Molecular mechanism of KAP1-dependent transcriptional silencing

Guido Alexander Stoll

Department of Medicine

University of Cambridge

Hughes Hall

September 2019

This dissertation is submitted for the degree of

Doctor of Philosophy

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration, except as specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit of 60,000 words.

Guido Stoll

Cambridge, September 2019

ABSTRACT

Retroviruses can integrate their DNA into the host-cell genome. Inherited retroviral DNA and other transposable elements account for at least half of the human genome. Transcription of transposable elements is tightly regulated to restrict their proliferation and prevent toxic gene expression. A major factor contributing to the repression of potentially harmful retroelements is KRAB-associated protein 1 (KAP1, also known as TRIM28 or TIF1β). Following its recruitment to retrotransposons by sequence-specific KRAB domain-containing zinc finger proteins (KRAB-ZFPs), KAP1 induces the assembly of an epigenetic silencing complex, with chromatin remodelling activities that repress transcription of the targeted retrotransposon and adjacent genes.

To understand the molecular basis of KAP1-dependent transcriptional regulation I determined the crystal structure of the RBCC domain of KAP1. My structural and biophysical data demonstrate that 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 and resulted in significant defects in KAP1-dependent transcriptional repression, suggesting that self-assembly may contribute to KAP1 function.

Furthermore, I characterized the interaction of KAP1 with the KRAB domain of KRAB-ZFPs, which is crucial for recruitment of KAP1 to its genomic targets. My data show that each KAP1 dimer can only bind a single KRAB domain, resulting in a 2:1 stoichiometry. Moreover, my 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 domain responsible for self-association and KRAB binding and establishes their role in retrotransposon silencing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ABC	Borane dimethylamine complex
ALS	Amyotrophic lateral sclerosis
APC	Adenomatous polyposis coli
APOBEC	Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like
Arc	Activity-regulated cytoskeleton-associated protein
ATM	Ataxia-telangiectasia mutated
CBP	cAMP response element-binding protein (CREB)-binding protein
CC	Coiled coil
cDNA	Complementary DNA
CTCF	CCCTC-binding factor
CTFR	Cystic fibrosis transmembrane conductance regulator
CV	Column volume
CXXC1	CXXC-type zinc finger protein 1
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
DSF	Differential scanning fluorimetry
DUF3669	Domain of unknown function 3669
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ERV	Endogenous retrovirus
ESC	Embryonic stem cell
EZH2	Enhancer of zeste homolog 2
FBS	Foetal bovine serum
GFP	Green fluorescent protein
GST	Glutathione S-transferase
H3K9	Histone H3 lysine 9
HDAC	Histone deacetylase
HEK239T	Human embryonic kidney cells expressing the SV40 large T-antigen
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HERV	Human endogenous retrovirus
hESC	Human embryonic stem cell

HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HMM	Hidden Markov Model
HP1	Heterochromatin protein 1
HUSH	Human silencing hub
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
KAP1	KRAB-associated protein 1
KO	Knockout
KRAB	Krüppel-associated box
LINE	Long interspersed nuclear element
lncRNA	Long non-coding RNA
LTR	Long terminal repeat
MAGE	Melanoma-associated antigen
MBP	Maltose binding protein
MeCP2	Methyl-CpG-binding protein 2
MHC	Major histocompatibility complex
MIR	Mammalian-wide interspersed repeats
miRNA	microRNA
MLV	Murine leukemia virus
MORC2	Microchidia CW-type zinc finger 2
NuRD	Nucleosome remodelling and deacetylase
OD_{600}	Optical density at 600 nm wavelength
ORF	Open reading frame
PBS	Primer binding site *or* phosphate-buffered saline
PBS-T	Phosphate-buffered saline with 0.1% Tween-20
PEG	Polyethylene glycol
PHD	Plant homeodomain
PIKK	Phosphatidylinositol 3-kinase-related kinase
piRISC	piRNA-induced silencing complex
piRNA	PIWI-interacting RNA
PML	Promyelocytic leukemia protein
PMSF	Phenylmethane sulfonyl fluoride
Pol	Polymerase

poly(A)	polyadenine *or* polyadenylation
PP1	Protein phosphatase 1
PRDM9	PR domain-containing protein 9
RAG	Recombination activating gene
RING	Really interesting new gene
RBCC	RING, B-box, coiled-coil
RNA	Ribonucleic acid
RNaseL	Ribonuclease L
RNP	Ribonucleoprotein
rRNA	Ribosomal RNA
SAD	Single anomalous dispersion
SAMHD1	SAM domain and HD domain containing protein 1
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE-AUC	Sedimentation equilibrium analytical ultracentrifugation
SEC-MALS	Size-exclusion chromatography coupled with multi-angle light scattering
SENP	Sentrin-specific peptidase
SETDB1	SET Domain Bifurcated 1
SILAC	Stable isotope labelling with amino acids in cell culture
SIM	SUMO interacting motif
SINE	Short interspersed nuclear elements
siRNA	Small interfering RNA
SLE	Systemic lupus erythematosus
SPR	Surface plasmon resonance
STAT	Signal transducer and activator of transcription
STORM	Stochastic optical reconstruction microscopy
SUMO	Small ubiquitin-like modifier
SVA	SINE-R/VNTR/Alu
T4L	Bacteriophage T4 lysozyme
TAD	Topologically associating domains
TCEP	Tris(2-carboxyethyl)phosphine
TE	Transposable element
TEV	Tobacco etch virus

TIF1	Transcription intermediary factor 1
TIR	Terminal inverted repeat
TLR4	Toll-like receptor 4
TPRT	Target-primed reverse transcription
Trex1	Three-prime repair exonuclease 1
TRIM	Tripartite motif
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
TSD	Target site duplication
UTR	Untranslated region
VNTR	Variable number of tandem repeats
WT	Wild type
ZAP	Zinc finger CCCH-type antiviral protein 1
ZFP	Zinc finger protein

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Yorgo Modis, for his great support, encouragement and guidance during the past three years, while also giving me much freedom regarding the direction of my research.

I am grateful to past and present members of the Modis group for help with new experimental techniques, for insightful discussions and for making this lab a great place to work. In particular, I would like to thank Christopher Douse for his invaluable advice, Zheng-Shan Chong for optimizing the luciferase reporter assay, generating many of the mammalian expression plasmids and performing the first set of transcriptional silencing assays and Shun-ichiro Oda for help with structure refinement.

This project was greatly aided by the excellent infrastructure provided by the MRC Laboratory of Molecular Biology. In particular, I would like to thank my collaborator Stephen McLaughlin from the MRC LMB biophysics facility, who performed the SE-AUC and SPR experiments. I thank Minmin Yu from the MRC LMB X-ray crystallography facility for help with X-ray diffraction data collection. Furthermore, I am grateful to Christopher Johnson from the MRC LMB biophysics facility for training on SEC-MALS and ITC instruments.

I thank Diamond Light Source for beamtime (proposals mx15916 and mx21426), and the staff of beamlines I03 and I04 for assistance with crystal testing and data collection.

I would like to thank Helen Rowe (University College London), who kindly provided the KAP1 KO HEK293T cells used in this project, as well as the luciferase reporter plasmids. The plasmids were provided with permission from David Haussler (University of California, Santa Cruz).

Finally, I am grateful to the Wellcome Trust for funding my research.

PUBLICATIONS

Parts of this thesis have been published in:

Stoll, G. A., Oda, S. I., Chong, Z. S., Yu, M., McLaughlin, S. H., & Modis, Y. (2019). Structure of KAP1 tripartite motif identifies molecular interfaces required for retroelement silencing. Proceedings of the National Academy of Sciences of the United States of America, 116(30), 15042–15051. doi:10.1073/pnas.1901318116

CHAPTER 1

INTRODUCTION

1.1 Transposable elements

Transposable elements (TEs), mobile DNA sequences capable of moving from one location in the genome to another, have accumulated to strikingly high numbers in most higher organisms. While only 1.5 % of the human genome code for protein, almost half our DNA is derived from TEs (**Fig. 1)** [1]. It is becoming increasingly clear that TEs represent major drivers of evolution with the potential to generate new transcriptional networks, as well as novel genes [2, 3]. At the same time, the mutagenic effects of TEs can cause disease and, as a consequence, their activity has to be tightly controlled by the host [4]. Several classes of TEs are present in the human genome, which are distinguished based on their replication mechanism.

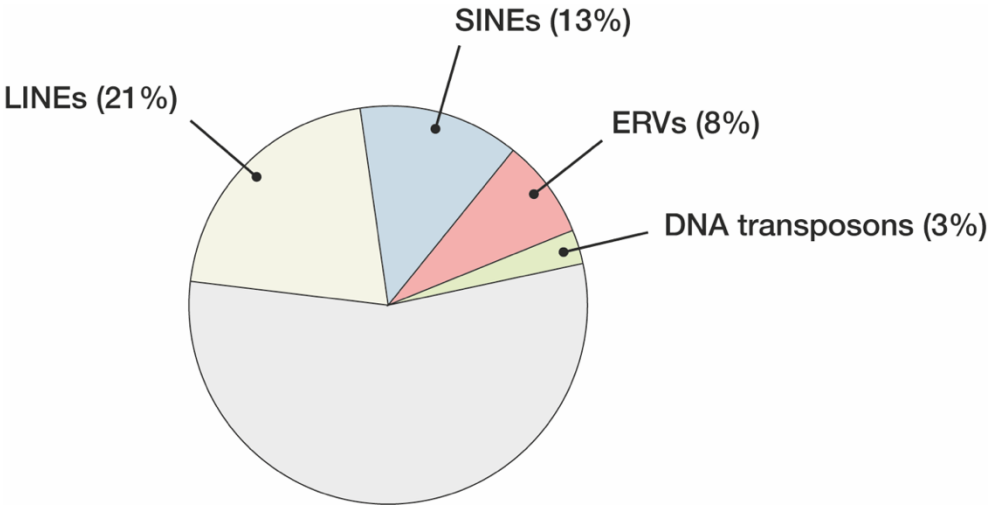

Fig. 1. Transposon-derived sequences in the human genome. Approximately 45 % of the human genome are clearly identifiable as transposon-derived. The most prevalent class of TEs in humans are non-LTR retrotransposons, including long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), which account for 34 % of our DNA. An additional 8 % are derived from LTR-retrotransposons, also referred to as endogenous retroviruses (ERVs), while 3 % of the human genome consist of remnants of DNA transposons [1].

1.1.1 DNA transposons

Intact DNA transposons are approximately 1-5 kb in length and are flanked by short terminal inverted repeats (TIRs; **Fig. 2**) [5]. These TIRs are recognized by the transposon-encoded transposase enzyme, which mediates excision of the TE sequence, followed by re-insertion at a different site in the genome [5]. Approximately 3% of the human genome is derived from DNA transposons [1]. While DNA transposons appear to have been highly active during early primate evolution, these elements have since lost their ability to mobilize in the genome, with the last transposition-competent members in the primate lineage becoming extinct approximately 37 million years ago [6]. Bats are the only mammals that show signs of ongoing DNA transposon activity [7].

Fig. 2. Structure of DNA transposons. Intact DNA transposons encode a transposase enzyme flanked on either end by terminal inverted repeats (TIRs).

1.1.2 Retrotransposons

In contrast to the 'cut-and-paste' mechanism utilized by most DNA transposons, retrotransposons replicate via an RNA intermediate which is reverse transcribed into DNA and reintegrated into the host genome [4]. This amplifying transposition mechanism allowed retrotransposon-derived sequences to accumulate and such elements now account for at least 40 % of our genome [1]. However, this figure may be considerably higher, as evolutionarily old transposon sequences are often heavily mutated and may not be recognized by most computational tools. Accordingly, more recent studies estimate that up to 69 % of the human genome consist of TE-derived repeats [8]. While DNA transposons are no longer mobile, retrotransposons are still active in humans and cause approximately one new integration event per 20 live births [9].

Retrotransposons are further classified into autonomous and non-autonomous elements. Autonomous retrotransposons encode the enzymatic activities required for retrotransposition, whereas non-autonomous retroelements, such as short interspersed nuclear elements (SINEs), rely on the replication machinery provided by their

autonomous counterparts [4]. Autonomous retroelements include long terminal repeat (LTR) retrotransposons, also called endogenous retroviruses (ERVs), and long interspersed nuclear elements (LINEs), which are the only non-LTR retrotransposon family capable of autonomous retrotransposition [10].

1.1.2.1 Endogenous Retroviruses

Endogenous retroviruses (ERVs) are LTR retrotransposons, which represent remnants of past germline infections by retroviruses. ERVs share the genomic structure of their exogenous counterparts, comprising *gag*, *pol* and *env* genes flanked on either end by LTRs, and can retain the ability to replicate [11]. Over time, however, ERVs lose the ability to form virus particles and adapt to intracellular replication [12]. Frequently, this is accompanied by a loss of the *env* gene, which appears to enhance the ability of the ERV to spread in the genome [13]. As a result, ERVs have accumulated and now account for approximately 8% of the human genome [1].

Intact ERVs are approximately 7-9 kb in length and replicate via a mechanism resembling the life cycle of exogenous retroviruses [4, 10]. Following transcription of the integrated provirus and translation of the encoded proteins, the viral RNA intermediate is packaged into virus-like particles in the cytoplasm, where it is converted into DNA by the viral reverse transcriptase. Subsequently, the DNA copy is integrated into the host genome by the viral integrase enzyme [10, 14] (**Fig. 3**).

Most human ERVs (HERVs) have acquired large numbers of mutations rendering them replication-incompetent. Many of these elements only exist as solitary LTRs generated by homologous recombination between the 5' and 3' LTRs, which deletes the internal genes of the provirus [1]. While full-length HERV copies are present in the human genome, none of these elements are thought to be currently active [12]. Notably, however, a completely intact HERV-K element was recently identified in some individuals, which may be replication-competent [15]. In contrast, ERVs remain highly mobile in other mammals including mice, where approximately 10 % of spontaneous mutations can be traced back to ERV integration events [16].

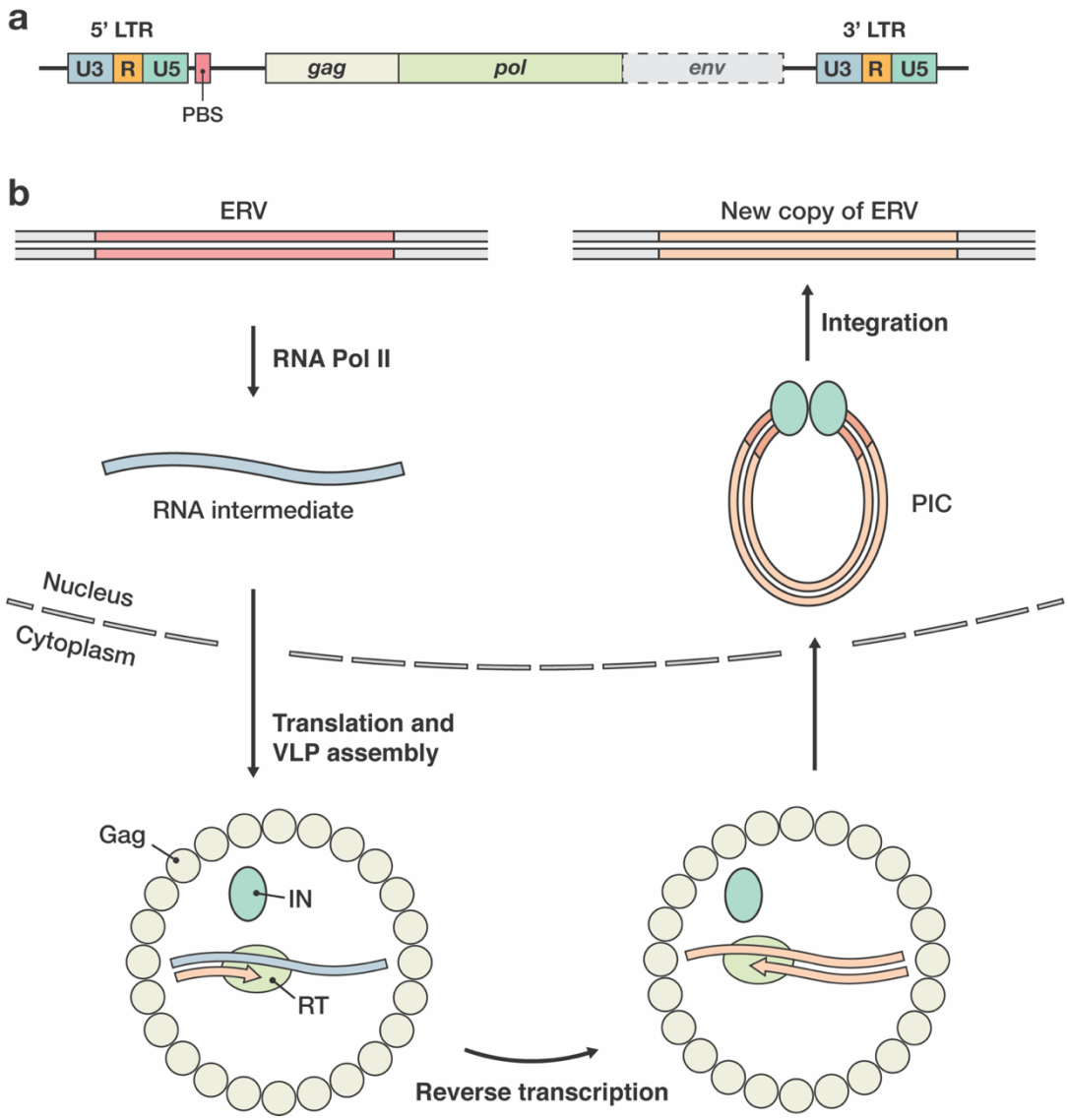

Fig. 3. Structure and replication of ERVs. (**a**) Structure of intact ERVs. Like exogenous retroviruses, ERVs contain a *gag* gene encoding structural components of the virus particle, a *pol* gene encoding the viral protease, integrase (IN) and reverse transcriptase (RT) enzymes. Some ERVs also contain an *env* gene encoding the viral envelope protein. However, the *env* gene is not required for intracellular replication and frequently lost. The internal genes of the provirus are flanked by long terminal repeats (LTRs). The 5' LTR is followed by a primer binding site (PBS), which recruits cellular tRNAs to the viral RNA intermediate to serve as primers for reverse transcription. (**b**) Replication of ERVs. Following transcription of the integrated provirus and production of the viral proteins, the viral RNA intermediate is packaged into a virus-like particle (VLP). Here, the viral reverse transcriptase generates a cDNA copy of the virus, which then associates with the viral integrase enzyme to form the pre-integration complex (PIC). The PIC subsequently enters the nucleus, resulting in integration of the new ERV copy into the host genome.

1.1.2.2 Long Interspersed Nuclear Elements (LINEs)

Long interspersed nuclear elements (LINEs) are an evolutionarily old family of autonomous retrotransposons that has accumulated to more than 800,000 copies in the human genome [1]. LINE-1 elements account for approximately 17 % of our DNA, while an additional 4 % are derived from LINE-2 and LINE-3 elements [1]. Whereas the LINE-2 and LINE-3 lineages are both extinct, approximately 100 LINE-1 elements are still active in humans and are estimated to cause one new insertion in 100 births [9]. Replication-competent LINE-1 elements are approximately 6 kb in size and consist of a 5' untranslated region (UTR), followed by two open reading frames (ORF1 and ORF2) and a 3' UTR (**Fig. 4a**). ORF1 encodes a 40 kDa RNA-binding protein (ORF1p), while the 150 kDa protein encoded by ORF2 (ORF2p) possesses reverse transcriptase and endonuclease activities [17–19]. In addition to ORF1 and ORF2, which are both essential for replication, the 5' UTR of some LINE-1 elements contains a short, primate-specific ORF (ORF0) on the antisense strand. While the precise function of ORF0 is unclear, it has been shown to be expressed and promote LINE-1 mobility [20]. Transcription of LINE-1 elements is driven by a bidirectional, CpG-rich RNA polymerase II (RNA Pol II) promoter region located in its 5' UTR and terminated by a polyadenylation (poly(A)) signal in the 3' UTR [21, 22]. However, RNA Pol II frequently reads through this poly(A) signal, resulting in transposition of adjacent genomic sequences (3' transduction) in up to 24 % of LINE-1 retrotransposition events [23, 24]. The bicistronic LINE-1 RNA is subsequently translated, producing ORF1p and ORF2p, which associate with RNA to form ribonucleoprotein (RNP) particles. While ORF1p and ORF2p usually bind to the RNA from which they were translated [25], they occasionally interact with cellular mRNAs, which can result in the integration of a cDNA copy of this mRNA into the genome [26]. Following nuclear import of the RNP, the LINE-1 RNA intermediate is reverse transcribed and integrated at a new genomic locus via a process known as target-primed reverse transcription (TPRT; **Fig. 4b**), which shows notable mechanistic differences to the replication of retroviruses [27]. The N-terminal endonuclease domain of ORF2p cleaves the bottom-strand of the target locus at a 5′-TTTT/AA-3′ consensus sequence, creating a single-strand break. This cleavage event liberates an oligo(T) stretch, which is thought to basepair with the poly(A) tail of the LINE-1 RNA and serve as primer for ORF2p-mediated reverse transcription [18, 28]. During reverse transcription, the top-strand of the integration site is cleaved by an as

yet unidentified nuclease, followed by second-strand cDNA synthesis by ORF2p. This replication mechanism generates short target site duplications (TSDs) at either end of the newly integrated LINE-1 element. TSDs are usually 7-20 bp in length, depending on the distance between the first and second cleavage event during TPRT. Notably, retrotransposition of LINE-1 elements is highly prone to 5' truncations and only a small proportion of new LINE-1 integrants are full-length [21].

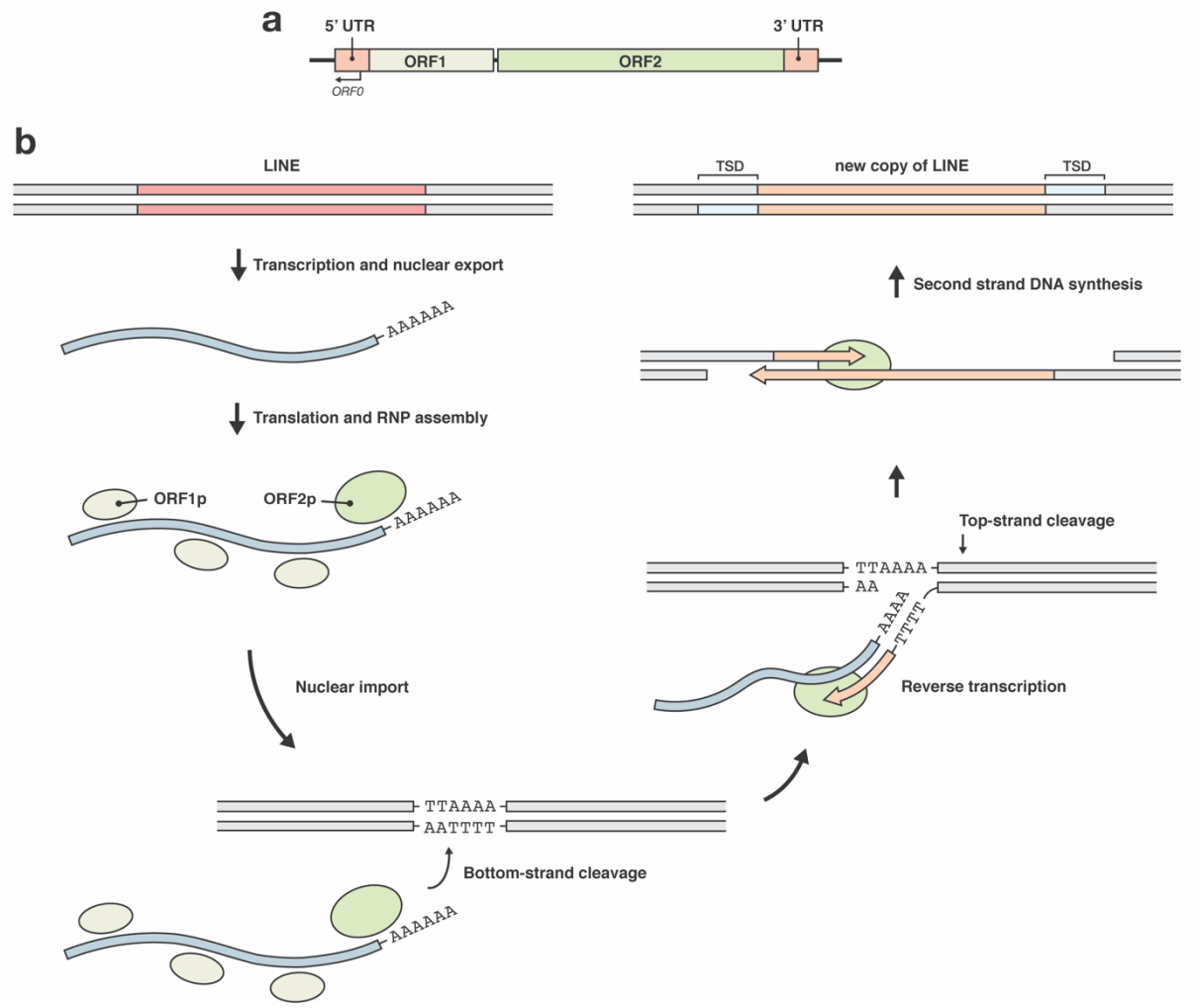

Fig. 4. LINE-1 replication. (**a**) Structure of a full-length LINE-1 element. (**b**) Retrotransposition mechanism of LINE-1. Following transcription of the LINE-1 element by RNA Pol II and production of the encoded proteins, ORF1p and ORF2p associate with the RNA from which they translated, forming LINE-1 ribonucleoprotein (RNP) particles. The LINE-1 RNA intermediate is reverse transcribed and integrated into the genome via a mechanism known as target-primed reverse transcription (TPRT): First, ORF2p generates a single-strand break at the target site, liberating an oligo(T) stretch, which basepairs with the poly(A) tail of the LINE-1 RNA and serves as a primer for ORF2p-mediated reverse transcription. Subsequently, the top strand at the target site is cleaved, followed by synthesis of the second cDNA strand. The new LINE-1 integrant is flanked by short target site duplications (TSDs).

1.1.2.3 Short Interspersed Nuclear Elements (SINEs)

SINEs are derived from small cellular RNAs. In contrast to autonomous retrotransposons, they do not encode functional proteins and rely on the reverse transcriptase and endonuclease activities provided by LINE-1 ORF2p for their replication. The most prevalent SINEs in humans are the primate specific *Alu* elements, which arose approximately 65 million years ago and have since accumulated to more than one million copies, accounting for at least 10% of the human genome [1, 29, 30]. Members of the *Alu* family remain active in humans and are estimated to cause one new integration events in every 20 live births [9]. Replication-competent *Alu* elements are approximately 300 bp long and have a bipartite structure consisting of two monomers derived from the signal recognition particle RNA (7SL RNA) gene separated by an A-rich linker. The 3' end of *Alu* elements is formed by a poly(A) tail [29, 31] (**Fig. 5**). *Alu* elements are transcribed by RNA Pol III from a promoter located in the 5' monomer. Notably, no terminator sequence is present in *Alu*s, resulting in transcription to continue into adjacent genes until an oligo(T) stretch terminates RNA Pol III. The RNA intermediate is exported into the cytoplasm, where it is thought to associate with ribosomes and compete with LINE-1 RNA for binding to ORF2p. Reverse transcription and integration of *Alu* elements subsequently occurs via ORF2p-mediated TPRT, generating TSDs on either side of the *Alu* sequence [29, 30]. However, in the case of *Alu*s, the poly(A) tail required for TPRT is encoded directly in the transposon sequence, rather than being added post-transcriptionally via a poly(A) signal as in LINE elements (**Fig. 5**). Their high copy number and repetitive nature make *Alu* elements prone to recombination events, which can result in large-scale genomic rearrangements, including deletions and duplications and potentially lead to disease [30].

Besides *Alu* elements, other families of SINEs exist in the human genome, including approximately 500,000 copies of the tRNA-derived mammalian-wide interspersed repeats (MIRs), which account for 2.5 % of our DNA. However, in contrast to *Alu*s, these elements are no longer replication-competent [1, 32].

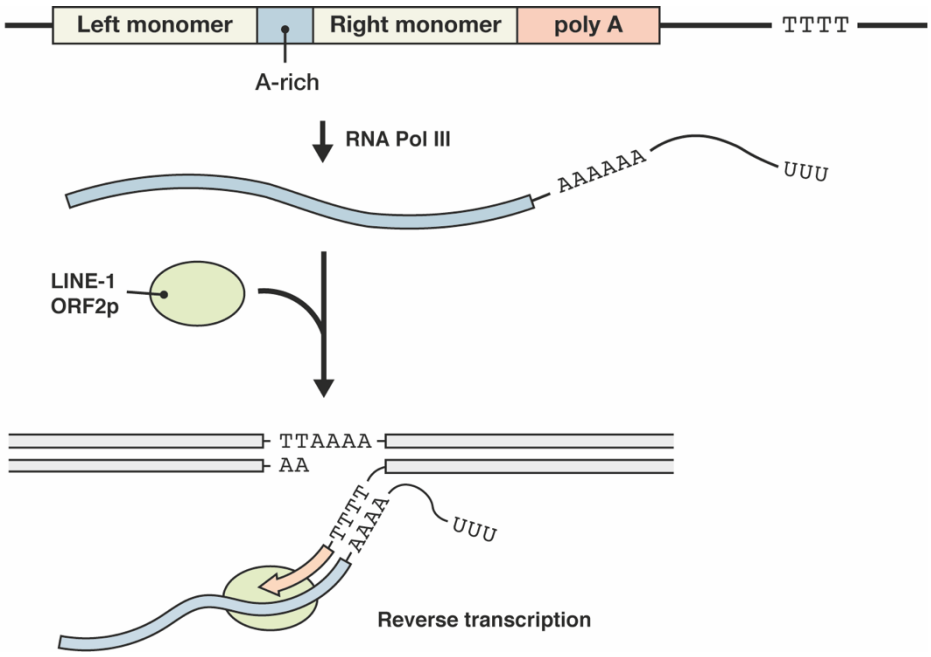

Fig. 5. Structure and replication of *Alu* elements. Transcription of *Alu* elements by RNA Pol III is initiated from an internal promoter located in the left monomer but continues into adjacent genes until a terminator sequence (TTTT) is encountered. The *Alu* RNA intermediate associates with ribosomes where it competes with LINE-1 RNA for ORF2p binding. Reverse transcription and integration of the *Alu* subsequently occurs via ORF2p-mediated TPRT.

1.1.2.4 SINE-R/VNTR/Alu (SVA) elements

SINE-R/VNTR/Alu (SVA) elements are a relatively young family of hominid-specific retrotransposons, which is estimated to have emerged approximately 25 million years ago. The human genome contains ~ 2700 SVA elements, which account for 0.2 % of our DNA [17, 33, 34]. Approximately 50 of these SVA elements are currently active and responsible for one new retrotransposition event in 920 births [4, 9].

Fig. 6. Structure of SVA elements. Intact SVA elements consist of: A variable number of CCCTCT repeats; two inverted *Alu*-like sequences; a variable number of GC-rich tandem repeats (VNTR); and a HERV-K-derived fragment (SINE-R).

With an average size of 2 kb, SVA elements are considerably larger than SINEs and are thought to be transcribed by RNA Pol II rather than RNA Pol III [35]. Like SINEs,

however, SVA elements do not encode functional proteins and require the enzymatic activities of LINE-1 ORF2p for transposition. Structurally, SVA elements are composites consisting of a hexameric repeat of variable length ($(CCCTCT)_n$), followed by two *Alu*-like repeats in antisense orientation, a variable number of GC-rich tandem repeats (VNTR) and a fragment derived from the *env* gene and 3' LTR of a HERV-K element (SINE-R) (**Fig. 6**). The 3' end of SVA elements contains a poly(A) signal [35]. As in LINE-1 elements, however, this poly(A) signal is relatively weak, leading to 3' transductions in approximately 10 % of SVA retrotransposition events [34]. In addition, SVA elements are frequently transcribed from a promoter located upstream of the TE itself. Accordingly, 8 % of the SVA copies found in the human genome appear to have transduced 5' flanking sequences to a new locus [36]. Following transcription of the SVA element by RNA Pol II, reverse transcription and integration of new SVA copies occurs via TPRT mediated by LINE-1 ORF2p [35].

1.1.3 Evolutionary impact

1.1.3.1 Novel protein-coding genes

TEs are one of the most powerful forces driving the evolution of higher species. The most obvious contribution of TEs in this context is the introduction of novel protein-coding genes that may be co-opted by the host. TE-derived proteins have been implicated in a variety of processes ranging from placenta formation to brain function. Prominent examples include syncytin-1 and syncytin-2, fusion proteins derived from HERV-W and HERV-FRD *env* genes, which are crucial for syncytiotrophoblast formation during placenta development [37–39]. In addition to their fusogenic properties, many retroviral envelope proteins contain immunosuppressive domains. In the case of syncytins, the immunomodulatory properties of this domain are thought to protect the foetus from the maternal immune system [40–42]. Besides syncytins, several other proteins of viral origin have been implicated in placenta formation. Suppressyn, a secreted protein derived from the *env* gene of a HERV-F elements, competes with syncytin-1 for receptor binding and regulates the extent of syncytin-mediated cell-cell fusion [43]. In addition, multiple proteins derived from the *gag* gene of ERVs, including Peg10 and Peg11, appear to be essential for normal placenta development [3, 44–47]. Finally, the HERV-K accessory protein Rec has been shown to stimulate IFITM1 expression in human

embryos, possibly by stabilizing its mRNA, which may provide increased resistance to viral infections [48].

TEs have also significantly contributed to the evolution of the immune response. RAG1 and RAG2, the proteins mediating V(D)J recombination, an essential mechanism in the vertebrate immune system to generate diversity in the B- and T-cell repertoire, derived from transposases of *Transib* DNA transposons [49–52]. Furthermore, co-opted ERV proteins have the potential to act as potent antiviral restriction factors. For instance, the mouse protein Fv1 inhibits murine leukemia virus (MLV) infection by interfering with uncoating of the viral genome [53]. Another example of a ERV-derived restriction factor is enJSRV, a mutated endogenous *gag* protein found in sheep, that exerts a dominant negative effect on particle assembly of exogenous retroviruses [53]. Finally, humans carry an ERV *env* gene that is thought to have been able to block entry of the now extinct retrovirus HERV-T by depleting its receptor from the cell surface. This property may have been a decisive factor contributing to the disappearance of the exogenous form of this virus [54].

Moreover, TE-derived proteins have been shown to play crucial roles in cognitive processes such as learning and memory. Activity-regulated cytoskeleton-associated protein (Arc), for instance, is an essential regulator of synaptic plasticity, which evolved from the capsid protein of Ty3/Gypsy LTR retrotransposons [55–57]. Notably, Arc has been reported to assemble into virus-like capsids and transfer mRNA between cells [57].

In the previous examples, transposons directly introduced novel genes that were co-opted by the host. Additionally, however, retrotransposons can also indirectly contribute to the evolution of new genes. As noted above, transcription of LINE and SVA elements frequently continues into adjacent cellular genes. Retrotransposition of these sequences to a different genomic locus can generate novel proteins via exon shuffling [10, 21, 58–60].

1.1.3.2 Evolution of transcriptional networks

In addition to providing new protein-coding sequences, TEs can drive the evolution of transcriptional networks by spreading transcription factor binding sites, promoters and other regulatory elements. ERV LTRs in particular are rich sources of transcription factor binding sites, which likely represent remnants of their previous exogenous life cycle and

allowed these viruses to respond to various stimuli [2]. The primate-specific ERV MER41, for example, sprinkled the genomes of our ancestors with STAT1 binding sites, thereby shaping the transcriptional program activated in response to interferon stimulation [61]. Most intriguingly, however, it was proposed that waves of high transposon activity may have facilitated major evolutionary developments by placing previously separately functioning genes under common regulation [2]. For instance, TE-mediated remodelling of transcriptional networks was likely crucial for the evolution of pregnancy in mammals. The DNA transposon MER20 is thought to have dispersed progesterone-responsive enhancer elements in the genome, thereby integrating numerous genes involved in differentiation of endometrial stromal cells into a novel, progesterone-responsive network [62]. Furthermore, a promoter derived from the ERV MER39 has been reported to cooperate with MER20 to drive prolactin production in endometrial cells [63]. In addition to the endometrium, transposon-derived regulatory elements are of particular importance in the placenta, where global hypomethylation alleviates the repression of many TEs. Enhancer sequences introduced by ERVs regulate placenta-specific gene expression and have been a major factor fuelling the rapid evolution of the placenta [64]. Equally, transcription of HLA-G, a major histocompatibility complex (MHC) molecule, which is specifically expressed in the placenta and is critical for protection of the foetus from the mother's immune system, has been shown to be driven by an LTR-derived enhancer element [65].

Additional examples of for the impact of TE-derived regulatory elements include the activation of β-globin expression in human red blood cells by an ERV LTR, which mediates the switch from foetal γ-globin to adult β-globin [66]. Furthermore, insertion of an LTR retrotransposon allowed expression of the enzyme amylase in the salivary gland in addition to the pancreas [67]. TEs not only affect gene expression by acting as promoters or enhancers, however. In addition, transposons can significantly influence the three-dimensional organization of chromatin. SINEs, for instance, have spread CCCTC-binding factor (CTCF) binding sites in mammalian genomes that can establish long-range chromatin contacts and act as insulator elements separating transcriptionally active and inactive regions [68]. Furthermore, HERV-H elements were recently shown to contribute to the formation of topologically associating domains (TADs) in human pluripotent stem cells [69]. Interestingly, these TADs appear to be

dependent on active transcription of HERV-K and are largely lost following HERV-K silencing during differentiation [69].

Finally, transposon-derived sequences are frequently targeted for epigenetic silencing by the host cell. Heterochromatin formation induced at TEs may spread to neighbouring genes, thereby influencing their expression [70].

1.1.3.3 TE-derived non-coding RNAs

It is becoming increasingly clear that TEs have been co-opted to regulate the development of early embryonic cells. HERV-H elements in particular are highly active in embryonic stem cells and are a rich source of regulatory long non-coding RNAs (lncRNAs), which have been shown to be essential for the maintenance of pluripotency in embryonic stem cells (ESCs). Accordingly, knockdown of HERV-H transcripts leads to loss of pluripotency causes ESCs to differentiate [71, 72]. The HERV-H derived lncRNA lnc-RoR, for instance, alleviates miRNA-mediated repression of pluripotency-associated transcription factors such as Oct4 and Nanog [73, 74]. Moreover, a significant proportion of Oct4 and Nanog binding sites is located in HERV-H elements [75, 76]. Upon activation by Oct4, HERV-H LTRs are thought act as enhancers and induce expression of adjacent pluripotency-associated genes [71]. Interestingly, this enhancer function of HERV-H LTRs appears to be further increased by HERV-H derived lncRNAs, which have been reported to serve as a scaffold promoting association of Oct4 with transcriptional activators including p300 and CBP [71]. Furthermore, HERV-H activation in embryonic cells generates ESC-specific chimeric transcripts, some of which have been shown to be essential for the maintenance of pluripotency [72].

