Structure of the tripartite motif of KAP1/TRIM28 identifies molecular interfaces required for transcriptional silencing of retrotransposons

Guido A. Stoll[1], Shun-ichiro Oda[1], Zheng-Shan Chong[1], Minmin Yu[2], Stephen H. McLaughlin[2] and Yorgo Modis[1,*]

[1] Molecular Immunity Unit, Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge, CB2 0QH, UK

[2] MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge, CB2 0QH, UK

* Correspondence: ymodis@mrc-lmb.cam.ac.uk (Y.M.)

Abstract

Transcription of transposable elements is tightly regulated to prevent damage to the genome. The family of KRAB domain-containing zinc finger proteins (KRAB-ZFPs) and KRAB-associated protein 1 (KAP1/TRIM28) play a key role in regulating retrotransposons. KRAB-ZFPs recognize specific retrotransposon sequences and recruit KAP1, which controls the assembly of an epigenetic silencing complex including histone H3K9 methyltransferase SETDB1. The chromatin remodeling activities of this complex repress transcription of the targeted transposable element and any adjacent genes. Here, we use biophysical and structural approaches to show that the tripartite motif (TRIM) of KAP1 forms antiparallel dimers, which further assemble into tetramers and higher-order oligomers in a concentration-dependent manner. Structure-based mutations in the B-box 1 domain prevented higher-order oligomerization without a significant loss of retrotransposon silencing activity in a cell-based assay, indicating that, in contrast to other TRIM family members, self-assembly is not essential for the function of KAP1. The crystal structure of the KAP1 RBCC dimer identifies the KRAB domain binding site, in the coiled-coil domain near the dyad. Mutations at this site abolished KRAB binding and transcriptional silencing activity of KAP1. This work identifies the interaction interfaces in the KAP1 RBCC motif responsible for self-association and KRAB binding and establishes their role in retrotransposon silencing.

Keywords: Transcriptional repressor; epigenetic silencing; histone H3 lysine 9 methylation (H3K9me3); endogenous retrovirus; retrotransposon; transposable element; tripartite motif (TRIM); SUMO; ubiquitin; E3 ligase

Introduction

Retrovirus genomes that integrate into the genome of germline cells are inherited by future generations. These endogenous retroviruses (ERVs) can retain the ability to replicate by transcriptional amplification and expression of the viral reverse transcriptase and integrase, which convert the genome transcripts into DNA and reintegrate it into the host genome. This amplifying retrotransposition mechanism has allowed ERVs, and other retrotransposons such as LINEs (long interspersed nuclear elements), to accumulate, accounting for over half of the human genome[1]. Approximately 100 human LINEs are still replication-competent and cause new integration events in 2-5% of the population[2,3].

Some ERVs and other transposable elements (TEs) have evolved to fulfill important cellular functions. TEs drive the evolution of transcriptional networks by spreading transcription factor binding sites, promoters and other regulatory elements[1,4,5]. TE-derived regulatory elements are particularly important in embryogenesis, when global hypomethylation promotes transcription. A significant fraction of pluripotency-associated transcription factor binding sites is located in TEs, which hence contribute to maintaining the pluripotency of embryonic cells[1]. TEs also serve as a reservoir of genes that can be co-opted by the host. For example, TE-derived proteins catalyze V(D)J recombination[6,7] and syncytiotrophoblast fusion in placental development[1,8].

At the same time, transcription of TEs must be tightly regulated to prevent pathogenesis. Disruption of protein coding sequences by transposition events can cause genetic disorders such as hemophilia and cystic fibrosis[9,10]. TE reactivation in somatic cells is associated with cancer, through disruption of tumor suppressor genes or enhanced transcription of oncogenes[10,11]. Accumulation of TE-derived nucleic acids is associated with autoimmune diseases including geographic atrophy, lupus and Sjögren's syndrome[2,12]. A key source of retrotransposon repression is the family of Krüppel-associated box zinc-finger proteins (KRAB-ZFPs) and KRAB-associated protein 1 (KAP1, also known as TRIM28 or TIF1β)[13]. KRAB-ZFPs, the largest family of mammalian transcription factors, recognize retrotransposons with a variable C-terminal array of zinc fingers[14,15]. The conserved N-terminal KRAB domain recruits KAP1[13], which serves as a platform for the assembly of a transcriptional silencing complex of repressive chromatin-modifying enzymes including SETDB1, a histone H3K9 methyltransferase, and the nucleosome remodeling and deacetylase (NuRD) complex[16].

The repressive histone modifications established by the KAP1 complex in early embryos promote cytosine methylation, a more permanent silencing mark[17,18]. However, certain retrotransposons

and genes adjacent to them only remain repressed if the cognate KRAB-ZFP remains bound[19]. Variations in KRAB-ZFP expression with cell type or during development can then generate tissue- or developmental stage-specific transcriptional regulation rather than simple silencing[20].

KAP1 is an 835-amino acid, 89-kDa protein from the tripartite motif (TRIM) family (**Fig. 1a**). Residues 57-413 contain the defining feature of the TRIM family: an RBCC motif consisting of a RING domain, two B-box-type zinc fingers and a coiled-coil domain. The RING domain has ubiquitin E3 ligase activity[21]. This activity can be directed to tumor suppressors AMPK and p53 by the MAGE proteins, which are overexpressed in human cancers[21,22]. The resulting proteasomal degradation of AMPK and p53 has been implicated in tumorigenesis[21,22]. The central region of KAP1 contains a PxVxL motif that recruits Heterochromatin Protein 1 (HP1) and is essential for transcriptional silencing[23]. The C-terminal region of KAP1 contains a PHD-bromodomain tandem (residues 624-812). The PHD domain recruits Ubc9, a SUMO E2 ligase, to SUMOylate several lysines in the bromodomain, in an intramolecular SUMO E3 ligase reaction[24,25]. KAP1 SUMOylation is required for recruitment and activation of SETDB1 and recruitment of NuRD[23-25]. KAP1 has also been reported to SUMOylate IRF7[26]. The KAP1 RING domain was found to be necessary (but not sufficient), and the PHD domain dispensable for SUMOylation of IRF7[26]. Similarly, the RING domain of PML/TRIM19 is required (but not sufficient) for the SUMO E3 ligase activity of PML[27,28].

The KAP1 RBCC motif has been reported to form homotrimers and bind KRAB domains with a stoichiometry of 3:1 KAP1:KRAB[29,30]. However, this is inconsistent with more recent reports that TRIM5, TRIM25 and TRIM69 form antiparallel dimers, a property predicted to be conserved across the TRIM family[31-34]. Moreover, various TRIMs (TRIM5, PML/TRIM19, TRIM32) further assemble into tetramers and higher-order oligomers, including two-dimensional lattices and molecular scaffolds (as seen in PML bodies), and these higher-order assemblies are important for their biological activities[33,35-37]. RING domains, including those of TRIM5α and TRIM32, form dimers, which contribute to E3 ligase activity by priming the associated E2 ubiquitin ligase for ubiquitin transfer[36,38,39]. The B-box 1 domain of TRIM19[37] and B-box 2 domain of KAP1 (PDB: 2YVR) both also form dimers. Here, we use biophysical and structural approaches to show that KAP1 forms antiparallel dimers, which further assemble into tetramers and higher-order oligomers in a concentration-dependent manner. Point mutants defective in higher-order oligomerization were not significantly impaired in their retrotransposon silencing activity in a cell-based assay, indicating that self-assembly is not essential for the silencing function of KAP1. In contrast, mutation of conserved residues in the coiled-coil domain inhibited KRAB binding and

transcriptional silencing. This work identifies the interaction interfaces in the KAP1 RBCC motif responsible for self-association and KRAB binding and establishes their role in retrotransposon silencing.

