCTTTTTCGAGATTTCCA and reverse, TGATGCCTGGTTACTCTCA; LEO1 promoter forward, ATACCAGGAGAACAGGT and reverse, GAGCACTGCGCAAAAGATGG; MED12 forward, TCTAGGACAAAAACCGGCTA and reverse, GATGCA AACCGGGAGAAG; CTR9 promoter forward, GGAGTGTGACCG CCAGATG and reverse, GAGGCTTTGTTGGCGTGTC; ACTB (β-ACTIN) forward, GGCGAACCAGGCGGGGTCTTT and reverse, AGCGATTTAGCGCC ACAAAG; U2 snrRNA untranscribed region forward, CAGCTGTGGCTG ACGGAGTGGCTG and reverse, CAGGGCTTCGGCAGCACTCATCA. The yields of ChIP DNA were quantified with the SYBR Green kit (Applied Biosystems) with 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. 5c and Supplementary Fig. 5c) in duplicates. PCRs consisting of 40 cycles of 95 °C for 15 s and 55 °C for 30 s were performed. Qt values were compared with a standard curve, the copy number was calculated, the amount of DNA precipitated by an irrelevant antibody to GAL4 or IgG subtracted, and the percentage ChIP DNA relative to input chromatin calculated.

Isothermal titration calorimetry (ITC) measurements. ITC measurements were recorded at 25 °C with a VP-ITC microcalorimeter (MicroCal). 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 s and a reference power of 250 µl. Binding isotherms were plotted and analyzed with Origin Software (MicroCal). The ITC measurements were fitted to a one-site binding model.

Dynamic light scattering (DLS). DLS experiments were performed with a DynaPro Titan instrument (Wyatt Technologies Corporation) equipped with a 488-nm laser. 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 × 10⁶ counts/s, and data were collected with acquisition time of 10 s and averaged for ten acquisitions. Analysis of the data was performed with Dynamics software.

In vitro DIFMUP phosphatase assay. Recombinant proteins with or without the inhibitors were incubated with 10 µM 6,8-difluoro-4-methylumbelliferophosphate (DIFMUP, Molecular Probes) at 37 °C in black, flat-bottom, polystyrol 96-well plates in 100 µl reaction buffer (50 mM Na, 3,dimethylglutaric acid, pH 6.5, and 150 mM NaCl), and fluorescence was measured every 20 min for 2 h with a 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, 31000)-coated Maxisorp plates (Thermo Scientific, 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 MgCl₂, 20 mM KCl and 5 mM DTT at 37 °C. After washing with PBS plus 0.5% Tween-20, antibodies were added and incubated at room temperature for 1 h, and this was followed by incubation with horseradish peroxidase–conjugated goat anti-mouse or mouse anti-rabbit secondary antibodies (Jackson Lab, cat. no. 115-035-174 and 211-032-171, respectively), the addition of TMB peroxide substrate (Thermo Scientific, 34021) and the measurement of absorbance signals at 450 nm.

Size-exclusion chromatography. Size-exclusion chromatography was carried out with a Superdex 75 HR 26/60 or HiloLoad 16/60-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, 151-1901) was used. All experiments were performed at 4 °C. Proteins were detected by absorbance measured at 280 nm.

GST pulldown. GST-and histidine-tagged proteins expressed in E. coli. were sonicated and incubated overnight at 4 °C. This was followed by pulldown with glutathione beads. The beads were then washed four times with suspension buffer (20 mM Tris, pH 7.5, 400 mM NaCl, and 5% glycerol) and subsequently eluted with elution buffer (20 mM Tris, pH 7.5, 400 mM NaCl, and 10 mM reduced glutathione). Eluted materials were resolved on SDS-PAGE gels, and proteins were detected by anti-histidine antibodies or anti-RPRD1A or anti-RPRD1B antibodies.

Protein cross-linking. Suberic acid-bis-(3-sulfo-N-hydroxy succinimide ester) (BS3, ProteoChem) 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 h, and this was followed by quenching of the unreacted BS3 with 40 mM Tris for 15 min at room temperature. Proteins were separated by Tris 4–20% SDS-PAGE gels (BioRad) and stained with the Coomassie blue reagent Instant Blue (Expecedone).

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 with GraphPad Prism 6.0 for Windows, in which the P value was determined with 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.