In addition to the prominent role of HERV-H-derived transcripts in embryonic development, functions for non-LTR retrotransposons in early embryos have also been described. For example, a recent study found that knockdown of LINE-1 in mouse embryos results in arrest at the 2-cell stage [77]. LINE-1-derived RNA was further proposed to serve as a scaffold for the assembly of a complex comprising Nucleolin and KAP1. This complex appeared to simultaneously induce rRNA synthesis and silence expression of Dux, a transcription factor crucial for maintaining the transcriptional program characteristic of the 2-cell state, thereby promoting progression to the blastocyst stage [77].

1.1.3.4 Somatic retrotransposition in the brain

While TE replication is generally thought to be repressed in most differentiated cell types, several studies indicate that somatic retrotransposition may be occurring in the brain. Neuronal progenitor cells express reduced levels of the transcription factor Sox2, which appears to favour LINE-1 reactivation and mobilization in both humans and mice [78–80]. Accordingly, *de novo* LINE-1 insertions are detectable in neuronal cells, although the exact frequency of retrotransposition events in neurons is still debated and may vary between individuals [80–83]. In particular hippocampal neurons, however, have been found to exhibit signs of high LINE-1 activity [80, 82]. While the physiological relevance of LINE-1 mobilization in neurons is still largely unclear, somatic LINE-1 insertions have the potential to alter neuronal gene expression patterns [84]. Furthermore, environmental stimuli appear to increase the rate of neuronal LINE-1 retrotransposition in mice [85–87]. Consequently, it has been proposed to that LINE-1 mobilization may contribute to neuronal plasticity and memory formation [78, 85, 87].

1.1.4 Implications in disease

Due to their mutagenic nature, TEs have the potential to cause disease if they escape repression. Retrotransposon activity is associated with several different types of diseases in humans, depending on whether it occurs in somatic cells or the germ line. Transposition events in germ cells may lead to genetic disorders, for example by disrupting protein coding sequences. To date, more than 120 examples of genetic diseases caused by retrotransposon insertions have been described, including cases of haemophilia caused by integration of LINE-1 into exons of coagulation factor genes and cases of cystic fibrosis caused by insertion *Alu* elements into exons of the CFTR gene [21, 88, 89]. In addition to directly inserting into protein coding sequences, TEs can disrupt gene function by providing alternative splice sites. For example, the most common mutation associated with the autosomal recessive disease Fukuyama muscular dystrophy is caused by the integration of an SVA element into the 3' UTR of the *fukutin* gene [90]. This has been shown to lead to aberrant mRNA splicing, resulting in replacement of the C-terminus of the protein with an SVA-derived sequence. The new C-terminus interferes with the function of the protein by changing it subcellular localization [91].

TE reactivation in somatic cells, on the other hand, may contribute to the development and progression of cancer, either by disrupting tumour suppressor genes or by inducing transcription of oncogenes located in the vicinity of the integration site [21, 92]. For example, insertions of LINE-1 elements into the tumour suppressor gene adenomatous polyposis coli (*APC*) have been described in several cases of colon cancer and likely played a driving role in the development of these malignancies [93, 94]. Furthermore, epigenetic mechanisms, such as DNA methylation, which repress TEs in most healthy somatic cells are dysregulated in cancer cells. Accordingly, retrotransposons are frequently reactivated in malignancies, which is thought to contribute to genome instability and has been linked to poor prognosis in multiple types of cancer [22]. In addition to the potential mutagenic effects of retrotransposon activity, several TE gene products have been described to directly exhibit oncogenic properties. The HERV-K accessory proteins Np9 and Rec, for instance, induce expression of the c-Myc proto-oncogene and promote cell proliferation [95]. Furthermore, the envelope proteins of many ERVs display immunosuppressive properties, which may protect cancer cells from the immune response [96].

In addition to their reported roles in tumorigenesis, transposon-derived nucleic acids and proteins resemble gene products produced during exogenous viral infections, and accordingly, elevated levels of TE transcripts have been linked to several autoinflammatory conditions including geographic atrophy, systemic lupus erythematosus (SLE) and Sjögren's syndrome [7, 97]. In mouse models, lupus pathogenesis is associated with an immune response against the ERV envelope protein gp70 [98]. Moreover, genes for transcription factors specifically targeting the corresponding ERV for epigenetic silencing have been identified as major lupus-susceptibility loci in mice [99]. Accumulation of HERV-derived gene products is also detectable in human SLE patients, accompanied by decreased expression of transcription factors involved in retrotransposon repression, suggesting that ERV reactivation may contribute to lupus pathogenesis in humans [99]. Equally, there is mounting evidence that TE dysregulation plays an important role in the progression of multiple sclerosis [100]. In particular, aberrant expression of the HERV-W envelope protein has been reported to activate Toll-like receptor 4 (TLR4) on oligodendral precursor cells, resulting in an inflammatory response that inhibits remyelination [101]. Monoclonal antibodies targeting the envelope protein of HERV-W showed therapeutic

potential in a mouse model of multiple sclerosis and are currently in clinical trials [102]. Aberrant retrotransposon activity may also contribute to the pathogenesis of several other neurological disorders including amyotrophic lateral sclerosis (ALS) and Rett syndrome. HERV-K env is detectable in neurons of ALS patients has been implicated in disease progression since HERV-K env expression is toxic to cultured human neurons and transgenic mice expressing HERV-K env in their neurons displayed a loss of motor neurons, accompanied by progressive motor dysfunction [103]. Rett syndrome, on the other hand, is caused by mutations in the methyl-CpG-binding protein 2 (MeCP2) gene. Defects in MeCP2 have been shown to result in elevated LINE-1 retrotransposition activity, which may contribute to the disease phenotype [104].

Notably, increased retrotransposon activity is also detectable in several psychiatric conditions including schizophrenia, although the role of TEs in these disorders is currently unclear [105, 106].

1.2 KAP1-dependent restriction of transposable elements

TEs represent major drivers of evolution and can fulfil important functions, particularly in early embryonic development. At the same time, they have the potential to severely damage the host genome if allowed to replicate unchecked. Accordingly, the host has evolved multiple mechanisms to tightly control retrotransposon activity. These include epigenetic mechanisms that specifically recognize integrated TE copies in the genome and repress transcription of these elements. This transcriptional silencing is complemented by an array of restriction factors that inhibit reverse transcription and reintegration of transposons that escape epigenetic repression.

1.2.1 The KAP1 / KRAB-ZFP system

One of the major mechanisms targeting TE sequences for transcriptional silencing is the KAP1/KRAB-ZFP system (**Fig. 7**). Krüppel-associated box (KRAB) domain zinc finger proteins (KRAB-ZFPs), the largest family of mammalian transcription factors, recognize retroelements with a variable C-terminal array of zinc fingers [107]. The conserved N-terminal KRAB domain recruits KRAB-associated protein 1 (KAP1, also known as TRIM28 or TIF1β), which serves as a platform for the assembly of an epigenetic silencing complex comprising the histone H3K9 methyltransferase SETDB1, heterochromatin

Protein 1 (HP1) and the nucleosome remodelling and deacetylase (NuRD) complex [108–113]. Together, the chromatin remodelling activities of the effectors recruited by KAP1 repress transcription of the targeted retrotransposon and adjacent genes.

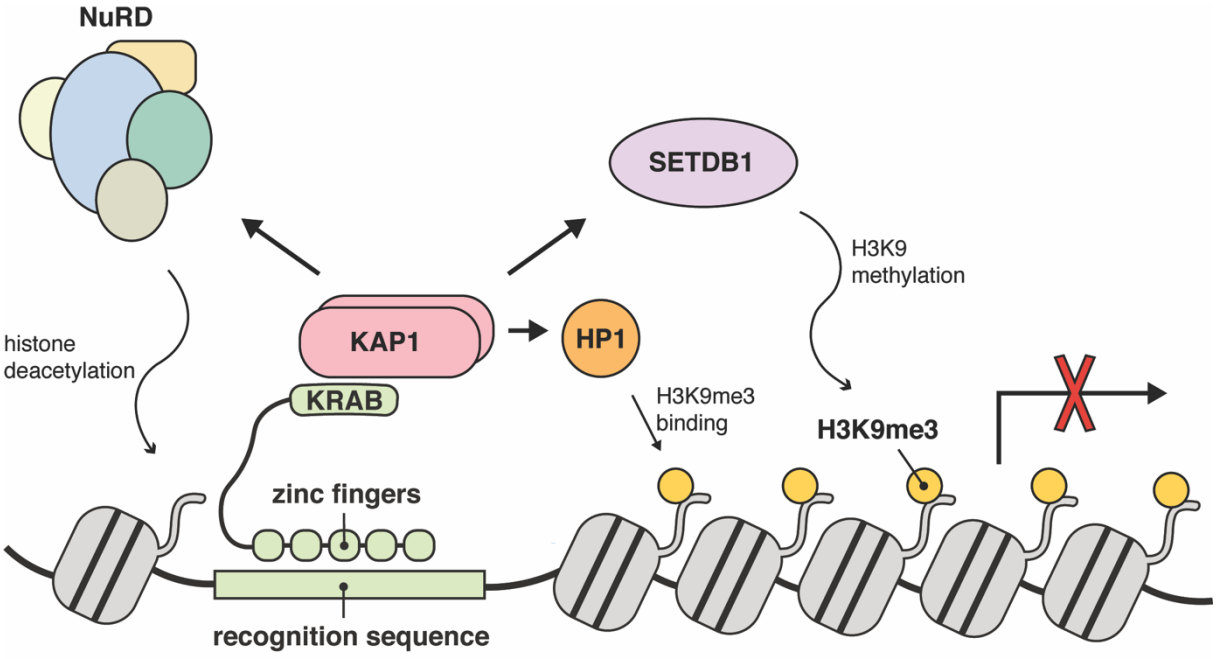

Fig. 7. KAP1-mediated transcriptional silencing of retrotransposons. KRAB-ZFPs specifically recognize transposon-derived sequences with their C-terminal zinc finger array. The N-terminal KRAB domain recruits KAP1, which serves as a platform for the assembly of a transcriptional silencing complex including SETDB1, HP1 and the NuRD complex [110, 111, 113]. The NuRD complex and SETDB1 remove activating histone marks and deposit repressive histone H3K9me3 modifications. HP1, on the other hand, binds H3K9me3 marks and has been implicated in heterochromatin formation and spreading [114]. Together, the chromatin modifiers recruited by KAP1 silence transcription of the target locus.

1.2.2 KRAB zinc finger proteins (KRAB-ZFPs)

KRAB-ZFPs first emerged in the common ancestor of coelacanths and tetrapods almost 420 million years ago [115, 116]. Since then, KRAB-ZFPs have been rapidly expanding and evolving, likely in response to invading TEs, and now represent the largest family of factors in mammals, with more than 400 members encoded in the human genome [117]. Approximately two-thirds of these proteins are thought to specifically target transposon-derived sequences for transcriptional repression [115]. This sequence-specific DNA binding is mediated by a variable array of C2H2 zinc fingers located in the C-terminus of the protein (**Fig. 7**). An average human KRAB-ZFP contains 12 zinc fingers, although the number of zinc fingers can vary from two to 40 [117]. This allows

KRAB-ZFPs to recognize relatively long stretches of DNA, thereby minimizing off-target effects. Each zinc finger in the array is usually 28 residues in size and adopts a ββα fold stabilized by coordination of a zinc ion. [118, 119]. A typical zinc finger binds in the major groove of DNA, recognizing three adjacent nucleotides. Subsequent zinc fingers in the tandem array are thought to wrap around the DNA and establish contacts in a similar manner. However, this canonical binding model is likely only correct for relatively short zinc finger arrays. In longer arrays with ten or more zinc fingers, the mode of DNA binding is significantly more complex and difficult to predict by current computational tools [118]. On the one hand, not all zinc fingers are necessarily contacting DNA. ZNF91, for example, contains 36 zinc fingers, 12 of which do not appear to contribute to DNA binding [107, 120]. Moreover, each zinc finger may affect the binding properties of neighbouring zinc fingers in the array [118]. The difficulties in predicting the binding mode of long zinc finger arrays are illustrated by a recent crystal structure of the eleven zinc fingers of the mouse KRAB-ZFP ZFP568 bound to DNA [121]. One of the zinc fingers in this array does not contact DNA but instead interacts with two of the other zinc fingers, while two more zinc fingers primarily bind the phosphate backbone. The remaining zinc fingers are involved in sequence-specific DNA recognition, but contact varying numbers of nucleotides, ranging from two to four, and therefore do not conform to the canonical 'one finger-three bases' rule. Moreover, two zinc fingers primarily contact the complementary DNA strand [121].

The majority of human KRAB-ZFP genes are concentrated in a six large cluster on chromosome 19 [122]. Due to their highly repetitive nature, these clusters are relatively unstable and prone to recombination events, resulting in frequent duplications or deletions. This is thought to facilitate rapid evolution of new KRAB-ZNF genes targeting currently active transposons [117, 119, 122]. Gene duplications may generate a redundant pair of KRAB-ZFPs, one of which is free to evolve new binding specificities. Conversely, deletions may remove KRAB-ZFP genes that are no longer required as the retrotransposon it originally targeted has decayed [118, 123]. Occasionally, however, such ancient KRAB-ZFPs are retained because they have been co-opted to regulate other cellular processes. As a result, most young KRAB-ZFPs found in the human genome target TEs for transcriptional silencing, while evolutionarily old family members often fulfil different functions [115, 116, 118]. For example, ZNF445 and ZNF57, which are highly conserved across mammals, are essential for maintaining genomic imprinting

[124–127]. By recruiting KAP1 to imprinting control regions, they protect these loci from the waves of genome-wide demethylation occurring during embryonic development. While the targets of most of these ancient KRAB-ZFPs are not readily recognizable as TE-derived, it seems likely that they originally recognized a transposon, which has since decayed, with only the KRAB-ZFP binding site remaining [118, 128].

Following recognition of their target sequence, KRAB-ZFPs induce epigenetic silencing of this locus by recruiting the co-repressor KAP1 via their N-terminal KRAB domain (**Fig. 7**). KRAB domains are approximately 50–75 amino acids in size and are composed of two modules known as KRAB-A and KRAB-B boxes [129]. While the KRAB-A box is required for transcriptional silencing, the KRAB-B box is not absolutely necessary and some KRAB-ZFPs only contain a KRAB-A box [130–132]. If present, however, the KRAB-B box enhances the repressive activities of the KRAB-ZFP [128, 133].

The KRAB domains found in KRAB-ZFPs are thought to have evolved from the KRAB domain of PR domain-containing protein 9 (PRDM9), which can be traced back more than 520 million years to the last common ancestor of vertebrates and echinoderms [134]. PRDM9, a key factor regulating meiotic recombination, contains an N-terminal KRAB domain, a central PR/SET domain with histone H3K4 and H3K36 methyltransferase activity and a C-terminal C2H2 zinc finger array, which facilitates sequence-specific DNA binding [135–137]. While precise role of the PRDM9 KRAB domain is still not fully understood, it is essential for PRDM9 function and appears to be mediating interactions with partner proteins including CXXC-type zinc finger protein 1 (CXXC1) [138]. Notably, however, the ancestral KRAB domain found in PRDM9 does not interact with KAP1 or induce transcriptional silencing, indicating that the ability to recruit the KAP1 co-repressor evolved at a later stage [116, 138, 139]. Most likely, duplication of the PRDM9 gene, accompanied by mutations in the KRAB domain that enabled interaction with KAP1 and loss of the PR/SET domain gave rise to the KRAB-ZFP family of proteins, which rapidly expanded and now constitutes the largest family of mammalian transcription factors [116, 118, 140]. The first evidence of such KAP1-recruiting KRAB domains can be found in coelacanth, indicating that the KAP1/KRAB-ZFP system emerged more than 400 million years ago, prior to the last common ancestor of tetrapods and coelacanths [115, 116].

The KRAB domains of most human KRAB-ZFPs, in particular the KRAB-A box, display a high level of sequence conservation [116] (**Fig. 8a**). The KRAB-B box, in contrast, is more variable than KRAB-A (**Fig. 8b**) and is absent in some KRAB-ZFPs [129]. Notably, however, the KRAB-A boxes of a small subset of approximately 40 KRAB-ZFPs significantly deviate from the consensus sequence. In contrast to most other KRAB-ZFP genes, which are relatively young, these diverging KRAB-ZFPs are evolutionarily older and do not appear to recruit KAP1 [116]. These KRAB domain variants have been proposed to originate from standard, KAP1-recruiting, KRAB-ZFPs, which over the course of evolution have been coopted to fulfil KAP1-independent functions distinct from TE repression. As a result, these variant KRAB-ZFP genes were retained, whereas standard KRAB-ZFPs are typically lost as the TE they originally targeted decays over time and replaced by new KRAB-ZFPs to respond to respond to current waves of TE invasion, as discussed earlier [116]. The precise functions of this KAP1-independent group of KRAB-ZFPs, however, are largely unexplored.

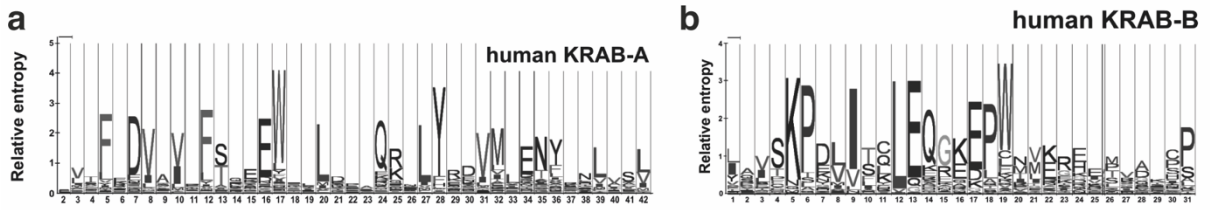

Fig. 8. Consensus sequence of human KRAB domains. (**a**) Hidden Markov Model (HMM) logo of human KRAB-A. (**b**) HMM logo of human KRAB-B. The HMM logos were taken from [141].

In addition to the KRAB domain and zinc fingers characteristic of this protein family, certain KRAB-ZFPs, in particular evolutionarily old family members with non-canonical KRAB domains, contain additional DUF3669 or SCAN domains at their N-terminus, but the role of these domains is still unclear [128]. Notably, however, the SCAN domain shows similarity to the capsid protein of Gypsy/Ty3 retrotransposons and can mediate homo- and heterodimerisation between SCAN domain containing proteins [142–146]. Further, it has been hypothesized that interactions with gag proteins may recruit these KRAB-ZFPs to ERV capsids, thereby targeting them to newly synthesized cDNA [145].

The transcriptional activities of KRAB-ZFPs are thought to be of particular importance during early development, where global hypomethylation favours reactivation of TEs [147–149]. The repressive histone modifications established by the KRAB-ZFP/KAP1 system in embryonic cells are then followed by cytosine methylation as a more

permanent form of silencing, which is maintained during development [150, 151]. While the DNA methylation induced by KRAB-ZFPs in embryos was long thought to result in irreversible silencing, it has now become apparent that repression of at least a subset of KRAB-ZFP targets requires continued presence of their cognate repressors. For example, the mouse KRAB-ZFPs ZFP932 and Gm15446 continue to regulate their ERV targets in differentiated tissues. Moreover, tissues naturally low in these transcription factors showed increased expression of adjacent genes driven from ERV promoters [152]. Consequently, it has been proposed KRAB-ZFPs may facilitate incorporation of TEs into cell-type or developmental stage specific transcriptional networks rather than simply silencing them [153]. In particular, tissue-specific variations in KRAB-ZFPs expressions, may allow a defined set of TEs to become derepressed and serve as cell-type specific enhancers [128, 154].

1.2.3 KRAB-associated protein 1 (KAP1)

KAP1, an 835 amino acid protein with a molecular weight of 89 kDa, is a member of the tripartite motif (TRIM) family (**Fig. 9**). Its N-terminus (amino acids 57-405) contains the defining feature of this protein family – an RBCC motif consisting of a RING domain, two B-boxes and a coiled-coil domain [155].

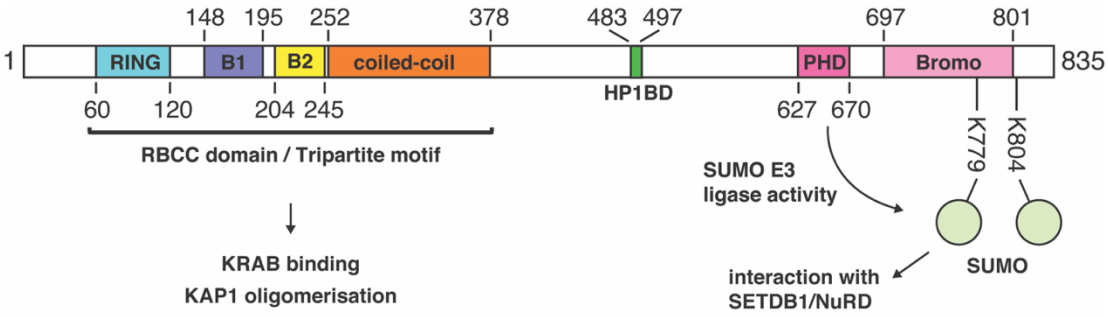

Fig. 9. Domain organization of KAP1. The N-terminus of KAP1 contains an RBCC motif consisting of a RING domain, two B-boxes (B1; B2) and a coiled-coil domain. The RBCC domain of KAP1 has been reported to mediate homotrimerization and interact with the KRAB-domain of KRAB-ZFPs. The central region of KAP1 contains a HP1 binding motif (HP1BD). The C-terminus of KAP1 contains a tandem PHD-bromodomain. The PHD domain acts as an intramolecular SUMO E3 ligase and mediates SUMO modification of KAP1 at 6 potential acceptor lysines. The two functionally most relevant SUMOylation sites K779 and K804, which are located in the bromodomain, are shown in the figure. SUMOylation of KAP1 is required for recruitment of SETDB1 and the NuRD complex. Domain boundaries shown in the figure were predicted by SMART [156]. The HP1 binding motif in KAP1 was described by [110, 157].

The primary function of the RBCC domain of KAP1 in silencing is to bind KRAB domains of KRAB-ZFPs and hence recruit KAP1 to its genomic targets. KAP1 has been reported to bind to KRAB domains with a 3:1 stoichiometry [158]. Furthermore, the RING domain, B-box 2 and coiled-coil domain of KAP1 all appeared to contribute to KRAB binding [159].

In addition to its role in KRAB binding, the RING domain of KAP1 has ubiquitin E3 ligase activity [160]. The physiological function of the ubiquitin ligase activity of KAP1 is largely unclear, although there is some evidence that KAP1 may ubiquitinate KRAB-ZNFs under certain circumstances, thereby modulating repression of the target gene [161, 162]. Importantly, however, the ubiquitin ligase activity of KAP1 has been implicated in tumorigenesis. MAGE-C2, expression of which is normally restricted to the male germline, but which is aberrantly expressed in many human cancers, directly binds to the coiled-coil domain of KAP1. This appears to increase the ubiquitin E3 ligase activity of KAP1 and direct it towards the tumour suppressor protein p53, resulting in its proteasome-mediated degradation [160]. Similarly, the normally germline-specific proteins MAGE-A3 and MAGE-A6 have been reported to bridge KAP1 with the tumour suppressor AMPK, thereby inducing its ubiquitination and subsequent degradation [163]. Interestingly, it has been proposed that the RING domain of KAP1 may also act as a small ubiquitin-like modifier (SUMO) E3 ligase. SUMOylation of the KAP1 substrate IRF7 appeared to require the RING domain, while the C-terminal domains of KAP1 were largely dispensable [164]. Similarly, SUMOylation of CDK9 by KAP1 has been reported to be dependent on its RING domain [165]. Moreover, KAP1 RING interacted with the SUMO E2 ligase Ubc9 in co-immunoprecipitation experiments [164].

The proline-, glycine- and serine-rich central part of KAP1 is predicted to be largely unstructured. Importantly, however, this region contains a PXVXL motif, which mediates recruitment of HP1 and is essential for KAP1-mediated transcriptional silencing [166]. The C-terminus of KAP1 comprises a tandem plant homeodomain (PHD)-bromodomain (amino acids 624-812), which cooperate to recruit repressive chromatin-modifying enzymes such as SETDB1 and the NuRD complex. The PHD domain binds to the SUMO E2 ligase Ubc9, thereby acting as intramolecular E3 ligase to direct SUMO modification of the bromodomain [112, 167, 168]. While many

bromodomains recognize acetylated lysines, the bromodomain of KAP1 has lost its ability to bind histone tails. Instead, its main function seems to be the recruitment SETDB1 and NuRD, which bind to the bromodomain of KAP1 following its SUMOylation by the PHD domain [167].

The C-terminus is the most variable region of TRIM proteins and accordingly, the nature of their C-terminal domains is used to sort the more than 70 members of the TRIM family into 11 classes [169, 170]. Together with TRIM24/TIF1α and TRIM33/TIF1γ, KAP1/TIF1β forms the class VI subfamily of TRIM proteins, which is also known as the transcription intermediary factor 1 (TIF1) family [171]. This subfamily is characterized by the presence of C-terminal PHD-bromodomains and like KAP1, the other class VI TRIM proteins appear to function primarily as transcriptional regulators. Furthermore, physical interactions and functional cooperation between the TIF1 family members have been described [172–174].

Initial biophysical studies on KAP1 reported that its RBCC domain self-associates to form homotrimers [158, 159]. However, this is inconsistent with more recent work on other member of the TRIM family. Most importantly, crystal structures of the coiled-coil domains of TRIM25, TRIM5, TRIM69 and TRIM20 show that these proteins are antiparallel dimers and, based on sequence conservation, coiled-coil mediated dimerization is likely conserved across the entire TRIM family [175–178]. The oligomeric state of KAP1, as well as the composition of KRAB:KAP1 complexes, therefore remain unclear.

1.2.4 Regulation of the transcriptional activities of KAP1 by post-translational modifications

1.2.4.1 SUMOylation

The transcriptional activities of KAP1 strongly depend on its posttranslational modifications (PTMs). Most notably, it has been demonstrated that only SUMO-conjugated KAP1 is able to recruit the repressive chromatin modifiers SETDB1 and NuRD [112, 168]. Accordingly, SUMOylation is essential for KAP1-dependent transcriptional silencing. SUMO modification of KAP1 is mediated by its own PHD domain, which possesses SUMO E3 ligase activity and is able to direct transfer of SUMO moieties to 6 potential acceptor lysines in KAP1 (K554, K575, K676, K750, K779 and

K804), with K779 and K804 being the most relevant SUMOylation sites [112] (**Fig. 9**). The extent of KAP1 SUMOylation is controlled by sentrin-specific peptidase 7 (SENP7), which removes poly-SUMO chains from KAP1 and has been shown to be essential for chromatin relaxation and efficient DNA repair in response to DNA damage [179].

SUMO interacting motifs (SIMs) in SETDB1 and CHD3, a subunit of the NuRD complex, have been shown to be essential for the interaction with SUMO-modified KAP1. However, the nature of these protein contacts is likely more complex than a simple SUMO-SIM interaction, as SETDB1 does not bind to all SUMOylated proteins, suggesting that additional contacts between KAP1 and SETDB1 may provide specificity [112]. Furthermore, SUMOylated KAP1, but not free SUMO, enhances the methyltransferase activity of SETDB1 [112]. Of the different SUMO paralogs encoded in the human genome, KAP1 appears to function primarily with SUMO2, while SUMO1 and SUMO3 are dispensable for KAP1-mediated silencing [180].

1.2.4.2 S824 phosphorylation

In addition to SUMOylation, KAP1 can be subjected to several other PTMs that modulate its transcriptional activities. Notably, phosphorylation at serine 824 (S824) alleviates KAP1-mediated repression. S824 of KAP1 can be phosphorylated by ataxia-telangiectasia mutated (ATM), as well as several other phosphatidylinositol 3-kinase-related kinases (PIKKs) [181]. This is particularly relevant in the context of a DNA damage response, where KAP1 S824 phosphorylation promotes chromatin relaxation to allow access of the DNA repair machinery to lesions located in heterochromatin [182]. Furthermore, KAP1 S824 phosphorylation in response to DNA damage induces expression of several genes controlling cell cycle arrest and apoptosis, such as p21 [155, 183]. The precise mechanism, by which S824 phosphorylation results in heterochromatin relaxation, however, remains controversial. Several studies reported, that KAP1 SUMOylation and S824 phosphorylation interfere with each other. Phosphorylation of KAP1 in response to DNA damage inducing agents resulted in reduced SUMOylation. Equally, phosphomimetic KAP1 mutants appeared to be significantly hypoSUMOylated, while SUMO-KAP1 fusion proteins displayed impaired S824 phosphorylation [183]. Dephosphorylation of KAP1 by PP1, on the other hand, promoted KAP1 SUMOylation [184]. In contrast to these reports, a different group found that KAP1 phosphorylation at S824 has no effect on its SUMOylation status.

Instead, their results suggested that upon phosphorylation, the extreme C-terminus of KAP1 resembles a SIM and competes with CHD3 for binding to SUMO [185].

Interestingly, Seki et al. found that a significant proportion of KAP1 is phosphorylated at S824 in murine embryonic stem cells (mESCs) [186]. Moreover, S824-P KAP1 appeared to be essential for the maintenance of pluripotency by cooperating with the transcription factor Oct4 to enhance expression of pluripotency-associated genes. This effect was dependent on the ability of S824-P KAP1 to form a complex with Oct4 and the transcriptional activator CBP on the respective promoter regions. Unphosphorylated KAP1, in contrast, lacked the ability to interact with Oct4 and CBP [186].

1.2.4.3 S473 phosphorylation

In addition to S824, KAP1 is also phosphorylated at S473 in response to DNA damage [187–189]. It has been reported that phosphorylation at this site, which is located in close proximity to the PXVXL motif of KAP1 inhibits its interaction with HP1 [190]. A later study, however, found that phosphomimetic KAP1 mutants showed no defects in HP1 binding. Instead, KAP1 S473 phosphorylation in combination with HP1 threonine 51 (T51) phosphorylation was proposed to reduce the ability of HP1 to bind to H3K9me3 marks. This in turn may release HP1 from chromatin and increase the accessibility of the DNA damage site, thereby contributing to efficient repair of lesions located in heterochromatin [187].

Furthermore, KAP1 has recently been described as an important regulator of muscle cell differentiation. In myoblasts, KAP1 is recruited to the promoters of numerous muscle-specific genes via a direct interaction with MyoD. Interestingly, the KAP1 complexes formed at these promoter regions comprise activators like p300 as well as repressors such as histone deacetylase 1 (HDAC1) and G9a. The activities of these transcriptional regulators counteract each other resulting in no expression of the target genes. During differentiation, however, KAP1 phosphorylation at S473 disrupts its interaction with HDAC1 and G9a without affecting its ability to recruit p300. This phosphorylation switch consequently leads to transcriptional activation of MyoD target genes and is a crucial regulator of myogenesis [191].

1.2.4.4 Other modifications

In addition, to the previously discussed serine phosphorylation, KAP1 is also phosphorylated at several tyrosine residues (Y449, Y458 and Y517) by Src family kinases. These phosphorylation events have been proposed to interfere with the interaction between KAP1 and HP1, thereby weakening the association of both proteins with chromatin [192]. Mono-ADP ribosylation of KAP1 by Sirtuin 6 (SIRT6), in contrast, has been shown to promote interaction with HP1, thereby enhancing the repressive activities of KAP1 [193]. Finally, KAP1 has been reported to be acetylated and this modification may be linked to its transcriptional activities since deacetylation by HDAC10 dampened KAP1-dependent repression of melanogenesis [194]. Further, KAP1 deacetylation by SIRT1 appears to play a role in the DNA damage response by promoting non-homologous end joining [195].

1.3 Additional mechanism regulating transposon activity

Several complementary silencing mechanisms have been shown to cooperate with the KRAB-ZFP/KAP1 system to target TEs for repression. These mechanisms are thought to be particularly important for evolutionarily young transposons that are not yet recognized by a KRAB-ZFP [196].

1.3.1 RNA-based silencing mechanisms

RNA-based silencing mechanisms are one of the oldest host defences limiting the activity of transposons and are thought to be employed by the majority of eukaryotic organisms [197–199]. In mammals, PIWI-interacting RNAs (piRNAs) are crucial for the repression of TEs in the male germline and defects in the piRNA pathway result in TE activation and sterility in mice [200, 201]. In addition to its established role in the germline, the piRNA pathway also appears to be active in human pluripotent stem cells, where it has been reported to repress young LINE-1 elements [202]. piRNA precursors are transcribed as long RNAs from genomic piRNA clusters, which are highly enriched in TE-derived sequences and rapidly evolving to adapt to newly emerging families of transposons [199, 201]. These transcripts are subsequently processed in a Dicer-independent manner into mature piRNAs, which are approximately 26-30 nucleotides in length. These small RNAs associate with PIWI-like proteins – members of the AGO

family that are specific to the germline – to form the piRNA-induced silencing complex (piRISC) [199]. Due to the high content of transposon sequences in the piRNA clusters they are produced from, piRNAs are able to guide this complex to TE transcripts, inducing their degradation. Notably, cleavage of the targeted RNA amplifies piRISC activity, since the sequence complementary to the initial piRNA is itself processed and loaded onto a PIWI protein [199]. This ping pong mechanism further serves to tailor the piRNA response to specific transposon families currently active in the cell. In addition to repressing TEs on a post-transcriptional level, the piRISC complex also appears to induce epigenetic silencing of transposons, by guiding the DNA methylation machinery to TEs [199].

Besides piRNAs, other classes of small RNAs also contribute to the control of transposons. For example, double-stranded RNA generated by bidirectional transcription from the LINE-1 5' UTR has been shown to be processed by Dicer into siRNAs that repress LINE-1 activity [203]. Further, the Microprocessor complex, which is involved in the generation of miRNAs recognizes and cleaves secondary structure elements in the 5' UTR of LINE-1 RNA as well as in *Alu*s, thereby reducing their retrotransposition activity [204].

1.3.2 Human silencing hub (HUSH)

The recently identified human silencing hub (HUSH) complex, consisting of the proteins TASOR, MPP8, and periphilin [205], has also been implicated in the repression of retrotransposons, particularly young LINE-1 elements, in human cells [206, 207]. Like KAP1, the HUSH complex is thought to mediate epigenetic silencing of its targets by recruiting the histone H3K9 methyltransferase SETDB1 [205, 208]. In addition, HUSH has been reported to interact with the chromatin remodeller microchidia CW-type zinc finger 2 (MORC2), thereby inducing chromatin compaction [209].

1.3.3 Post-transcriptional repression of transposons by antiviral restriction factors

Due to the similarities in their replication mechanisms, many antiviral factors that restrict exogenous retroviruses are also effective inhibitors of ERVs and other retrotransposons. Indeed, it has been speculated that some of these proteins originally evolved to repress TEs, before being adapted to fight exogenous viruses [7].

Members of the apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3 (APOBEC3) family exert antiretroviral activity by catalysing cytidine deamination in nascent cDNA, thereby causing hypermutation of the viral genome [210]. Similarly, APOBEC3 proteins have been reported to edit reverse transcripts produced during replication of various retrotransposon families. [211–213] Interestingly, however, APOBEC3 proteins appear to restrict retrotransposition via additional, deamination-independent mechanisms, which may involve inhibition of reverse transcription and sequestration of LINE-1 RNPs to cytoplasmic stress granules [214, 215].

Three-prime repair exonuclease 1 (Trex1) degrades cytoplasmic cDNA and has been shown to repress retrotransposition of both ERVs and LINE-1 elements [216]. However, the mechanism by which Trex1 inhibits LINE-1 appears to be independent of its exonuclease activity. Instead, Trex1 has been proposed to act by sequestering ORF1p and inducing its proteasomal degradation [217].

SAM domain and HD domain containing protein 1 (SAMHD1) is a known HIV restriction factor, which interferes with retroviral replication by depleting dNTPs in non-dividing cells. Similar to Trex1, however, SAMHD1 appears to inhibit LINE-1 elements via a non-canonical mechanism independent of its dNTPase activity, since catalytically inactive SAMHD1 variants are only marginally affected in their ability to repress LINE-1 retrotransposition [218]. The precise mechanism by which SAMHD1 restricts TEs is still debated but may involve sequestration of LINE-1 RNPs to stress granules and/or inhibition of ORF2p [218–220].

In addition to SAMHD1, Trex1 and the APOBEC3 family, multiple other antiviral factors including zinc finger CCCH-type antiviral protein 1 (ZAP) and RNAseL have been shown to inhibit retrotransposition of TEs that escape epigenetic repression [221–223].

THESIS OBJECTIVES

1. Determine the oligomeric state of KAP1

Reports on the oligomeric state of KAP1 were conflicting. While several studies suggested that KAP1 trimerizes [158, 159], more recent work on other TRIM proteins indicates that most, if not all, members of the TRIM family form antiparallel dimers through their coiled-coil domains [175]. In addition, certain members of the TRIM family have the ability to further oligomerize via their RING or B-box domains to form higher-order structures, which can be essential for their cellular functions [224–227]. However, whether this applies to KAP1 is unclear. I therefore re-examined the oligomeric state of KAP1 and assessed its potential to self-assemble into higher-order oligomers.

2. Determine the structure of the RBCC domain of KAP1

The RBCC domain of KAP1 is essential for its function as a transcriptional repressor by mediating interactions with numerous partner proteins including KRAB-ZNFs, SMARCAD1 and EZH2 [158, 228, 229]. However, when this project began, almost no structural information on the RBCC motif of KAP1 was available, except for a structure of the isolated B-box 2 domain (PDB: 2YVR). Moreover, no complete RBCC structure of any TRIM protein had been determined. To aid our understanding of the molecular basis of KAP1-dependent transcriptional silencing, I therefore aimed to determine the structure of the RBCC domain of KAP1.

3. Characterize the interaction of KAP1 with KRAB-ZNFs

The interaction of the RBCC domain of KAP1 with KRAB domains of KRAB-ZNFs is crucial for recruitment of KAP1 to its genomic targets. However, many aspects of this important interaction, including the stoichiometry of the complex and the location of the KRAB binding site in KAP1, remained unclear. Here, I investigated the composition

of KAP1:KRAB complexes and located the interaction surface in KAP1 mediating KRAB binding.

4. Determine the structure of a KRAB:KAP1 complex

Available biophysical and structural data on KRAB domains suggest that this domain is largely disordered in isolation but folds upon binding to KAP1 [158]. To identify the residues in KAP1 and the KRAB domain involved in the KAP1-KRAB binding interface and understand the conformational changes that may occur upon binding, I attempted to determine the structure of KAP1 in complex with a KRAB domain.