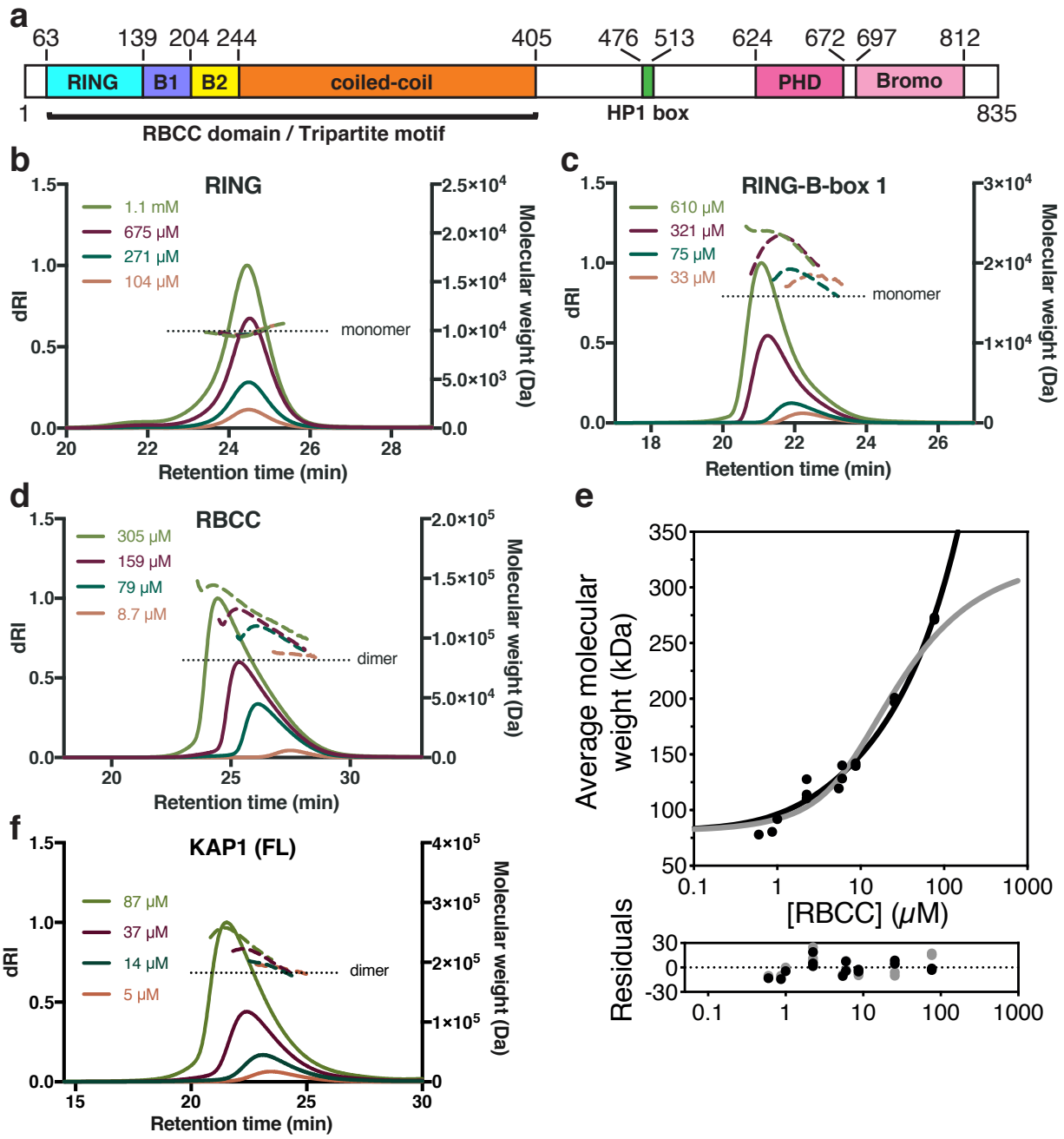

Fig. 1. Self-assembly in solution of the KAP1 RBCC motif and its subcomponent domains.

(**a**) Domain organization of KAP1. B-Boxes 1 and 2 are labeled B1 and B2, respectively.

(**b**) Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) of KAP1 RING at various concentrations.

(**c**) SEC-MALS of the KAP1 RING-B-Box 1 fragment at various concentrations.

(**d**) SEC-MALS of the KAP1 RBCC motif at various concentrations.

(**e**) Sedimentation equilibrium analytical ultracentrifugation (SE-AUC) analysis of KAP1 RBCC molecular weight as a function of protein concentration. The average molecular weight isotherm from individual fits of analytical centrifugation at different concentrations was fitted to an isodesmic self-association model (black line) yielding dissociation constants of $K_{d2,4}$ = 9 μM and K_{diso} = 19 μM for the dimer-tetramer and isodesmic equilibria, respectively. Alternatively, the average molecular weight isotherm from individual fits of analytical centrifugation at different concentrations could be fitted to a dimer-tetramer-octamer model (grey line) giving dissociation constants of $K_{d2,4}$ = 12.6 μM and $K_{d4,8}$ = 6 μM for the dimer-tetramer and tetramer-octamer equilibria, respectively.

(**f**) SEC-MALS of full-length KAP1 at various concentrations.

Results

KAP1 forms dimers that self-assemble into higher-order oligomers

Each of the domains within the RBCC motif (RING, B-box 1, B-box 2 and coiled coil) has been reported to independently form dimers or oligomers in TRIM-family proteins[31-37]. The presence of two or more domains capable of oligomerization independently can lead to polymerization of TRIMs into lattices or scaffolds[33,35,36], but it remains unclear whether this applies to KAP1. To assess the self-assembly potential of KAP1, we purified the KAP1 RBCC motif and some of its constituent domains to determine their hydrodynamic properties. Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) showed that the RING domain was monomeric at concentrations ranging from 0.1 to 1.1 mM (1-10.5 g L^{-1}) (**Fig. 1b**). A fragment containing the RING and B-box 1 domains was mostly monomeric at 33 μM (0.5 g L^{-1}) but as the protein concentration was increased to 610 μM (9.7 g L^{-1}) the average hydrodynamic radius and apparent molecular weight of the protein increased by up to 33% (**Fig. 1c**), indicating that monomers and dimers (or higher order oligomers) were in dynamic equilibrium with each other in

solution. We conclude that B-box 1 drives assembly of RING-B-box 1 fragment into weakly associated dimers, with a dissociation constant in the low micromolar range.

SEC-MALS data for the whole RBCC fragment showed unambiguously that it was dimeric at low protein concentrations (8.7 μM, 0.35 g L^{-1}), and that its hydrodynamic radius and apparent molecular weight increased by up to 68% as the protein concentration was increased to 0.3 mM (12.4 g L^{-1}) (**Fig. 1d**). The isolated KAP1 B-box 2 crystallized as a dimer with a surface area of 1,160 $Å^2$ buried at the dimer interface (PDB: 2YVR). Together, these data suggest that the coiled-coil domain of KAP1 forms tight homodimers, like most if not all other members of the TRIM family, and that KAP1 dimers can self-assemble through further weak homotypic interactions between the B-boxes to form higher-order oligomers.

To obtain a more direct and quantitative model of KAP1 self-association we performed sedimentation equilibrium analytical ultracentrifugation (SE-AUC) on KAP1 RBCC at concentrations from 1 to 200 μM. The equilibrium sedimentation profiles were consistent with a dynamic equilibrium between dimeric and oligomeric KAP1 (**Supplementary Fig. 1**), in agreement with the SEC-MALS data. The average molecular mass at the lowest concentration was approximately 80 kDa (**Fig. 1e**), consistent with the formation of a dimer with a subnanomolar affinity. Further oligomerization was evident as the concentration increased. The observed increase in average molecular weight of KAP1 oligomers with increasing protein concentration at sedimentation equilibrium could be explained with two alternative models of self-association. The model with the best fit was an isodesmic self-association model in which KAP1 dimer to tetramer association is followed by unlimited consecutive additions of dimers (**Fig. 1e**). A simpler $4R_2 \rightarrow 2R_4 \rightarrow R_8$ model with dimers, tetramers and octamers in dynamic equilibrium produced a fit of similar quality (**Fig. 1e**). In support of the isodesmic model, the weight-average fit residuals were slightly lower at the highest protein concentrations than for the dimer-tetramer-octamer model (**Fig. 1e**). However, the improved fit of the isodesmic model could stem from the greater number of parameters versus the dimer-tetramer-octamer model. Both models yielded a dimer-tetramer dissociation constant $K_{d2,4}$ and higher-order dissociation constants ($K_{d4,8}$ and K_{diso}) of the order of 10 μM (6-19 μM). Full-length KAP1 self-assembled in a similar manner indicating that higher-order oligomerization is not an artifact of isolating the RBCC domain (**Fig. 1f**). We conclude that KAP1 forms tight dimers, which can associate into tetramers and octamers at high local concentration of KAP1. KAP1 may also form higher-order species but our SE-AUC data cannot definitively confirm or rule out the presence of KAP1 species larger than octamers.

Crystal structure of the KAP1 RBCC tripartite motif (TRIM)

To understand the molecular basis of KAP1 self-assembly and identify its oligomerization interfaces, we determined the crystal structure of KAP1 RBCC. Although crystals of KAP1 RBCC were readily obtained, they initially diffracted X-rays poorly. Crystals suitable for structure determination were obtained by fusing bacteriophage T4 lysozyme (T4L) to the N-terminus of KAP1 RBCC and methylating primary amines in the purified protein prior to crystallization (see **Materials and Methods**). Diffraction was anisotropic, with data up to 2.63 Å resolution but with overall completeness falling below 90% at 3.9 Å resolution (**Table 1**). The structure was determined with the single anomalous dispersion (SAD) method using the anomalous scattering signal from the zinc atoms in the RING and B-boxes. The asymmetric unit contained two KAP1 RBCC molecules. The atomic models for T4L and KAP1 B-box 2 were docked into the phased electron density and the rest of the atomic model was built *de novo* using available structures of other TRIMs[31,33,34,36,37] as guides (**Supplementary Fig. 2**).

Table 1. Crystallographic data collection and refinement statistics.