CHAPTER 2

ROLE OF KAP1 HIGHER-ORDER OLIGOMERIZATION IN TRANSCRIPTIONAL SILENCING

2.1 Introduction

KAP1 is a member of the TRIM family of proteins. Its N-terminus contains the defining feature of this protein family - an RBCC motif consisting of a RING domain, two B-boxes and a coiled-coil domain. The RBCC domain is essential for the cellular functions of KAP1 by mediating its interaction with KRAB domains of KRAB-ZNFs, which is crucial for recruitment of KAP1 to its genomic targets. Moreover, contacts with numerous other binding partners of KAP1, including PP1, SMARCAD1, EZH2 and members of the MAGE protein family are also formed via the RBCC domain [160, 184, 228–230]. However, at the start of this project, no structural information was available on the RBCC motif of KAP1 except for a crystal structure of the isolated B-box 2 domain (PDB: 2YVR). Equally, biophysical data on KAP1 was limited. Several early studies reported that the RBCC domain of KAP1 mediates homotrimerization of the protein. However, this is inconsistent with more recent work on other members of the TRIM family. Crystal structures of the coiled-coil domains of TRIM25, TRIM5, TRIM69 and TRIM20 revealed that all of these proteins are antiparallel dimers – a feature that, based on sequence conservation, is predicted to be conserved across the entire TRIM family [175–178]. This antiparallel arrangement places the RING and B-box domains at either end of the coiled-coil. In some TRIM family members, these domains are themselves capable of oligomerization independently of the coiled-coil domain, which can lead to polymerization of TRIMs into lattices or scaffolds. The B-box 2 domain of TRIM5, for instance, has been shown to trimerize [227, 231], thereby mediating assembly of TRIM5 dimers into hexagonal nets around viral capsids entering the cell [232, 233]. The RING

domain equally has the potential to mediate further oligomerization of TRIM dimers. The RING domain of TRIM32, for example, forms stable homodimers at either end of the coiled-coil domain, thereby inducing the assembly of TRIM32 into tetramers [234]. The RING domain of PML/TRIM19, on the other hand, is able to tetramerize, a property that has been reported to contribute to the formation of PML nuclear bodies [226]. The assembly of TRIM proteins into such higher-order structures can be essential for their cellular function, as is well established for both TRIM5 and PML [224, 235]. However, whether KAP1 has the potential to form higher-order oligomers is unclear. Here, I used a combination of biophysical and structural approaches to confirm that KAP1 forms antiparallel dimers like other members of the TRIM family rather than trimers as previously reported. Further, my results show that KAP1 dimers self-assemble into higher-order oligomers via homodimerization of its B-box 1 domain. Structure-based mutations in B-box 1 abolished higher-order oligomerization and reduced KAP1-dependent transcriptional repression in luciferase reporter assays suggesting a potential role for higher-order oligomer formation in KAP1 function.

2.2 Results

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

To determine the oligomeric state of KAP1 and assess its self-assembly potential, I purified full-length KAP1, as well as its RBCC motif (residues 50-413) (**Fig. 10a**). Protein of sufficient quality and quantity for structural and biophysical studies was readily obtained using a bacterial expression system (**Fig. 10b and c**). Notably, however, purification of the isolated RBCC domain required buffers containing relatively high salt concentrations (0.3 - 0.5 M NaCl), with lower salt concentrations resulting in significant protein precipitation. Full-length KAP1, in contrast, was stable at all NaCl concentrations tested (0.15 M – 0.5 M).

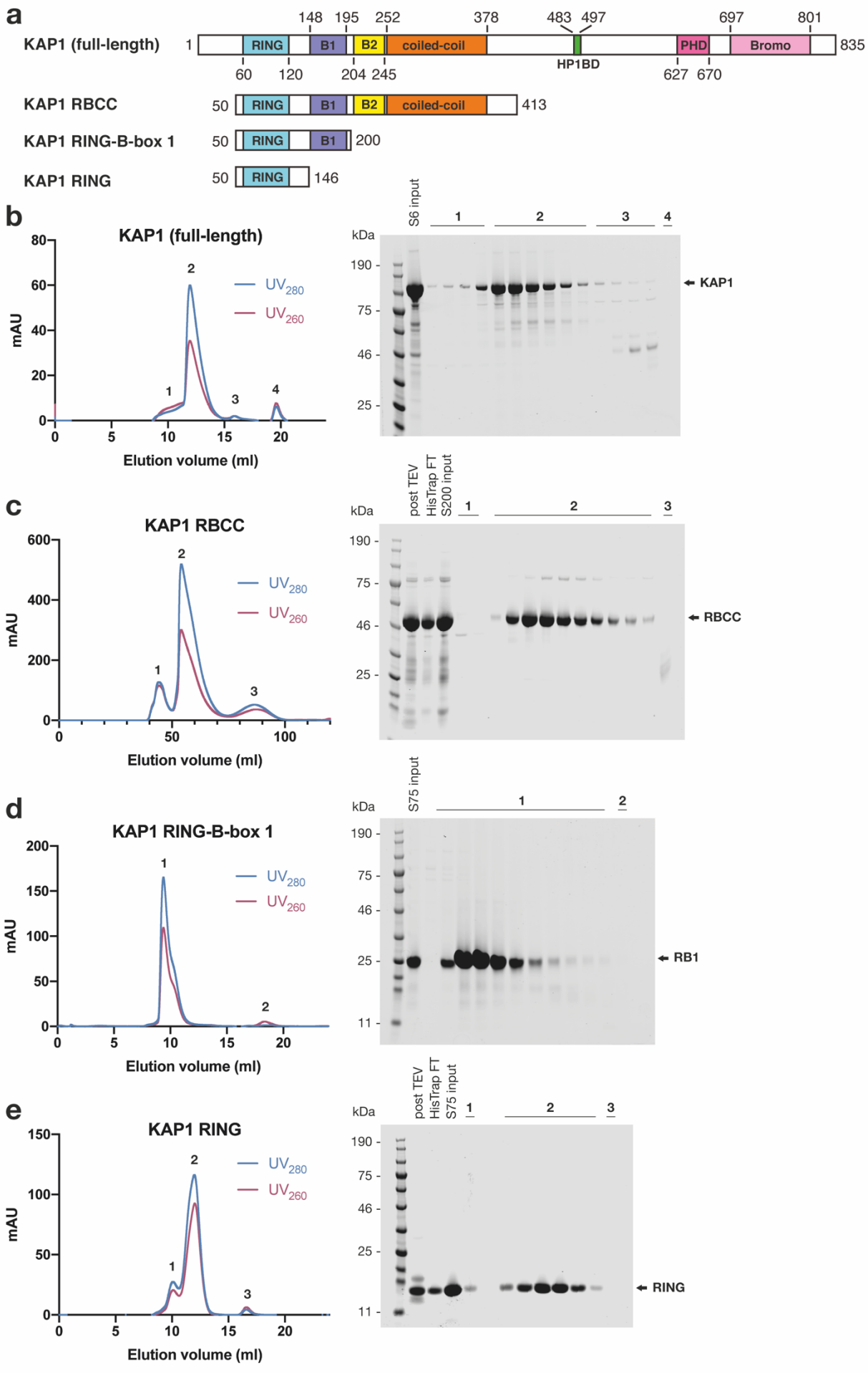
a
KAP1 (full-length)
KAP1 RBCC
KAP1 RING-B-box 1
KAP1 RING
RING
B1
B2
coiled-coil
PHD
Bromo
HP1BD
b
KAP1 (full-length)
mAU
Elution volume (ml)
UV280
UV260
S6 input
kDa
KAP1
c
KAP1 RBCC
post TEV
HisTrap FT
S200 input
RBCC
d
KAP1 RING-B-box 1
S75 input
RB1
e
KAP1 RING
RING

Fig. 10. Purification of KAP1 constructs. (**a**) Schematic representations of the protein constructs purified to analyse the oligomerization behaviour of KAP1. Domain boundaries shown in the figure were predicted by SMART [156]. The HP1 binding motif in KAP1 is annotated as described in [110]. (**b-e**) Purification of full-length KAP1, KAP1 RBCC, KAP1 RING-Box 1, and KAP1 RING constructs. Left Panels: Elution profiles of the respective constructs following affinity purification and removal of the tag with tobacco etch virus (TEV) protease. The following columns were used: Superose 6 Increase 10/300 (full-length KAP1), HiLoad 16/600 Superdex 200 pg column (RBCC) and Superdex 75 10/300 column (RING-B-box1 and RING). Right Panel: SDS-PAGE gels of the major peaks in the elution profiles. post TEV: sample after overnight incubation with TEV protease. HisTrap FT: sample after removal of protease and uncleaved protein.

The apparent molecular weight of KAP1 in solution was subsequently analysed using size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS). Due to the instability of the KAP1 RBCC construct at low salt concentrations, all experiments involving the isolated RBCC domain were performed at 0.5 M NaCl. Since full-length KAP1 was less sensitive to salt concentration, its oligomerization behaviour was analysed both at 0.5 M NaCl and at a more physiological salt concentration of 0.2 M.

At low protein concentration (8.7 µM, 0.35 g L^{-1}), the average molecular weight derived from SEC-MALS (86.8 kDa) for KAP1 RBCC closely matched the theoretical mass of a homodimer (81.6 kDa), showing that KAP1 forms dimers like other members of the TRIM family rather than trimers (**Fig. 11a**). Interestingly, however, when the protein was injected at higher concentrations, asymmetric tailing peaks characteristic of weakly self-associating proteins were observed for KAP1 RBCC, accompanied by a shift to increasingly early elution volumes. Consistently, the apparent molecular weight increased by up to 68% as the protein concentration was increased to 0.3 mM (12.4 g L^{-1}), indicating that KAP1 RBCC dimers have the potential to self-assemble into higher-order oligomers (**Fig. 11a**).

A similar behaviour was observed when full-length KAP1 was analysed in a buffer containing 0.2 M NaCl confirming that higher-order oligomerization is not an artefact of isolating the RBCC domain (**Fig. 12a**). Notably, however, high salt concentrations appeared to weaken the interactions between full-length KAP1 dimers. While full-length KAP1 readily self-assembled at 0.2 M NaCl, this was almost completely abolished by increasing the salt concentration to 0.5 M NaCl (**Fig. 12b**). The isolated RBCC

domain, in contrast, retained the ability to assemble into higher-order oligomers even at 0.5 M NaCl (**Fig. 11a**). A salt concentration of 0.5 M NaCl is unphysiologically high, however, and consequently, self-assembly of KAP1 is expected to occur in a cellular setting.

To obtain a more direct and quantitative model of KAP1 self-association sedimentation equilibrium analytical ultracentrifugation (SE-AUC) on KAP1 RBCC at concentrations from 1 to 200 μM was performed in collaboration with Dr Stephen McLaughlin (MRC Laboratory of Molecular Biology). As for the SEC-MALS experiments, a buffer containing 0.5 M NaCl was used to stabilize the RBCC domain. The equilibrium sedimentation profiles were consistent with a dynamic equilibrium between dimeric and oligomeric KAP1 (**Fig. 11b**), in agreement with my SEC-MALS data. The average molecular mass at the lowest concentration was approximately 80 kDa (**Fig. 11c**), 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. 11c**). 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. 11c**). 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. 11c**). 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. Together, these data indicate 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 the SE-AUC data cannot definitively confirm or rule out the presence of KAP1 species larger than octamers.

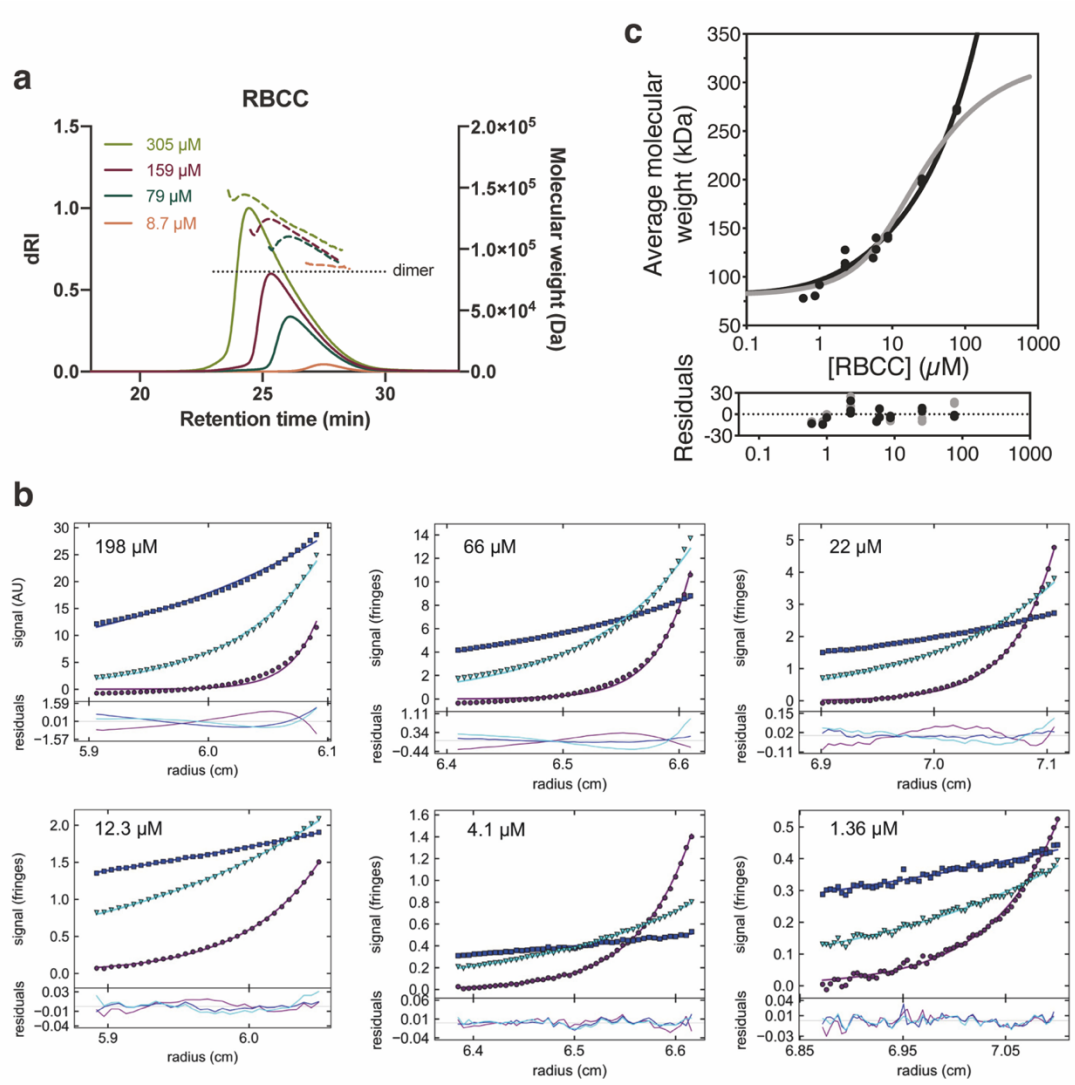

Fig. 11. Self-assembly of the KAP1 RBCC domain in solution. (**a**) SEC-MALS analysis of KAP1 RBCC in a buffer containing 0.5 M NaCl. (**b**) Sedimentation equilibrium analytical centrifugation (SE-AUC) of KAP1 RBCC. The equilibrium distributions of KAP1 RBCC, at different concentrations in each inset, during centrifugation at 5,000 rpm (blue squares), 8,500 rpm (cyan triangles) and 15,000 rpm (magenta circles) were fitted to a single species model (solid lines) to obtain average molecular weights. (**c**) SE-AUC analysis of RBCC molecular weight as a function of protein concentration. The average molecular weight isotherm from individual fits at different concentrations was fitted to an isodesmic self-association model (black line) yielding dissociation constants of $K_{d2,4}$ = 9 μM [with a 1-σ (68.3%) confidence interval of 7-11 μM] and K_{diso} = 19 μM [1-σ confidence interval 16-23 μM] for the dimer-tetramer and isodesmic equilibria, respectively. An alternative fit to a dimer-tetramer-octamer model (grey line) yielded dissociation constants of $K_{d2,4}$ = 12.6 μM [1-σ confidence interval 7.4-23.5 μM] and $K_{d4,8}$ = 6 μM [1-σ confidence interval 3-11 μM] for the dimer-tetramer and tetramer-octamer equilibria, respectively. The SE-AUC experiments were performed by Dr Stephen McLaughlin (MRC Laboratory of Molecular Biology).

RING, B-box 1 and B-box 2 domains have each been reported to dimerize or oligomerize in TRIM family proteins [226, 227, 234, 236]. To map, which of its constituent domains drives self-assembly of the RBCC dimer in the case of KAP1, I assessed the oligomerization potential of some of these domains. As for full-length KAP1 and the RBCC domain, highly pure protein was readily obtained using a bacterial expression system (**Fig. 10d and e**). In contrast to the full RBCC domain, KAP1 RING and KAP1 RING-B-box 1 constructs did not require high salt concentrations for stability and consequently, all experiments involving these constructs were performed at 0.2 M NaCl.

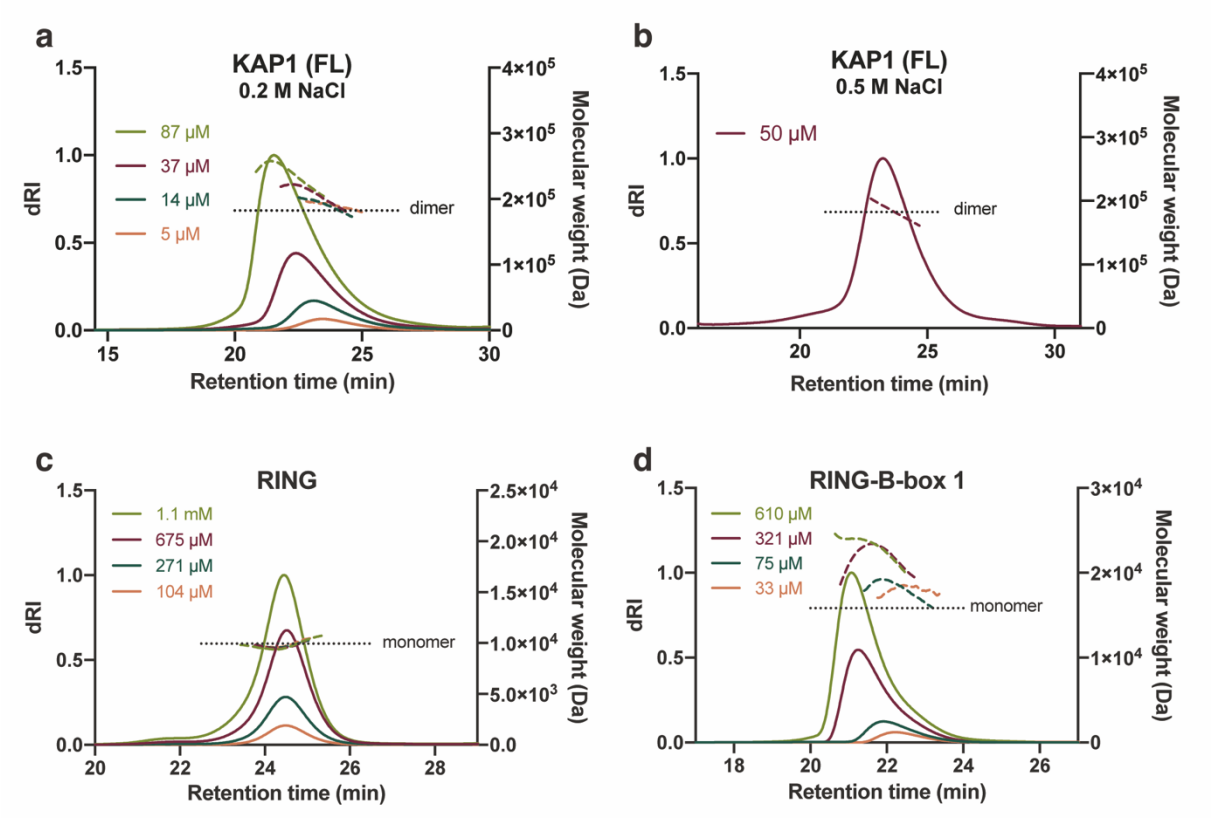

Fig. 12. Self-assembly of full-length KAP1 in solution and mapping of the domain mediating higher-order oligomerization of KAP1. (**a**) SEC-MALS analysis of full-length KAP1 in a buffer containing 0.2 M NaCl. (**b**) SEC-MALS analysis of full-length KAP1 in a buffer containing 0.5 M NaCl. (**c**) SEC-MALS analysis the RING domain of KAP1 in a buffer containing 0.2 M NaCl. (**d**) SEC-MALS analysis of KAP1 RING-B-box 1 in a buffer containing 0.2 M NaCl.

SEC-MALS analysis of the isolated RING domain of KAP1 showed that it remained strictly monomeric even at protein concentrations as high as 1.1 mM (10.5 g L^{-1}) (**Fig. 12c**). In contrast, a fragment containing the RING and B-box 1 domains readily self-

associated with increasing protein concentration (**Fig. 12d**), suggesting that B-box 1 drives assembly of the RING-B-box 1 fragment into weakly associated dimers. While this indicates that B-box 1 is involved in KAP1 higher-order oligomerization, a contribution of B-box 2 cannot be ruled out. Notably, the isolated B-box 2 domain of KAP1 crystallized as a dimer with a surface area of 1,160 $Å^2$ buried at the dimer interface (PDB: 2YVR), suggesting that it may also contribute to KAP1 higher-order-oligomerization. Evaluation of the relative contributions of B-box 1 and B-box 2 was difficult since B-box 1-B-box 2-coiled-coil and B-box 2-coiled-coil constructs were both insoluble. However, comparison of the B-box 2 domains of KAP1 and TRIM5α makes a role for KAP1 B-box 2 in higher-order oligomerization appear unlikely. In the crystal structure of KAP1 B-box 2, the two monomers associate via hydrophobic interactions involving L227 and L245 (**Fig. 13a**). TRIM5α is known to assemble into hexagonal lattices, which is mediated by trimerization of its B-box 2. The molecular basis of TRIM5α higher-order oligomerization was recently elucidated by Wagner et al., who could recapitulate the trimeric arrangement of TRIM5α B-box 2 using monovalent BCC "miniTRIM" constructs containing the B-box 2 domain and a truncated coiled-coil hairpin [227]. This approach eliminated further assembly driven by coiled-coil dimerization and allowed to determine the structure of the B-box 2 trimer interface. This crystal structure revealed that trimerization of B-box 2 is mediated by an interaction network consisting of hydrophobic contacts involving L105, L118 and W117, as well as polar interactions including a salt bridge formed by R121 and E120 [227] (**Fig. 13b**).

Notably, this interface is located on the opposite side of the B-box 2 domain compared to the interaction surface observed in the KAP1 B-box 2 crystal structure. Moreover, superposition of KAP1 B-box 2 on the TRIM5α "miniTRIM" structure reveals that the corresponding residues in TRIM5α are in contact with the coiled-coil domain and not accessible for protein-protein interactions (**Fig. 13c**). Consequently, the interface observed in the KAP1 B-box 2 crystal structure is likely an artefact resulting from removal of the coiled-coil domain, which exposed hydrophobic residues on the backside of the B-box 2 domain.

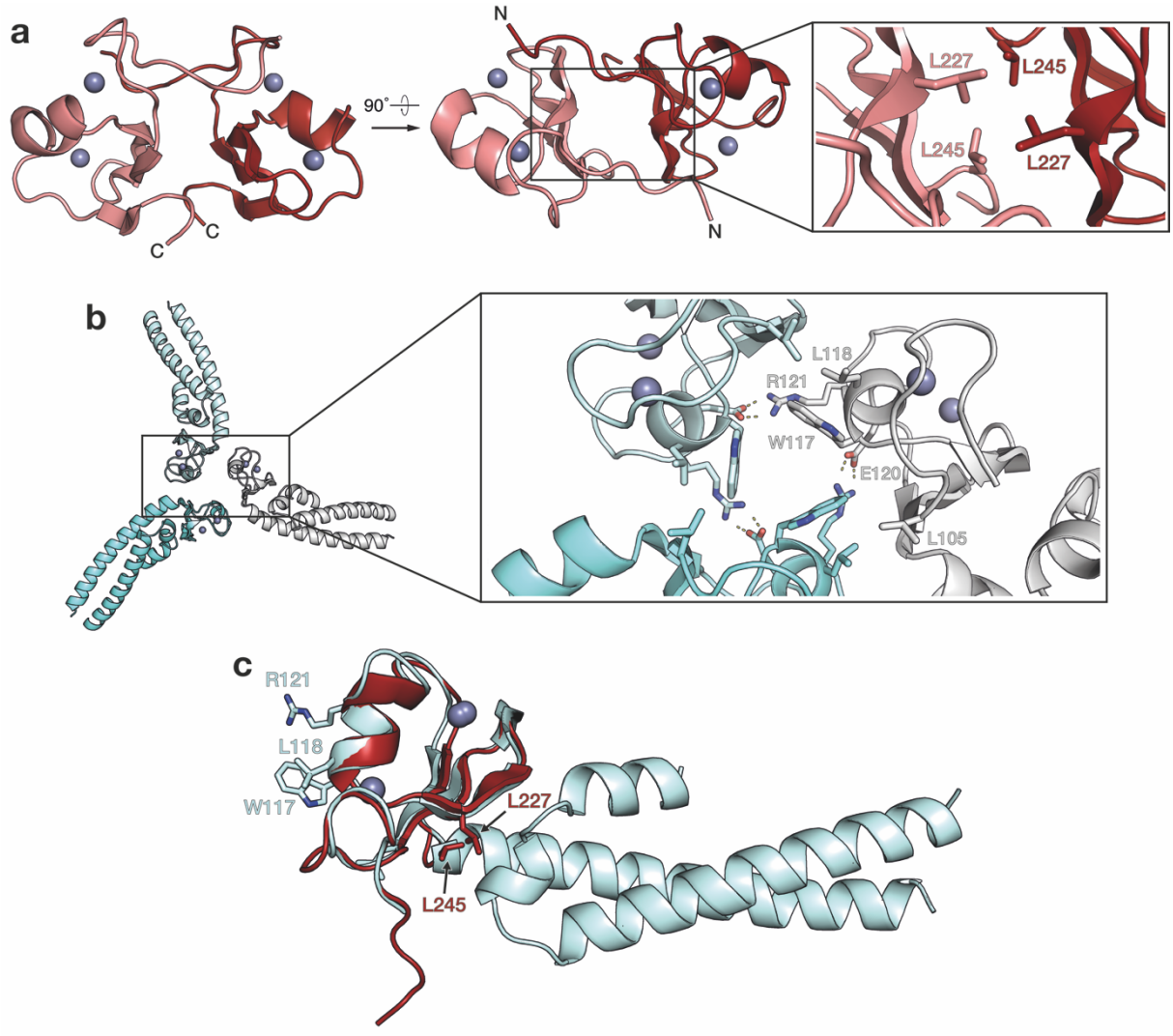

Fig. 13. Comparison of the oligomerization interfaces of the B-box 2 domains of KAP1 and TRIM5α. (**a**) Structure of the KAP1 B-box 2 dimer (PDB: 2YVR). (**b**) Trimeric assembly of the B-box 2 domain of TRIM5α (PDB: 5IEA) [227] (**c**) Superposition of KAP1 B-box 2 (red) on the TRIM5α "miniTRIM" structure (cyan). Residues involved in the respective interaction interfaces are highlighted.

Together, the data presented in this chapter show 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.

2.2.2 Crystal structure of the RBCC domain of KAP1

To understand the molecular basis of KAP1 self-assembly and identify its oligomerization interfaces, I attempted to determine the crystal structure of the RBCC domain of KAP1. Although crystals of KAP1 RBCC were readily obtained (**Fig. 14**), they initially diffracted X-rays poorly, to a resolution of approximately 9 Å.

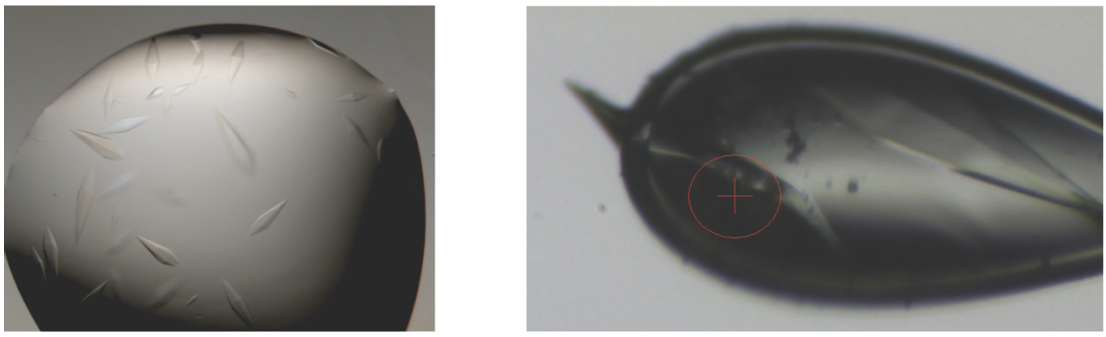

Fig. 14. Crystallization of KAP1 RBCC. Left Panel: crystals of KAP1 RBCC grown by sitting-drop vapor diffusion against a reservoir solution containing 7 % PEG 4,000, 0.1 M NaCl, 20% glycerol, 0.1 M HEPES pH 7. Right panel: KAP1 RBCC crystals following freezing in liquid nitrogen and mounting on the beamline.

Multiple approaches to improve the diffraction quality of the crystals were tested. Initially, the effects of a range of different additives were assessed using the Morpheus and ANGSTROM additive screens (Molecular Dimensions) [237]. To this end, the initial crystallization condition (7% PEG 4,000, 0.1 M NaCl, 20% glycerol, 0.1 M HEPES pH 7) was supplemented with 10% of the additive solutions from these screens before crystallization experiments were set up. However, the tested additives failed to promote crystallization in a different crystal form or improve diffraction. To rule out crystal damage during flash-cooling, data collection at room temperature was attempted, but no improvement in diffraction quality was observed.

As efforts to improve the crystals by optimizing the crystallization condition were unsuccessful, I subsequently tested whether modifications to the KAP1 RBCC construct would result in crystals of better quality. Fusion of bacteriophage T4 lysozyme (T4L) to the C-terminus of a TRIM5 B-box 2-coiled-coil fragment has been successfully utilized to facilitate crystallization of this construct [177]. I therefore tested a similar strategy and expressed KAP1 RBCC as an N-terminal T4L fusion. The resulting T4L-RBCC fusion protein was soluble and could be purified from *Escherichia coli* (*E. coli)* with high yield (**Fig. 15a**). However, while the fusion protein crystallized in a variety of conditions (**Fig. 15b**), these crystals only diffracted to a resolution of approximately 8 Å, precluding structure determination.

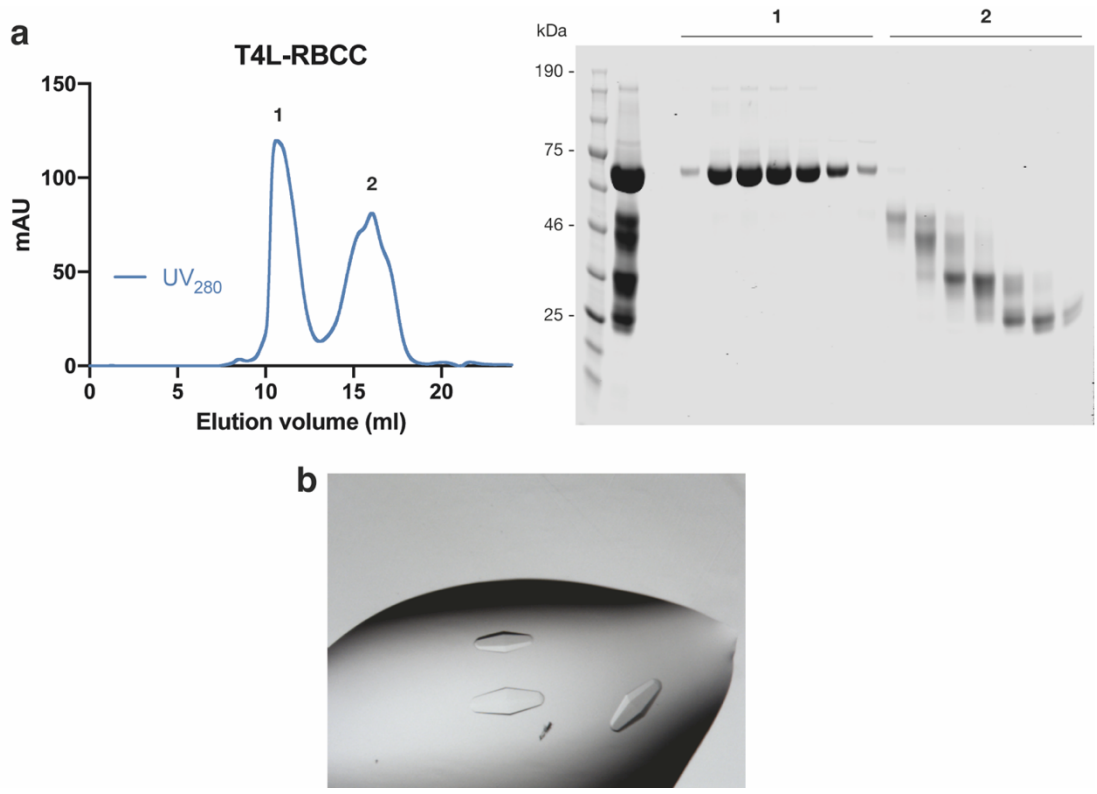

Fig. 15. Purification and crystallization of T4L-RBCC fusion protein. (**a**) Left panel: Elution profile of the T4L-RBCC fusion protein on a Superdex 200 Increase 10/300 column following nickel-affinity purification. Right panel: SDS-PAGE gel of the major peaks in the elution profile. (**b**) Crystals of the T4L-RBCC fusion protein grown by sitting-drop vapor diffusion against a reservoir containing 1.75 M NaCl, 0.1 M HEPES pH 7.

Notably, the morphology of the T4L-RBCC crystals strongly resembled that of the RBCC crystals, suggesting that the addition of T4L was not sufficient to disrupt the crystal contacts leading to the previously observed, weakly diffracting crystal form. I hence tested whether a larger fusion partner such as maltose binding protein (MBP) could change the crystallization behaviour of the RBCC domain. Since the N-terminus of KAP1 RBCC is predicted to be α-helical, I aimed to continue the C-terminal α-helix of MBP with a helical linker to produce a more rigid fusion protein and increase its crystallization propensity [238]. If using helical linkers, the exact length of the spacer may be critical for crystallization success, as every additional amino acid shifts the orientation of MBP relative to the RBCC domain by approximately 100º [239]. Consequently, I designed six N-terminal MBP fusion constructs with linkers ranging from one to six alanine residues. While all of these fusion proteins could be produced in *E. coli* with high yields (**Fig. 16a-i**), no crystal formation was observed. Surface entropy

reduction [240] on the background of the T4L-RBCC fusion protein was tested but the resulting K365A/K366A variant was insoluble in *E. coli*.

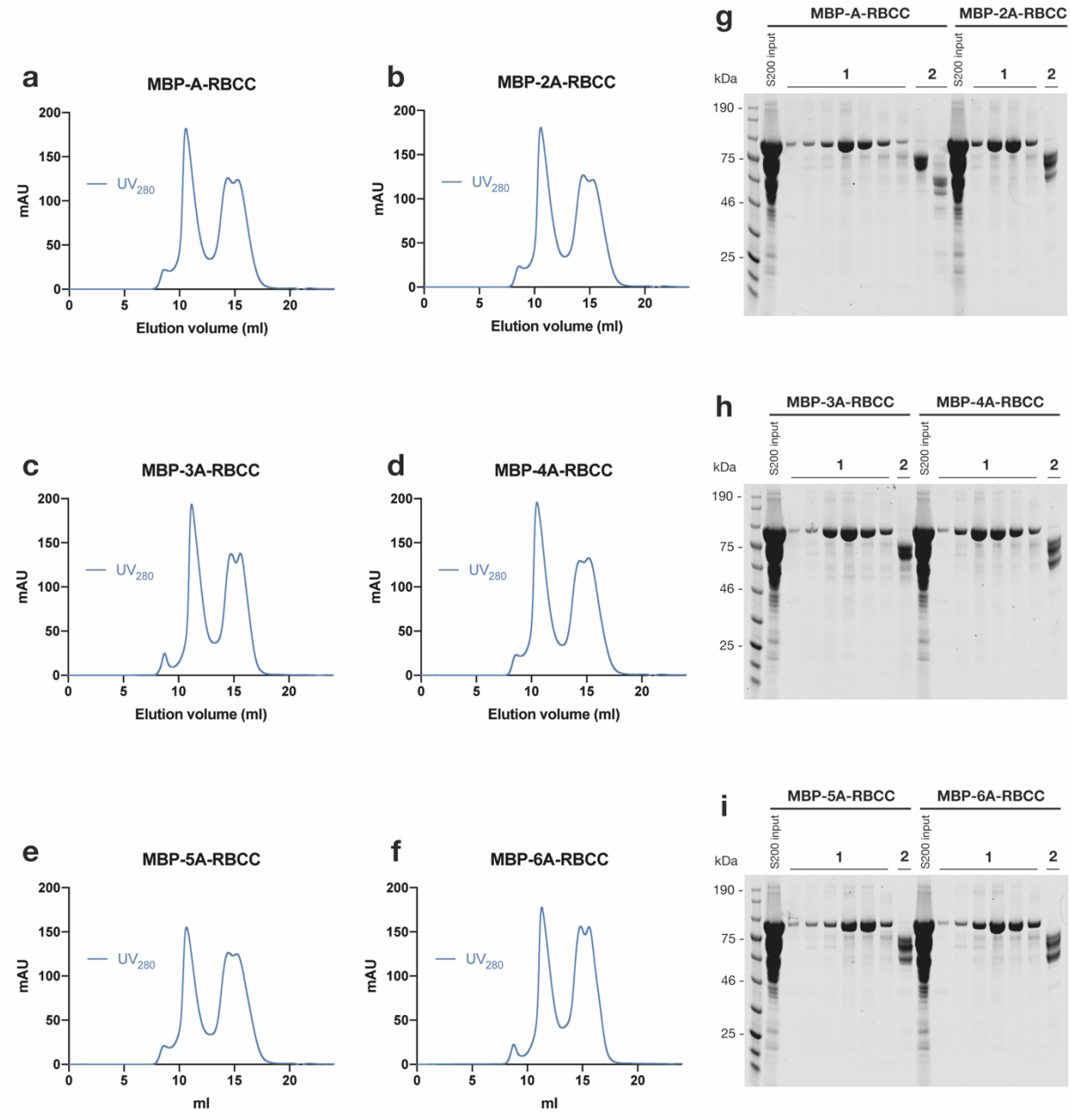

Fig. 16. Purification of MBP-RBCC fusion constructs. (**a-e**) Elution profiles of MBP-RBCC fusion constructs with linkers ranging from one (MBP-A-RBCC) to six alanine residues (MBP-6A-RBCC) on a Superdex 200 Increase 10/300 column following nickel-affinity purification. (**g-i)** SDS-PAGE gels of the major peaks in the elution profiles.

I subsequently tested whether the diffraction quality of the KAP1 RBCC crystals could be improved by reductive methylation of lysine residues in the protein [241–243]. Methylation was well tolerated and did not cause aggregation or precipitation of the protein (**Fig. 17a**). Successful dimethylation of 27 amino groups in KAP1 RBCC (all 26

lysine residues and the N-terminus of the protein) was confirmed by mass spectrometry (**Fig. 17b**). However, the modified protein failed to crystallize.

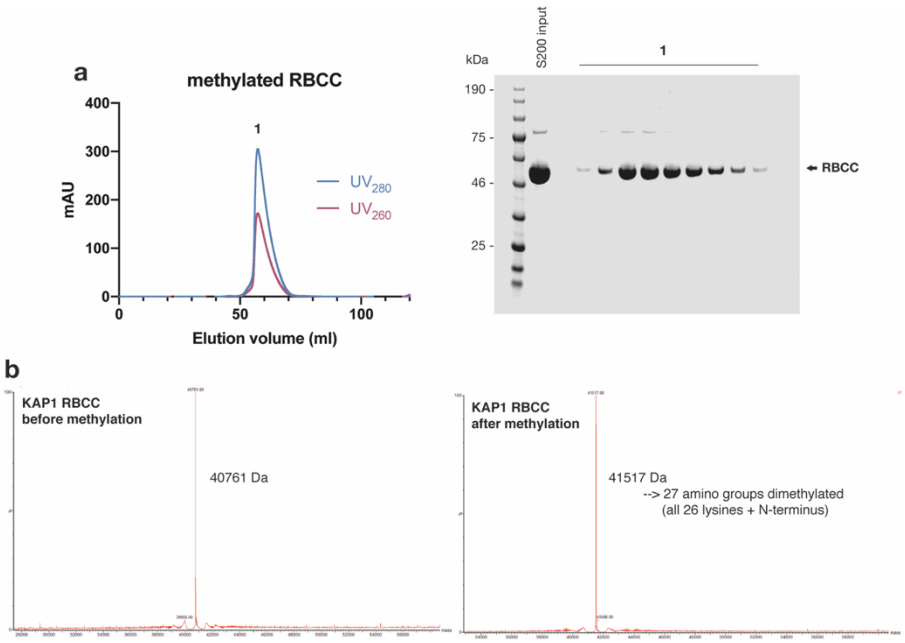

Fig. 17. Reductive methylation of KAP1 RBCC. (**a**) Left Panel: Elution profile of methylated KAP1 RBCC on a HiLoad 16/600 Superdex 200 pg column. Right Panel: SDS-PAGE gel of the major peak in the elution profile. (**b**) Mass spectra of the RBCC domain of KAP1 before and after methylation of primary amines (lysines and N-terminus) with dimethylamine borane complex (ABC).

Crystals suitable for structure determination were eventually obtained by combining several of the previously described approaches: T4L was fused to the N-terminus of KAP1 RBCC followed by reductive methylation of primary amines in the purified protein prior to crystallization (**Fig. 18a-c**). 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 by single anomalous dispersion (SAD) phasing using the anomalous scattering signal from the zinc atoms in the RING and B-boxes. The asymmetric unit contained two molecules. The atomic model was refined at 2.9 Å resolution. Representative samples of the electron density map are shown in **Fig. 20**.

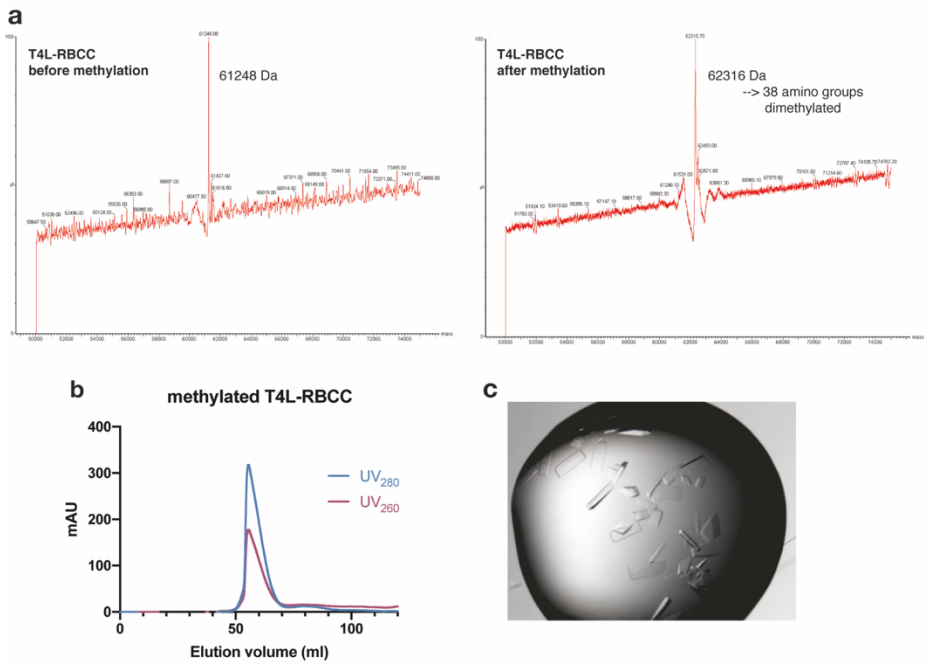

Fig. 18. Reductive methylation and crystallization of T4L-RBCC. (**a**) Mass spectra of the RBCC domain of KAP1 before and after methylation of primary amines. (**b**) Elution profile of methylated T4L-RBCC on a HiLoad 16/600 Superdex 200 pg column. (**c**) Crystals of methylated T4L-RBCC grown by sitting-drop vapor diffusion against a reservoir containing 17% PEG 3,350, 100 mM $MgCl_2$, 0.1 M HEPES pH 7.5.

The overall structure of the KAP1 RBCC dimer resembles a dumbbell (**Fig. 19**). The coiled-coil domain forms a 16 nm-long antiparallel coiled coil that contains all the dimer contacts. Each end of the coiled coil is 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 (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 51-55 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. 19b**).

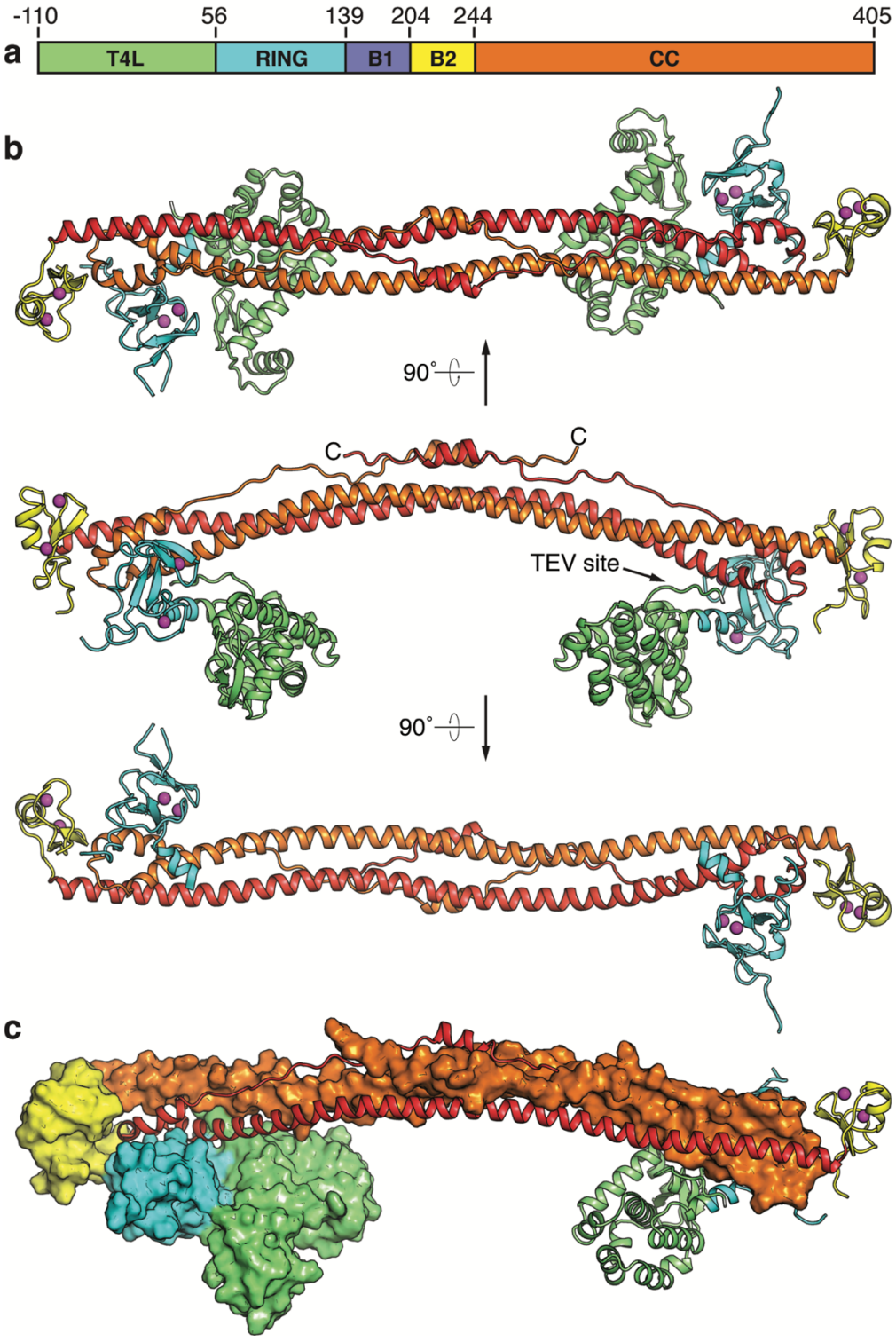

Fig. 19. Crystal structure of KAP1 RBCC. (**a**) Domain organization of the crystallized construct. T4L, T4 lysozyme; B1, B-Box 1; B2, B-box 2; CC, coiled-coil. (**b**) Overall structure of the RBCC homodimer. Three views along or perpendicular to the dyad are shown. The components are coloured as in (*a*). Zn atoms are shown as magenta spheres. (**c**) View of the RBCC dimer perpendicular to the dyad, with one subunit shown as a cartoon and the other as a surface.

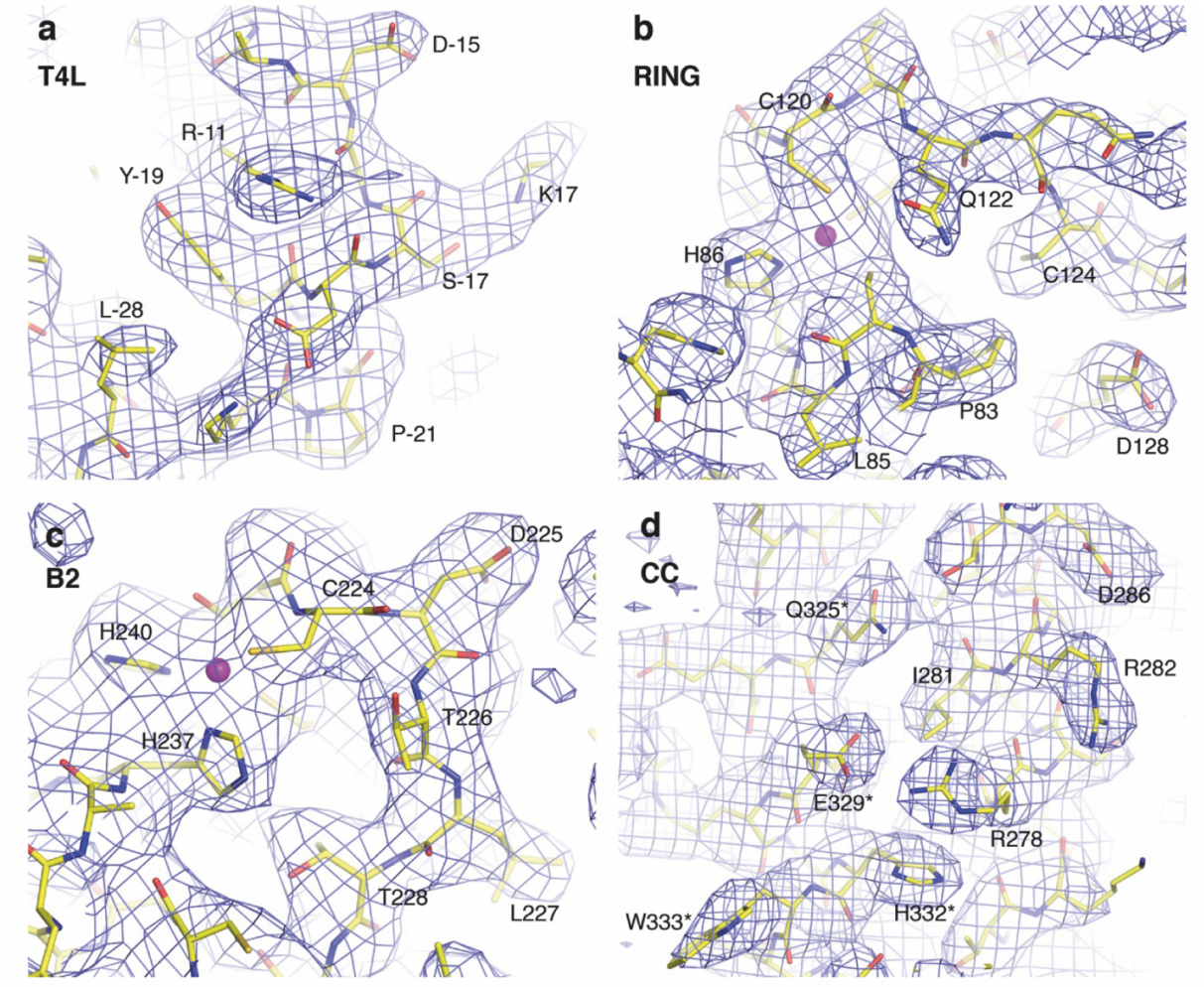

Fig. 20. Representative samples of electron density from the KAP1 RBCC crystal structure. (**a**) Electron density from the T4 lysozyme (T4L) domain. (**b**) Electron density from the RING domain. (**c**) Electron density from the B-box 2 domain. (**d**) Electron density from the coiled-coil (CC) domain. An isomesh contour level of 1.0 σ in PyMol was used for all panels.

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.

Table 1. Crystallographic data collection and refinement statistics for T4L-RBCC.

Data collection	**T4 lysozyme-KAP1 RBCC**		
X-ray source	DLS I03		
Space group	*C222₁*		
Cell dimensions			
a, b, c (Å)	59.77, 169.3, 374.5		
α = β = γ (°)	90		
Wavelength (Å)	1.28189		
Resolution (Å)	187–2.63	84.1–3.9[a]	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			
% favoured	93.2		
% allowed	6.45		
% outliers	0.35		
PDB code	PDB: 6QAJ		

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

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

[c]$R_{pim} = \Sigma_{hkl} (N_{hkl} - 1)^{-1/2} \times \Sigma_i \lvert I_{hkl,i} - \langle I \rangle_{hkl} \rvert / \Sigma_{hkl}\Sigma_i \lvert I_{hkl,i} \rvert$, 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 coiled-coil domain forms a helical hairpin consisting of a 15 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. 19**), 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 [175, 244] (RMSD 2.6 Å), which forms a dimeric antiparallel coiled coil with the same fold and similar length and curvature (**Fig. 21a**). TRIM5α forms a dimeric coiled coil with the same fold and length but lower curvature [177] (RMSD 3.9 Å; **Fig. 21b**), and TRIM69 forms a dimeric coiled coil with different secondary structure [176] (RMSD 4.0 Å, **Fig. 21c**).

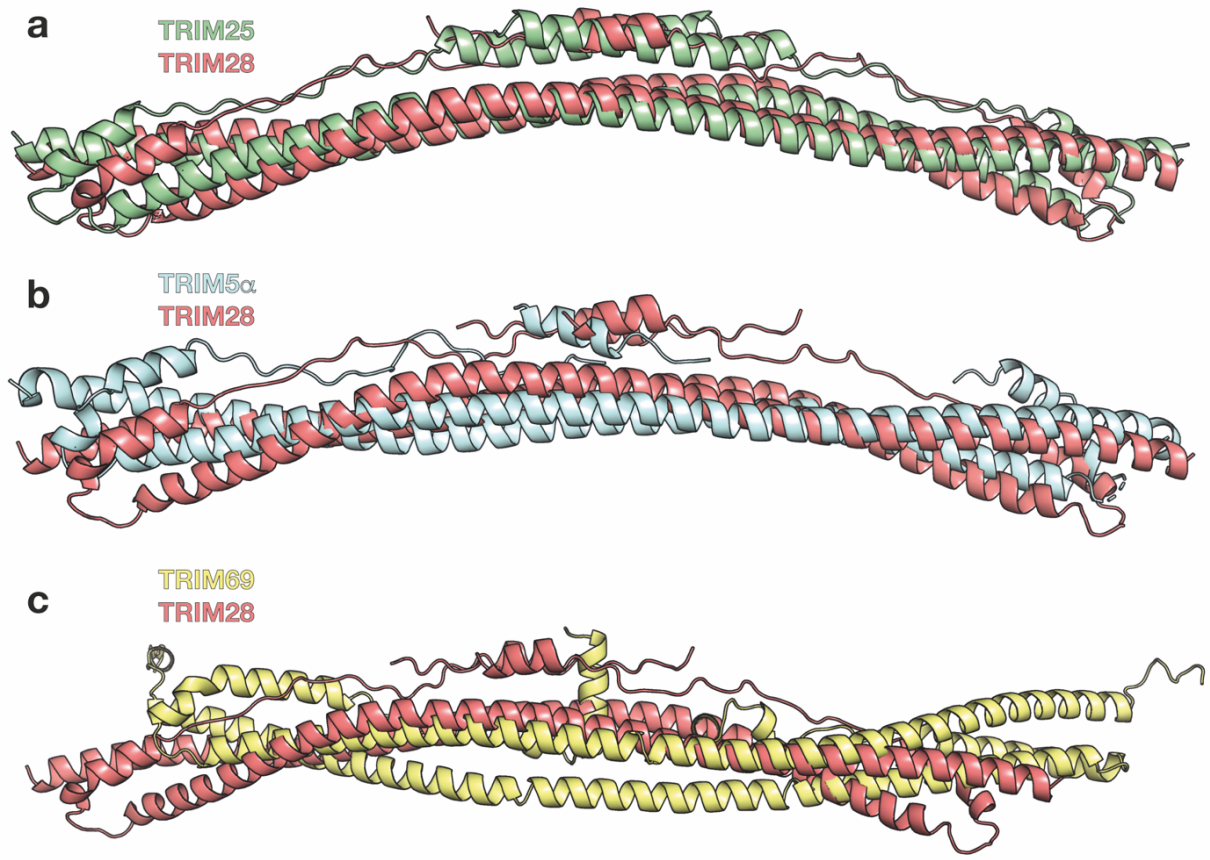

Fig. 21. Comparison of the coiled-coil domains of different TRIM proteins. (**a**) Alignment of the coiled-coil domains of KAP1 (red) and TRIM25 (PDB: 4LTB; green) [175]. (**b**) Alignment of the coiled-coil domains of KAP1 (red) and TRIM5α (PDB: 4TN3; blue) [177]. (**c**) Alignment of the coiled-coil domains of KAP1 (red) and TRIM69 (PDB: 4NQJ; yellow) [176].

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

My biophysical data indicate that KAP1 dimers have the ability to self-assemble into higher-order oligomers through one or both of the B-boxes. Higher-order oligomerization has been described for other members of the TRIM family and can be essential for the cellular functions of the protein as is well established for TRIM5 and PML/TRIM19. To be able to assess the relevance of higher-order oligomerization in KAP1-mediated transcriptional repression, I attempted to identify the interaction interfaces responsible for KAP1 self-assembly and designed structure-based mutations aimed at disrupting potential dimer contacts. B-box 2 forms crystal contacts with B-box 2 domains from two different neighbouring RBCC dimers in the crystal lattice and as noted above, a contribution of B-box 2 to KAP1 self-assembly was possible based on my SEC-MALS data. To assess the physiological significance of these interfaces, I mutated a cluster of residues in B-box 2 involved in homotypic crystal contacts, yielding the variant N235A/A236D/K238A/D239A/F244A/L245A (**Fig. 22a**). The RING domain of KAP1 is strictly monomeric in solution and does not form homotypic crystal contacts in the RBCC structure. However, it does form crystal contacts with the coiled-coil domain of a neighbouring RBCC dimer. Since RING-coiled-coil interactions would be compatible with my SEC-MALS data, RING domain residues in the contact region were mutated in a second variant, V114A/Q123A/F125A/K127A (**Fig. 22a**). My SEC-MALS data strongly suggested a contribution of B-box1 to KAP1 self-assembly. Since this domain was disordered in the RBCC structure, a structural model of KAP1 B-box 1 was generated from the PML/TRIM19 B-box1 dimer structure (PDB: 2MVW) [236] using Phyre2 [245]. Based on this model, I mutated residues predicted to be involved in KAP1 B-box 1 dimerization, yielding the variant A160D/T163A/E175R (**Fig. 22b**). The thermal stability of these three KAP1 variants as assessed by differential scanning fluorimetry (DSF) was comparable to that of wild-type KAP1 indicating that the introduced mutations did not disrupt the overall fold of the protein (**Fig. 22c**). 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. 23a and b**). 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. 23c and d**).

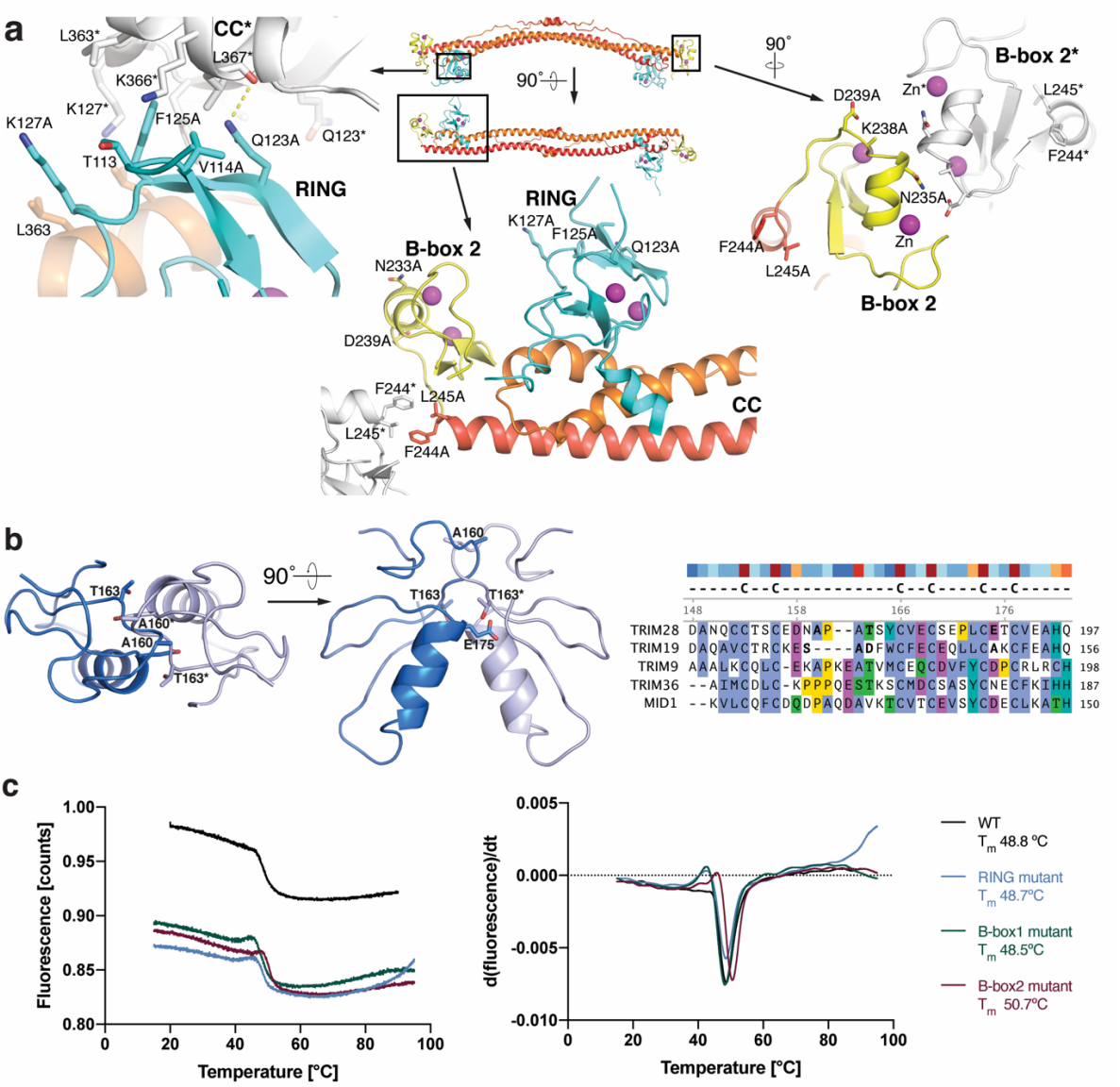

Fig. 22. Design of KAP1 mutants to disrupt higher-order oligomerization. (**a**) Positions of the mutations in the RING and B-box 2 domains. A reference RBCC dimer is coloured as in **Fig. 10**. Adjacent RBCC dimers forming crystal packing contacts are shown in grey with their residue numbers followed by an asterisk. (**b**) Model of a KAP1/TRIM28 B-box 1 dimer based on the TRIM19 B-box 1 dimer structure with selected residues forming dimer contacts shown. An alignment of B-box 1 sequences (right), the TRIM28 B-box 1 model, and the TRIM19 B-box 1 structure were used to identify mutations in KAP1 B-box 1 likely to disrupt dimer contacts. Residues known or predicted to participate in dimer contacts are shown in bold typeface in the sequence alignment. (**C**) Differential scanning fluorimetry (DSF) of the KAP1 RBCC (WT and variants). Intrinsic protein fluorescence at 330 nm and 350 nm was monitored between 15 and 95°C. Melting temperatures (T_m) were calculated from the turning point of the first derivative of the F_{350}:F_{330} ratio as a function of temperature.

At the highest concentration tested (179 μM), the apparent molecular weight derived from SEC-MALS for the RBCC B-box 1 mutant (97 kDa) was slightly above the expected

mass of a homodimer (87 kDa), indicating that the mutant may retain residual oligomerization potential. However, the level of self-assembly observed in the B-box 1 mutant is negligible compared to WT RBCC domain (**Fig. 23e**). Together, these data indicate that the assembly of KAP1 RBCC dimers observed at high protein concentration occurs primarily through dimerization of B-box 1 domain.

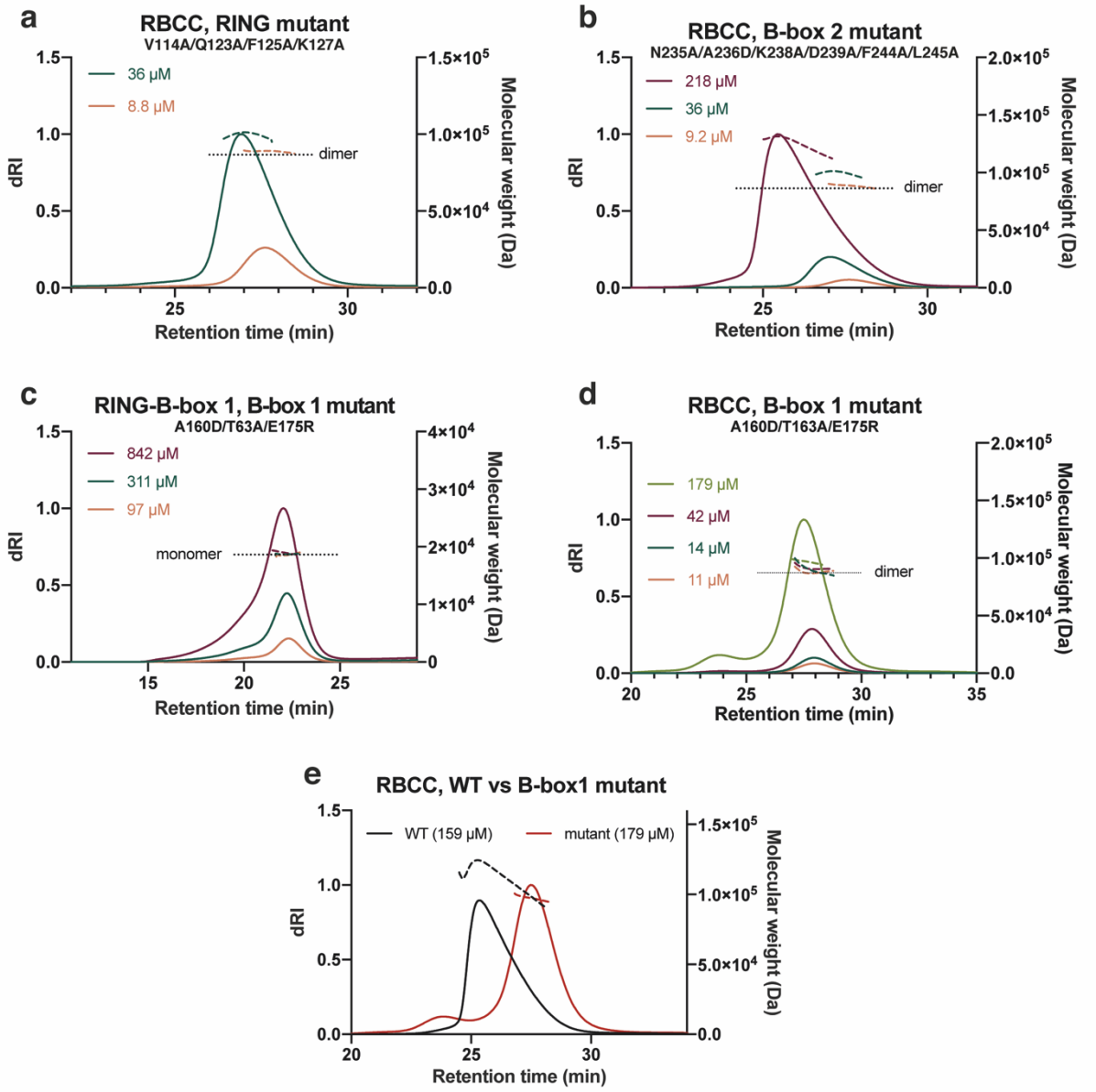

Fig. 23. Self-assembly properties of KAP1 mutants. (**a-d**) SEC-MALS data for (**a**) RBCC with RING domain mutations, (**b**) RBCC with B-box 2 mutations, (**c**) RING-B-box 1 with B-box 1 mutations, (**d**) RBCC with B-box 1 mutations. (**e**) Comparison of wild-type RBCC (black curve) and B-Box 1 mutant (red curve).

2.2.4 Crystal structure of KAP1 B-box1

My mutagenesis studies indicated that higher-order oligomerization of KAP1 is primarily mediated by homodimerization of its B-box 1 domain. However, this domain was disordered in the KAP1 RBCC structure and consequently no precise structural information was available on the interaction interface. I therefore attempted to determine the crystal structure of the isolated B-box 1 domain of KAP1. KAP1 B-box 1 (residues 147 - 200) could be produced in *E. coli* with high yields (**Fig. 24a**) and readily crystallized, forming small but well-diffracting crystals (**Fig. 24b**).

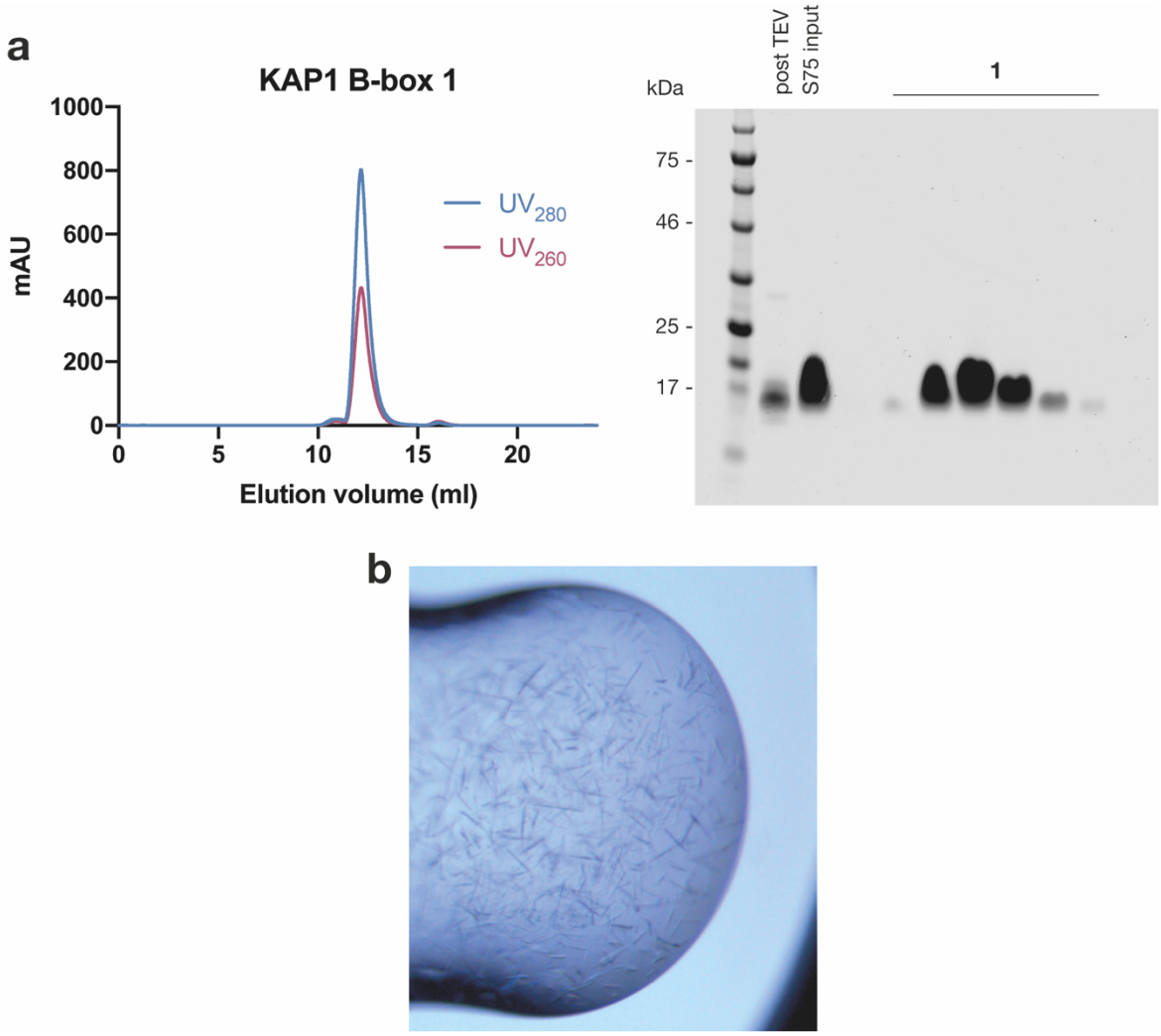

Fig. 24. Purification and crystallization of KAP1 B-box 1. (**a**) Purification of KAP1 B-box 1. Left Panel: Elution profile of KAP1 B-box 1 on a Superdex 75 10/300 column following nickel-affinity purification and removal of the His_6 tag. Right Panel: SDS-PAGE gel of the major peak in the elution profile. Also shown is the sample following overnight incubation with TEV protease (post TEV). (**b**) KAP1 B-box1 crystals grown in 1.6 M sodium citrate, pH 6.5.

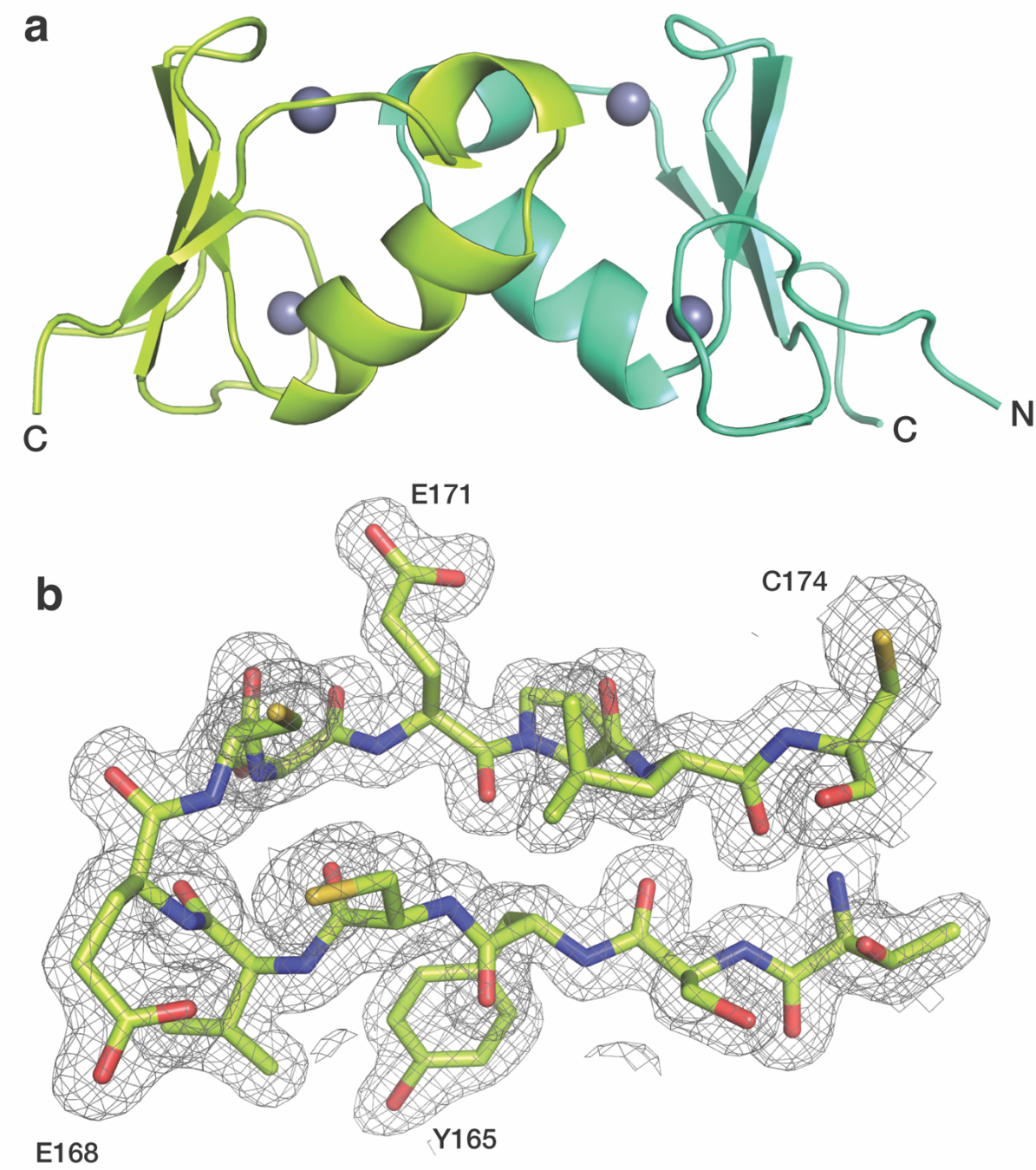

Fig. 25. Crystal structure of KAP1 B-box 1. (**a**) Overall structure of B-box 1 homodimer. (**b**) Representative sample of electron density from the KAP1 B-box1 crystal structure. An isomesh contour level of 1.0 σ in PyMol was used for the figure.

Since molecular replacement using other TRIM B-box 1 structures as search models was unsuccessful, the structure was determined by SAD phasing using the anomalous scattering signal from the zinc atoms. The asymmetric unit contained two molecules. The atomic model (**Fig. 25a**) was refined at 1.67 Å (**Table 2**). A sample of the electron density map is shown in **Fig. 25b**. As expected, the KAP1 B-box1 monomer is structurally similar to the B-box 1 domain of PML (**Fig. 27a**) and accordingly, the overall

fold of the protein was well predicted by the Phyre2 homology model (**Fig. 27b**). Surprisingly, however, the dimer interface in the KAP1 B-box1 crystal structure is markedly different from that observed in the PML B-box 1 structure. Instead of A160, T163 and E175 as predicted by the Phyre2 model, KAP1 B-box 1 dimer contacts involve residues D158, V183 and R184.

Table 2. Crystallographic data collection and refinement statistics for KAP1 B-box1.