Data collection	T4 lysozyme-KAP1 RBCC		
X-ray source	DLS I03		
Space group	$C222_1$		
Cell dimensions a, b, c (Å) α = β = γ (°)	59.77, 169.3, 374.5 90		
Wavelength (Å)	1.28189		
Resolution (Å)	187–2.63	84.1–3.9[b]	3.17–2.63
Observations	156,326	125,590	5,596
Unique reflections	33,980	15,929	924
R_{merge} [b]	0.078	0.088	0.832
R_{pim} [c]	0.029	0.034	0.363
$\langle I \rangle / \sigma I$	13.4	8.14	2.2
Spherical completeness (%)	32.4	89.2	3.9
Ellipsoidal completeness (%)	86.6	–	61.7
Multiplicity	8.5	7.9	6.1
CC(1/2)	0.998	0.994	0.820
SAD Phasing			
CC_{ano}	0.479		
$\lvert D_{ano} \rvert / \sigma D_{ano}$	0.918		
Overall figure of merit	0.67		
Refinement			
Resolution (Å)	63–2.9		

R_{work} / R_{free} [d]	0.261 / 0.291
No. of non-H atoms	
Protein	6955
Zn^{2+} Ions	8
Solvent	0
No. riding H atoms	6908
Mean B-factor (Å^2)[e]	111
MolProbity Clashscore	7.35
RMS[f] deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.859
Ramachandran plot	
% favored	93.2
% allowed	6.45
% outliers	0.35
PDB code	PDB: 6QAJ

[b]Dataset reprocessed at 84.1-3.9 Å resolution with CCP4 (MOSFLM, AIMLESS)

[b]$R_{sym} = \Sigma_{hkl}\Sigma_i |I_{hkl,i} - \langle I\rangle_{hkl}| / \Sigma_{hkl}\Sigma_i|I_{hkl,i}|$, where I_{hkl} is the intensity of a reflection and $\langle I\rangle_{hkl}$ is the average of all observations of the reflection.

[c]$R_{pim} = \Sigma_{hkl} (N_{hkl} - 1)^{-1/2} \times \Sigma_i |I_{hkl,i} - \langle I\rangle_{hkl}| / \Sigma_{hkl}\Sigma_i|I_{hkl,i}|$, where I_{hkl} is the intensity of a reflection and $\langle I\rangle_{hkl}$ is the average of all observations of the reflection.

[d]R_{free}, R_{work} with 5% of F_{obs} sequestered before refinement.

[e]Residual B-factors after TLS refinement. See PDB entry for TLS refinement parameters.

[f]R.M.S., root mean square.

The overall structure of the KAP1 RBCC dimer resembles a dumbbell (**Fig. 2**). The coiled-coil domain forms a 16 nm long antiparallel coiled coil that contains all the dimer contacts. The ends of the coiled coil are capped by a B-box 2 domain. The RING domains (residues 63-138) are bound to one side of the coiled coil, close to but not in contact with the B-box 2 domain (residues 204-243) from the same subunit. Unexpectedly, there was no interpretable electron density for B-box 1 (residues 139-203), indicating that its position relative to the other domains is variable and does not obey the crystallographic symmetry. The T4L is rigidly linked to the RING domain via a continuous fused α-helix consisting of residues 158-162 from T4L (numbered 51 to 55 in the structure) and residues 56-62 from KAP1. The only other significant contacts between the N-terminal fusion region and the KAP1 RBCC are through the tobacco etch virus (TEV) protease cleavage site, which precedes the T4L and, atypically, is mostly ordered in the structure (**Fig. 2**). The TEV cleavage sequence is sandwiched in an extended conformation between the T4L and the coiled coil, forming multiple polar and hydrophobic contacts with both domains. Although not

physiologically relevant, these contacts appear to stabilize the crystal lattice by constraining the orientation of the T4L relative to KAP1 RBCC. The T4L also forms extensive crystal packing contacts, consistent with the improved diffraction properties of the T4L-RBCC crystals versus crystals of the RBCC alone.

The coiled-coil domain forms a helical hairpin consisting of a 15.3 nm long α-helix (residues 244-348) followed by a turn and a shorter partially helical segment (residues 357-405). The first helical segment contains the majority of the dimer contacts, mostly hydrophobic leucine zipper-type coiled-coil interactions with the first helical segment from the other subunit. The second segment packs against the first to form a four-helix bundle around the twofold axis of the dimer, where the second segments from the two subunits overlap (**Fig. 2**), and a three-helix bundle at the distal ends of the dimer, where the second segments do not overlap. The central portion of the second segment has poor electron density indicating a relatively high level of conformational flexibility. The coiled-coil domain is structurally most similar to the coiled-coil domain of TRIM25[33,34] (Rmsd 2.6 Å), which forms a dimeric antiparallel coiled coil with the same fold and similar length and curvature. TRIM5α forms a dimeric coiled coil with the same fold and length but lower curvature[31] (Rmsd 3.9 Å), and TRIM69 forms a dimeric coiled coil with different secondary structure[32] (Rmsd 4.0 Å).

The KAP1 RING and B-box 2 do not form dimers in the RBCC crystal structure

RING domains of E3 ubiquitin ligases recruit ubiquitin-conjugated E2 ligases to the substrate and prime ubiquitin transfer from the E2 ligase to the substrate by stabilizing the E2-ubiquitin intermediate in a closed state competent for transfer[39,40]. This stabilization has been proposed to be dependent on RING dimerization in TRIM5α, TRIM25, TRIM32 and BIRC-family E3 ligases[36,38,39]. The most similar RING domain structure to the KAP1 RING domain is that of TRIM32[36] (Rmsd 1.9 Å). However, in contrast to the TRIM32 RING, which forms dimers in solution with α-helices flanking the core RING domain[36], the KAP1 RING domains in the RBCC dimer are located on opposite ends of the coiled-coil domain and do not form any homotypic contacts. The KAP1 RING domains form crystal contacts with the coiled-coil domain (but not the RING domain) of a neighboring RBCC dimer. Similarly, the KAP1 B-box 2 domain also does not form homotypic contacts in the RBCC dimer. The B-box 2 domain does form crystal contacts with B-box 2 domains from two different neighboring RBCC dimers in the crystal lattice, but these homotypic contacts are distinct from those formed by the isolated B-box 2 in solution. The latter are moreover

incompatible with the RBCC dimer structure as the coiled-coil domain blocks the B-box 2 surface that mediates dimerization of the isolated B-box 2.

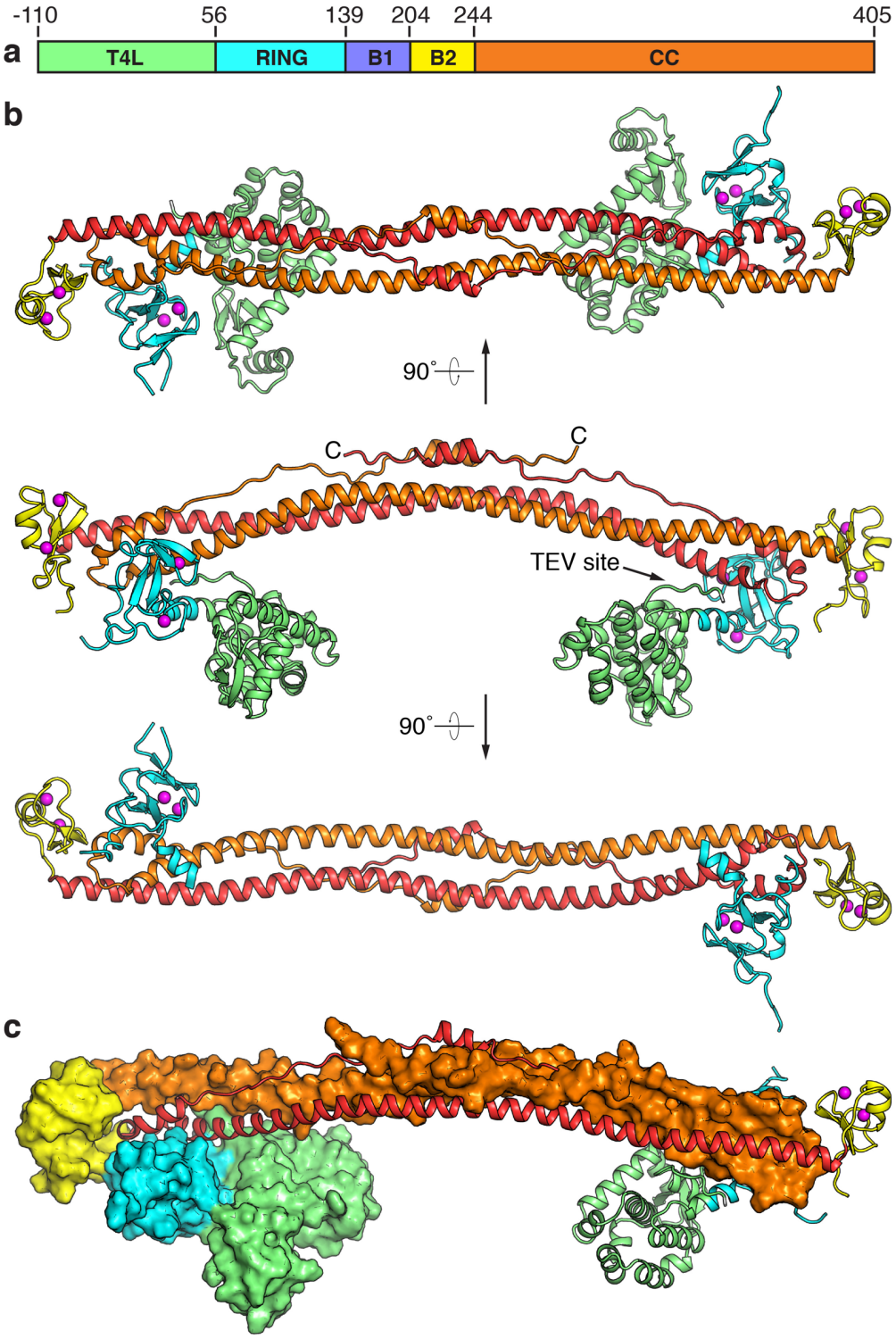

Fig. 2. Crystal structure of KAP1 RBCC.