Data collection	**KAP1 B-box1**
X-ray source	DLS I03
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	32.778, 51.466, 73.227
α = β = γ (°)	90
Wavelength (Å)	1.28180
Resolution (Å)	42.11–1.67 (1.73–1.67)
Observations	147967 (4515)
Unique reflections	12750 (793)
R_{merge}	0.1452 (0.9101)
R_{pim}	0.04338 (0.4036)
$\langle I \rangle / \sigma I$	13.12 (1.25)
Completeness (%)	85.15 (54.25)
Multiplicity	11.6 (5.7)
CC(1/2)	0.998 (0.745)
SAD Phasing	
CC_{ano}	0.627
$\lvert D_{ano} \rvert / \sigma D_{ano}$	1.276
FOM	0.419
Refinement	
Resolution (Å)	42.11–1.67
R_{work} / R_{free}	0.170/0.193
No. of non-H atoms	
Protein	717
Zn^{2+} Ions	4
Solvent	135
Mean B-factor (Å^2)	25.28
MolProbity Clashscore	1.46
RMS[f] deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.36
Ramachandran plot	
% favoured	100.00

To assess the relevance of this interface for KAP1 self-assembly, key residues at the interaction site were mutated yielding the variant R183A/V184D (**Fig. 26a**). A KAP1 RBCC construct bearing these mutations remained strictly dimeric at concentrations as high as 0.3 mM (**Fig. 26b**), confirming that these mutations completely abolished the

self-assembly potential of KAP1. These data show that self-assembly of KAP1 dimers is exclusively mediated by B-box 1 and rule out a contribution of B-box 2. The previously tested KAP1 B-box 1 mutant A160D/T163A/E175R was also deficient in higher-order oligomerization. However, these mutations are not located at the dimer interface (**Fig. 27c**) and consequently, their effect on KAP1 self-assembly is likely indirect. While the KAP1 B-box 1 crystal structure cannot fully explain the mechanism by which the A160D/T163A/E175R mutations interfere with self-assembly, it is possible that they induce structural changes in B-box 1 that are incompatible with dimerization.

Together, the data presented in this chapter demonstrate that KAP1 higher-order oligomerization is mediated by homodimerization of its B-box 1 domain using an interaction surface that includes residues D158, V183 and R184.

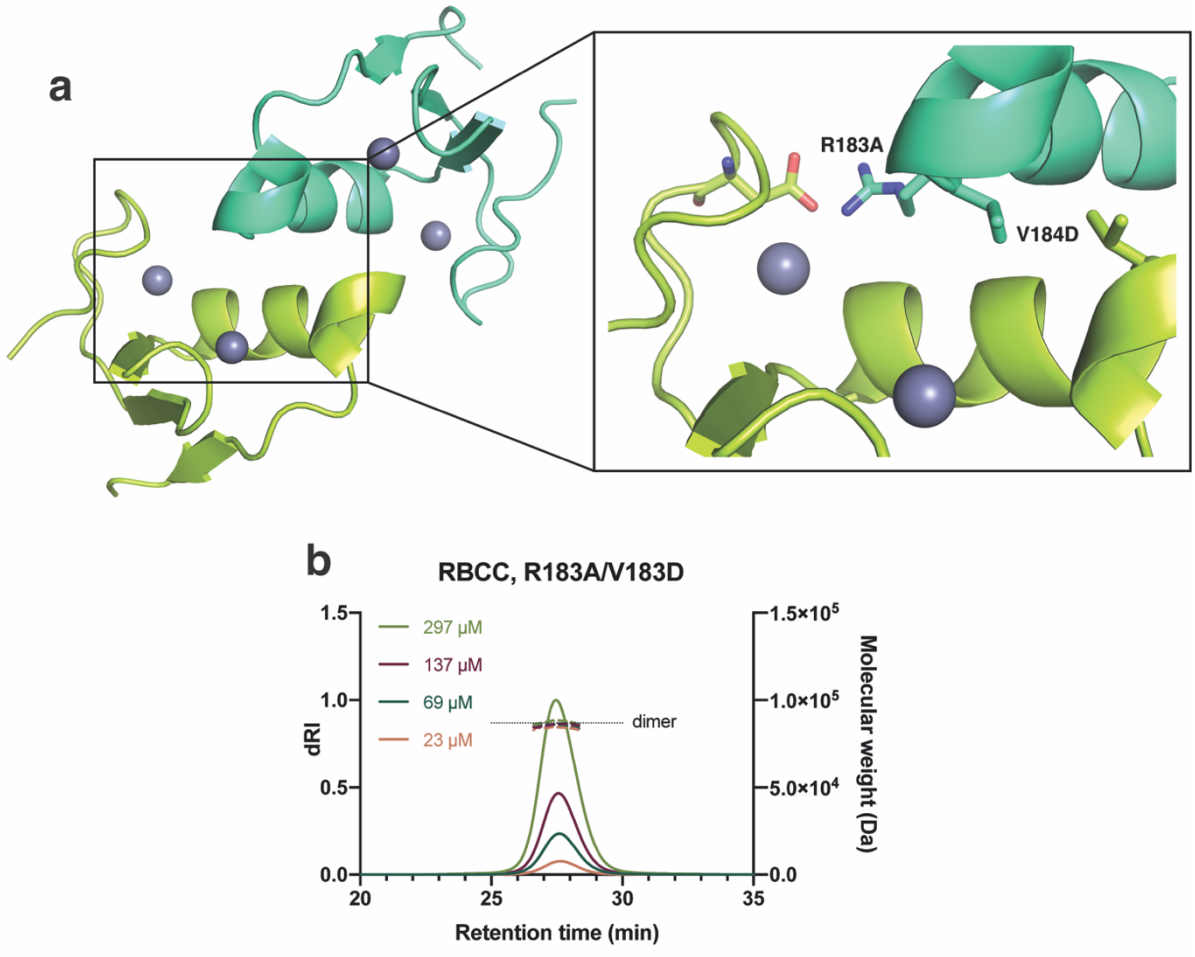

Fig. 26. Mutation of B-box 1 dimer interface abolishes KAP1 higher-order oligomerization. (**a**) Position of mutations aimed at disrupting dimer contacts. (**c**) SEC-MALS analysis of KAP1 RBCC bearing R183A/V184D mutation.

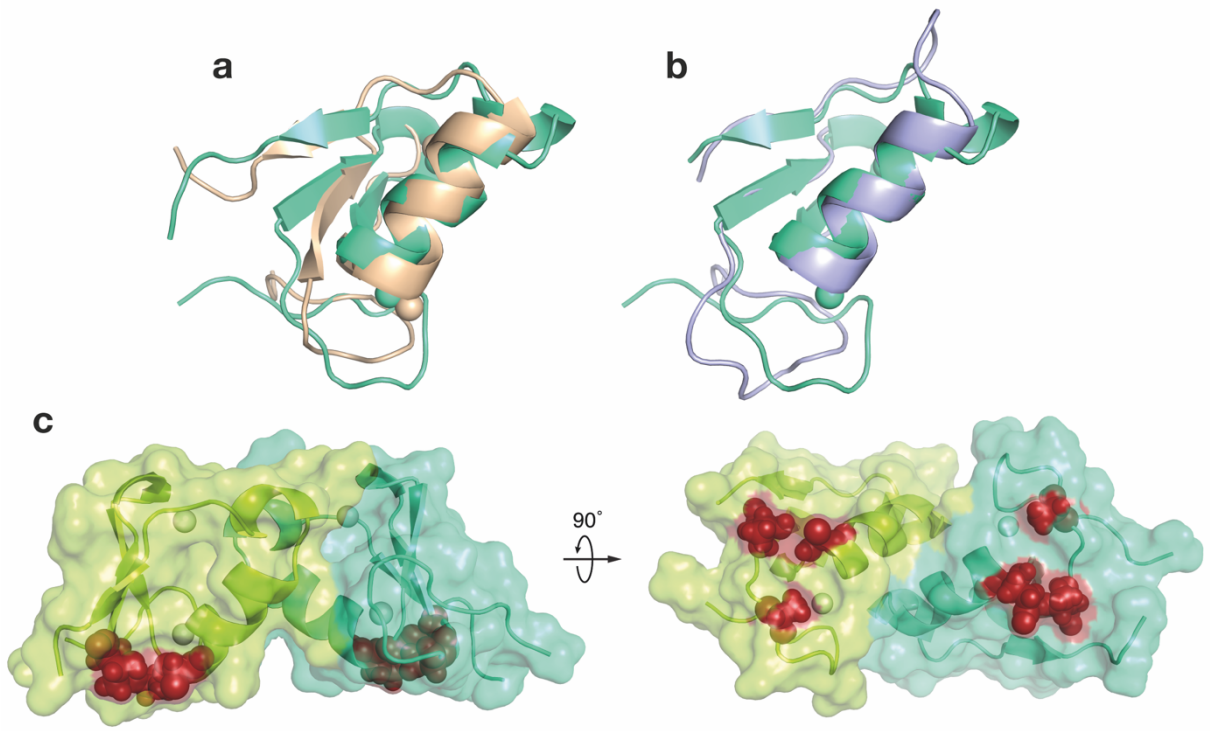

Fig. 27. Comparison of KAP1 B-box 1 crystal structure and Phyre2 homology model. (**a**) Alignment of the structures of KAP1 B-box 1 (green) and TRIM19 B-box 1 (PDB: 2MVW; yellow). (**b**) Alignment of the KAP1 B-box 1 crystal structure (green) and the Phyre2 homology model of KAP1 B-box 1 (yellow). (**c**) Location of the A160D/T163A/E175R mutations in the KAP1 B-box 1 crystal structure. The dimer interface predicted by the Phyre2 homology model is shown in red.

2.2.5 Mutation of B-box 1 affects KAP1-dependent transcriptional silencing

Various TRIM proteins assemble into higher-order oligomers, two-dimensional lattices or molecular scaffolds that are important for their physiological function. To determine whether self-assembly of KAP1 into higher-order oligomers is required for its retroelement repression, I compared the transcriptional silencing activities of wild-type KAP1 and the oligomerization-deficient B-box 1 mutants A160D/T163A/E175R and R183A/V184D. I used previously described reporter constructs in which sequences from an SVA-D (SINE–Variable number tandem repeat–Alu, type D) retroelement (recognized by ZNF91) or a LINE-1 retroelement (recognized by ZNF93) cloned upstream of a minimal SV40 promoter strongly enhance firefly luciferase activity unless the respective KRAB-ZFP and KAP1 are both present to repress the reporter [120] (**Fig. 28a**). This reporter assay was originally established in embryonic stem cells but had recently been adapted for use in KAP1-knockout (KO) HEK 293T cells [207], which were cotransfected with the reporter plasmid and plasmids encoding ZNF91 or ZNF93, KAP1

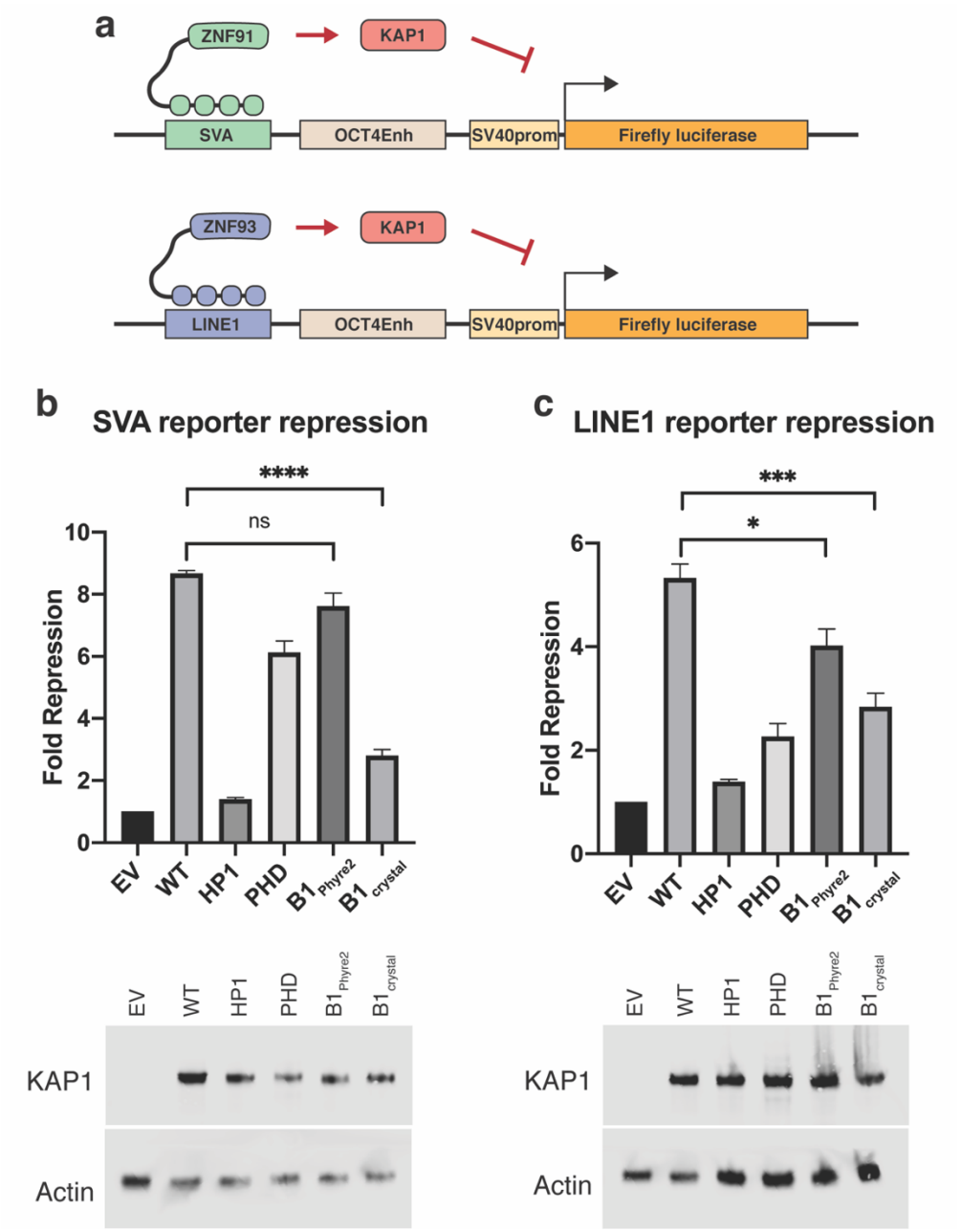

Fig. 28. Transcriptional silencing assays with oligomerization-deficient KAP1 mutants. (**a**) Schematics of the reporter systems used to measure KAP1 activity. (**b**) SVA reporter repression with the KAP1 B-box1 variants A160D/T163A/E175R ($B1_{Phyre2}$) and R183A/V184D ($B1_{crystal}$), an HP1-box mutant (HP1) and a PHD mutant (PHD). (**c**) LINE-1 reporter repression with the same set of mutants as in (*b*). Data were normalized to KAP1 KO cells transfected with an empty vector (EV). Data are presented as fold-repression of reporter luciferase luminescence. Error bars represent standard error of the mean between measurements (n = 4). Statistical significance was assigned as follows: not significant (ns), $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ***; $P < 0.0001$. Lower panels: Western blots of cell lysates from KAP1 KO HEK293T cells transfected with each of the variants (WT, $B1_{Phyre2}$, $B1_{crystal}$, HP1, PHD) or empty vector (EV).

(WT or mutant) and *Renilla* luciferase under a constitutive promoter. Firefly luciferase luminescence from the reporter was normalized against the cotransfected *Renilla* luciferase to control for transfection efficiency. Two well-characterized KAP1 mutants were used as references to validate the assay and gauge the effects of the oligomerization-deficient mutants: Mutations in the HP1-binding motif of KAP1 (R487E/V488E) abolished its transcriptional activity, while disruption of the PHD domain (C651A) resulted in moderate derepression of the reporter (**Fig. 28b and c**), consistent with previous reports [112, 246]. The structure-based B-box 1 mutant R183A/V184D showed significant defects in silencing in both the SVA ($p = 1.7 \times 10^{-7}$) and the LINE1 ($p = 6 \times 10^{-4}$) reporter assay (**Fig. 28b and c**). In contrast, the A160D/T163A/E175R variant, which was designed based on a Phyre2 homology model of B-box1 had a similar repression activity as WT KAP1 on the SVA reporter (**Fig. 28b**). In the LINE-1 reporter assay, the A160D/T163A/E175R variant showed slightly reduced silencing activity compared to WT KAP1 ($p = 0.021$), but the defect was significantly less pronounced than in the R183A/V184D mutant (**Fig. 28c**). These differences in the transcriptional activity of the two B-box1 mutants were unexpected as both variants are deficient in dimer-dimer assembly (**Fig. 23e and 26b**).

Previously, no function for the B-box 1 of KAP1 had been described. The data presented here show that this domain plays an important role in KAP1-dependent transcriptional regulation and suggest that its contribution to retrotransposon silencing may be linked to its ability to induce assembly of KAP1 dimers into higher-order oligomers. However, further experiments will be required to understand why two distinct mutations in B-box 1 that both inhibit higher-order oligomerization *in vitro*, behave differently *in vivo*.

2.3 Discussion

Initial biophysical studies on KAP1 reported that it forms trimers, both in isolation and in complex with a KRAB domain [158, 159]. 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 [175–178]. The SEC-MALS and SE-AUC data presented here 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 (**Fig. 29**).

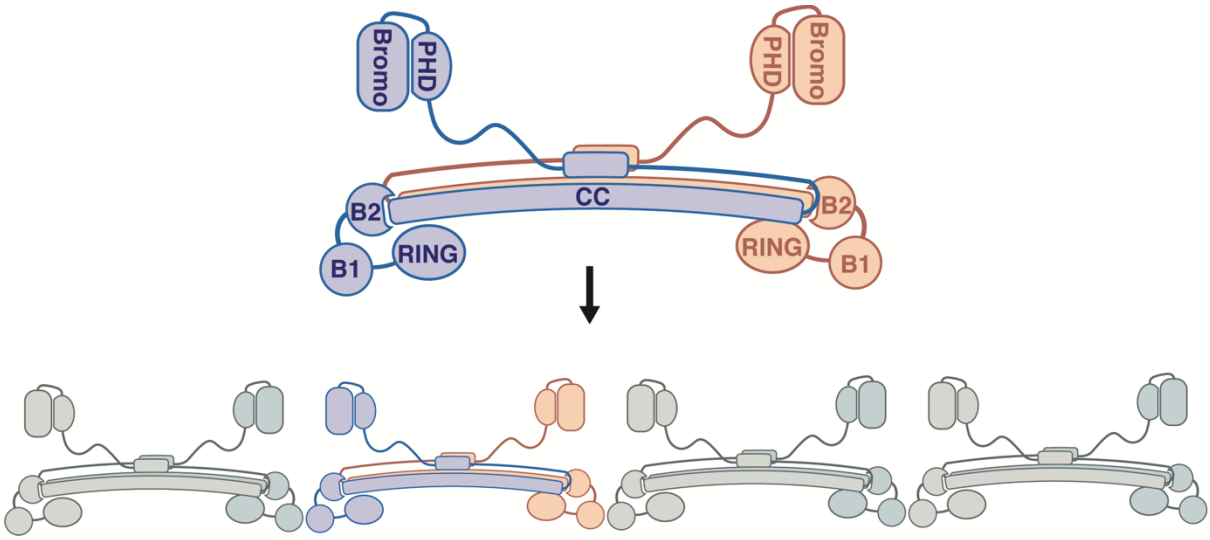

Fig. 29. Model for B-box 1-mediated higher-order oligomerization of KAP1.

This concentration-dependent self-association of KAP1 is encoded by B-box 1 and is relatively weak with a dissociation constant in the low micromolar range. This suggested that the RBCC domain of KAP1 may be able to self-associate into polymeric chains or molecular scaffold as observed for other TRIMs. I therefore investigated the possibility that KAP1 self-assembly may contribute to its function as a transcriptional repressor. 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. This model is supported by the observation that the oligomerization-deficient R183A/V184D variant shows pronounced defects in its transcriptional silencing activity. However, a second set of mutations (A160D/T163A/E175R) also interfered with higher-order oligomerization *in vitro*, but only marginally affected the activity of KAP1 in luciferase reporter assays. As noted earlier, the apparent molecular weight determined for the A160D/T163A/E175R variant at high protein concentrations was slightly above the expected mass of a homodimer indicating that this mutant may be able to self-assemble with very low affinity. In contrast, the R183A/V184 mutant was strictly dimeric at all concentrations tested. Potentially, this residual oligomerization potential of the A160D/T163A/E175R mutant is sufficient for KAP1 to retain near wild-type activity in the luciferase reporter assay. Alternatively, the observation that only one of two distinct oligomerization-

deficient KAP1 mutants shows clear defects in silencing may indicate that self-assembly is not be the primary function of the B-box 1 domain. Conceivably, B-box 1 might instead function by recruiting an as yet unidentified partner protein of KAP1 using an interaction interface that is disrupted by the R183/V184D mutations but largely unaffected by the A160D/T163A/E175R substitutions.

Several other studies were recently published supporting the results described here. Analysis of KAP1 by SEC-MALS and small-angle X-ray scattering (SAXS) confirmed that it forms antiparallel dimers rather than trimers [247]. Moreover, studies on mouse KAP1 revealed that the murine KAP1 homolog also has the ability to self-assemble in a B-box1-dependent manner, showing that this property of KAP1 is conserved across both species [248]. Indeed, comparison of the amino acid sequences of KAP1 B-box 1 domains from different vertebrate species reveals a strikingly high degree of conservation (**Fig. 30**), including critical residues in the B-box 1 dimer interface such as R183, V184 and D158 (which forms a salt bridge with R183). This suggests that these KAP1 homologs likely share the ability to self-assemble into higher-order oligomers via homodimerization of their B-box 1 domains.

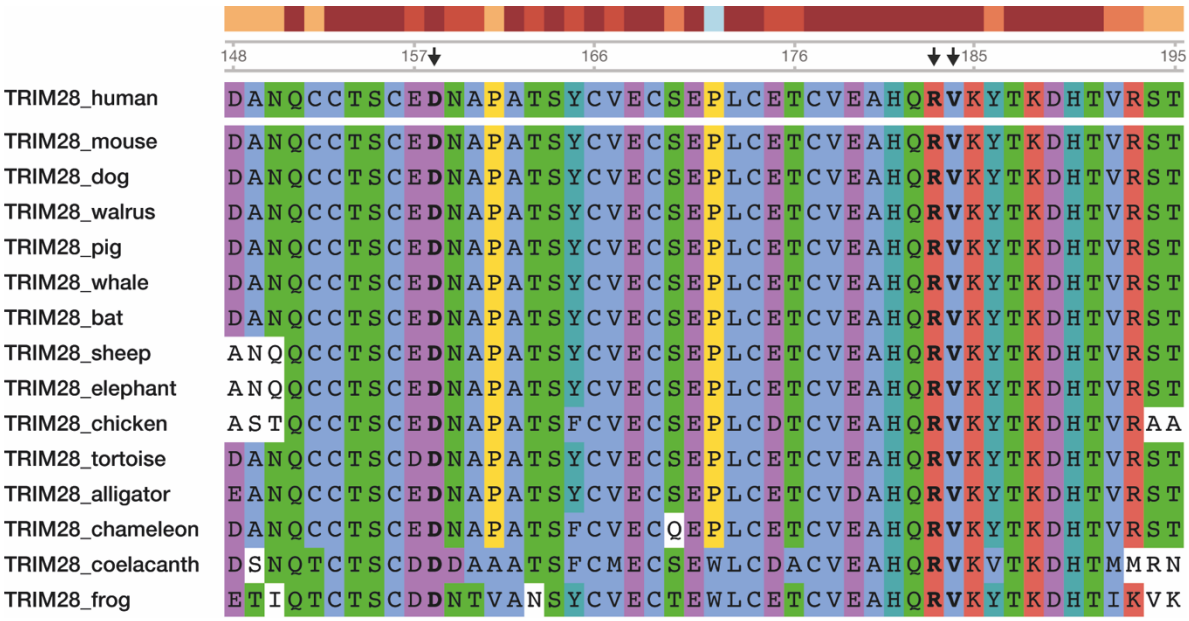

Fig. 30. Multiple sequence alignment of KAP1 B-box1 domains. KAP1 B-box 1 domains from various vertebrate species were aligned using MUSCLE [249]. Conservation of key residues in the B-box 1 homodimer interface are highlighted with arrows.

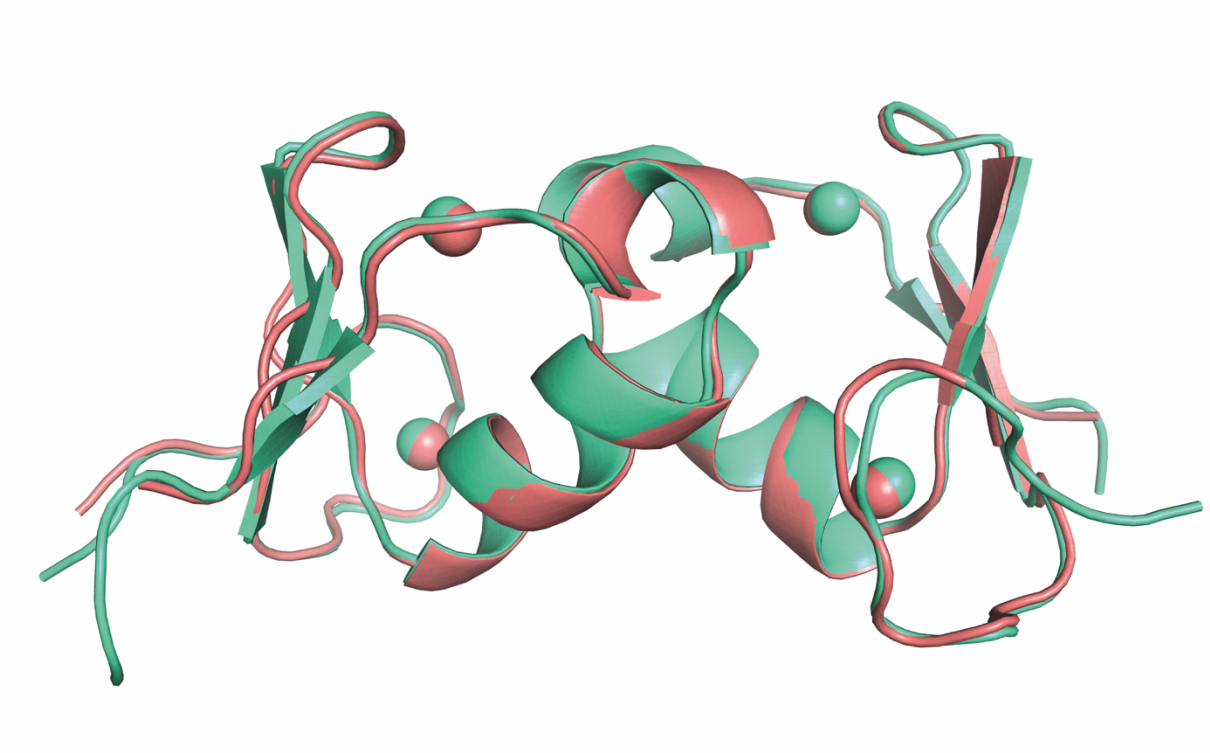

Fig. 31. Alignment of two KAP1 B-box 1 crystal structures. The KAP1 B-box 1 structure described here (green) and the B-box 1 structure reported by Sun et al. [248] (PDB 6O5K; red) were aligned using PyMol.

Sun et al. further reported the crystal structure of B-box 1 of murine KAP1 and identified structure-based mutations that abolish self-assembly of KAP1 dimers [248]. The B-box 1 domains of human and murine B1 are completely conserved in their amino acid sequence (**Fig. 30**) and accordingly, both structures are virtually identical (**Fig. 31**). To investigate the functional significance of KAP1 higher-order oligomerization in transcriptional silencing, Sun et al. introduced the corresponding mutations into human KAP1. The repression activity of these variants was subsequently assessed in human embryonic stem cells (hESCs) using a reporter construct in which a fragment of HERV-K known to be targeted by KAP1 is cloned upstream of green fluorescent protein (GFP). However, KAP1 bearing R183D or V184D substitutions was able to repress the HERV-K GFP reporter with similar efficiency as wild-type KAP1 [248]. This is surprising since in my hands, a very similar KAP1 variant, an R183A/V184D double mutant, showed clear defects in silencing in both an SVA and an LINE1 luciferase reporter assay (**Fig. 28**). Several explanations are possible for this discrepancy in our findings: Notably, the reporter used by Sun et al. was based on an LTR retrotransposon, while I used reporters based on non-LTR retrotransposons. This opens up the intriguing possibility that KAP1

higher-order oligomerization may only be required for silencing of certain classes of transposons, while being dispensable for repression of others. In addition, the assays were performed in different cell lines – my reporter assays were performed in HEK293T cells, whereas Sun et al. used hESCs. Conceivably, the requirement for KAP1 higher-order oligomerization for efficient transposon silencing may be cell-type dependent. To explore these possibilities, it will be necessary to clone sequences derived from SVA, LINE-1 and HERVK retrotransposons into the same reporter backbone (GFP or luciferase) and perform assays in both hESCs and HEK293T cells.

Alternatively, the apparent discrepancy in our results could simply be a consequence of the different experimental design. My experiments were performed in a KAP1 KO cell line, while Sun et al. used KAP1 knockdown cells, which still contained significant levels of endogenous KAP1 as assessed by western blot. The presence of WT KAP1 likely severely limits the dynamic range of the assay, making it challenging to assess the effect of KAP1 mutations that do not result in complete loss of function but only show a moderate decrease in activity. Another factor that may be contributing to the differences in our findings is the time point at which the reporter activity was measured. In my assays, luciferase activity was measured 48 h after transfection, while Sun et al. monitored GFP fluorescence over a period from 3 days to 14 days post-transduction. Conceivably, KAP1 self-assembly is particularly important at the early stages of silencing. As noted above, higher-order oligomerization of KAP1 may contribute to silencing by concentrating higher numbers of chromatin-modifying enzymes at the target site, thereby enabling rapid deposition of repressive histone marks. Oligomerization-deficient KAP1, in contrast, would still be able to recruit these factors and induce silencing, but potentially with slower kinetics. Indeed, in the time course presented by Sun et al. oligomerization-deficient KAP1 variants appear to have an intermediate phenotype, which is most pronounced at approximately 10 days post-transduction, before reaching near wild-type levels of GFP expression at 14 days post-transduction. However, assessment such effects is difficult since no previously characterized KAP1 mutants with moderate defects in silencing were included in the assay as references.

The RING domain of KAP1 has been reported to have ubiquitin E3 ligase activity [160, 163]. This activity has been implicated in cancer development, where aberrant

expression of MAGE family proteins expression induces KAP1-mediated ubiquitination of the tumour suppressors p53 and AMPK, resulting in their proteasomal degradation [160, 163]. E3 ligase activity is thought to be dependent on RING dimerization in most TRIM family member, including TRIM5α, TRIM23, TRIM25 and TRIM32 [234, 250–252]. In the case of TRIM25, it has been proposed that the N-terminus of the protein folds back onto the coiled-coil thereby allowing RING dimerization in the centre of the coiled-coil domain [234]. The crystal structure of KAP1 RBCC, however, is incompatible with RING interactions within a single RBCC dimer, as the RING domains are positioned at either end of the coiled-coil (**Fig. 19**). Furthermore, the KAP1 RINGs do not form homotypic crystal contacts and do not dimerize or self-associate in solution. The absence of homotypic contacts between the RINGs suggests that KAP1 may be amongst the minority of RING E3 ligases that can promote ubiquitin transfer from E2 to substrate without dimerization, perhaps using structural elements from outside the core RING domain as seen for example in CBL-B [253].

The observation that the RING domain of KAP1 is monomeric was recently confirmed by another study [254]. Moreover, neither full-length KAP1 nor its isolated RING domain displayed ubiquitin E3 ligase activity *in vitro* and, in contrast to previous reports, addition of MAGE-C2 did not enhance KAP1 activity. Together, this may indicate that the RING domain of KAP1 requires heterodimerization with the RING domain of a partner protein [254].

Alternatively, ubiquitination may not be the primary function of the RING domain of KAP1. The RING domain of PML, for instance, has been implicated in SUMOylation [255, 256]. While the SUMO E3 ligase activity of KAP1 is thought to be mediated by its PHD domain, there is increasing evidence supporting a role for KAP1 RING in SUMOylation [164, 165, 247, 248]. Notably, KAP1 appears to retain some auto-SUMOylation activity *in vitro* even after deletion of its PHD domain and deletion of the RING domain impaired KAP1 SUMOylation [247, 248]. Moreover, SUMOylation of the KAP1 substrate IRF7 was reported to be independent of the PHD domain and although no direct interaction between KAP1 RING and the SUMO E2 ligase Ubc9 has yet been demonstrated, KAP1 RING and Ubc9 were found to co-immunoprecipitate from cell lysates [164].

Together, the data presented here demonstrate that the B-box 1 domain KAP1 mediates self-assembly of KAP1 into higher-order oligomers and suggest a previously unknown role for B-box 1 in KAP1-dependent transcriptional silencing of retrotransposons. However, further studies are required to confirm whether the functional defects observed in the KAP1 B-box 1 mutant are indeed a direct consequence of its inability to form higher-order oligomers.

CHAPTER 3

INTERACTION OF KAP1 WITH KRAB DOMAINS

3.1 Introduction

KAP1 is not able to bind to DNA directly. Instead, it is recruited to its genomic targets by KRAB-ZFPs, which constitute the largest family of mammalian transcription factors. KRAB-ZFPs recognize retrotransposon sequences with a variable C-terminal array of zinc fingers, while the conserved N-terminal KRAB domain recruits KAP1 to induce epigenetic silencing of the targeted transposon and adjacent genes. KRAB domains are approximately 70 amino acids in size and largely disordered in isolation but are thought to fold upon binding to KAP1. KRAB:KAP1 complexes were previously reported as containing one KRAB molecule and three KAP1 molecules [158], but this stoichiometry is incompatible with the dimeric nature of KAP1. The composition of KRAB:KAP1 complexes therefore remains unclear. Moreover, while KRAB binding was mapped to the RBCC domain of KAP1 and has been reported to require intact RING, B-box 2 and coiled-coil domains, the exact interaction interface remains unknown [159]. Here, I re-examined the composition of KRAB:KAP1 complexes, demonstrating that each KAP1 dimer binds a single KRAB domain. Further, I show that the KRAB binding site is located in the central region of the coiled-coil domain near the dyad. Mutations at this site abolished KRAB binding and transcriptional silencing activity of KAP1.

3.2 Results

3.2.1 Stoichiometry of KRAB:KAP1 complexes is 1:2

Complexes of KAP1 and KRAB domains were successfully reconstituted by co-expressing the proteins in *E.coli* (**Fig. 32a**). SEC-MALS analysis of KAP1 RBCC bound to the KRAB domain of ZFP93, a KRAB-ZFP that binds to a LINE-1 element known to

be silenced by KAP1, showed that the complex retained the ability to self-assemble into higher-order oligomers (**Fig. 32b**). Notably, the RBCC:KRAB appeared to oligomerize more readily than KAP1 RBCC alone since signs of self-assembly were evident even at protein concentrations as low as 0.4 g L^{-1} (**Fig. 32b**). The isolated RBCC domain, in contrast, was dimeric at comparable concentrations (**Fig. 11a**).

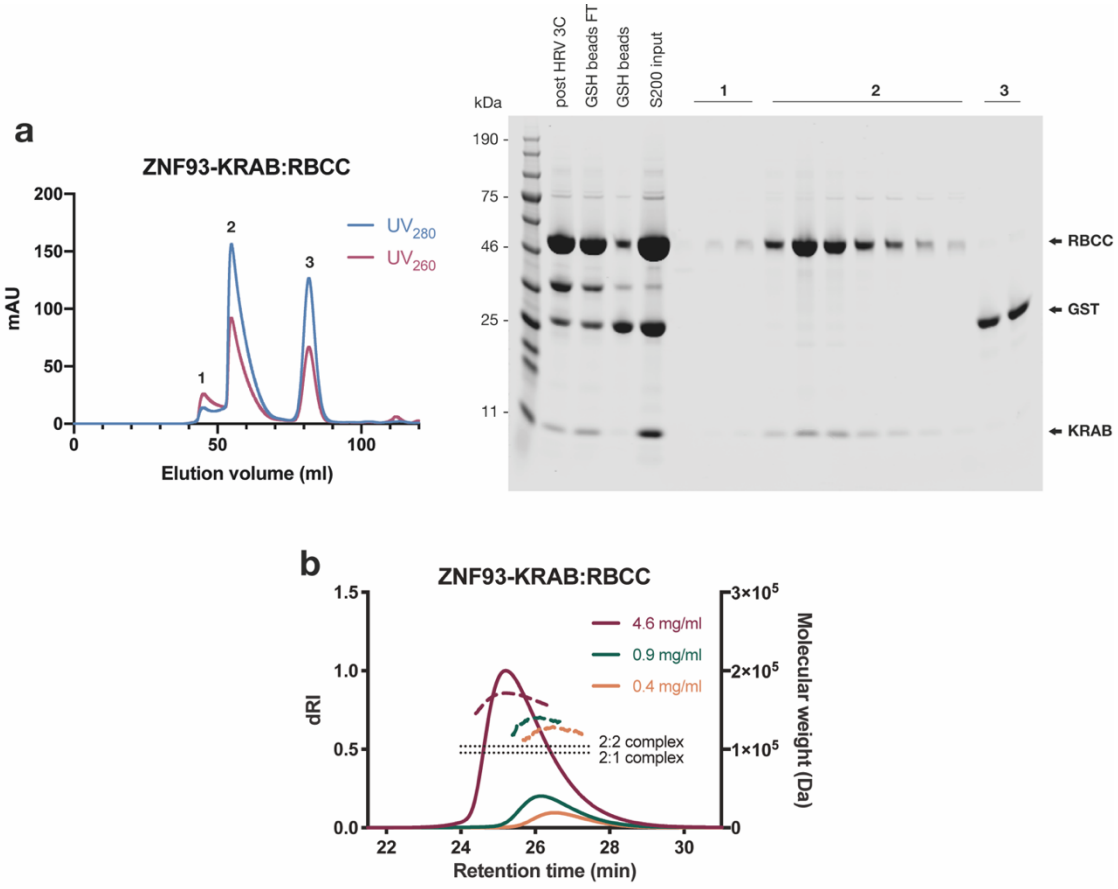

Fig. 32. Purification and SEC-MALS analysis of KRAB:RBCC complex. (**a**) Purification of ZNF93-KRAB:RBCC complex. Left panel: Elution profile of the KRAB:RBCC complex on a HiLoad 16/600 Superdex 200 pg column following glutathione-affinity purification and removal of the GST tag. Right Panel: SDS-PAGE gel of the major peaks in the elution profile. post HRV 3C: sample after overnight incubation with human rhinovirus (HRV) 3C protease. GSH beads FT: sample after removal of protease and uncleaved protein. GSH beads: protein that remained bound after the second glutathione-affinity purification step. (**b**) SEC-MALS analysis of the RBCC:KRAB at various concentrations. The expected molecular weights of 2:1 and 2:2 RBCC:KRAB complexes are indicated with dashed lines.