(**a**) Domain organization of the crystallized KAP1 RBCC construct. T4L, T4 lysozyme; B1, B-Box 1; B2, B-box 2; CC, coiled-coil.

(**b**) Overall structure of the RBCC homodimer. Three views along or perpendicular to the dyad are shown. The components are colored as follows in (**a**). Zinc atoms are shown as magenta spheres.

(**c**) View of the RBCC dimer perpendicular to the dyad, with one subunit shown in cartoon representation and the other in surface representation.

Higher-order assembly of KAP1 dimers is dependent on B-box 1 interactions

Our hydrodynamic data indicate that KAP1 dimers self-assemble into tetramers, octamers and higher-order oligomers through one or both of the B-boxes. To identify the sites responsible for higher-order oligomerization of KAP1, we designed structure-based mutations in the B-boxes aimed at disrupting potential dimer contacts (**Fig. 3**). In B-box 2, a cluster of residues involved in homotypic crystal contacts was mutated, yielding the variant N235A/A236D/K238A/D239A/F244A/L245A (**Fig. 3a**). Residues in the RING domain forming crystal contacts (with the coiled-coil domain) were mutated in a second variant, V114A/Q123A/F125A/K127A (**Fig. 3a**). Since B-box 1 was disordered in the RBCC structure, we mutated residues predicted to be involved in B-box 1 dimerization based on a structural model of the KAP1 B-box 1 dimer generated from the TRIM19 B-box 1 dimer structure (PDB: 2MVW)[37], yielding the variant A160D/T163A/E175R (**Fig. 3f**). The oligomerization potential of each of these variants was then assessed by SEC-MALS. Mutations in the RING and B-box 2 domains did not alter the self-assembly properties of KAP1 RBCC (**Fig. 3b,c**). The latter was unexpected since a crystal structure of the isolated KAP1 B-box 2 was dimeric (PDB: 2YVR). In contrast, the B-box 1 mutations abolished oligomerization of the RING-Box-1 fragment and almost completely inhibited higher-order oligomerization of KAP1 RBCC dimers (**Fig. 3d,e**). We conclude that the assembly of KAP1 RBCC dimers observed at high protein concentration occurs primarily through dimerization of B-box 1 domain.

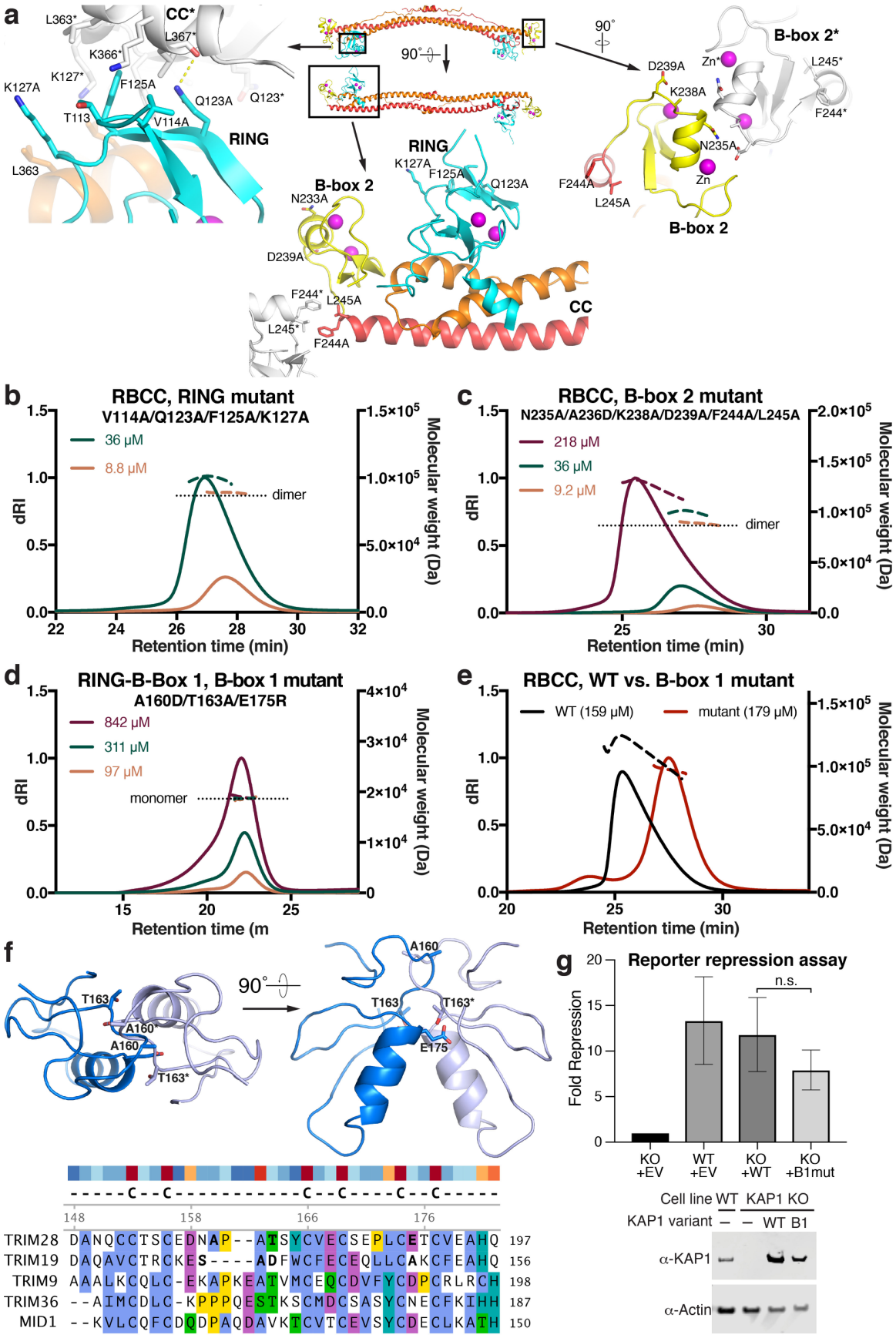

Fig. 3. Self-assembly properties and KRAB binding activity of site-directed mutants of KAP1 RBCC.

(**a**) Position of the mutations RING and B-box 2 domains within the RBCC dimer structure. Any residues that were mutated are shown in ball-and-stick representation. A reference RBCC dimer is colored as in **Fig 2**. Adjacent RBCC dimers forming crystal packing contacts are shown in light grey with their residue numbers followed by an asterisk.

(**b**) SEC-MALS of KAP1 RBCC with RING domain mutations V114A/Q123A/F125A/K127A at two protein concentrations.

(**c**) SEC-MALS of KAP1 RBCC with the B-box 2 mutations N235A/A236D/K238A/D239A/F244A/L245A at various concentrations.

(**d**) SEC-MALS of the RING-B-box 1 fragment with B-Box 1 mutations A160D/T163A/E175R at various concentrations.

(**e**) SEC-MALS of WT KAP1 RBCC (black curve) and of RBCC with B-Box 1 mutations A160D/T163A/E175R (red curve).

(**f**) Model of a TRIM28 B-box 1 dimer based on the TRIM19 B-box 1 dimer structure (PDB: 2MVW)[37] with selected residues near the twofold axis forming dimer contacts shown in ball-and-stick representation. Residue labels in the second protomer are followed by an asterisk. A multiple sequence alignment of selected B-box 1 sequences (bottom) was used together with the TRIM28 B-box 1 model and the TRIM19 B-box 1 structure to identify mutations in the KAP1 B-box 1 likely to participate in homodimer contacts (A160D/T163A/E175R). Residues known or predicted to participate in dimer contacts are shown in bold in the sequence alignment.

(**g**) Cell-based transcriptional silencing assay with the oligomerization-deficient B-box 1 variant. Data are presented as fold-repression of reporter luciferase luminescence expressed from an SVA retrotransposon in KAP1 KO HEK293T cells transfected with WT KAP1 or with variants bearing the B-box 1 domain mutations, A160D/T163A/E175R (B1mut), as in panels **e,f**. Data were normalized to KAP1 KO cells transfected with an empty vector (EV), which were defined as having fold-repression = 1. Error bars represent standard error of the mean between measurements; n = 5. Bottom: Western blots of cell lysates from WT or KAP1 KO HEK293T cells transfected with WT KAP1 (WT), B-box 1 mutant KAP1 (B1), or empty vector (-). The blots were probed with anti-KAP1 or anti-actin primary antibodies.

Self-assembly of KAP1 RBCC dimers is not required for retrotransposon silencing

Various TRIM proteins assemble into higher-order oligomers, two-dimensional lattices or molecular scaffolds that are important for physiological function[33,35-37]. To determine whether self-assembly of KAP1 into higher-order oligomers is required for its retrotransposon repression, we assayed the transcriptional silencing activities of wild type KAP1 and the oligomerization-deficient mutant. We used a reporter assay in which an SVA-D (SINE–Variable number tandem repeat–Alu, type D) retrotransposon cloned upstream of a minimal SV40 promoter strongly enhances firefly luciferase activity unless the KRAB-ZFP ZNF91 and KAP1 are both present to repress the SVA-D element[15]. The assay was adapted for use in KAP1-knockout (KO) HEK 293T cells[41], which were co-transfected with the reporter plasmid and plasmids encoding ZNF91, KAP1 (WT or mutant), and *Renilla* luciferase under a constitutive promoter. Firefly luciferase luminescence from the reporter was normalized against the co-transfected *Renilla* luciferase. Unexpectedly, the A160D/T163A/E175R B-box 1 variant had a similar repression activity as WT, despite being deficient in dimer-dimer assembly (**Fig. 3g**). We therefore conclude that self-assembly of KAP1 RBCC dimers into higher-order oligomers is not required for KAP1-dependent transcriptional silencing of the SVA retrotransposon under the assay conditions tested.

Conserved residues in the CC domain bind a KRAB domain and are required for silencing

The primary function of the RBCC domain of KAP1 in silencing is to bind the KRAB domains of KRAB-ZFPs and hence recruit KAP1 to its genomic targets. KRAB:KAP1 complexes were previously reported as containing one KRAB molecule and three KAP1 molecules[29,30]. However, this seemed unlikely given that KAP1 is dimeric so we decided to reexamine the composition of KRAB:KAP1 complexes. To this end, a complex of KAP1 and the KRAB domain from ZNF93, a KRAB-ZFP that binds to a LINE-1 element known to be silenced by KAP1[15] was reconstituted by coexpressing both proteins in *E.coli*. SEC-MALS analysis showed that KAP1 and the KRAB domain formed a stable complex, which retained the same ability to self-assemble into higher-order oligomers as KAP1 alone (**Supplementary Fig. 3a**). The average molecular mass derived from SEC-MALS for the KAP1-KRAB complex of 226 kDa was inconsistent with a 1:3 KRAB:KAP1 stoichiometry and suggested instead the stoichiometry of the complex was 1:2 KRAB:KAP1 (236 kDa theoretical molecular weight; **Fig. 4a**).

Having established that each RBCC dimer binds a single KRAB domain, we reasoned that the interaction interface must be located on the dyad, in the central region of the KAP1 coiled-coil domain, as every other location would result in two equivalent binding sites (and a 2:2

stoichiometry). Intriguingly, examination of our KAP1 RBCC crystal structure revealed a cluster of solvent-exposed hydrophobic residues near the twofold axis (V293, M297, L300; **Fig. 4b**). Moreover, these amino acids are conserved in KAP1 but not present in other TRIM family members (**Supplementary Fig. 4**). To determine whether this region of the coiled-coil domain mediates KRAB binding, we designed the variant V293S/K296A/M297A/L300S (**Fig. 4b**). KAP1 RBCC domain bearing these mutations failed to bind to MBP-KRAB in a pulldown assay (**Fig. 4c**). Other properties of KAP1 such as dimerization and higher-order oligomerization were unaffected (**Supplementary Fig. 3b**), indicating that the mutation did not interfere with the overall fold of the RBCC domain. Notably, the thermal stability of this variant as assessed by differential scanning fluorimetry (DSF) was significantly higher than that of WT RBCC domain, further supporting a functional role of these residues (**Supplementary Fig. 3c**). A second variant with mutations on the dyad on opposite face of the coiled-coil domain, F391A/L395S/W398A, was mostly insoluble. The lack of binding of the V293S/K296A/M297A/L300S variant to MBP-KRAB was confirmed using surface plasmon resonance (SPR). While WT RBCC domain bound MBP-tagged KRAB domain with high affinity (8 nM K_d), binding was completely lost in the coiled-coil mutant (**Fig. 4d**).

The effect of the coiled-coil mutations on the transcriptional silencing activity of KAP1 was measured using the reporter assay described above. Consistent with its inability to bind KRAB the V293S/K296A/M297A/L300S variant lost all repression activity (**Fig. 4e**), while being expressed at similar concentrations as WT KAP1 (**Fig. 4f**). These data indicate that KRABs bind KAP1 on the twofold axis of the RBCC dimer on a surface that includes residues V293/K296/M297/L300 and that this binding surface is required for KAP1 repression activity.

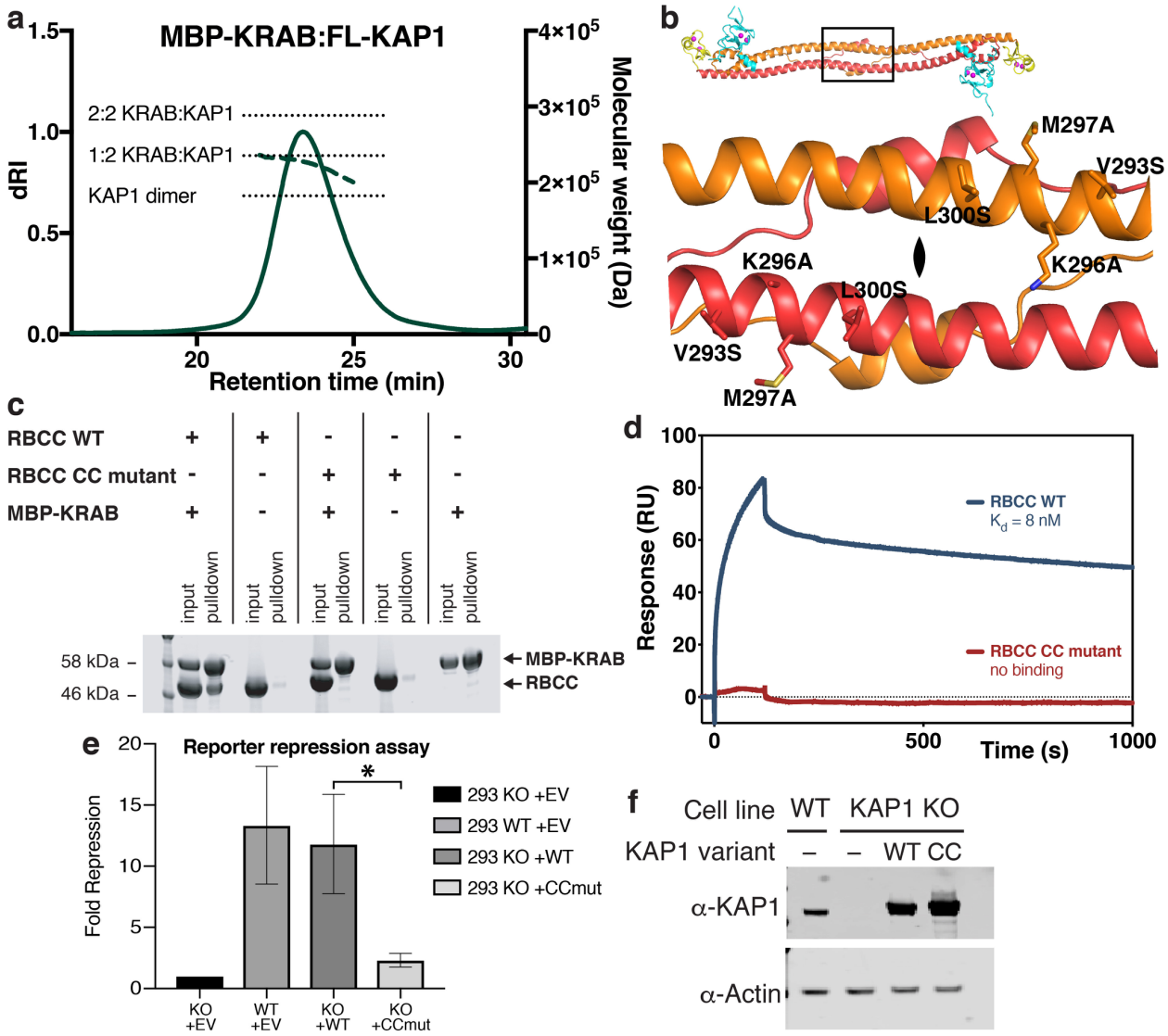

Fig. 4. Formation of a 2:1 KAP1:KRAB complex and identification of KRAB binding residues in the KAP1 coiled-coil domain required for silencing.

(**a**) SEC-MALS of full-length KAP1 in complex with MBP-KRAB from ZNF93. The expected molecular weights of a KAP1 dimer, a 2:1 KAP1:KRAB complex and a 2:2 KAP1:KRAB complex are indicated with dashed lines. The total protein concentration of the loaded sample was 1.2 g L^{-1}.

(**b**) Closeup of the cluster of solvent-exposed hydrophobic residues near the twofold axis (V293, M297, L300). The variant V293S/K296A/M297A/L300S was generated to test for KRAB binding.

(**c**) Pulldown KAP1-KRAB binding assay. KAP1 RBCC was incubated with Twin-StrepII-MBP-ZNF93 KRAB. The StrepII-tagged KRAB domain was immobilized as bait on Strep-Tactin

Sepharose resin. The resin was washed and KAP1 RBCC binding was assayed by SDS-PAGE and Coomassie staining.

(**d**) Surface plasmon resonance (SPR) KAP1-KRAB binding assay. MBP-KRAB was immobilized onto the chip. WT KAP1 and the V293S/K296A/M297A/L300S coiled-coil variant were flowed over the chip in the mobile phase. Binding kinetics of WT RBCC: k_{on} = 3.6 x 10^4 M^{-1} s^{-1}; k_{off} = 2.7 x 10^{-4} s^{-1}.

(**e**) Transcriptional silencing assay with the coiled-coil variant. Data are presented as fold-repression of reporter luciferase luminescence expressed from an SVA retrotransposon in KAP1 KO HEK293T cells transfected with WT KAP1 or with the V293S/K296A/M297A/L300S coiled coil variant (CCmut). Data were normalized relative to KAP1 KO cells transfected with an empty vector (EV). The first three data points (293 KO +EV, 293 WT +EV, 293 KO +WT) are the same as the first three data points in **Fig. 3g**. Error bars represent standard error of the mean; n = 5.