A possible explanation for this apparent difference in affinity may be that both experiments were performed in the presence of different salt concentrations. While the buffer used for SEC-MALS analysis of the RBCC:KRAB complex contained 0.2 M NaCl, the buffer used for the isolated RBCC domain contained 0.5 M NaCl. As noted above,

high salt concentrations inhibited self-assembly (at least in the context of full-length KAP1). Alternatively, binding of the KRAB domain may increase the propensity of KAP1 to form higher-order oligomers. Due to the small size of KRAB domains (~ 8 kDa) and the observed self-assembly it was not possible to determine the stoichiometry of the complex. To address these problems, the ZFP93 KRAB domain was expressed as an N-terminal MBP fusion protein to produce more distinct shifts in molecular weight (**Fig. 33a**). Additionally, the experiment was performed with full-length KAP1 at a NaCl concentration of 0.5 M to reduce KAP1 self-assembly. The average molecular mass derived from SEC-MALS for the KAP1:MBP-KRAB complex of 226 kDa was inconsistent with a 1:3 KRAB:KAP1 stoichiometry (327 kDa theoretical molecular weight) and suggested instead that the stoichiometry of the complex was 1:2 KRAB:KAP1 (236 kDa theoretical molecular weight; **Fig. 33b**).

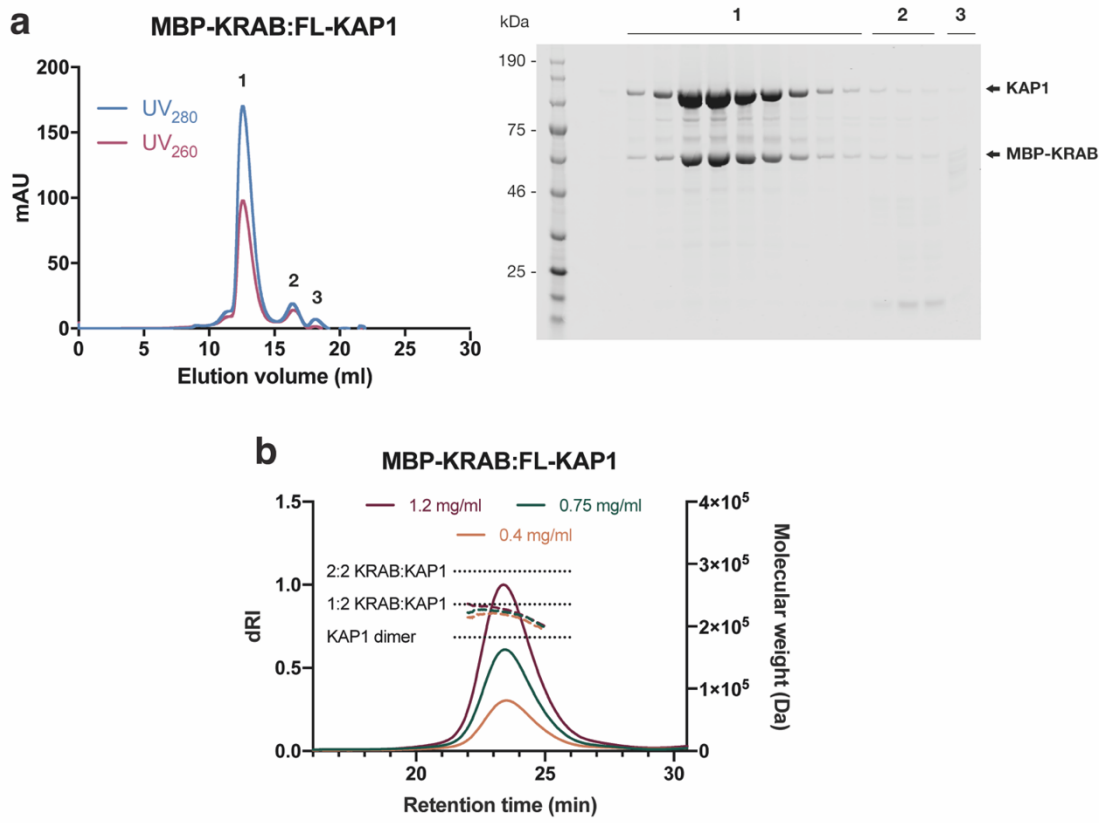

Fig. 33. Purification and SEC-MALS analysis of MBP-KRAB:KAP1 complex. (**a**) Purification of MBP-KRAB:KAP1 complex. Left panel: Elution profile of the MBP-KRAB:KAP1 complex on a Superose 6 Increase 10/300 column following Strep-Tactin affinity purification. Right panel: SDS-PAGE gel of the major peaks in the elution profile. (**b**) SEC-MALS analysis of full-length KAP1 bound to ZNF93 MBP-KRAB at various concentrations. The expected molecular weights of a KAP1 dimer, and for 2:1 and 2:2 KAP1:KRAB complexes are indicated with dashed lines.

3.2.2 Conserved residues in the coiled-coil domain of KAP1 mediate KRAB binding

The observation that each RBCC dimer binds a single KRAB domain, suggested 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).

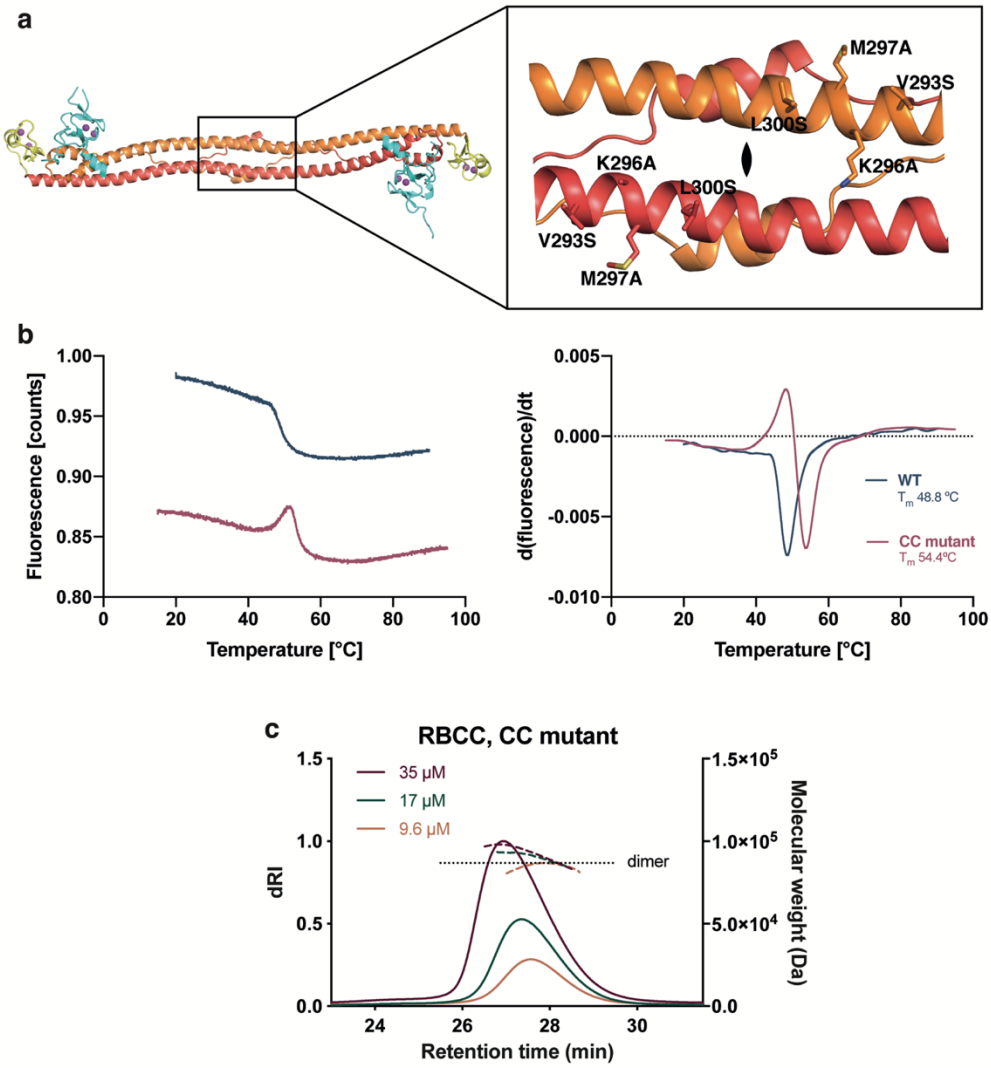

Fig. 34. Mutation of potential KRAB binding residues in the KAP1 coiled-coil domain. (**a**) Closeup of the cluster of solvent-exposed hydrophobic residues near the dyad. The variant V293S/K296A/M297A/L300S (CC mutant) was generated to test for KRAB binding. (**b**) DSF analysis of KAP1 RBCC (WT and CC mutant). Intrinsic protein fluorescence at 330 nm and 350 nm was monitored between 15 and 95°C. Melting temperatures (T_m) were calculated from the turning point of the first derivative of the F_{350}:F_{330} ratio as a function of temperature. (**c**) SEC-MALS analysis of KAP1 RBCC bearing V293S/K296A/M297A/L300S mutation at various concentrations.

Intriguingly, examination of my KAP1 RBCC crystal structure revealed a cluster of solvent-exposed hydrophobic residues near the twofold axis (V293, M297, L300; **Fig. 34a**). Moreover, these amino acids are conserved in KAP1 but not present in other TRIMs (**Fig. 35**). To determine whether this region of the coiled-coil domain mediates KRAB binding, I designed the variant V293S/K296A/M297A/L300S (**Fig. 34a**). KAP1 RBCC domain bearing these mutations failed to bind to MBP-KRAB in a pulldown assay (**Fig. 36a**). Other properties of KAP1 such as dimerization and higher-order oligomerization were unaffected (**Fig. 34c**), indicating that the mutations did not interfere with the overall fold of the RBCC domain. Notably, the thermal stability of this variant as assessed by DSF was significantly higher than that of WT RBCC domain, further supporting a functional role of these residues (**Fig. 34b**). 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 with surface plasmon resonance (SPR) measurements performed in collaboration with Dr Stephen McLaughlin (MRC Laboratory of Molecular Biology). To this end, MBP-KRAB was covalently coupled to the sensor chip before multiple concentrations of either WT RBCC or the coiled-coil mutant were passed over the chip. The binding kinetics of WT RBCC were biphasic and consequently fitted with two sets of curves, resulting in two dissociation constants (K_{d1} and K_{d2}). The major binding event is characterised by a fast association rate (k_{on1} = 3.6 ± 0.96 x 10^4 M^{-1} s^{-1}) and a slow dissociation rate (k_{off} = 2.7 ± 0.1 x 10^{-4} s^{-1}), resulting in a dissociation constant K_{d1} of 8 ± 2 nM. The second binding event is weak (K_{d2} = 118 μM) and is unlikely to represent a specific interaction (**Fig. 36b**).

In contrast to the high affinity interaction observed between WT RBCC and MBP-tagged KRAB domain, no binding was detectable with the coiled-coil mutant (**Fig. 36 c**).

Fig. 35. Multiple sequence alignment of TRIM coiled-coil domains. The coiled-coil domains of KAP1/TRIM28 from various vertebrate species (upper sequence cluster), and of other human TRIMs (lower sequence cluster) were aligned using Clustal X [257]. Conservation of residues involved in KRAB binding is highlighted with red columns and arrows.

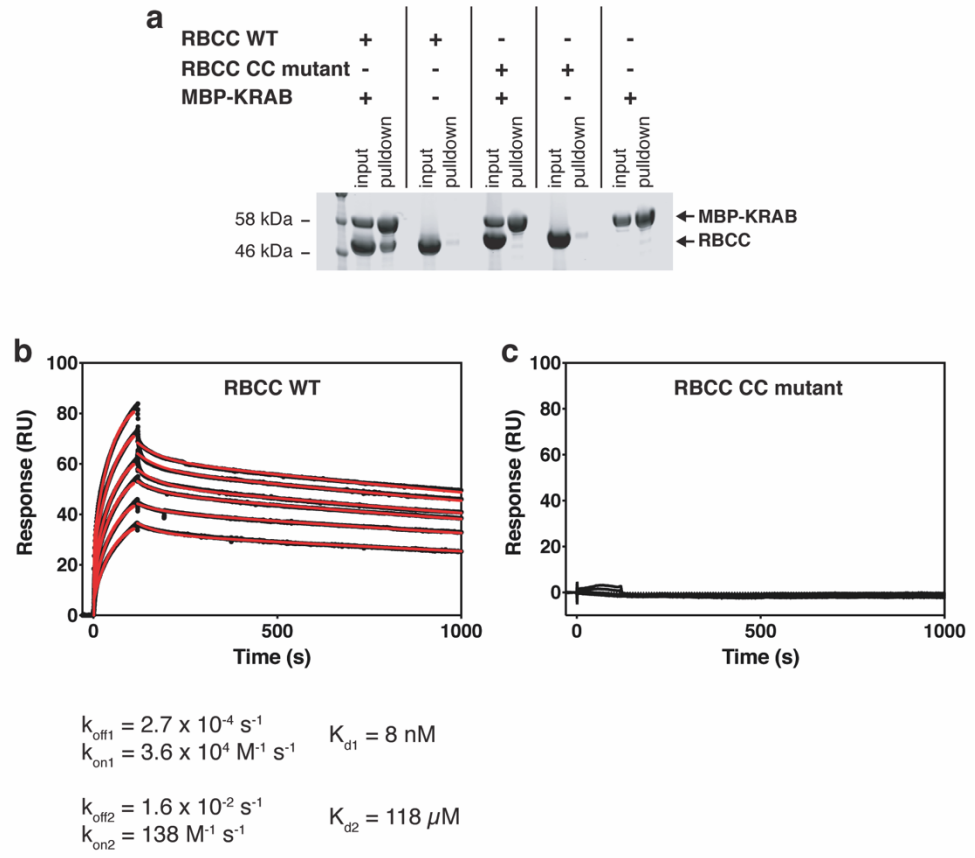

Fig. 36. Mutations in the coiled-coil domain of KAP1 abolish KRAB binding. (**a**) Pulldown KAP1-KRAB binding assay. KAP1 RBCC was incubated with Twin-StrepII-MBP-ZNF93 KRAB and the mixture loaded on Strep-Tactin Sepharose. Bound proteins were detected by SDS-PAGE/Coomassie. (**b-c**) Surface plasmon resonance (SPR) analysis of KAP1 RBCC. Analysis of the interactions of wild-type RBCC, (**a**), and of the coiled-coil mutant (V293S/K296A/M297A/L300S), (**b**), with MBP-KRAB immobilized on the sensor chip. Sensograms are shown for 1:2 dilution series starting from 34 µM for wild-type RBCC and 35 µM for mutant RBCC. The fits for the association and dissociation kinetics for wild-type KAP1 are shown in red with the corresponding rate constants and derived dissociation constants. The SPR experiment was performed by Stephen McLaughlin (MRC Laboratory of Molecular Biology).

3.2.3 Mutation of the KRAB binding site abolishes KAP1-dependent transcriptional silencing

To further validate the results of the *in vitro* interaction assays, the effect of the coiled-coil mutations on the transcriptional silencing activity of KAP1 was measured using the luciferase reporter assay described above (**Fig. 37**).

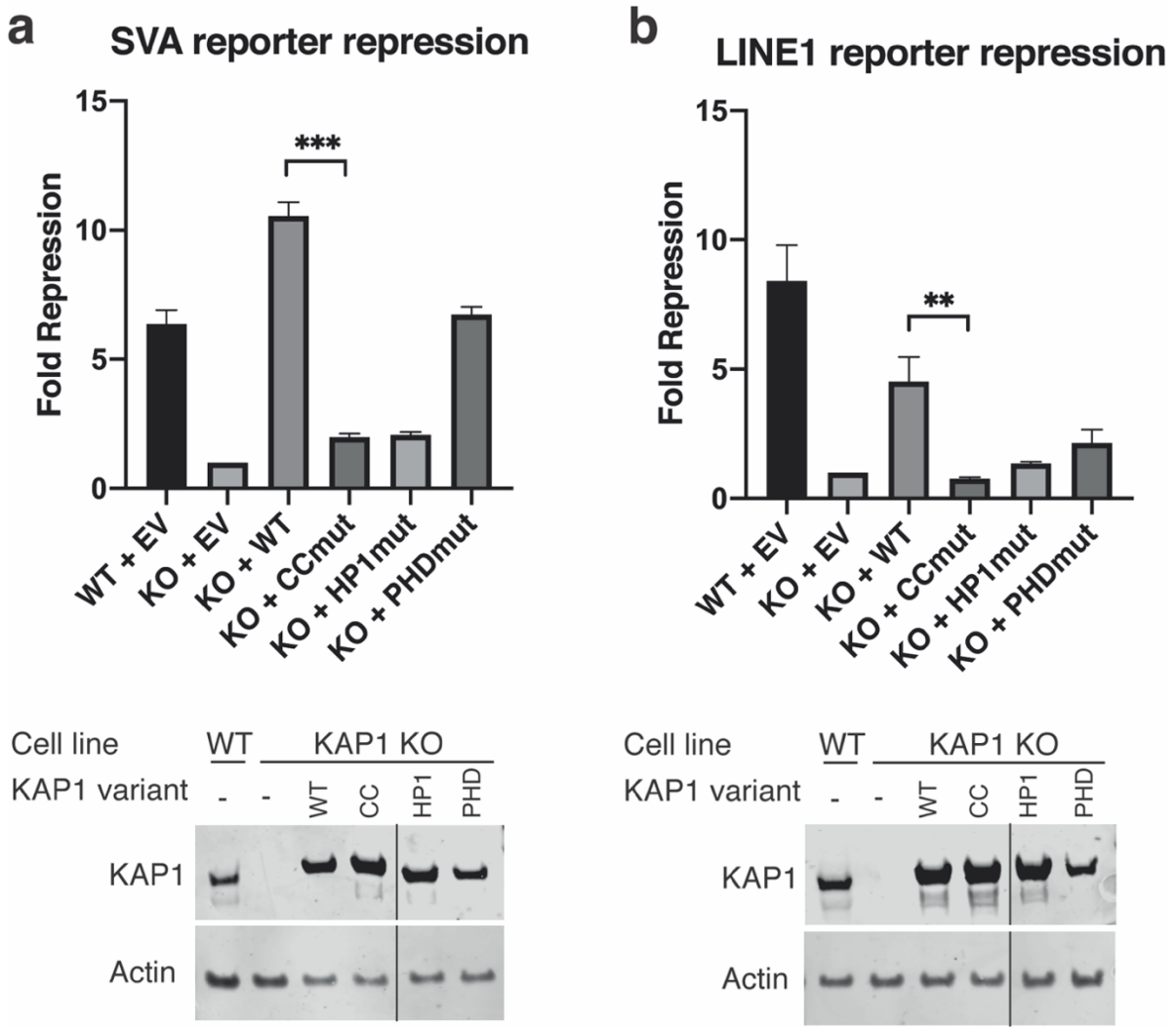

Fig. 37. Transcriptional silencing assays with KRAB binding site mutant. (**a**) SVA reporter repression with the KRAB binding-deficient coiled-coil mutant (CCmut), an HP1-box mutant (HP1mut) and a PHD mutant (PHDmut). (**b**) LINE-1 reporter repression with the same set of mutants as in (*a*). Data were normalized to KAP1 KO cells transfected with an empty vector (EV). Error bars represent standard error of the mean between measurements (n = 3). Statistical significance was assigned as follows: not significant (ns), $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Lower panels: Western blots of cell lysates from WT or KAP1 KO HEK293T cells transfected with each of the variants (WT, CC, HP1, PHD) or empty vector (-). The assays shown in this figure were performed by Zheng-Shan Chong (University of Cambridge).

As the interaction of KAP1 with the KRAB domains of KRAB-ZFPs is critical for recruitment of KAP1 to its genomic targets, KAP1 variants defective in KRAB-binding would be expected to lose all repression activity. Consistent with its inability to bind KRAB domains *in vitro,* introduction of the V293S/K296A/M297A/L300S mutations resulted in profound de-repression of the reporter to levels comparable to the empty vector control and the KAP1 HP1-box mutant, which was previously reported to result in complete loss of function [246] (**Fig. 37a and b**). The effects of the V293S/K296A/M297A/L300S mutations were consistent across reporter constructs containing sequences from two unrelated retrotransposons. Both of these elements are recognized by distinct KRAB-ZFPs (ZNF93 and ZNF91, respectively), indicating that different KRAB domains utilize the same binding site on KAP1. Together, these data provide further support that KRABs bind KAP1 on the twofold axis of the RBCC dimer on a surface that includes residues V293/K296/M297/L300 and show that this binding surface is required for KAP1 repression activity.

All previous experiments were perfomed using a KAP1 variant bearing four amino acid substitutions. To more finely map the KRAB-binding site of KAP1 and assess the contribution of each of these residues to KRAB binding, I performed transcriptional silencing assays with the individual point mutants V293S, K296A, M297A and L300S (**Fig. 38a and b**). The K296A, M297A and L300S variants each showed significant defects in silencing of both the SVA and LINE1 reporter constructs, while being expressed at levels comparable to WT KAP1. The V293S variant, in contrast, was not expressed at detectable levels in transfected KAP1 KO HEK293T cells. This was unexpected since a KAP1 variant combining all four of these point mutations was expressed at WT levels. Consequently, the role of V293 in KRAB binding and transcriptional silencing could not be assessed. These data indicate that K296, M297 and L300 are all required for KRAB binding, while the importance of V293 remains uncertain.

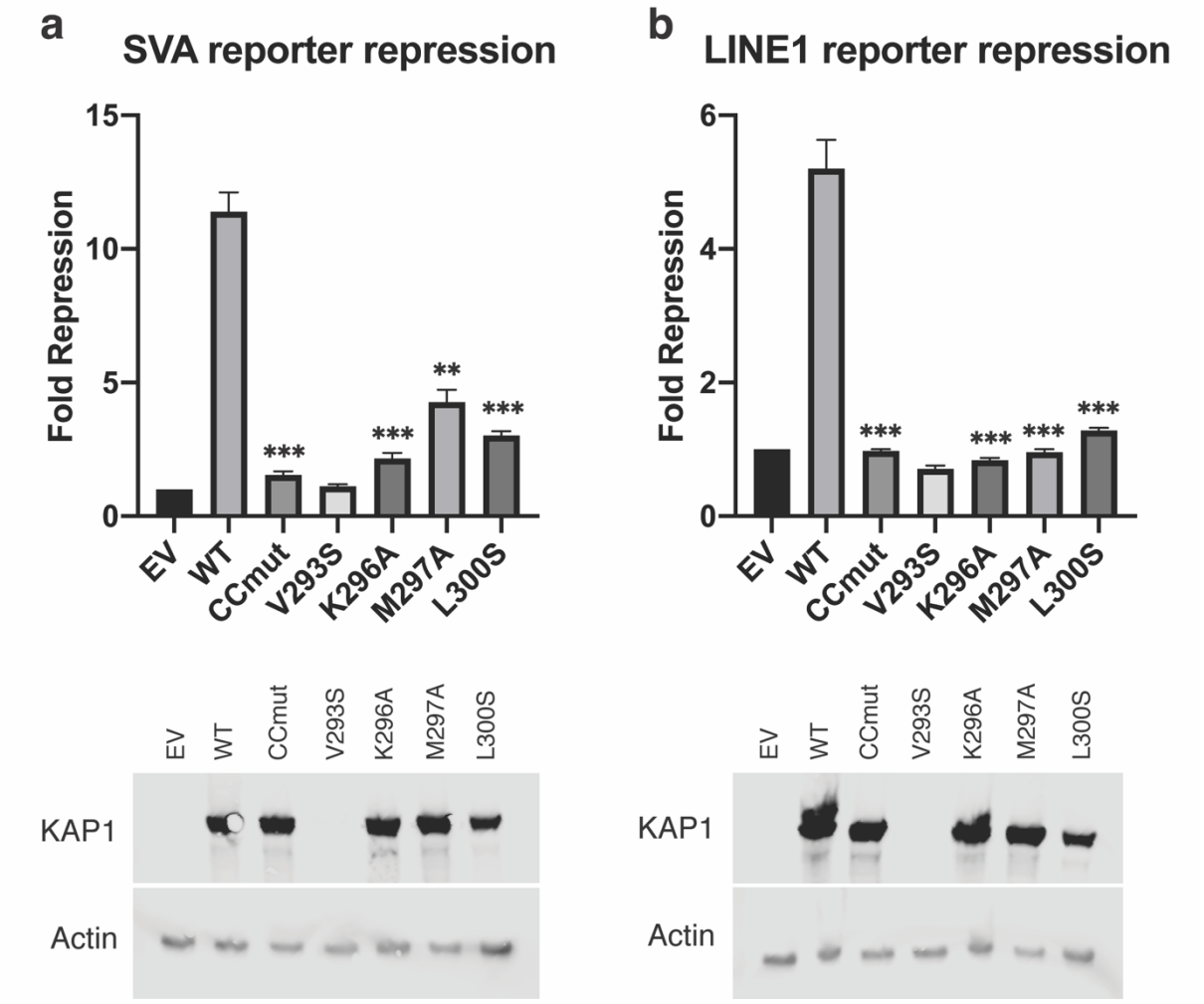

Fig. 38. Effects of single point mutations in the KRAB binding site. (**a**) SVA reporter repression with the KAP1 variant V293S/K296A/M297A/L300S (CCmut) and individual point mutants (**b**) LINE-1 reporter repression with the same set of mutants as in (*a*). Data were normalized to KAP1 KO cells transfected with an empty vector (EV). Data are presented as fold-repression of reporter luciferase luminescence. Error bars represent standard error of the mean between measurements (n = 3). Statistical significance was assigned as follows: not significant (ns), $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Lower panels: Western blots of cell lysates from KAP1 KO HEK293T cells transfected with each of the variants or empty vector.

3.2.4 Crystallization of a KRAB:KAP1 complex

To obtain further insights on the interaction of KAP1 with KRAB-ZFPs, I attempted to determine the structure of KAP1 in complex with a KRAB domain. As shown above, KRAB:KAP1 complexes were successfully produced by coexpressing both proteins in *E. coli*. Initial crystallization trials with the RBCC domain of KAP1 bound to various ZFP93 KRAB domain constructs yielded no or only poorly diffracting crystals. However, inspection of the T4L-RBCC crystal lattice revealed unoccupied space around the central region of the coiled-coil domain large enough to potentially accommodate a 70 amino acid KRAB domain. I therefore applied the same optimization strategy to the KRAB:RBCC complex that proved successful for the isolated RBCC domain. A T4L-RBCC fusion construct was coexpressed with the KRAB domain of ZFP93, followed by reductive methylation of primary amines (**Fig. 39a**).

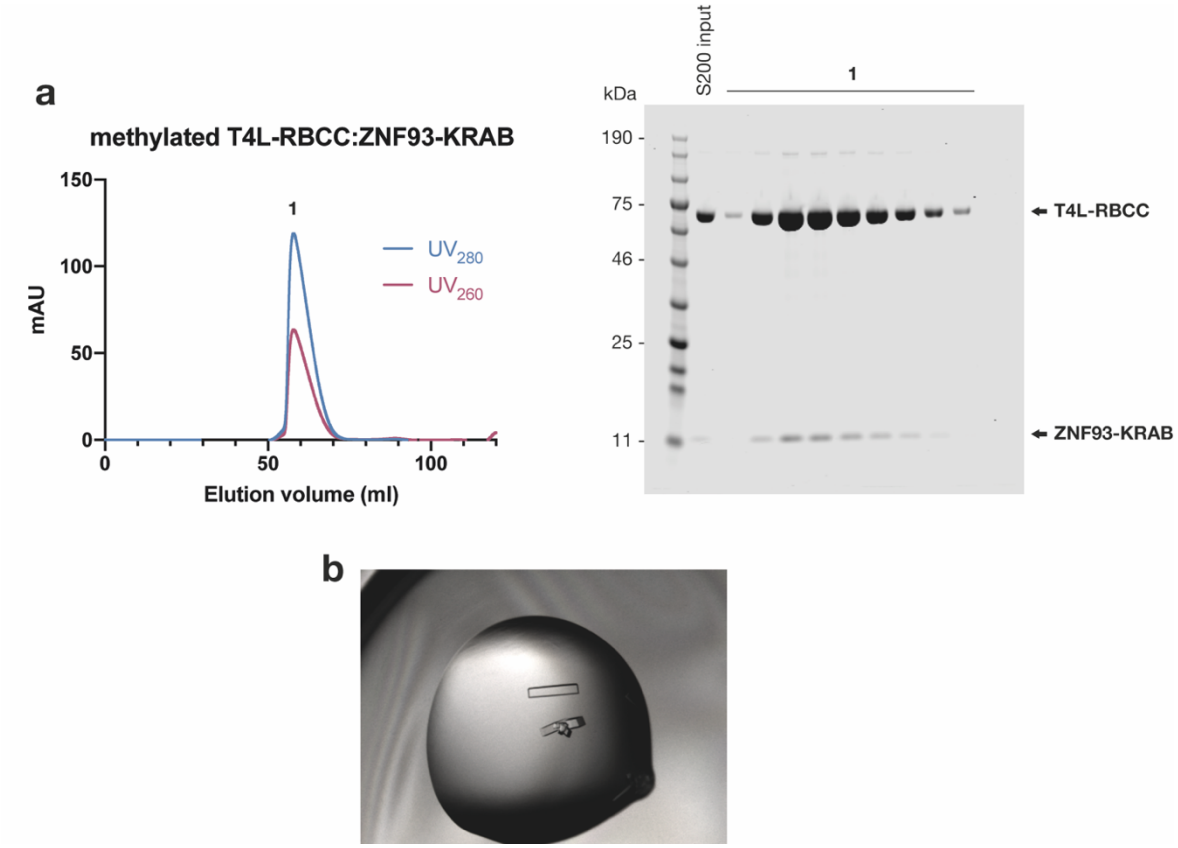

Fig. 39. Purification and crystallization of T4L-RBCC in complex with ZNF93-KRAB. (a) Right Panel: Elution profile of the T4L-RBCC:ZNF93-KRAB complex on a HiLoad 16/600 Superdex 200 pg column following reductive methylation of primary amines. Right panel: SDS-PAGE gel of the major peak in the elution profile. (b) Crystals of methylated T4L-RBCC:ZNF93-KRAB grown in 13% PEG 4,000, 0.25 M $MgCl_2$, 0.1 M Tris pH 8.5.

The resulting sample crystallized in very similar conditions as T4L-RBCC alone (**Fig. 39b**). Equally, the morphology of the crystals strongly resembled the previously obtained T4L-RBCC crystals. However, the T4L-RBCC:KRAB crystals diffracted X-rays poorly (**Table 3**), precluding structure determination.

Table 3. Crystallographic data collection statistics for T4L-RBCC:ZNF93-KRAB complex.

Data collection	**T4L-RBCC:ZNF93-KRAB complex**
X-ray source	DLS I03
Space group	*C222*$_1$
Cell dimensions	
a, b, c (Å)	59.68, 167.4, 376.38
α = β = γ (°)	90
Wavelength (Å)	0.968620
Resolution (Å)	55.97–4.39 (4.47–4.39)
Observations	76,267 (3,228)
Unique reflections	12508 (539)
R_{merge} [b]	0.1778 (1.6182)
R_{pim} [c]	0.082 (0.709)
$\langle I \rangle / \sigma I$	3.35 (1.02)
Completeness (%)	99.5 (88.9)
Multiplicity	6.1 (6.0)
CC(1/2)	0.944 (0.698)

As noted previously, no interpretable electron density for B-box 1 of KAP1 was visible in the crystal structure of the RBCC motif indicating a high degree of flexibility. Since the C-terminus of the RING domain and the N-terminus of B-box 2 were located in close proximity, I attempted to improve the diffraction quality of the T4L-RBCC:KRAB crystals by deleting B-box 1 (residues 141-202). Indeed, an identical strategy was soon after reported to have been successful in optimizing crystals of the RBCC domain of KAP1 in complex with the CUE domain of SMARCAD1 [258]. The resulting T4L-RBCCΔB1 construct was soluble and deletion of B-box 1 did not affect its ability to interact with KRAB domains (**Fig. 40a**). As before, free amines in the T4L-RBCCΔB1:ZNF93-KRAB complex were methylated prior to crystallization (**Fig. 40b**). Deletion of B-box 1 caused the complex to crystallize in a different space group ($P2_12_12_1$) and improved the diffraction quality of the crystals (**Table 4**). These crystals diffracted to a resolution of approximately 3.5 Å and electron density corresponding to the KRAB domain was clearly present in the central region of the KAP1 coiled-coil domain. However, the map quality in this region was poor and it was not possible to confidently build an atomic model of the KRAB domain (**Fig. 41**).

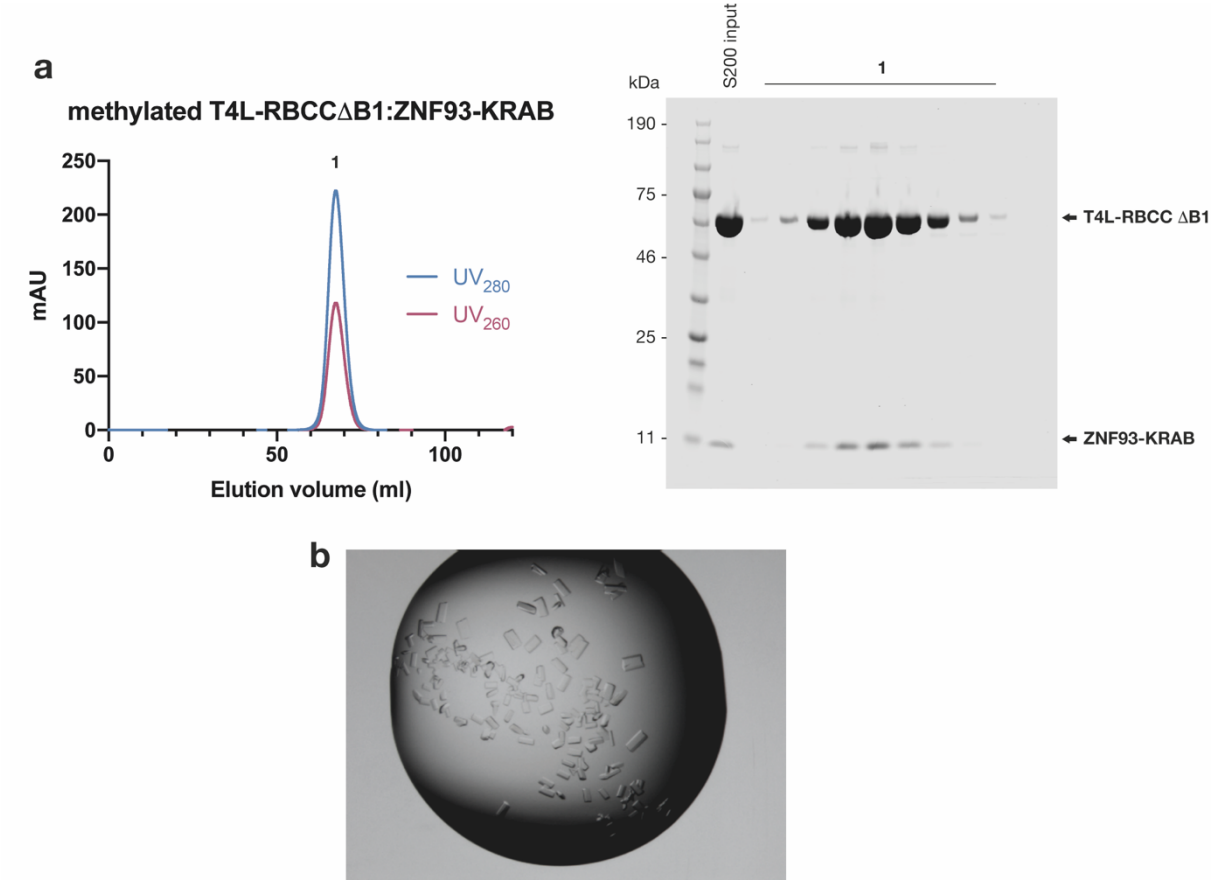

Fig. 40. Purification and crystallization of T4L-RBCCΔB1 in complex with ZNF93-KRAB. (**a**) Left panel: Elution profile of the T4L-RBCCΔB1:ZNF93-KRAB complex on a HiLoad 16/600 Superdex 200 pg column following reductive methylation of primary amines. Right panel: SDS-PAGE gel of the major peak in the elution profile. (**b**) Crystals of T4L-RBCCΔB1:ZNF93-KRAB grown in 18% PEG 8,000, 0.2 M $MgCl_2$, 0.1 M Tris pH 9.

Table 4. Crystallographic data collection statistics for T4L-RBCCΔB1:ZNF93-KRAB complex.

Data collection	**T4L-RBCCΔB1:ZNF93-KRAB**
X-ray source	DLS I04
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	58.00, 121.23, 378.78
α = β = γ (°)	90
Wavelength (Å)	0.956313
Resolution (Å)	63.13 - 3.52 (3.58 - 3.52)
Observations	446,342 (22,198)
Unique reflections	34,386 (1680)
R_{merge} [a]	0.193 (2.100)
R_{pim} [b]	0.056 (0.596)
$\langle I \rangle / \sigma I$	7.3 (1.3)
Completeness (%)	100.00 (99.41)
Multiplicity	13.0 (13.2)
CC(1/2)	1.000 (0.930)

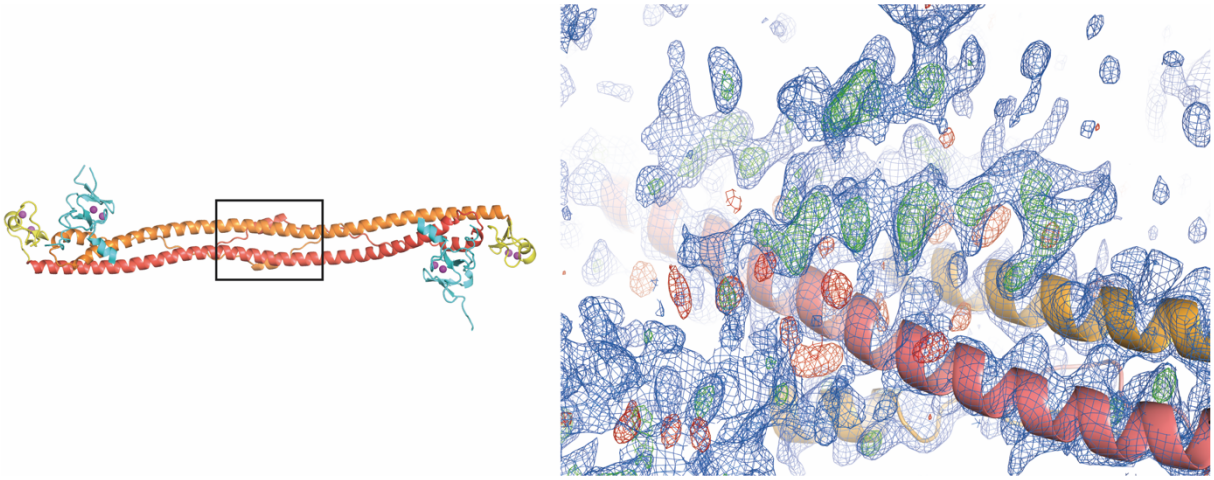

Fig. 41. Electron density for the KRAB domain of ZNF93. Left panel: Location of the electron density corresponding to the ZNF93 KRAB domain in the context of the KAP1 RBCC dimer. Right Panel: Electron density for the ZNF93 KRAB domain. Initial phases were obtained by molecular replacement using the T4L-RBCC structure as search model, followed by several rounds of refinement with REFMAC and PHENIX (R_{work}/R_{free}: 0.35/0.37). The 2Fo-Fc map (blue) is contoured at 1.0 σ. The Fo-Fc map (green/red) is contoured at 3.0 σ.