(**f**) Western blots of cell lysates from WT HEK293T cells (WT) or KAP1 KO HEK293T cells (KO) transfected with WT KAP1 (WT), coiled-coil mutant KAP1 (CC) or with an empty vector (-). The blots were probed with anti-KAP1 or anti-actin primary antibodies.

Discussion

KAP1 is critical for preserving genome integrity by repressing potentially harmful transcription of retrotransposons. Following its recruitment by KRAB-ZFPs to retrotransposons, KAP1 induces their epigenetic silencing by coordinating the assembly of a large repressor complex comprising the histone methyltransferase SETDB1, HP1 and NuRD. Initial biophysical studies on the KAP1 RBCC domain suggested that it formed trimers, both in isolation and in complex with a KRAB domain[29,30]. More recent work on other TRIMs, in contrast, demonstrated that the coiled-coil domains of various TRIMs mediate formation of antiparallel dimers and suggests that this property is conserved across the entire TRIM family[31-34]. Our SEC-MALS and SE-AUC data establish that KAP1 is dimeric rather than trimeric, and that dimers self-assemble into tetramers, octamers and possibly higher-order species at high local concentration of KAP1. Notably, the cooperativity of self-assembly implies that octamers are slightly more stable than tetramers, which could in principle promote the formation of large oligomeric assemblies of KAP1 in the nucleus. This concentration-dependent self-association of KAP1 is primarily encoded by B-box 1 and is relatively weak with a dissociation constant in the low micromolar range. The lack of a discrete

endpoint oligomeric state (e.g. tetramers) led us to hypothesize that the RBCC may be able to self-associate into a lattice or molecular scaffold as observed for other TRIMs[33,35-37]. We therefore investigated the possibility that KAP1 self-assembly may contribute to its transcriptional silencing function. Conceivably, formation of large oligomeric assemblies of KAP1 at its genomic target loci might amplify silencing activity by increasing the number of recruited repressive chromatin-modifying molecules such as SETDB1. However, KAP1 mutants deficient in self-assembly repressed transcription from an SVA reporter retrotransposon with an efficiency similar to WT KAP1, indicating that at least in this setting higher-order oligomerization is not absolutely required for KAP1 function.

The KAP1 RING domain has ubiquitin E3 ligase activity[21]. RING domains of E3 ubiquitin ligases recruit E2 ligases to the substrate and promote ubiquitin transfer from the E2 ligase to the substrate[39,40]. This E3 ligase activity is thought to be dependent on dimerization of the RING (in the context of RBCC dimerization) in TRIM5α, TRIM25, TRIM32 and BIRC-family E3 ligases[36,38,39]. Unexpectedly, the crystal structure of KAP1 RBCC is incompatible with RING dimerization as the RING domains are positioned on opposite ends of the RBCC dimer (**Fig. 2b**). Furthermore, the KAP1 RINGs do not form any significant crystal contacts and do not dimerize or self-associate in solution. The absence of homotypic contacts between the RING domains (or B-box 2 domains) in the RBCC dimer suggests that KAP1 may be amongst the minority of E3 ligases that can promote ubiquitin transfer from E2 to substrate without forming RING dimers, perhaps using structural elements from outside the core RING domain as seen for example in CBL-B[40]. Alternatively, the primary function of the RING domain may be to contribute to SUMO E3 ligase activity, consistent with reports that the RING domain of KAP1[26] and PML[27,28] is required for SUMOylation of specific substrates.

The interaction between the RBCC domain of KAP1 and KRAB domains is critical for recruitment of KAP1 to retrotransposons. However, the stoichiometry of KAP1-KRAB complexes remained unclear and the location of the KRAB binding surface on KAP1 unknown. Our SEC-MALS data strongly suggests that each RBCC dimer binds a single KRAB domain. Moreover, we show that KRAB binding occurs on the dyad of the RBCC dimer on a surface that includes residues V293/K296/M297/L300 and that this interface is required for KAP1 repression activity. It was previously proposed that RING domain, B-box 2 and coiled-coil domain of KAP1 all contribute to KRAB binding[26,27]. However, these results were inferred from mutations in residues essential for maintaining the structural integrity of the KAP1 RBCC dimer that would likely cause misfolding of the protein. Based on the distance of the B-box 2 from the KRAB-binding residues in the coiled-

coil domain and the small size of KRAB domains (~70 amino acids), direct involvement of B-box 2 in KRAB binding can be ruled out. Similarly, a contribution of the RING domain, which is also distant from the dyad, appears highly unlikely. We note that since a single KRAB domain binds to the antiparallel coiled-coil dimer on (or near) the dyad, the contacts formed by the KRAB domain with each of the KAP1 protomers in the dimer are necessarily different, lending an inherent asymmetry to the KRAB:KAP1 interaction. Hence, KRAB the loss of twofold symmetry in the KRAB:KAP1 complex is a consequence of KRAB binding to dyad.

We have mapped the specific molecular features within the KAP1 RBCC domain that are responsible for KAP1 self-assembly and KRAB binding. We show that binding of a single KRAB domain to the dimeric antiparallel coiled-coil domain of KAP1 is the first step in KAP1-depedendent epigenetic silencing of retrotransposon transcription. Further studies are needed to complete our understanding of how KAP1 recognizes KRAB-ZFPs, how it activates chromatin-modifying enzymes in subsequent steps of transcriptional silencing, and how it promotes ubiquitination of specific substrate proteins without RING dimerization.

Materials and Methods

Expression vectors

Synthetic genes encoding KAP1 RING (residues 50-146), RING-B-box 1 (RB1; residues 50-200), RBCC (residues 50-413) and full-length KAP1 (residues 1-835; UniProt: Q13263) codon-optimized for *Escherichia coli* (*E. coli*) were cloned into the first multiple cloning site (MCS) of the pETDuet plasmid (Novagen), with N-terminal hexahistidine purification (His_6) tag followed by a TEV protease cleavage site (ENLYFQG). The T4L-RBCC fusion construct was constructed by inserting a synthetic gene encoding the RBCC motif (residues 56-413) of human KAP1 codon-optimized for *E.coli* into the first MCS of pETDuet. The gene was preceded by sequences encoding: a His_6 tag; a TEV protease cleavage site; and bacteriophage T4 lysozyme (T4L) with the N-terminal methionine deleted and the last three residues replaced by a single alanine residue. A synthetic gene encoding the KRAB domain (residues 1-71) from ZNF93 (UniProt: P35789) codon-optimized for *E.coli* was expressed from the pET20 plasmid (Novagen) with N-terminal Twin-StrepII and maltose binding protein (MBP) affinity tags (the MBP tag was also required for protein solubility). For coexpression with KAP1, the same ZNF93 KRAB construct was subcloned into MCS1 of pCDFDuet.

For the transcriptional silencing assay, expression constructs containing WT or mutant full-length human KAP1 preceded by a triple FLAG tag and linker sequence (MDYKDHDGDYKDHDIDYKDDDDKGSGG) were assembled in pLEXm[42] with the NEBuilder HiFi DNA Assembly Cloning Kit (New England BioLabs).

Protein expression and purification

E. coli BL21 (DE3) cells (New England BioLabs) were transformed with the respective expression construct and starter cultures were grown overnight at 30˚C in 2×TY medium. Starter cultures were used to inoculate 2×TY medium and cells were incubated at 37˚C and 220 rpm to an optical density (OD_{600}) of 0.4-0.5. For the expression of KAP1 constructs, cultures were then supplemented with 50 µM $ZnSO_4$ and the temperature of the incubator was set to 18˚C. In the case of KRAB domain constructs, the incubator temperature was lowered to 16˚C. Protein expression was induced at OD_{600} = 0.8 with 0.2 mM isopropyl-β-D- thiogalactopyranoside (IPTG). All subsequent steps were performed at 4˚C. After 16 h cells were harvested by centrifugation, and stored at -80˚C.

To purify the RBCC domain of KAP1 cells were resuspended in lysis buffer containing 50 mM Tris pH 8, 0.3 M NaCl, 20 mM imidazole, 0.5 mM TCEP, 1:10,000 (v/v) benzonase solution (Sigma), 1×cOmplete EDTA-free protease inhibitors (Roche). The cells were lysed by sonication immediately after addition of 1 mM phenylmethane sulfonyl fluoride (PMSF). The lysate was clarified by centrifugation (30 min, 40,000×g). The supernatant was applied to a 5-ml HisTrap HP nickel-affinity column (GE Healthcare) preequilibrated in wash buffer (50 mM Tris pH 8, 0.3 M NaCl, 20 mM imidazole, 0.5 mM TCEP). The column was washed with 30 column volumes (CV) of wash buffer before elution with elution buffer (50 mM Tris pH 8, 0.3 M NaCl, 0.25 M imidazole, 0.5 mM TCEP). Subsequently, the buffer was exchanged to 50 mM Tris pH 8, 0.3 M NaCl, 0.5 mM TCEP and the His_6 tag was removed by incubating the protein overnight at 4˚C with 1:50 (w/w) TEV protease. Following a second nickel-affinity chromatography step to remove uncleaved protein and protease, the sample was further purified by size-exclusion chromatography using a HiLoad (16/600) Superdex 200 pg column (GE Healthcare) preequilibrated in 20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP.

RING and RB1 constructs were purified as described above, except that a Superdex 75 (10/300) column (GE Healthcare) equilibrated in 20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP was used for the final size-exclusion chromatography step. T4L-RBCC fusion protein was purified as the RBCC domain, except that the His_6 tag was not removed. Full-length KAP1 was purified as

T4L-RBCC, except that a Superose 6 increase (10/300) column (GE Healthcare) preequilibrated in 20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP was used for the final size-exclusion chromatography step.

To purify MBP-tagged ZNF93 KRAB domain, bacteria pellets were resuspended in lysis buffer (50 mM Tris pH 8, 0.15 M NaCl, 0.5 mM TCEP, 1:10,000 (v/v) benzonase solution (Sigma), 1×cOmplete EDTA-free protease inhibitors (Roche)) and lysed by sonication. The lysate was clarified by centrifugation (30 min, 40,000×g). The supernatant was applied to a 5 ml StrepTrap column (GE Healthcare) preequilibrated in wash buffer (50 mM Tris pH 8, 0.15 M NaCl, 0.5 mM TCEP). The column was washed with 30 CV of wash buffer, before the protein was eluted with wash buffer supplemented with 2.5 mM D-desthiobiotin and further purified by size-exclusion chromatography using a HiLoad (16/600) Superdex 200 pg column (GE Healthcare) preequilibrated in 20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP.

KAP1:MBP-KRAB complex was purified as the isolated KRAB domain, except that lysis and wash buffer contained 0.2 M NaCl and a Superose 6 increase (10/300) column was used for the final size-exclusion chromatography step.

Size-exclusion chromatography and multi-angle light scattering (SEC-MALS) analysis

100 µl of protein sample was subjected to size-exclusion chromatography (SEC) at 293 K using a Superdex 200 (10/300) column (GE Healthcare) preequilibrated in 20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP (for KAP1 RBCC), a Superose 6 (10/300) column preequilibrated in 20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP (for full-length KAP1), a Superose 6 (10/300) column preequilibrated in 20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP (for KAP1:MBP-KRAB complex) or a Superdex 75 (10/300) column preequilibrated in 20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP (for RING and RB1 constructs) with a flow rate of 0.5 ml min^{-1}. The SEC system was coupled to multi-angle light scattering (MALS) and quasi-elastic light scattering (QELS) modules (DAWN-8+, Wyatt Technology). The protein was also detected as it eluted from the column with a differential refractometer (Optilab T-rEX, Wyatt Technology) and a UV detector at 280 nm (Agilent 1260 UV, Agilent Technology). Molar masses of peaks in the elution profile were calculated from the light scattering and protein concentration, quantified using the differential refractive index of the peak assuming a dn/dc of 0.186, using ASTRA6 (Wyatt Technology).

Sedimentation-equilibrium analytical ultracentrifugation (SE-AUC) analysis

KAP1 RBCC samples at approximately 200 μM (8 g L^{-1}) and 12 μM (0.5 g L^{-1}) were diluted in a 1:3 series in 20 mM HEPES pH 8.0, 0.5 M NaCl, 0.5 mM TCEP. 110 μL samples were loaded in 12 mm 6-sector cells and centrifuged at 5,000, 8,500 and 15,000 rpm at 20˚C in an An50Ti rotor using an Optima XL-I analytical ultracentrifuge (Beckmann). At each speed, comparison of several scans was used to judge whether equilibrium had been reached. The data were analyzed in SEDPHAT 13b[43]. Equilibrium sedimentation distributions were fit to obtain average masses. An SE-AUC average mass isotherm compiled from fits to the data was analyzed in SEDPHAT using isodesmic and dimer-tetramer-octamer oligomerization models. The partial-specific volumes (v-bar), solvent density and viscosity were calculated using Sednterp (www.rasmb.org/sednterp). Data were plotted with the program GUSSI[44] or PRISM 8 (GraphPad).

X-ray crystallography

Prior to crystallization, free amines in T4L-RBCC were methylated by incubating 15 ml of protein solution (~1 g L^{-1}) in 20 mM HEPES pH 8, 0.5 M NaCl with 300 μl of 1 M dimethylamine borane complex (ABC; Sigma-Aldrich) and 600 μl of 1 M formaldehyde for 2 h at 4˚C. An additional 300 μl of 1 M ABC and 600 μl of formaldehyde were then added. After further 2 h at 4˚C, 150 μl of ABC was added and the sample was incubated overnight at 4˚C. The reaction was then quenched with 1.875 ml of 1 M Tris pH 8. The sample was supplemented with 2 mM DTT and purified with a HiLoad (16/600) Superdex 200 pg column preequilibrated in 20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP[45]. Crystals were grown at 18˚C by sitting drop vapor diffusion. Methylated T4L-RBCC at 4.5 g L^{-1} (72 μM) was mixed with an equal volume of reservoir solution optimized from the Index screen (Hampton Research): 15% (w/v) PEG 3350, 75 mM $MgCl_2$, 0.1 M HEPES pH 7.5. Plate-shaped crystals appeared after 2 days and were frozen in liquid nitrogen with 33% ethylene glycol as a cryoprotectant. X-ray diffraction data were collected at 100 K at Diamond Light Source (beamline I03) and processed with autoPROC[46] and STARANISO (Global Phasing, Ltd). The X-ray energy was tuned to 9,672 eV, corresponding to the zinc L-III edge, for data collection. Phases were determined with the single anomalous dispersion (SAD) method in PHENIX[47] using zinc as the anomalously scattering heavy atom. The atomic model was built with COOT[45] and iteratively refined with REFMAC[48] and PHENIX at 2.9 Å resolution. The atomic models for T4 lysozyme (PDB: 1LYD, 2LZM) and KAP1 B-box 2 (PDB: 2YVR) were docked into the phased electron density and the rest of the atomic model was built *de novo* using available structure of other TRIMs as guides (PDB: 4LTB[33], 4TN3[31], 5FEY[36], 5NT1[34]). See **Table 1** for data collection and refinement statistics. Structure figures were generated with PyMOL (Schrodinger, LLC). Portions of the final refined electron density maps are shown in **Supplementary Fig. 2**.

Methylation of 38 amino groups in the T4L-RBCC fusion protein was confirmed by mass spectrometry (**Supplementary Fig. 6**), but none of the dimethyl-amine groups on the methylated lysines were resolved in the electron density.

Atomic model of the KAP1 B-box 1 domain

A structural model of the KAP1 B-box 1 domain was generated from the TRIM19 B-box 1 structure (PDB: 2MVW)[37] using the Phyre 2 server (www.sbg.bio.ic.ac.uk/phyre2). The KAP1 B-box 1 domain was then superimposed onto each protomer of the TRIM19 B-box 1 dimer to generate a model of the KAP1 B-box 1 dimer.

Pulldown assay

8 nmol of purified KAP1 RBCC were incubated with 2 nmol of Twin-StrepII-MBP-KRAB for 45 min on ice. StrepII-tagged bait protein was subsequently captured with 100 µl of Strep-Tactin Sepharose (IBA) for 1 h at 4˚C. After four washes with 1 ml of wash buffer (20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP), the beads were boiled in 100 µl of 2×SDS-PAGE loading buffer and bound proteins were analyzed by SDS-PAGE and Coomassie staining.

Surface plasmon resonance (SPR)

SPR was performed using a Biacore T200 with dextran coated CM5 chips (GE Healthcare). Reference control and analyte CM5 chips were equilibrated in SPR buffer (20 mM HEPES pH 8.0, 0.5 M NaCl, 0.5 mM TCEP) at 20˚C. MBP-KRAB was immobilized onto the CM5 chips until a response unit (RU) value of approximately 600 was reached. SPR runs were performed with analytes injected for 120 s followed by a 900 s dissociation in 1:2 dilution series with initial concentrations of 34 µM for WT KAP1 and 35 µM for mutant KAP1. The sensor surface was regenerated after each injection cycle with 20 mM NaOH for 30 s with a 120-s post-regeneration stabilization period. Data were fitted using a biphase kinetic model with KaleidaGraph (Synergy Software) and PRISM 8 (GraphPad) to determine k_{on}, k_{off} and K_d.

Transcriptional silencing assay

Retrotransposon silencing activity of WT and mutant KAP1 was measured with a luciferase reporter assay in which a SINE-VNTR-Alu type D (SVA-D) retrotransposon cloned upstream of a minimal SV40 promoter strongly enhances firefly luciferase activity unless the KRAB-ZFP ZFN91 and KAP1 are both present to repress the SVA element[15]. The assay was adapted for use in HEK 293T cells, as described[41]. KAP1 KO 293T cells in 24-well plates were co-transfected with 40 ng ml^{-1} of firefly luciferase reporter plasmid (pGL4cp_VNTR-OCT4Enh_E2)[41], 400 ng ml^{-1} plasmid

encoding ZNF91 (pCAG_ZNF91_HA), 400 ng ml^{-1} plasmid encoding the KAP1 variant of interest (pLEXm_KAP1), and 4 ng ml^{-1} plasmid encoding *Renilla* luciferase under a constitutive promoter (pRTTK_Renilla). Cells were lysed 48 h post-transfection and luciferase activity was measured using the Dual Luciferase assay kit (Promega) and a Pherastar FS platereader (BMG Labtech). All transfections were performed with polyethylenimine (Sigma-Aldrich) at a PEI:DNA ratio of 2:1. Replicates were performed on separate days. Raw luciferase values were normalized to *Renilla* luciferase values to control for transfection efficiency. Data are presented as fold-repression of reporter luciferase luminescence normalized to KAP1 KO cells transfected with an empty vector.