3.3 Discussion

The interaction between KAP1 and KRAB-ZFPs is critical for recruitment of KAP1 to retroelements, but the stoichiometry of KAP1-KRAB complexes has remained unclear and the location of the KRAB binding surface on KAP1 unknown. My SEC-MALS data shows unambiguously that the KAP1 RBCC dimer can only bind a single KRAB domain. Moreover, the data presented here demonstrate 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 [159]. 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, direct involvement of B-box 2 in KRAB binding can be ruled out. Similarly, a contribution of the RING, which is also distant from the dyad, appears highly unlikely. A refined model for recruitment of KAP1 to its genomic targets is presented in **Fig. 42**.

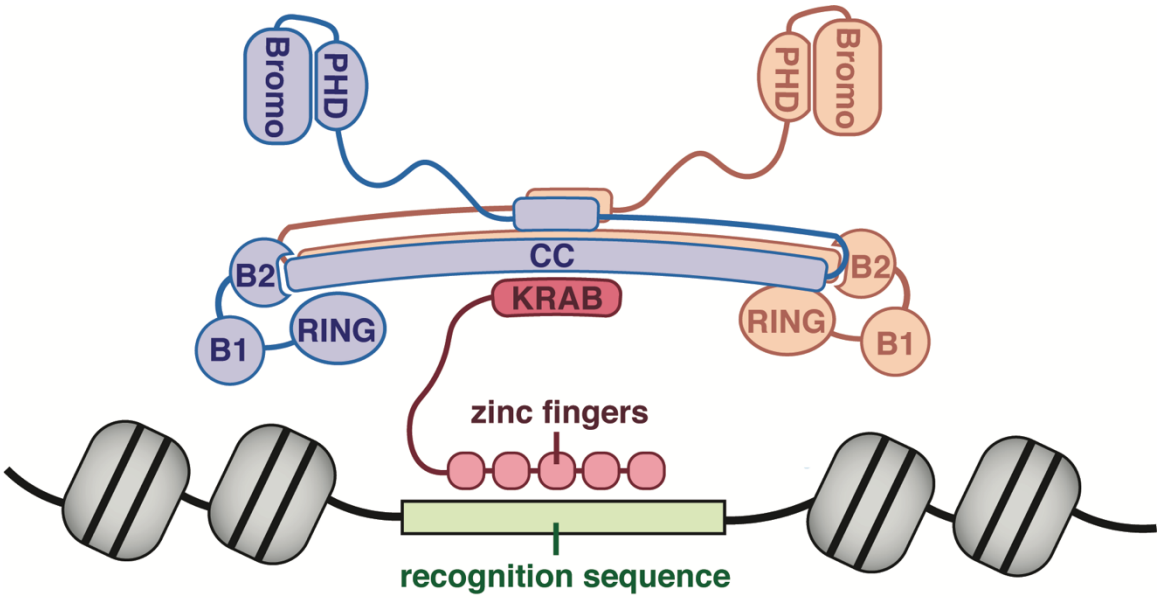

Fig. 42. Model for binding of KAP1 dimers to KRAB-ZFPs.

The results presented here are further supported by a recently published study examining the interaction of murine KAP1 with KRAB-ZFPs. Consistent with my data, each KAP1 dimer was found to bind only a single KRAB domain with the KRAB binding site being located in the central region of the coiled-coil domain [248]. The interaction site was further mapped to a region including V294, D295, K297A, M298A. These residues correspond to V293, D294, K296, M297 of human KAP1 and overlap with the interaction interface I identified.

In addition, I successfully crystallized a complex of KAP1 RBCC and the KRAB domain of ZNF93. Initial T4L-RBCC:ZNF93-KRAB crystals diffracted to a resolution of 4.4 Å. The resolution could be improved to 3.5 Å by deleting the flexible B-box 1 domain of KAP1. Electron density for the KRAB domain was clearly visible in the central region of the coiled-coil domain of KAP1, supporting my previous results, which indicated that the KRAB binding site of KAP1 is located near the dyad of the RBCC dimer. However, the quality of the map was not sufficient to build an atomic model of the ZNF93 KRAB domain. The KRAB binding site is located on the two-fold axis of the KAP1 dimer and the KRAB domain does not appear to be involved in any significant crystal contacts. Consequently, it is possible that the KRAB domain is present in two different orientations in the crystal lattice, which may be a major factor contributing to the poor map quality in this region. A possible approach to optimize these crystals might

therefore be to increase the space occupied by the KRAB domain, which might result in only one orientation being compatible with the crystal lattice. This could potentially be achieved either by extending the C-terminus of the construct or by fusing the KRAB domain to another protein, such as T4L or MBP.

Together, the data presented here demonstrate that each KAP1 dimer only binds a single KRAB domain and show that the KRAB binding site is located in the central region of the coiled-coil domain. Point mutation in the interaction surface abolished KRAB binding and transcriptional silencing activity of KAP1. In addition, I successfully crystallized KAP1 bound to a KRAB domain, and although the crystals require further optimization to allow structure determination, these results lay the foundation for future structural studies on KRAB:KAP1 complexes.

CHAPTER 4

CONCLUSIONS AND OUTLOOK

KRAB-ZFPs and KAP1/TRIM28 are essential regulators of retrotransposon transcription. KRAB-ZFPs specifically recognize TE-derived sequences and recruit KAP1, inducing the assembly of an epigenetic silencing complex. The data presented in this thesis significantly increase our understanding of the molecular basis of this process. I demonstrate that the RBCC domain of KAP1 forms antiparallel dimers and show that each KAP1 dimer can only bind a single KRAB domain using an interaction surface located in the coiled-coil domain of KAP1. My results further indicate that the B-box 1 domain of KAP1 induces self-assembly of KAP1 dimers into higher-order oligomers. This B-box 1 mediated oligomerization occurs in free KAP1, as well as KAP1 bound to a KRAB domain, suggesting that KAP1 may assemble into large structures at its genomic targets. Conceivably, this may amplify the silencing activity of KAP1 by concentrating higher numbers of repressive chromatin-modifying enzymes, such as SETDB1, at the target locus. Consistent with this model, mutations in the B-box 1 dimer interface, which prevented higher-order oligomerization, resulted in significant defects in KAP1-dependent transcriptional silencing. However, additional experiments are required to conclusively show whether self-assembly indeed contributes to KAP1 function. Notably, a recent study found that KAP1 higher-order oligomerization was not required for the repression of a HERV-K reporter in human ESCs [248]. Initially, it will therefore be crucial to validate the results of my luciferase reporter assays. In particular, it would be informative to establish whether the silencing defects of the KAP1 B-box 1 mutant in the luciferase assay are reflected by reduced levels of repressive histone modifications on known genomic targets of KAP1. To this end, differences in the genomic distribution of H3K9me3 marks in cells expressing the oligomerization-deficient B-box1 mutant could be assessed using chromatin immunoprecipitation sequencing (ChIP-seq) or Cleavage Under Targets and Release Using Nuclease (CUT&RUN) sequencing [259]. Moreover, this approach would reveal whether distinct classes of transposons are differentially affected by mutations in KAP1 B-box1. Potential transposon-specific differences could then be further investigated by inserting sequences derived from these

elements into the luciferase reporter backbone I used for my transcriptional silencing assays [120]. Furthermore, the hypothesis that higher-order oligomerization contributes to KAP1 function would be strengthened considerably, if evidence for B-box 1 mediated self-assembly of KAP1 *in vivo* could be obtained to complement my biophysical data. A recent analysis of the subcellular localization of KAP1 by stochastic optical reconstruction microscopy (STORM) revealed that KAP1 is arranged in small clusters in the nucleus [165]. Consequently, it would be interesting to establish whether the formation of these clusters is dependent on B-box 1 dimerization by comparing the size of KAP1 clusters in the nuclei of cells expressing WT KAP1 or the oligomerization deficient B-box 1 mutant. Alternatively, the primary function of B-box 1 may not be to mediate KAP1 self-assembly, but to recruit a partner protein of KAP1. To explore this possibility, stable isotope labelling with amino acids in cell culture (SILAC)-based proteomics could be used to screen for proteins that bind to WT KAP1 but not to the B-box 1 mutant. Finally, the observation that B-box1 mutations did not affect the silencing activity of KAP1 in ESCs might indicate that the requirement for KAP1 higher-order oligomerization for efficient transcriptional repression is developmental stage specific. It would therefore be informative to perform the experiments described above in human ESCs as well as differentiated cell types.

Together, my results suggest a previously unappreciated role for B-box 1 in KAP1-mediated silencing, although further studies are required to clarify the mechanism, by which B-box 1 contributes to repression. Most importantly, however, my crystal structure of the RBCC domain of KAP1 allowed me to identify the KRAB domain binding site, which is located in the central region of the coiled-coil domain near the dyad. Mutations at this site abolished KRAB binding and transcriptional silencing activity of KAP1. The KRAB binding deficient KAP1 point mutants I described will be valuable tools for future research aiming to dissect the mechanisms of KAP1-dependent transcriptional regulation. In particular, they will allow to confidently differentiate between KRAB-dependent and potential KRAB-independent functions of KAP1, of which there is increasing evidence. Notably, APOBEC3A was proposed to interact with KAP1 and recruit it to the HIV LTR, thereby inducing silencing of the integrated provirus [260]. Moreover, several recent reports described cellular functions for KAP1 that appear to be distinct from its established role in as a repressor of retroelements. For example, multiple studies have implicated KAP1 in the regulation of RNA Pol II pausing

[165, 261, 262]. Furthermore, KAP1 has been reported to cooperate with the transcription factor MyoD to regulate muscle cell differentiation [191]. In such cases, KAP1 variants deficient in KRAB binding could be utilized to determine whether KAP1 acts via an as yet unidentified KRAB-ZFP or whether genomic recruitment of KAP1 is mediated by a different mechanism.

In conclusion, the work presented in this thesis provides insights into the molecular basis of KAP1-dependent transcriptional silencing, and tools to further expand our mechanistic understanding of this process.

CHAPTER 5

MATERIALS AND METHODS

5.1 Plasmids

5.1.1 Bacterial expression vectors

Synthetic genes encoding KAP1 RING (residues 50-146), B-box1 (residues 147-200), RING-B-box 1 (RB1; residues 50-200), RBCC (residues 50-413) and full-length KAP1 (residues 1-835; UniProt: Q13263) codon-optimized for *E. coli* were cloned into the first multiple cloning site (MCS) of the pETDuet plasmid (Novagen), with an N-terminal hexahistidine purification (His_6) tag followed by a TEV protease cleavage site (ENLYFQG). Design of the boundaries of these KAP1 expression constructs was guided by secondary structure prediction using PSIPRED [263], structure prediction using Phyre2 [245] and available structures of related TRIM proteins (PDB: 2MVW [236], 4LTB [175], 4TN3 [177], 5FER [234], 5FEY [234])

An alternative expression plasmid for full-length KAP1 was constructed by replacing the His_6 tag with a glutathione S-transferase (GST) tag. The T4L-RBCC fusion construct was generated 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. The T4L-RBCCΔB1 construct was derived from the T4L-RBCC plasmid by deleting residues 141-202 of KAP1.

MBP-RBCC fusion constructs were generated by inserting a synthetic gene encoding the RBCC domain of KAP1 (residues 56-413) into the first MCS of pETDuet. The gene was preceded by sequences encoding: a His_6 tag; a TEV protease cleavage site; MBP; and a variable linker sequence consisting of one to six alanine residues.

The expression construct for the KRAB domain of ZNF93 was generated by inserting a synthetic gene encoding residues 1-71 of ZNF93 (UniProt: P35789) codon-optimized

for *E.coli* into the pET20 plasmid (Novagen) with N-terminal Twin-StrepII and maltose binding protein (MBP) affinity tags followed by a human rhinovirus (HRV) 3C protease cleavage site (LEVLFQGP). The boundaries of this KRAB construct were chosen based on domain prediction by SMART [156] and secondary structure prediction using PSIPRED [263]. The MBP tag was also required for protein solubility. For co-expression with KAP1, the same ZNF93 KRAB construct was subcloned into the first MCS of pCDFDuet.

An alternative approach used to produce complexes of KAP1 RBCC and the KRAB domain of ZNF93 was to co-express both proteins from the pETDuet plasmid. A synthetic gene encoding residues 50-413 of KAP1 codon-optimized for *E.coli* was inserted into MCS1 adding an N-terminal His_6 tag followed by a TEV protease cleavage site. A synthetic gene encoding residues 1-71 of ZNF93 codon-optimized for *E.coli* was cloned into MCS2 adding an N-terminal glutathione S-transferase (GST) followed by a HRV 3C protease cleavage site.

Mutations were introduced by amplifying the whole vector with 5' phosphorylated primers carrying the desired mutation, followed by circularization of the PCR product using T4 DNA ligase (New England BioLabs). All constructs were verified by sequencing.

5.1.2 Mammalian expression vectors

For expression in mammalian cells, full-length human KAP1 (WT or mutant) was inserted into pLEXm [264] with an N-terminal triple FLAG tag followed by a linker sequence (MDYKDHDGDYKDHDIDYKDDDDKGSGG). All KAP1 pLEXm plasmids except the R183A/V184D mutant were generated by Zheng-Shan Chong (University of Cambridge).

SVA and LINE1 luciferase reporter plasmids have been previously described and were kind gifts from Helen Rowe with permission from David Haussler [120, 207]:

Table 5. Luciferase reporter plasmids.

Plasmid	Description
pGL4cp-VNTR-OCT4Enh-E2 **pGL4cp-L1PA4-OCT4Enh-E2**	Firefly luciferase reporter plasmids
pCAG_ZNF91_HA	ZNF91
pCAG_ZNF93	ZNF93
pRTTK_Renilla	*Renilla* luciferase under a constitutive promoter

5.2 Mammalian cell culture

KAP1 KO HEK293T and WT HEK293T cells were kind gifts from Helen Rowe. The KAP1 KO cell line was generated by CRISPR/Cas9-mediated genome editing followed by isolation of a monoclonal cell population. KAP1 knockout was confirmed by western blotting and in luciferase reporter assays [207, 265].

Both cell lines were grown at 37ºC in 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM; Sigma-Aldrich) supplemented with 10% foetal bovine serum (FBS; Gibco) and 1x penicillin/streptomycin. Cells were regularly tested for mycoplasma contamination.

5.3 Protein expression and purification

5.3.1 Protein expression

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). After 16 h cells were harvested by centrifugation (15 min, 6,000×*g*) and either used immediately or stored at -80˚C until required. All subsequent steps were performed at 4˚C.

5.3.2 Protein purification

5.3.2.1 KAP1 RBCC domain and T4L-RBCC fusion construct

Cell pellets were resuspended in lysis buffer containing 50 mM Tris pH 8, 0.3 M NaCl, 20 mM imidazole, 0.5 mM tris(2-carboxyethyl)phosphine (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.

T4L-RBCC fusion protein was purified as the RBCC domain, except that the His_6 tag was not removed.

5.3.2.2 KAP1 RING and RING-B-box1 constructs

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.

5.3.2.3 Full-length KAP1

Full-length KAP1 was expressed with either an N-terminal His_6 or an N-terminal GST tag. His_6 tagged KAP1 was purified as described for 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 GST-tagged KAP1, cell pellets were resuspended in lysis buffer (50 mM Tris pH 8, 0.2 M NaCl, 0.5 mM TCEP, 1:10,000 (v/v) benzonase solution, 1×cOmplete EDTA-free protease inhibitors). The cells were lysed by sonication immediately after addition of 1 mM PMSF. The lysate was clarified by centrifugation (30 min, 40,000×*g*). The supernatant was applied to a 5-ml GSTrap column (GE Healthcare) preequilibrated in wash buffer (50 mM Tris pH 8, 0.2 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 25 mM reduced glutathione. Subsequently, the buffer was exchanged to 50 mM Tris pH 8, 0.2 M NaCl, 0.5 mM TCEP and the GST tag was removed by incubating the protein overnight at 4°C with 1:50 (w/w) TEV protease. Following a second glutathione-affinity chromatography step to remove uncleaved protein and protease, the sample was

further purified by size-exclusion chromatography using a Superose 6 Increase 10/300 column (GE Healthcare) preequilibrated in 20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP.

5.3.2.4 KRAB domain of ZNF93

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, 1×cOmplete EDTA-free protease inhibitors) 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 (Sigma-Aldrich) and further purified by size-exclusion chromatography using a HiLoad (16/600) Superdex 200 pg column preequilibrated in 20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP.

5.3.2.5 KAP1:MBP-KRAB complex

Full-length KAP1:MBP-KRAB complexes were 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.

5.3.2.6 RBCC:KRAB complex

KAP1 RBCC and GST-tagged ZNF93-KRAB were co-expressed in *E. coli* BL21 (DE3) cells. Complexes of KAP1 RBCC and the KRAB domain of ZNF93 were subsequently purified by glutathione-affinity chromatography as described for GST-KAP1, except that the GST tag was removed with 1:200 (w/w) HRV 3C protease and a HiLoad (16/600) Superdex 200 pg column preequilibrated in 20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP was used for the final size-exclusion chromatography step.

5.3.2.7 T4L-RBCC:KRAB and T4L-RBCCΔB1:KRAB complexes

T4L-RBCC or T4L-RBCCΔB1 was co-expressed with ZNF93-KRAB carrying N-terminal Twin-StrepII and MBP tags in *E. coli* BL21 (DE3) cells. To purify the complex, bacteria pellets were resuspended in lysis buffer (50 mM Tris pH 8, 0.2 M NaCl, 0.5 mM TCEP, 1:10,000 (v/v) benzonase solution, 1×cOmplete EDTA-free protease inhibitors) 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 preequilibrated in wash buffer (50 mM Tris pH 8, 0.2 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. Subsequently, the buffer was exchanged to 50 mM Tris pH 8, 0.2 M NaCl, 0.5 mM TCEP and the Twin-StrepII-MBP-tag was removed by incubating the protein overnight at 4˚C with 1:200 (w/w) HRV 3C protease. Following a second Strep-Tactin-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 preequilibrated in 20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP.

5.4 X-ray crystallography

5.4.1 T4L-KAP1 RBCC fusion protein

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 [243]. Crystals were grown at 18˚C by sitting drop vapor diffusion. 1 µl of 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 [266] 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 [267] using zinc as the anomalously scattering heavy atom. The atomic model was built with COOT [268] and iteratively refined with REFMAC [269] and PHENIX at 2.9 Å resolution. The atomic models for T4 lysozyme (PDB: 1LYD [270], 2LZM [271])

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 structures of other TRIMs as guides (PDB: 4LTB [175], 4TN3 [177], 5FEY [234], 5NT1 [244]). See **Table 1** for data collection and refinement statistics. See **Fig. 17** for sample electron density. Structure figures were generated with PyMOL (Schrodinger, LLC). Methylation of T4L-RBCC at 38 sites was confirmed by mass spectrometry (**Fig. 15a**), but none of the dimethyl-amine groups were visible in the map.

5.4.2 B-box 1 domain of KAP1

Crystals were grown at 18°C by sitting drop vapor diffusion. 1 µl of KAP1 B1-box 1 at 10 g L^{-1} (1.7 mM) was mixed with an equal volume of reservoir solution optimized from the JCSG screen (Qiagen), containing 1.7 M sodium citrate pH 6.5. Thin rod-shaped crystals appeared after 1-2 days and were frozen in liquid nitrogen without additional cryoprotection. X-ray diffraction data were collected at 100 K at Diamond Light Source (beamline I03) and processed with DIALS [272] and AIMLESS [273] using the xia2 automated data processing pipeline [274] in the CCP4 software suite [275]. The X-ray energy was tuned to 9,673 eV, corresponding to the zinc L-III edge, for data collection. Phases were determined with the SAD method in PHENIX using zinc as the anomalously scattering heavy atom. An initial model for KAP1 B-box 1 was built using AutoBuild in PHENIX, manually completed with COOT and iteratively refined with PHENIX at 1.67 Å resolution. See **Table 2** for data collection and refinement statistics. See **Fig. 21b** for sample electron density. Structure figures were generated with PyMOL.

5.4.3 T4L-RBCC:ZNF93-KRAB complex

Prior to crystallization, free amines in the complex were methylated as described above. Crystals were grown at 18°C by sitting drop vapor diffusion. Methylated T4L-RBCC:KRAB at 4.5 g L^{-1} was mixed with an equal volume of reservoir solution optimized from the JBScreen Classic 2 screen (Jena Bioscience): 13% PEG 4,000, 0.25 M $MgCl_2$, 0.1 M Tris pH 8.5. Crystals appeared within 2-3 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 xia2 (DIALS, AIMLESS). See **Table 3** for data collection statistics.

5.4.4 T4L-RBCCΔB1:ZNF93-KRAB complex

Prior to crystallization, free amines in the complex were methylated as described above. Crystals were grown at 18˚C by sitting drop vapor diffusion. Methylated T4L-RBCC at 4.5 g L^{-1} was mixed with an equal volume of reservoir solution optimized from the JCSG screen (Qiagen): 18% PEG 8,000, 0.2 M $MgCl_2$, 0.1 M Tris pH 9. Crystals appeared within 2-3 days and were frozen in liquid nitrogen with 20% glycerol as a cryoprotectant. X-ray diffraction data were collected at 100 K at Diamond Light Source (beamline I03) and processed with xia2 (DIALS, AIMLESS). See **Table 4** for data collection statistics.

5.5 Biophysical assays

5.5.1 Differential Scanning Fluorimetry (DSF)

10 µl samples containing KAP1 RBCC (wild-type or mutant) at 25-50 µM (1-2 g L^{-1}) in 20 mM HEPES pH 8.0, 0.5 M NaCl, 0.5 mM TCEP were loaded into glass capillaries (Nanotemper) by capillary action. Intrinsic protein fluorescence at 330 nm and 350 nm was monitored between 15 and 95°C in a Prometheus NT.48 instrument (Nanotemper), and the T_m values calculated within the accompanying software by taking the turning point of the first derivative of the F_{350}:F_{330} ratio as a function of temperature.

5.5.2 Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS)

100 µl of protein sample was subjected to SEC at 293 K using the columns detailed in **Table 6** preequilibrated in the indicated buffer 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). Protein in the eluate was also detected with a differential refractometer (Optilab T-rEX, Wyatt Technology) and a UV detector at 280 nm (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).

Table 6. Columns and buffers used for SEC-MALS.

Sample	SEC column	Buffer
KAP1 (full-length)	Superose 6 (10/300)	20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP or 20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP
KAP1:MBP-KRAB complex	Superose 6 (10/300)	20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP
KAP1 RBCC (WT and mutants)	Superdex 200 (10/300)	20 mM HEPES pH 8, 0.5 M NaCl, 0.5 mM TCEP
RBCC:KRAB complex	Superdex 200 (10/300)	20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP
KAP1 RING	Superdex 75 (10/300)	20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP
KAP1 RING-B-box 1 (WT and mutants)	Superdex 75 (10/300)	20 mM HEPES pH 8, 0.2 M NaCl, 0.5 mM TCEP

5.5.3 Sedimentation-equilibrium analytical ultracentrifugation (SE-AUC)

KAP1 RBCC samples at 0.2 mM (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, 0.5 M NaCl, 0.5 mM TCEP. 110 µL samples were loaded in 12 mm 6-sector cells and centrifuged at 5, 8.5 and 15 krpm at 20°C in an An50Ti rotor in 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 analysed in SEDPHAT 13b [276]. Equilibrium sedimentation distributions were fit to obtain average masses. An SE-AUC average mass isotherm compiled from fits to the data was analysed in SEDPHAT using isodesmic and dimer-tetramer-octamer oligomerization models. The partial-specific volumes (v-bar), solvent density and viscosity were calculated with Sednterp (www.rasmb.org/sednterp). The experiment was performed by Dr Stephen McLaughlin (MRC Laboratory of Molecular Biology).

5.5.4 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 20 mM HEPES pH 8.0, 0.5 M NaCl, 0.5 mM TCEP at 20°C. MBP-KRAB was immobilized onto the chips until a response unit 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

biphasic kinetic model with KaleidaGraph (Synergy Software) and PRISM 8 (GraphPad) to determine k_{on}, k_{off} and K_d. The experiment was performed by Dr Stephen McLaughlin (MRC Laboratory of Molecular Biology).

5.6 Atomic model of KAP1 B-box 1

An atomic model of the KAP1 B-box 1 domain was generated from the TRIM19 B-box 1 structure [236] (PDB: 2MVW) with Phyre2 [245]. The KAP1 B-box 1 was then superimposed onto each protomer of the TRIM19 B-box 1 dimer to generate a model of the KAP1 B-box 1 dimer.

5.7 Molecular biology

5.7.1 Pulldown assays

8 nmol of KAP1 RBCC was incubated with 2 nmol of Twin-StrepII-MBP-KRAB for 45 min on ice. StrepII-tagged bait protein was then captured with 100 µl of Strep-Tactin Sepharose (IBA) for 1 h at 4˚C. After four washes with 1 ml of 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 analysed by SDS-PAGE followed by staining with Quick Coomassie Stain (Generon).

5.7.2 Transcriptional silencing assays

KAP1 silencing activity was measured with a reporter assays in which a SINE-VNTR-Alu (SVA) type D or LINE-1 sequence upstream of a minimal SV40 promoter enhances firefly luciferase activity unless KAP1 and the cognate KRAB-ZFP (ZNF91 and ZNF93, respectively) are present to repress the reporter [120]. The assay was adapted for use in HEK293T cells as described [207]. KAP1 KO 293T cells were seeded in 24-well plates at a density of 3 x 10^4 cells per well. The next day, the cells in each well were cotransfected with 20 ng firefly luciferase reporter plasmid, 0.2 µg plasmid encoding ZNF91 or ZNF93, 0.2 µg pLEXm plasmid encoding WT or mutant KAP1 and 0.4 ng plasmid encoding *Renilla* luciferase using 1.5 µl of FuGENE 6 (Promega) following the protocol provided by the manufacturer. Luciferase activity was measured 48 h post-transfection using the Dual Luciferase assay kit (Promega) with a Pherastar FSX platereader (BMG Labtech).

Replicates were performed on separate days. Firefly luciferase values were normalized to *Renilla* luciferase values to control for transfection efficiency. Statistical significance was assessed with an unpaired *t* test (assuming Gaussian distributions, without Welch's correction) with PRISM 8.

5.7.3 Western blotting

4 x 10^5 HEK293T cells were lysed in 100 µl of Passive Lysis Buffer (Promega). 10 µl of cell lysates were separated on a NuPAGE 4-12% Bis-Tris polyacrylamide gel (ThermoFisher). The samples were transferred onto nitrocellulose membranes using the iBlot2 Dry Blotting System (ThermoFisher). The membrane was blocked with 5% (w/v) skim milk powder (Sigma-Aldrich) in PBS for 1 h at room temperature before it was incubated overnight at 4˚C with primary antibody diluted in PBS-T (PBS with 0.1% Tween-20) containing 5% (w/v) skim milk powder. Subsequently, the membrane was washed four times with PBS-T and incubated with fluorescent secondary antibody diluted in PBS-T containing 5% (w/v) skim milk powder. After 30 min at room temperature, the membrane was washed four times with PBS-T, twice with PBS and once with ultrapure water. Blots were imaged using an Odyessy CLx gel scanner (LI-COR Biosciences).

The following antibodies were used:

Table 7. Primary antibodies.

Target	Antibody type	Manufacturer	Catalog #	Dilution
KAP1	Rabbit polyclonal IgG	Abcam	ab10484	1:10,000
β-actin	Rabbit monoclonal IgG	Abcam	ab219733	1:2,000

Table 8. Secondary antibodies.

Description	Manufacturer	Catalog #	Dilution
DyLight 800 goat anti-rabbit IgG	Cell Signaling Technology	5151	1:10,000

REFERENCES

1. Lander ES, Linton LM, Birren B, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921 . doi: 10.1038/35057062

2. Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: From conflicts to benefits. Nat Rev Genet 18:71–86 . doi: 10.1038/nrg.2016.139

3. Naville M, Warren IA, Haftek-Terreau Z, et al (2016) Not so bad after all: Retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. Clin Microbiol Infect 22:312–323 . doi: 10.1016/j.cmi.2016.02.001

4. Friedli M, Trono D (2015) The Developmental Control of Transposable Elements and the Evolution of Higher Species. Annu Rev Cell Dev Biol 31:429–51 . doi: 10.1146/annurev-cellbio-100814-125514

5. Munoz-Lopez M, Garcia-Perez J (2010) DNA Transposons: Nature and Applications in Genomics. Curr Genomics 11:115–128 . doi: 10.2174/138920210790886871

6. Pace JK 2nd, Feschotte C (2007) The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res 17:422–432 . doi: 10.1101/gr.5826307

7. Goodier JL (2016) Restricting retrotransposons: A review. Mob DNA 7: . doi: 10.1186/s13100-016-0070-z

8. de Koning APJ, Gu W, Castoe TA, et al (2011) Repetitive elements may comprise over Two-Thirds of the human genome. PLoS Genet 7: . doi: 10.1371/journal.pgen.1002384

9. Hancks DC, Kazazian HH (2012) Active human retrotransposons: Variation and disease. Curr Opin Genet Dev 22:191–203 . doi: 10.1016/j.gde.2012.02.006

10. Garcia-Perez JL, Widmann TJ, Adams IR (2016) The impact of transposable elements on mammalian development. Dev 143:4101–4114 . doi: 10.1242/dev.132639

11. Grandi N, Tramontano E (2018) Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. Front Immunol 9:1–16 . doi: 10.3389/fimmu.2018.02039

12. Römer C, Singh M, Hurst LD, Izsvák Z (2017) How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency. Curr Opin Virol 25:49–58 . doi: 10.1016/j.coviro.2017.07.001

13. Magiorkinis G, Gifford RJ, Katzourakis A, et al (2012) Env-less endogenous retroviruses are genomic superspreaders. Proc Natl Acad Sci U S A 109:7385–7390 . doi: 10.1073/pnas.1200913109

14. Gerdes P, Richardson SR, Mager DL, Faulkner GJ (2016) Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. Genome Biol 17:100 . doi: 10.1186/s13059-016-0965-5

15. Wildschutte JH, Williams ZH, Montesion M, et al (2016) Discovery of unfixed endogenous retrovirus insertions in diverse human populations. Proc Natl Acad Sci U S A 113:E2326–E2334 . doi: 10.1073/pnas.1602336113

16. Maksakova IA, Romanish MT, Gagnier L, et al (2006) Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet 2:e2 . doi: 10.1371/journal.pgen.0020002

17. Beck CR, Garcia-Perez JL, Badge RM, Moran J V. (2011) LINE-1 Elements in Structural Variation and Disease. Annu Rev Genomics Hum Genet 12:187–215 . doi: 10.1146/annurev-genom-082509-141802

18. Feng Q, Moran J V, Kazazian HHJ, Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905–916 . doi: 10.1016/s0092-8674(00)81997-2

19. Mathias SL, Scott AF, Kazazian HHJ, et al (1991) Reverse transcriptase encoded by a human transposable element. Science 254:1808–1810 . doi: 10.1126/science.1722352

20. Denli AM, Narvaiza I, Kerman BE, et al (2015) Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. Cell 163:583–593 . doi: 10.1016/j.cell.2015.09.025

21. Hancks DC, Kazazian HH (2016) Roles for retrotransposon insertions in human disease. Mob DNA 7: . doi: 10.1186/s13100-016-0065-9

22. Burns KH (2017) Transposable elements in cancer. Nat Rev Cancer 17:415–424 . doi: 10.1038/nrc.2017.35

23. Tubio JMC, Li Y, Ju YS, et al (2014) Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science (80-) 345: . doi: 10.1126/science.1251343

24. Goodier JL, Ostertag EM, Kazazian HHJ (2000) Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum Mol Genet 9:653–657 . doi: 10.1093/hmg/9.4.653

25. Kulpa DA, Moran J V. (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat Struct Mol Biol 13:655–660 . doi: 10.1038/nsmb1107

26. Esnault C, Maestre J, Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. Nat Genet 24:363–367 . doi: 10.1038/74184

27. Cost GJ, Feng Q, Jacquier A, Boeke JD (2002) Human L1 element target-primed reverse transcription in vitro. EMBO J 21:5899–5910 . doi: 10.1093/emboj/cdf592

28. Jurka J (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A 94:1872–1877 . doi: 10.1073/pnas.94.5.1872

29. Deininger P (2011) Alu elements: Know the SINEs. Genome Biol 12:1–12 . doi: 10.1186/gb-2011-12-12-236

30. Kim S, Cho C-S, Han K, Lee J (2016) Structural Variation of Alu Element and Human Disease . Genomics Inform 14:70 . doi: 10.5808/gi.2016.14.3.70

31. Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379 . doi: 10.1038/nrg798

32. Carnevali D, Conti A, Pellegrini M, Dieci G (2017) Whole-genome expression analysis of mammalian-wide interspersed repeat elements in human cell lines. DNA Res 24:59–69 . doi: 10.1093/dnares/dsw048

33. Richardson SR, Doucet AJ, Kopera HC, et al (2015) The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. Microbiol Spectr 3:MDNA3-0061–2014 . doi: 10.1128/microbiolspec.MDNA3-0061-2014

34. Wang H, Xing J, Grover D, et al (2005) SVA elements: a hominid-specific retroposon family. J Mol Biol 354:994–1007 . doi: 10.1016/j.jmb.2005.09.085

35. Hancks DC, Kazazian HHJ (2010) SVA retrotransposons: Evolution and genetic instability. Semin Cancer Biol 20:234–245 . doi: 10.1016/j.semcancer.2010.04.001

36. Damert A, Raiz J, Horn A V, et al (2009) 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome Res 19:1992–2008 . doi: 10.1101/gr.093435.109

37. Dupressoir A, Lavialle C, Heidmann T (2012) From ancestral infectious retroviruses to bona fide cellular genes: Role of the captured syncytins in placentation. Placenta 33:663–671 . doi: 10.1016/j.placenta.2012.05.005

38. Sha M, Lee X, Li X ping, et al (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403:785–789 . doi: 10.1038/35001608

39. Blaise S, De Parseval N, Bénit L, Heidmann T (2003) Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proc Natl Acad Sci U S A 100:13013–13018 . doi: 10.1073/pnas.2132646100

40. Mangeney M, Renard M, Schlecht-Louf G, et al (2007) Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. Proc Natl Acad Sci U S A 104:20534–20539 . doi: 10.1073/pnas.0707873105

41. Tolosa JM, Schjenken JE, Clifton VL, et al (2012) The endogenous retroviral envelope protein syncytin-1 inhibits LPS/PHA-stimulated cytokine responses in human blood and is sorted into placental exosomes. Placenta 33:933–941 . doi: 10.1016/j.placenta.2012.08.004

42. Lavialle C, Cornelis G, Dupressoir A, et al (2013) Paleovirology of "syncytins", retroviral env genes exapted for a role in placentation. Philos Trans R Soc B Biol Sci 368: . doi: 10.1098/rstb.2012.0507

43. Sugimoto J, Sugimoto M, Bernstein H, et al (2013) A novel human endogenous retroviral protein inhibits cell-cell fusion. Sci Rep 3:1–9 . doi: 10.1038/srep01462

44. Ono R, Nakamura K, Inoue K, et al (2006) Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nat Genet 38:101–106 . doi: 10.1038/ng1699

45. Henke C, Ruebner M, Faschingbauer F, et al (2013) Regulation of murine placentogenesis by the retroviral genes Syncytin-A, Syncytin-B and Peg10. Differentiation 85:150–160 . doi: 10.1016/j.diff.2013.02.002

46. Sekita Y, Wagatsuma H, Nakamura K, et al (2008) Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. Nat Genet 40:243–248 . doi: 10.1038/ng.2007.51

47. Ito M, Sferruzzi-Perri AN, Edwards CA, et al (2015) A trans-homologue interaction between reciprocally imprinted miR-127 and Rtl1 regulates placenta development. Dev 142:2425–2430 . doi: 10.1242/dev.121996

48. Grow EJ, Flynn RA, Chavez SL, et al (2015) Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature 522:221–246 . doi: 10.1038/nature14308

49. Feschotte C, Pritham EJ (2007) DNA Transposons and the Evolution of Eukaryotic Genomes. Annu Rev Genet 41:331–368 . doi: 10.1146/annurev.genet.40.110405.090448

50. Zhou L, Mitra R, Atkinson PW, et al (2004) Transposition of hAT elements links transposable elements and V(D)J recombination. Nature 432:995–1001 . doi: 10.1038/nature03157

51. Kapitonov V V., Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 3:0998–1011 . doi: 10.1371/journal.pbio.0030181

52. Zhang Y, Cheng TC, Huang G, et al (2019) Transposon molecular domestication and the evolution of the RAG recombinase. Nature 569:79–84 . doi: 10.1038/s41586-019-1093-7

53. Frank JA, Feschotte C (2017) Co-option of endogenous viral sequences for host cell function. Curr Opin Virol 25:81–89 . doi: 10.1016/j.coviro.2017.07.021

54. Blanco-Melo D, Gifford RJ, Bieniasz PD (2017) Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. Elife 6:1–19 . doi: 10.7554/eLife.22519

55. Plath N, Ohana O, Dammermann B, et al (2006) Arc/Arg3.1 Is Essential for the Consolidation of Synaptic Plasticity and Memories. Neuron 52:437–444 . doi: 10.1016/j.neuron.2006.08.024

56. Nikolaienko O, Patil S, Eriksen MS, Bramham CR (2018) Arc protein: a flexible hub for synaptic plasticity and cognition. Semin Cell Dev Biol 77:33–42 . doi: 10.1016/j.semcdb.2017.09.006

57. Pastuzyn ED, Day CE, Kearns RB, et al (2018) The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. Cell 172:275-288.e18 . doi: 10.1016/j.cell.2017.12.024

58. Oliver KR, Greene WK (2011) Mobile DNA and the TE-Thrust hypothesis: Supporting evidence from the primates. Mob DNA 2:1–17 . doi: 10.1186/1759-8753-2-8

59. Moran J V, DeBerardinis RJ, Kazazian HHJ (1999) Exon shuffling by L1 retrotransposition. Science 283:1530–1534 . doi: 10.1126/science.283.5407.1530

60. Xing J, Wang H, Belancio VP, et al (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci U S A 103:17608–17613 . doi: 10.1073/pnas.0603224103

61. Chuong EB, Elde NC, Feschotte C (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science (80-) 351:1083–1087 . doi: 10.1126/science.aad5497

62. Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet 43:1154–1159 . doi: 10.1038/ng.917

63. Emera D, Wagner GP (2012) Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. Proc Natl Acad Sci U S A 109:11246–11251 . doi: 10.1073/pnas.1118566109

64. Chuong EB, Rumi MAK, Soares MJ, Baker JC (2013) Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat Genet 45:325–329 . doi: 10.1038/ng.2553

65. Ferreira LMR, Meissner TB, Mikkelsen TS, et al (2016) A distant trophoblast-specific enhancer controls HLA-G expression at the maternal-fetal interface. Proc Natl Acad Sci U S A 113:5364–5369 . doi: 10.1073/pnas.1602886113

66. Pi W, Zhu X, Wu M, et al (2010) Long-range function of an intergenic retrotransposon. Proc Natl Acad Sci U S A 107:12992–12997 . doi: 10.1073/pnas.1004139107

67. Ting CN, Rosenberg MP, Snow CM, et al (1992) Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. Genes Dev 6:1457–1465 . doi: 10.1101/gad.6.8.1457