Western blotting of KAP1 in HEK293T cells

4 x 10^5 HEK293T cells were lysed in 100 µl Passive Lysis Buffer (Promega). 10 µl of cell lysates were boiled with 2.5 µl 4x gel loading buffer for 5 min at 98˚C and run on a NuPAGE 4-12% Bis-Tris polyacrylamide gel for 45 min at 200 V. Gels were blotted using an iBlot Gel Transfer Device and Transfer Stacks (ThermoFisher). Blots were blocked with 5% milk in PBS for 1 h at room temperature and incubated with primary antibody overnight at 4˚C. Rabbit anti-KAP1 antibody (Abcam ab10484) was diluted 1:10,000 whilst rabbit anti-actin antibody (Abcam ab219733) was diluted 1:2,000. Blots were incubated with fluorescent secondary antibodies for 30 min at room temperature. Secondary anti-rabbit antibodies were diluted 1:10,000. Blots were imaged using an Odyessy CLx gel scanner (LI-COR Biosciences).

Statistics

No statistical methods were used to predetermine sample size, experiments were not randomized, and the investigators were not blinded to experimental outcomes. For the KAP1 complementation assays, data are representative of five independent experiments and statistical significance was assessed using an unpaired t-test (assuming Gaussian distributions, without Welch's correction) in PRISM 8 (GraphPad). Repression activity data are represented as the mean ± s.e.m. Statistical significance was assigned as follows: n.s., $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, n = 5.

Data availability

The structure factors and atomic coordinates were deposited in the Protein Data Bank with code PDB: 6QAJ. The original experimental X-ray diffraction images were deposited in the SBGrid Data Bank (SBGrid.org), with Data ID 637. Other data are available from the corresponding author upon reasonable request.

References

1. Friedli M, Trono D. The developmental control of transposable elements and the evolution of higher species. *Annu Rev Cell Dev Biol* **31**, 429-451 (2015).
2. Goodier JL. Restricting retrotransposons: a review. *Mob DNA* **7**, 16 (2016).
3. Friedli M*, et al.* Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res* **24**, 1251-1259 (2014).
4. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083-1087 (2016).
5. Chuong EB, Rumi MA, Soares MJ, Baker JC. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* **45**, 325-329 (2013).
6. Kapitonov VV, Jurka J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* **3**, e181 (2005).
7. Zhou L, Mitra R, Atkinson PW, Hickman AB, Dyda F, Craig NL. Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature* **432**, 995-1001 (2004).
8. Dupressoir A, Lavialle C, Heidmann T. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta* **33**, 663-671 (2012).
9. Chen JM*, et al.* Detection of two Alu insertions in the CFTR gene. *J Cyst Fibros* **7**, 37-43 (2008).
10. Hancks DC, Kazazian HH, Jr. Roles for retrotransposon insertions in human disease. *Mob DNA* **7**, 9 (2016).
11. Lamprecht B*, et al.* Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med* **16**, 571-579, 571p following 579 (2010).
12. Hung T*, et al.* The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science* **350**, 455-459 (2015).

13. Ecco G, Imbeault M, Trono D. KRAB zinc finger proteins. *Development* **144**, 2719-2729 (2017).
14. Imbeault M, Helleboid PY, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550-554 (2017).
15. Jacobs FM*, et al.* An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242-245 (2014).
16. Cheng CT, Kuo CY, Ann DK. KAPtain in charge of multiple missions: Emerging roles of KAP1. *World J Biol Chem* **5**, 308-320 (2014).
17. Rowe HM*, et al.* De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**, 519-529 (2013).
18. Quenneville S*, et al.* The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell Rep* **2**, 766-773 (2012).
19. Ecco G*, et al.* Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Dev Cell* **36**, 611-623 (2016).
20. Playfoot CJ, Adams IR. KRABs RegulaTE Gene Expression beyond the Embryo. *Dev Cell* **36**, 591-592 (2016).
21. Doyle JM, Gao J, Wang J, Yang M, Potts PR. MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases. *Mol Cell* **39**, 963-974 (2010).
22. Pineda CT*, et al.* Degradation of AMPK by a cancer-specific ubiquitin ligase. *Cell* **160**, 715-728 (2015).
23. Iyengar S, Farnham PJ. KAP1 protein: an enigmatic master regulator of the genome. *J Biol Chem* **286**, 26267-26276 (2011).
24. Ivanov AV*, et al.* PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol Cell* **28**, 823-837 (2007).

25. Zeng L*, et al.* Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat Struct Mol Biol* **15**, 626-633 (2008).
26. Liang Q*, et al.* Tripartite motif-containing protein 28 is a small ubiquitin-related modifier E3 ligase and negative regulator of IFN regulatory factor 7. *J Immunol* **187**, 4754-4763 (2011).
27. Quimby BB, Yong-Gonzalez V, Anan T, Strunnikov AV, Dasso M. The promyelocytic leukemia protein stimulates SUMO conjugation in yeast. *Oncogene* **25**, 2999-3005 (2006).
28. Shen TH, Lin HK, Scaglioni PP, Yung TM, Pandolfi PP. The mechanisms of PML-nuclear body formation. *Mol Cell* **24**, 331-339 (2006).
29. Peng H*, et al.* The structurally disordered KRAB repression domain is incorporated into a protease resistant core upon binding to KAP-1-RBCC domain. *J Mol Biol* **370**, 269-289 (2007).
30. Peng H*, et al.* Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. *J Mol Biol* **295**, 1139-1162 (2000).
31. Goldstone DC*, et al.* Structural studies of postentry restriction factors reveal antiparallel dimers that enable avid binding to the HIV-1 capsid lattice. *Proc Natl Acad Sci U S A* **111**, 9609-9614 (2014).
32. Li Y*, et al.* Structural insights into the TRIM family of ubiquitin E3 ligases. *Cell Res* **24**, 762-765 (2014).
33. Sanchez JG, Okreglicka K, Chandrasekaran V, Welker JM, Sundquist WI, Pornillos O. The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer. *Proc Natl Acad Sci U S A* **111**, 2494-2499 (2014).
34. Koliopoulos MG*, et al.* Molecular mechanism of influenza A NS1-mediated TRIM25 recognition and inhibition. *Nat Commun* **9**, 1820 (2018).
35. Wang P*, et al.* RING tetramerization is required for nuclear body biogenesis and PML sumoylation. *Nat Commun* **9**, 1277 (2018).

36. Koliopoulos MG, Esposito D, Christodoulou E, Taylor IA, Rittinger K. Functional role of TRIM E3 ligase oligomerization and regulation of catalytic activity. *EMBO J* **35**, 1204-1218 (2016).
37. Huang SY*, et al.* The B-box 1 dimer of human promyelocytic leukemia protein. *J Biomol NMR* **60**, 275-281 (2014).
38. Yudina Z*, et al.* RING Dimerization Links Higher-Order Assembly of TRIM5alpha to Synthesis of K63-Linked Polyubiquitin. *Cell reports* **12**, 788-797 (2015).
39. Dou H, Buetow L, Sibbet GJ, Cameron K, Huang DT. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat Struct Mol Biol* **19**, 876-883 (2012).
40. Dou H, Buetow L, Sibbet GJ, Cameron K, Huang DT. Essentiality of a non-RING element in priming donor ubiquitin for catalysis by a monomeric E3. *Nat Struct Mol Biol* **20**, 982-986 (2013).
41. Robbez-Masson L*, et al.* The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res* **28**, 836-845 (2018).
42. Aricescu AR, Lu W, Jones EY. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta crystallographica Section D, Biological crystallography* **62**, 1243-1250 (2006).
43. Schuck P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal Biochem* **320**, 104-124 (2003).
44. Brautigam CA. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymol* **562**, 109-133 (2015).
45. Rayment I*, et al.* Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* **261**, 50-58 (1993).
46. Vonrhein C*, et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**, 293-302 (2011).

47. Adams PD*, et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).

48. Murshudov GN*, et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011).

Acknowledgements

KAP1 KO HEK 293T cells were a kind gift from Helen Rowe (University College London). We thank Helen Rowe for providing the plasmids for the transcriptional silencing assay with permission from David Haussler (University of California Santa Cruz). We thank Christopher Douse and other members of the Modis lab for insightful discussions. Crystallographic data were collected on beamline I03 at Diamond Light Source (DLS). Access to DLS (proposal MX15916) was supported by the Wellcome Trust, MRC, and BBSRC. This work was supported by Wellcome Trust Senior Research Fellowship 101908/Z/13/Z to Y.M. and Wellcome Trust PhD Studentship 205833/Z/16/Z to G.S.

Author Contributions

Conceptualization, G.S. and Y.M.; Methodology, G.S., S.O., M.Y., S.M. and Y.M.; Validation – crystal structure, G.S. and S.O.; Investigation, G.S., S.C., and S.O.; Writing – Original Draft, G.S. and Y.M.; Writing – Review & Editing, G.S., S.C., S.O. and Y.M.; Visualization, G.S., S.C., S.O. and Y.M.; Supervision, Y.M.; Project Administration, Y.M.; Funding Acquisition, G.S. and Y.M.

Conflict of Interest

The authors declare no conflict of interest.