68. Schmidt D, Schwalie PC, Wilson MD, et al (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148:335–348 . doi: 10.1016/j.cell.2011.11.058

69. Zhang Y, Li T, Preissl S, et al (2019) Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat Genet 51: . doi: 10.1038/s41588-019-0479-7

70. Rebollo R, Karimi MM, Bilenky M, et al (2011) Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS Genet 7: . doi: 10.1371/journal.pgen.1002301

71. Lu X, Sachs F, Ramsay L, et al (2014) The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity AbstractAbstract MainMain MethodsMethods AccessionsAccessions ReferencesReferences AcknowledgementsAcknowledgements Author informationAuthor information Integrate

72. Wang J, Xie G, Singh M, et al (2014) Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature 516:405–409 . doi: 10.1038/nature13804

73. Wang Y, Xu Z, Jiang J, et al (2013) Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal. Dev Cell 25:69–80 . doi: 10.1016/j.devcel.2013.03.002

74. Loewer S, Cabili MN, Guttman M, et al (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet 42:1113–1117 . doi: 10.1038/ng.710

75. Kunarso G, Chia NY, Jeyakani J, et al (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet 42:631–634 . doi: 10.1038/ng.600

76. Santoni FA, Guerra J, Luban J (2012) HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology 9:1 . doi: 10.1186/1742-4690-9-111

77. Percharde M, Lin CJ, Yin Y, et al (2018) A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell 174:391-405.e19 . doi: 10.1016/j.cell.2018.05.043

78. Richardson SR, Morell S, Faulkner GJ (2014) L1 Retrotransposons and Somatic Mosaicism in the Brain. Annu Rev Genet 48:1–27 . doi: 10.1146/annurev-genet-120213-092412

79. Muotri AR, Chu VT, Marchetto MCN, et al (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature 435:903–910 . doi: 10.1038/nature03663

80. Coufal NG, Garcia-Perez JL, Peng GE, et al (2009) L1 retrotransposition in human neural progenitor cells. Nature 460:1127–1131 . doi: 10.1038/nature08248

81. Evrony GD, Cai X, Lee E, et al (2012) Single-neuron sequencing analysis of l1 retrotransposition and somatic mutation in the human brain. Cell 151:483–496 . doi: 10.1016/j.cell.2012.09.035

82. Upton KR, Gerhardt DJ, Jesuadian JS, et al (2015) Ubiquitous L1 mosaicism in hippocampal neurons. Cell 161:228–239 . doi: 10.1016/j.cell.2015.03.026

83. Faulkner GJ, Billon V (2018) L1 retrotransposition in the soma: A field jumping ahead. Mob DNA 9:1–18 . doi: 10.1186/s13100-018-0128-1

84. Baillie JK, Barnett MW, Upton KR, et al (2011) Somatic retrotransposition alters the genetic landscape of the human brain. Nature 479:534–537 . doi: 10.1038/nature10531

85. Muotri AR, Zhao C, Marchetto MCN, Gage FH (2009) Environmental influence on L1 retrotransposons in the adult hippocampus. Hippocampus 19:1002–1007 . doi: 10.1002/hipo.20564

86. Bedrosian TA, Quayle C, Novaresi N, Gage FH (2018) Early life experience drives structural variation of neural genomes in mice. Science (80-) 359:1395–1399 . doi: 10.1126/science.aah3378

87. Bachiller S, del-Pozo-Martín Y, Carrión ÁM (2017) L1 retrotransposition alters the hippocampal genomic landscape enabling memory formation. Brain Behav Immun 64:65–70 . doi: 10.1016/j.bbi.2016.12.018

88. Kazazian HH, Wong C, Youssoufian H, et al (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332:164–166 . doi: 10.1038/332164a0

89. Chen JM, Masson E, Macek M, et al (2008) Detection of two Alu insertions in the CFTR gene. J Cyst Fibros 7:37–43 . doi: 10.1016/j.jcf.2007.04.001

90. Kobayashi K, Nakahori Y, Miyake M, et al (1998) An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. Nature 394:388–392 . doi: 10.1038/28653

91. Taniguchi-Ikeda M, Kobayashi K, Kanagawa M, et al (2011) Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. Nature 478:127–131 . doi: 10.1038/nature10456

92. Lamprecht B, Walter K, Kreher S, et al (2010) Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat Med 16:571–579 . doi: 10.1038/nm.2129

93. Kinzler KW, Vogelstein B, Horii A, et al (1992) Disruption of the APC Gene by a Retrotransposal Insertion of LI Sequence in a Colon Cancer. Cancer Res 52:643–645

94. Scott EC, Gardner EJ, Masood A, et al (2016) A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res 26:745–755 . doi: 10.1101/gr.201814.115

95. Denne M, Sauter M, Armbruester V, et al (2007) Physical and Functional Interactions of Human Endogenous Retrovirus Proteins Np9 and Rec with the Promyelocytic Leukemia Zinc Finger Protein. J Virol 81:5607–5616 . doi: 10.1128/jvi.02771-06

96. Mangeney M, Pothlichet J, Renard M, et al (2005) Endogenous retrovirus expression is required for murine melanoma tumor growth in vivo. Cancer Res 65:2588–2591 . doi: 10.1158/0008-5472.CAN-04-4231

97. Hung T, Pratt GA, Sundararaman B, et al (2015) The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science (80-) 350:455–459 . doi: 10.1126/science.aac7442

98. Ito K, Baudino L, Kihara M, et al (2013) Three Sgp loci act independently as well as synergistically to elevate the expression of specific endogenous retroviruses implicated in murine lupus. J Autoimmun 43:10–17 . doi: 10.1016/j.jaut.2013.01.014

99. Treger RS, Pope SD, Kong Y, et al (2019) The Lupus Susceptibility Locus Sgp3 Encodes the Suppressor of Endogenous Retrovirus Expression SNERV. Immunity 50:334-347.e9 . doi: 10.1016/j.immuni.2018.12.022

100. Gröger V, Cynis H (2018) Human endogenous retroviruses and their putative role in the development of autoimmune disorders such as multiple sclerosis. Front Microbiol 9: . doi: 10.3389/fmicb.2018.00265

101. Kremer D, Schichel T, Förster M, et al (2013) Human endogenous retrovirus type W envelope protein inhibits oligodendroglial precursor cell differentiation. Ann Neurol 74:721–732 . doi: 10.1002/ana.23970

102. Curtin F, Perron H, Kromminga A, et al (2015) Preclinical and early clinical development of GNbAC1, a humanized IgG4 monoclonal antibody targeting endogenous retroviral MSRV-Env protein. MAbs 7:265–275 . doi: 10.4161/19420862.2014.985021

103. Li W, Lee MH, Henderson L, et al (2015) Human endogenous retrovirus-K contributes to motor neuron disease. Sci Transl Med 7: . doi: 10.1126/scitranslmed.aac8201

104. Muotri AR, Marchetto MCN, Coufal NG, et al (2010) L1 retrotransposition in neurons is modulated by MeCP2. Nature 468:443–446 . doi: 10.1038/nature09544

105. Karlsson H, Bachmann S, Schröder J, et al (2001) Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. Proc Natl Acad Sci U S A 98:4634–4639 . doi: 10.1073/pnas.061021998

106. Bundo M, Toyoshima M, Okada Y, et al (2014) Increased L1 retrotransposition in the neuronal genome in schizophrenia. Neuron 81:306–313 . doi: 10.1016/j.neuron.2013.10.053

107. Wolf G, Greenberg D, Macfarlan TS (2015) Spotting the enemy within: Targeted silencing of foreign DNA in mammalian genomes by the Krüppel-associated box zinc finger protein family. Mob DNA 6:1–20 . doi: 10.1186/s13100-015-0050-8

108. Friedman JR, Fredericks WJ, Jensen DE, et al (1996) KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes Dev 10:2067–2078 . doi: 10.1101/gad.10.16.2067

109. Kim SS, Chen YM, O'Leary E, et al (1996) A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. Proc Natl Acad Sci U S A 93:15299–15304 . doi: 10.1073/pnas.93.26.15299

110. Lechner MS, Begg GE, Speicher DW, Rauscher FJ 3rd (2000) Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. Mol Cell Biol 20:6449–6465 . doi: 10.1128/mcb.20.17.6449-6465.2000

111. Schultz DC, Ayyanathan K, Negorev D, et al (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev 16:919–932 . doi: 10.1101/gad.973302

112. Ivanov A V., Peng H, Yurchenko V, et al (2007) PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain Required for Gene Silencing. Mol Cell 28:823–837 . doi: 10.1016/j.molcel.2007.11.012

113. Schultz DC, Friedman JR, Rauscher FJ 3rd (2001) Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. Genes Dev 15:428–443 . doi: 10.1101/gad.869501

114. Larson AG, Elnatan D, Keenen MM, et al (2017) Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature 547:236–240 . doi: 10.1038/nature22822

115. Imbeault M, Helleboid PY, Trono D (2017) KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature 543:550–554 . doi: 10.1038/nature21683

116. Helleboid P, Heusel M, Duc J, et al (2019) The interactome of <scp>KRAB</scp> zinc finger proteins reveals the evolutionary history of their functional diversification. EMBO J 1–16 . doi: 10.15252/embj.2018101220

117. Huntley S, Baggott DM, Hamilton AT, et al (2006) A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. Genome Res 16:669–677 . doi: 10.1101/gr.4842106

118. Bruno M, Mahgoub M, Macfarlan TS (2019) The Arms Race Between KRAB – Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. 1–24

119. Emerson RO, Thomas JH (2009) Adaptive evolution in zinc finger transcription factors. PLoS Genet 5: . doi: 10.1371/journal.pgen.1000325

120. Jacobs FMJ, Greenberg D, Nguyen N, et al (2014) An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature 516:242–5 . doi: 10.1038/nature13760

121. Patel A, Yang P, Tinkham M, et al (2018) DNA Conformation Induces Adaptable Binding by Tandem Zinc Finger Proteins. Cell 173:221-233.e12 . doi: 10.1016/j.cell.2018.02.058

122. Lukic S, Nicolas JC, Levine AJ (2014) The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. Cell Death Differ 21:381–387 . doi: 10.1038/cdd.2013.150

123. Yang P, Wang Y, Macfarlan TS (2017) The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. Trends Genet 33:871–881 . doi: 10.1016/j.tig.2017.08.006

124. Mackay DJG, Callaway JLA, Marks SM, et al (2008) Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet 40:949–951 . doi: 10.1038/ng.187

125. Li X, Ito M, Zhou F, et al (2008) A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell 15:547–557 . doi: 10.1016/j.devcel.2008.08.014

126. Quenneville S, Verde G, Corsinotti A, et al (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell 44:361–372 . doi: 10.1016/j.molcel.2011.08.032

127. Takahashi N, Coluccio A, Thorball CW, et al (2019) ZNF445 is a primary regulator of genomic imprinting. Genes Dev 33:49–54 . doi: 10.1101/gad.320069.118

128. Ecco G, Imbeault M, Trono D (2017) KRAB zinc finger proteins. Development 144:2719–2729 . doi: 10.1242/dev.132605

129. Urrutia R (2003) KRAB-containing zinc-finger repressor proteins. Genome Biol 4: . doi: 10.1186/gb-2003-4-10-231

130. Looman C, Abrink M, Mark C, Hellman L (2002) KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. Mol Biol Evol 19:2118–2130 . doi: 10.1093/oxfordjournals.molbev.a004037

131. Witzgall R, O'Leary E, Leaf A, et al (1994) The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. Proc Natl Acad Sci U S A 91:4514–4518 . doi: 10.1073/pnas.91.10.4514

132. Margolin JF, Friedman JR, Meyer WK, et al (1994) Kruppel-associated boxes are potent transcriptional repression domains. Proc Natl Acad Sci U S A 91:4509–4513 . doi: 10.1073/pnas.91.10.4509

133. Vissing H, Meyer WK, Aagaard L, et al (1995) Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. FEBS Lett 369:153–157 . doi: 10.1016/0014-5793(95)00728-r

134. Birtle Z, Ponting CP (2006) Meisetz and the birth of the KRAB motif. Bioinformatics 22:2841–2845 . doi: 10.1093/bioinformatics/btl498

135. Baudat F, Buard J, Grey C, et al (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327:836–840 . doi: 10.1126/science.1183439

136. Myers S, Bowden R, Tumian A, et al (2010) Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327:876–879 . doi: 10.1126/science.1182363

137. Powers NR, Parvanov ED, Baker CL, et al (2016) The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. PLoS Genet 12:e1006146 . doi: 10.1371/journal.pgen.1006146

138. Imai Y, Baudat F, Taillepierre M, et al (2017) The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. Chromosoma 126:681–695 . doi: 10.1007/s00412-017-0631-z

139. Patel A, Horton JR, Wilson GG, et al (2016) Structural basis for human PRDM9 action at recombination hot spots. Genes Dev 30:257–265 . doi: 10.1101/gad.274928.115

140. Leseva M, Knowles BB, Messerschmidt DM, Solter D (2015) Erase-Maintain-Establish: Natural Reprogramming of the Mammalian Epigenome. Cold Spring Harb Symp Quant Biol 80:155–163 . doi: 10.1101/sqb.2015.80.027441

141. Born N, Thiesen H-J, Lorenz P (2014) The B-subdomain of the Xenopus laevis XFIN KRAB-AB domain is responsible for its weaker transcriptional repressor activity compared to human ZNF10/Kox1. PLoS One 9:e87609 . doi: 10.1371/journal.pone.0087609

142. Williams AJ, Blacklow SC, Collins T (1999) The Zinc Finger-Associated SCAN Box Is a Conserved Oligomerization Domain. Mol Cell Biol 19:8526–8535 . doi: 10.1128/mcb.19.12.8526

143. Sander TL, Haas AL, Peterson MJ, Morris JF (2000) Identification of a novel SCAN box-related protein that interacts with MZF1B. The leucine-rich scan box mediates hetero- and homoprotein associations. J Biol Chem 275:12857–12867 . doi: 10.1074/jbc.275.17.12857

144. Schumacher C, Wang H, Honer C, et al (2000) The SCAN domain mediates selective oligomerization. J Biol Chem 275:17173–17179 . doi: 10.1074/jbc.M000119200

145. Emerson RO, Thomas JH (2011) Gypsy and the Birth of the SCAN Domain. J Virol 85:12043–12052 . doi: 10.1128/jvi.00867-11

146. Ivanov D, Stone JR, Maki JL, et al (2005) Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. Mol Cell 17:137–143 . doi: 10.1016/j.molcel.2004.12.015

147. Rowe HM, Jakobsson J, Mesnard D, et al (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463:237–240 . doi: 10.1038/nature08674

148. Rowe HM, Kapopoulou A, Corsinotti A, et al (2013) TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome Res 23:452–461 . doi: 10.1101/gr.147678.112

149. Matsui T, Leung D, Miyashita H, et al (2010) Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 464:927–931 . doi: 10.1038/nature08858

150. Rowe HM, Friedli M, Offner S, et al (2013) De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development 140:519–529 . doi: 10.1242/dev.087585

151. Quenneville S, Turelli P, Bojkowska K, et al (2012) 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 . doi: 10.1016/j.celrep.2012.08.043

152. Ecco G, Cassano M, Kauzlaric A, et al (2016) Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. Dev Cell 36:611–623 . doi: 10.1016/j.devcel.2016.02.024

153. Playfoot CJ, Adams IR (2016) KRABs RegulaTE Gene Expression beyond the Embryo. Dev Cell 36:591–592 . doi: 10.1016/j.devcel.2016.03.006

154. Pontis J, Planet E, Offner S, et al (2019) Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. Cell Stem Cell 24:724-735.e5 . doi: 10.1016/j.stem.2019.03.012

155. Cheng CT, Kuo CY, Ann DK (2014) KAPtain in charge of multiple missions: Emerging roles of KAP1. World J Biol Chem 5:308–320 . doi: 10.4331/wjbc.v5.i3.308

156. Letunic I, Bork P (2018) 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46:D493–D496 . doi: 10.1093/nar/gkx922

157. Ryan RF, Schultz DC, Ayyanathan K, et al (1999) KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. Mol Cell Biol 19:4366–4378 . doi: 10.1128/mcb.19.6.4366

158. Peng H, Gibson LC, Capili AD, et al (2007) 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 . doi: 10.1016/j.jmb.2007.03.047

159. Peng H, Begg GE, Schultz DC, et al (2000) 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 295:1139–62 . doi: 10.1006/jmbi.1999.3402

160. Doyle JM, Gao J, Wang J, et al (2010) MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases. Mol Cell 39:963–974 . doi: 10.1016/j.molcel.2010.08.029

161. Xiao TZ, Bhatia N, Urrutia R, et al (2011) MAGE I transcription factors regulate KAP1 and KRAB domain zinc finger transcription factor mediated gene repression. PLoS One 6: . doi: 10.1371/journal.pone.0023747

162. Xiao TZ, Suh Y, Longley BJ (2014) MAGE proteins regulate KRAB zinc finger transcription factors and KAP1 E3 ligase activity. Arch Biochem Biophys 563:136–144 . doi: 10.1016/j.abb.2014.07.026

163. Pineda CT, Ramanathan S, Fon Tacer K, et al (2015) Degradation of AMPK by a cancer-specific ubiquitin ligase. Cell 160:715–728 . doi: 10.1016/j.cell.2015.01.034

164. Liang Q, Deng H, Li X, et al (2011) 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 . doi: 10.4049/jimmunol.1101704

165. Ma X, Yang T, Luo Y, et al (2019) TRIM28 promotes HIV-1 latency by SUMOylating CDK9 and inhibiting P-TEFb. Elife 8: . doi: 10.7554/eLife.42426

166. Iyengar S, Farnham PJ (2011) KAP1 protein: An enigmatic master regulator of the genome. J Biol Chem 286:26267–26276 . doi: 10.1074/jbc.R111.252569

167. Zeng L, Yap KL, Ivanov A V., et al (2008) Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. Nat Struct Mol Biol 15:626–633 . doi: 10.1038/nsmb.1416

168. Mascle XH, Germain-Desprez D, Huynh P, et al (2007) Sumoylation of the transcriptional intermediary factor 1beta (TIF1beta), the Co-repressor of the KRAB Multifinger proteins, is required for its transcriptional activity and is modulated by the KRAB domain. J Biol Chem 282:10190–10202 . doi: 10.1074/jbc.M611429200

169. Short KM, Cox TC (2006) Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. J Biol Chem 281:8970–8980 . doi: 10.1074/jbc.M512755200

170. Esposito D, Koliopoulos MG, Rittinger K (2017) Structural determinants of TRIM protein function. Biochem Soc Trans 45:183–191 . doi: 10.1042/BST20160325

171. Ozato K, Shin D-M, Chang T-H, Morse HC 3rd (2008) TRIM family proteins and their emerging roles in innate immunity. Nat Rev Immunol 8:849–860 . doi: 10.1038/nri2413

172. Herquel B, Ouararhni K, Martianov I, et al (2013) Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. Nat Struct Mol Biol 20:339–346 . doi: 10.1038/nsmb.2496

173. Herquel B, Ouararhni K, Khetchoumian K, et al (2011) Transcription cofactors TRIM24, TRIM28, and TRIM33 associate to form regulatory complexes that suppress murine hepatocellular carcinoma. Proc Natl Acad Sci U S A 108:8212–8217 . doi: 10.1073/pnas.1101544108

174. Fong K-W, Zhao JC, Song B, et al (2018) TRIM28 protects TRIM24 from SPOP-mediated degradation and promotes prostate cancer progression. Nat Commun 9:5007 . doi: 10.1038/s41467-018-07475-5

175. Sanchez JG, Okreglicka K, Chandrasekaran V, et al (2014) The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer. Proc Natl Acad Sci 111:2494–2499 . doi: 10.1073/pnas.1318962111

176. Li Y, Wu H, Wu W, et al (2014) Structural insights into the TRIM family of ubiquitin E3 ligases. Cell Res 24:762–765 . doi: 10.1038/cr.2014.46

177. Goldstone DC, Walker PA, Calder LJ, et al (2014) Structural studies of postentry restriction factors reveal antiparallel dimers that enable avid binding to the HIV-1 capsid lattice. Proc Natl Acad Sci 111:9609–9614 . doi: 10.1073/pnas.1402448111

178. Weinert C, Morger D, Djekic A, et al (2015) Crystal structure of TRIM20 C-terminal coiled-coil/B30.2 fragment: Implications for the recognition of higher order oligomers. Sci Rep 5:1–10 . doi: 10.1038/srep10819

179. Garvin AJ, Densham RM, Blair-Reid SA, et al (2013) The deSUMOylase SENP7 promotes chromatin relaxation for homologous recombination DNA repair. EMBO Rep 14:975–83 . doi: 10.1038/embor.2013.141

180. Yang BX, El Farran CA, Guo HC, et al (2015) Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. Cell 163:230–245 . doi: 10.1016/j.cell.2015.08.037

181. White DE, Negorev D, Peng H, et al (2006) KAP1, a novel substrate for PIKK family members, colocalizes with numerous damage response factors at DNA lesions. Cancer Res 66:11594–11599 . doi: 10.1158/0008-5472.CAN-06-4138

182. Ziv Y, Bielopolski D, Galanty Y, et al (2006) Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. Nat Cell Biol 8:870–876 . doi: 10.1038/ncb1446

183. Li X, Lee YK, Jeng JC, et al (2007) Role for KAP1 serine 824 phosphorylation and sumoylation/desumoylation switch in regulating KAP1-mediated transcriptional repression. J Biol Chem 282:36177–36189 . doi: 10.1074/jbc.M706912200

184. Li X, Lin HH, Chen H, et al (2010) SUMOylation of the transcriptional co-repressor KAP1 is regulated by the serine and threonine phosphatase PP1. Sci Signal 3:ra32 . doi: 10.1126/scisignal.2000781

185. Goodarzi AA, Kurka T, Jeggo PA (2011) KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. Nat Struct Mol Biol 18:831–839 . doi: 10.1038/nsmb.2077

186. Seki Y, Kurisaki A, Watanabe-Susaki K, et al (2010) TIF1beta regulates the pluripotency of embryonic stem cells in a phosphorylation-dependent manner. Proc Natl Acad Sci 107:10926–10931 . doi: 10.1073/pnas.0907601107

187. Bolderson E, Savage KI, Mahen R, et al (2012) Krüppel-associated box (KRAB)-associated co-repressor (KAP-1) Ser-473 phosphorylation regulates heterochromatin protein 1β (HP1-β) mobilization and DNA repair in heterochromatin. J Biol Chem 287:28122–28131 . doi: 10.1074/jbc.M112.368381

188. Hu C, Zhang S, Gao X, et al (2012) Roles of Kruppel-associated box (KRAB)-associated co-repressor KAP1 Ser-473 phosphorylation in DNA damage response. J Biol Chem 287:18937–18952 . doi: 10.1074/jbc.M111.313262

189. White D, Rafalska-Metcalf IU, Ivanov A V., et al (2012) The ATM Substrate KAP1 Controls DNA Repair in Heterochromatin: Regulation by HP1 Proteins and Serine 473/824 Phosphorylation. Mol Cancer Res 10:401–414 . doi: 10.1158/1541-7786.MCR-11-0134

190. Chang C-JC-WC-JC-W, Chou H-Y, Lin Y-S, et al (2008) Phosphorylation at Ser473 regulates heterochromatin protein 1 binding and corepressor function of TIF1beta/KAP1. BMC Mol Biol 9:61 . doi: 10.1186/1471-2199-9-61

191. Singh K, Cassano M, Planet E, et al (2015) A KAP1 phosphorylation switch controls MyoD function during skeletal muscle differentiation. Genes Dev 29:513–525 . doi: 10.1101/gad.254532.114

192. Kubota S, Fukumoto Y, Aoyama K, et al (2013) Phosphorylation of KRAB-associated protein 1 (KAP1) at Tyr-449, Tyr-458, and Tyr-517 by nuclear tyrosine kinases inhibits the association of KAP1 and Heterochromatin Protein 1α (HP1α) with heterochromatin. J Biol Chem 288:17871–17883 . doi: 10.1074/jbc.M112.437756

193. Van Meter M, Kashyap M, Rezazadeh S, et al (2014) SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. Nat Commun 5:5011 . doi: 10.1038/ncomms6011

194. Lai IL, Lin TP, Yao YL, et al (2010) Histone deacetylase 10 relieves repression on the melanogenic program by maintaining the deacetylation status of repressors. J Biol Chem 285:7187–7196 . doi: 10.1074/jbc.M109.061861

195. Lin YH, Yuan J, Pei H, et al (2015) KAP1 deacetylation by SIRT1 promotes non-homologous end-joining repair. PLoS One 10:1–18 . doi: 10.1371/journal.pone.0123935

196. Castro-Diaz N, Ecco G, Coluccio A, et al (2014) Evolutionally dynamic L1 regulation in embryonic stem cells. Genes Dev 28:1397–1409 . doi: 10.1101/gad.241661.114

197. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM (2009) The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc B Biol Sci 364:99–115 . doi: 10.1098/rstb.2008.0168

198. Dumesic PA, Madhani HD (2014) Recognizing the enemy within: Licensing RNA-guided genome defense. Trends Biochem Sci 39:25–34 . doi: 10.1016/j.tibs.2013.10.003

199. Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: The vanguard of genome defence. Nat Rev Mol Cell Biol 12:246–258 . doi: 10.1038/nrm3089

200. Carmell MA, Girard A, van de Kant HJG, et al (2007) MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. Dev Cell 12:503–514 . doi: 10.1016/j.devcel.2007.03.001

201. Aravin AA, Sachidanandam R, Girard A, et al (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 316:744–747 . doi: 10.1126/science.1142612

202. Marchetto MCN, Narvaiza I, Denli AM, et al (2013) Differential L1 regulation in pluripotent stem cells of humans and apes. Nature 503:525–529 . doi: 10.1038/nature12686

203. Yang N, Kazazian HH (2006) L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol 13:763–771 . doi: 10.1038/nsmb1141

204. Heras SR, MacIas S, Plass M, et al (2013) The Microprocessor controls the activity of mammalian retrotransposons. Nat Struct Mol Biol 20:1173–1183 . doi: 10.1038/nsmb.2658

205. Tchasovnikarova IA, Timms RT, Matheson NJ, et al (2015) GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science 348:1481–1485 . doi: 10.1126/science.aaa7227

206. Liu N, Lee CH, Swigut T, et al (2018) Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. Nature 553:228–232 . doi: 10.1038/nature25179

207. Robbez-Masson L, Tie CHC, Conde L, et al (2018) The hush complex cooperates with trim28 to repress young retrotransposons and new genes. Genome Res 28:836–845 . doi: 10.1101/gr.228171.117

208. Timms RT, Tchasovnikarova IA, Lehner PJ (2016) Position-effect variegation revisited: HUSHing up heterochromatin in human cells. BioEssays 38:333–343 . doi: 10.1002/bies.201500184

209. Tchasovnikarova IA, Timms RT, Douse CH, et al (2017) Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. Nat Genet 49:1035–1044 . doi: 10.1038/ng.3878

210. Harris RS, Dudley JP (2015) APOBECs and virus restriction. Virology 479–480:131–145 . doi: 10.1016/j.virol.2015.03.012

211. Carmi S, Church GM, Levanon EY (2011) Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. Nat Commun 2:519 . doi: 10.1038/ncomms1525

212. Richardson SR, Narvaiza I, Planegger RA, et al (2014) APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. Elife 2014:1–20 . doi: 10.7554/eLife.02008

213. Armitage AE, Katzourakis A, de Oliveira T, et al (2008) Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences. J Virol 82:8743–8761 . doi: 10.1128/JVI.00584-08

214. Horn A V, Klawitter S, Held U, et al (2014) Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. Nucleic Acids Res 42:396–416 . doi: 10.1093/nar/gkt898

215. Bogerd HP, Wiegand HL, Hulme AE, et al (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A 103:8780–8785 . doi: 10.1073/pnas.0603313103

216. Stetson DB, Ko JS, Heidmann T, Medzhitov R (2008) Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. Cell 134:587–598 . doi: 10.1016/j.cell.2008.06.032

217. Li P, Du J, Goodier JL, et al (2017) Aicardi-Goutières syndrome protein TREX1 suppresses L1 and maintains genome integrity through exonuclease-independent ORF1p depletion. Nucleic Acids Res 45:4619–4631 . doi: 10.1093/nar/gkx178

218. Zhao K, Du J, Han X, et al (2013) Modulation of LINE-1 and Alu/SVA Retrotransposition by Aicardi-Goutières Syndrome-Related SAMHD1. Cell Rep 4:1108–1115 . doi: 10.1016/j.celrep.2013.08.019

219. Hu S, Li J, Xu F, et al (2015) SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation. PLoS Genet 11:e1005367 . doi: 10.1371/journal.pgen.1005367

220. Herrmann A, Wittmann S, Thomas D, et al (2018) The SAMHD1-mediated block of LINE-1 retroelements is regulated by phosphorylation. Mob DNA 9:1–17 . doi: 10.1186/s13100-018-0116-5

221. Goodier JL, Pereira GC, Cheung LE, et al (2015) The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. PLoS Genet 11:e1005252 . doi: 10.1371/journal.pgen.1005252

222. Moldovan JB, Moran J V (2015) The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. PLoS Genet 11:e1005121 . doi: 10.1371/journal.pgen.1005121

223. Zhang A, Dong B, Doucet AJ, et al (2014) RNase L restricts the mobility of engineered retrotransposons in cultured human cells. Nucleic Acids Res 42:3803–3820 . doi: 10.1093/nar/gkt1308

224. Lallemand-Breitenbach V, de The H (2018) PML nuclear bodies: from architecture to function. Curr Opin Cell Biol 52:154–161 . doi: 10.1016/j.ceb.2018.03.011

225. Ganser-Pornillos BK, Pornillos O (2019) Restriction of HIV-1 and other retroviruses by TRIM5. Nat Rev Microbiol 17:546–556 . doi: 10.1038/s41579-019-0225-2

226. Wang P, Benhenda S, Wu H, et al (2018) RING tetramerization is required for nuclear body biogenesis and PML sumoylation. Nat Commun 9:1–10 . doi: 10.1038/s41467-018-03498-0

227. Wagner JM, Roganowicz MD, Skorupka K, et al (2016) Mechanism of B-box 2 domain-mediated higher-order assembly of the retroviral restriction factor TRIM5alpha. Elife 5: . doi: 10.7554/eLife.16309

228. Ding D, Bergmaier P, Sachs P, et al (2018) The CUE1 domain of the SNF2-like chromatin remodeler SMARCAD1 mediates its association with KRAB-associated protein 1 (KAP1) and KAP1 target genes. J Biol Chem 293:2711–2724 . doi: 10.1074/jbc.RA117.000959

229. Li J, Xi Y, Li W, et al (2017) TRIM28 interacts with EZH2 and SWI/SNF to activate genes that promote mammosphere formation. Oncogene 36:2991–3001 . doi: 10.1038/onc.2016.453

230. Sachs P, Ding D, Bergmaier P, et al (2019) SMARCAD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. Nat Commun 10: . doi: 10.1038/s41467-019-09078-0

231. Keown JR, Yang JX, Douglas J, Goldstone DC (2016) Characterisation of assembly and ubiquitylation by the RBCC motif of Trim5alpha. Sci Rep 6:26837 . doi: 10.1038/srep26837

232. Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, et al (2011) Hexagonal assembly of a restricting TRIM5alpha protein. Proc Natl Acad Sci U S A 108:534–539 . doi: 10.1073/pnas.1013426108

233. Li Y-L, Chandrasekaran V, Carter SD, et al (2016) Primate TRIM5 proteins form hexagonal nets on HIV-1 capsids. Elife 5: . doi: 10.7554/eLife.16269

234. Koliopoulos MG, Esposito D, Christodoulou E, et al (2016) Functional role of TRIM E 3 ligase oligomerization and regulation of catalytic activity. EMBO J 35:1–15 . doi: 10.15252/embj.201593741

235. Ganser-Pornillos BK, Pornillos O (2019) Restriction of HIV-1 and other retroviruses by TRIM5. Nat Rev Microbiol 9: . doi: 10.1038/s41579-019-0225-2

236. Huang SY, Naik MT, Chang CF, et al (2014) The B-box 1 dimer of human promyelocytic leukemia protein. J Biomol NMR 60:275–281 . doi: 10.1007/s10858-014-9869-4

237. Gorrec F (2016) Protein crystallization screens developed at the MRC Laboratory of Molecular Biology. Drug Discov Today 21:819–825 . doi: 10.1016/j.drudis.2016.03.008

238. Waugh DS (2016) Crystal structures of MBP fusion proteins. Protein Sci 25:559–571 . doi: 10.1002/pro.2863

239. Jin T, Chuenchor W, Jiang J, et al (2017) Design of an expression system to enhance MBP-mediated crystallization. Sci Rep 7:1–11 . doi: 10.1038/srep40991

240. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D (2007) Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. Protein Sci 16:1569–1576 . doi: 10.1110/ps.072914007

241. Rayment I, Holden HM (1994) The three-dimensional structure of a molecular motor. Trends Biochem Sci 19:129–134 . doi: 10.1016/0968-0004(94)90206-2

242. Kim Y, Quartey P, Li H, et al (2008) Large-scale evaluation of protein reductive methylation for improving protein crystallization. Nat. Methods 5:853–854

243. Walter TS, Meier C, Assenberg R, et al (2006) Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. Structure 14:1617–1622 . doi: 10.1016/j.str.2006.09.005

244. Koliopoulos MG, Lethier M, Van Der Veen AG, et al (2018) Molecular mechanism of influenza A NS1-mediated TRIM25 recognition and inhibition. Nat Commun 9: . doi: 10.1038/s41467-018-04214-8

245. Kelley LA, Mezulis S, Yates CM, et al (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845

246. Sripathy SP, Stevens J, Schultz DC (2006) The KAP1 Corepressor Functions To Coordinate the Assembly of De Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-Mediated Transcriptional Repression. Mol Cell Biol 26:8623–8638 . doi: 10.1128/mcb.00487-06

247. Fonti G, Marcaida M, Bryan L, et al (2019) KAP1 is an antiparallel dimer with a natively functional asymmetry. bioRxiv 1:553511 . doi: 10.1101/553511

248. Sun Y, Keown JR, Black MM, et al (2019) A Dissection of Oligomerization by the TRIM28 Tripartite Motif and the Interaction with Members of the Krab-ZFP Family. J Mol Biol 431:2511–2527 . doi: 10.1016/j.jmb.2019.05.002

249. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797 . doi: 10.1093/nar/gkh340

250. Yudina Z, Roa A, Johnson R, et al (2015) RING Dimerization Links Higher-Order Assembly of TRIM5alpha to Synthesis of K63-Linked Polyubiquitin. Cell Rep 12:788–797 . doi: 10.1016/j.celrep.2015.06.072

251. Dawidziak DM, Sanchez JG, Wagner JM, et al (2017) Structure and catalytic activation of the TRIM23 RING E3 ubiquitin ligase. Proteins 85:1957–1961 . doi: 10.1002/prot.25348

252. Sanchez JG, Chiang JJ, Sparrer KMJ, et al (2016) Mechanism of TRIM25 Catalytic Activation in the Antiviral RIG-I Pathway. Cell Rep 16:1315–1325 . doi: 10.1016/j.celrep.2016.06.070

253. Dou H, Buetow L, Sibbet GJ, et al (2013) Essentiality of a non-RING element in priming donor ubiquitin for catalysis by a monomeric E3. Nat Struct Mol Biol 20:982–986 . doi: 10.1038/nsmb.2621

254. Stevens R V, Esposito D, Rittinger K (2019) Characterisation of class VI TRIM RING domains: linking RING activity to C-terminal domain identity. Life Sci alliance 2: . doi: 10.26508/lsa.201900295

255. Shen TH, Lin H-K, Scaglioni PP, et al (2006) The mechanisms of PML-nuclear body formation. Mol Cell 24:331–339 . doi: 10.1016/j.molcel.2006.09.013

256. Duprez E, Saurin AJ, Desterro JM, et al (1999) SUMO-1 modification of the acute promyelocytic leukaemia protein PML: implications for nuclear localisation. J Cell Sci 112 (Pt 3:381–393

257. Larkin MA, Blackshields G, Brown NP, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948 . doi: 10.1093/bioinformatics/btm404

258. Lim M, Newman JA, Williams HL, et al (2019) A Ubiquitin-Binding Domain that Binds a Structural Fold Distinct from that of Ubiquitin. Structure 27:1316-1325.e6 . doi: 10.1016/j.str.2019.05.003

259. Skene PJ, Henikoff S (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife 6: . doi: 10.7554/eLife.21856

260. Taura M, Song E, Ho YC, Iwasaki A (2019) Apobec3A maintains HIV-1 latency through recruitment of epigenetic silencing machinery to the long terminal repeat. Proc Natl Acad Sci U S A 116:2282–2289 . doi: 10.1073/pnas.1819386116

261. Bunch H, Zheng X, Burkholder A, et al (2014) TRIM28 regulates RNA polymerase II promoter-proximal pausing and pause release. Nat Struct Mol Biol 21:876–883 . doi: 10.1038/nsmb.2878

262. McNamara RP, Reeder JE, McMillan EA, et al (2016) KAP1 Recruitment of the 7SK snRNP Complex to Promoters Enables Transcription Elongation by RNA Polymerase II. Mol Cell 61:39–53 . doi: 10.1016/j.molcel.2015.11.004

263. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292:195–202 . doi: 10.1006/jmbi.1999.3091

264. Aricescu AR, Lu W, Jones EY (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr 62:1243–1250 . doi: 10.1107/S0907444906029799

265. Tie CH, Rowe HM (2017) Epigenetic control of retrotransposons in adult tissues: implications for immune regulation. Curr Opin Virol 25:28–33 . doi: 10.1016/j.coviro.2017.06.007

266. Vonrhein C, Flensburg C, Keller P, et al (2011) Data processing and analysis with the autoPROC toolbox. Acta Crystallogr D Biol Crystallogr 67:293–302 . doi: 10.1107/S0907444911007773

267. Adams PD, Afonine P V, Bunkoczi G, et al (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213–221 . doi: 10.1107/S0907444909052925

268. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132 . doi: 10.1107/S0907444904019158

269. Murshudov GN, Skubak P, Lebedev AA, et al (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67:355–367 . doi: 10.1107/S0907444911001314

270. Rose DR, Phipps J, Michniewicz J, et al (1988) Crystal structure of T4-lysozyme generated from synthetic coding DNA expressed in Escherichia coli. Protein Eng Des Sel 2:277–282 . doi: 10.1093/protein/2.4.277

271. Weaver LH, Matthews BW (1987) Structure of bacteriophage T4 lysozyme refined at 1.7 A resolution. J Mol Biol 193:189–199 . doi: 10.1016/0022-2836(87)90636-x

272. Winter G, Waterman DG, Parkhurst JM, et al (2018) DIALS: implementation and evaluation of a new integration package. Acta Crystallogr Sect D, Struct Biol 74:85–97 . doi: 10.1107/S2059798317017235

273. Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69:1204–1214 . doi: 10.1107/S0907444913000061

274. Winter G (2010) {\it xia2}: an expert system for macromolecular crystallography data reduction. J Appl Crystallogr 43:186–190 . doi: 10.1107/S0021889809045701

275. Winn MD, Ballard CC, Cowtan KD, et al (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67:235–242 . doi: 10.1107/S0907444910045749

276. Schuck P (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal Biochem 320:104–124 . doi: 10.1016/s0003-2697(03)00289-6